ELSEVIER

Computer Networks 36 (2001) 533–555

COMPUTER NETWORKS

www.elsevier.com/locate/comnet

Stochastic models for telecom commodity prices

Chris Kenyon *, Giorgos Cheliotis

IBM Research, Zurich Research Laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland

Abstract

Bandwidth is becoming commoditized and markets are starting to appear. Potential behaviors of these markets are not yet understood because these markets are still in the early stages of development. This is reflected in the lack of current research on the structure and dynamics of network commodity market prices. We present a method for constructing telecom commodity spot price processes as a first step for understanding these developing markets. Bandwidth, like electricity, is not storable so we draw inspiration from electricity prices and models. However, unique network features of telecommunications require specific inclusion. These are geographical substitution (arbitrage), quality of service (QoS), and the continuing pace of technological development. Developing liquidity acts as a further complication. Thus we model price development as a combination of link price processes modified by prices for equivalent QoS routes. We demonstrate our method on a simple triangular network topology and characterize a network contract graph derived from more than 10 major carrier backbones and new entrant networks. Our results cover the existence and value of arbitrage opportunities together with their effect on price development and network value (NPV). Application of this work ranges from network design to infrastructure valuation and construction of real options. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Bandwidth; Markets; Real options; Quality of service; Geographical arbitrage

1. Introduction

Bandwidth is becoming commoditized and markets are starting to appear, developing as an extension of phone-minute trading (see [15] for an introduction). Typical companies involved include Enron Broadband, Williams Communications and newer specialized companies such as LightTrade, RateXchange, Band-X, GTX, Arbinet, AIG, and Bandwidth Market. Estimates of potential market size range as high as a good fraction of $700B. [1]

We present here a model for spot price development of commodity bandwidth contracts. For concreteness the reader may imagine that a standard commodity point-to-point bandwidth contract is for T3 (45 Mb/s) capacity (with defined delay, jitter, packet loss, etc.) and several timescales of contracts are available with standardized starting times (e.g., every 15 min/h at:00/day at 00:00) and lengths. This model is applicable for the following uses: price forecasting in a network context; valuation of network infrastructure applicable to investment or swapping; as an input for pricing derivative contracts. We demonstrate the first two uses in a simple network. Price volatility is a key input to derivative pricing and we examine this as well.

* Corresponding author.
E-mail addresses: chk@zurich.ibm.com (C. Kenyon), gic@zurich.ibm.com (G. Cheliotis).

[1] http://www.derivativesstrategy.com/magazine, August 2000.

1389-1286/01/$ - see front matter
PII: S 1 3 8 9 - 1 2 8 6 (0 1) 0 0 1 8 6 - 4

Potential behaviors of bandwidth markets are not understood because these markets are still in the early stages of development. This is reflected in the lack of current research on the structure and dynamics of network commodity market prices. We present our expectations for the stylized features of these commodity prices and a method for constructing telecom commodity spot price processes. Bandwidth, like electricity, is not storable so we draw inspiration from electricity prices and models. However, we argue that unique features of telecommunications require specific inclusion in spot price process modeling. These are geographical substitution (arbitrage), quality of service (QoS), and the continuing pace of technological development. Developing liquidity acts as a further complication. Liquidity refers to the ease with which partners for trades at a given price can be found. Geographical arbitrage means that spot price development on point-to-point links cannot be understood in isolation from price development on alternative routes with equivalent QoS. This implies some form of price modification derived from load balancing across appropriately specified QoS-limited alternative routes. Technological development continually pushes prices down as new equipment is installed by competitors. So, unlike other commodities, the prices revert towards a mean whose drift is strongly structurally downwards. Market liquidity is quantified as the extent to which geographical arbitrage modifies link prices. Hence we model price development as a combination of link price processes modified by prices for equivalent QoS routes.

We present a method for constructing telecom commodity spot price processes as a first step for understanding these developing markets. We demonstrate our method on a simple triangular network topology and characterize a network contract graph derived from more than 10 major carrier (e.g., AT&T) backbones and new entrants (e.g., Level3, Enron). Our results cover the existence and value of arbitrage opportunities together with their effect on price development and network value (NPV). Application of this work ranges from network design to infrastructure valuation and construction of real options.

2. Previous work

Modeling of telecom prices is in its infancy, which reflects the immaturity of bandwidth as a commodity. Chiu et al. [8] proposed using Pilipović's pricing model (see [22,30]) from energy markets, which is very similar to that developed by Schwartz and Smith [28] for general commodities, which in turn was exactly equivalent to an earlier model [26] under a linear transformation of parameters. The observations of congestion on the Internet suggest, however, that even for single links, these models are insufficient because they do not include spikes or jumps in prices. Price spikes are particular features of electricity prices, for which some modeling has been done [10,11]. Jumps are a feature of oil prices and have been included in a model sponsored by an oil company [13] as well as in numerous academic studies going as far back as [19]. We take the view that a price spike will also represent a (local) change in the price equation itself and thus we can draw from the regime switching literature [1,12]. The latter paper is of special interest as it incorporates feedback. We also use feedback in the sense that the size of a spike sets the new mean until the spike has passed.

Quite apart from link price processes in isolation there is the question of network effects in the sense of interactions. We are neither referring here to the "more compatibility is better" notion nor to that of increasing returns "network effects" in economics (e.g., [14]). The most important network effect here is geographical arbitrage, i.e., the existence of many prices for end-to-end capacity at equivalent QoS. This has been observed in the market as reported by Cheliotis [6]. Detection of such arbitrage opportunities is in general an NP-complete problem based on shortest path algorithms with side constraints. However Cheliotis [5] described a variety of pseudo-polynomial time algorithms that result from quantization of these constraints as is to be expected under commoditization. Chiu and Crametz [7] is related in that they showed that geographical arbitrage opportunities could exist in the forward market even with a no-arbitrage situation in the spot market.

Lastly mention should be made of the vast computer science literature on price setting for network resources in order to achieve a given aim, e.g., social welfare maximization [25], cost allocation [9], congestion control [23], etc. [27]. We approach price dynamics from a completely different direction in that we start by modeling the price process rather than modeling supply and demand and then solving for the best price in some sense relative to a given network.

As far as the authors are aware there has as yet been no investigation of the expected telecom commodity price dynamics using plausible link price processes together with modifications due to network effects. This is the subject of the current paper.

3. Factors

Bandwidth is a geographically distributed commodity, and the major long-distance suppliers form an oligopoly. Here we list and discuss factors that may affect telecom commodity pricing in comparison with other commodities, primarily electricity. [2]

Geographical arbitrage. This means that, given the same QoS, the cheapest of all available routes will set the end-to-end price in a competitive (liquid) market. This is because the actual route is irrelevant with respect to data transport, as long as certain QoS requirements are met. The set of routes connecting two given geographical locations at the same QoS level are perfect substitutes.

Unlike in electrical networks – an electrical network flow being a whole – individual flows do not have alternative routes (except in the sense that aggregate flows can be divided logically in different ways). Thus an electricity network is a coherent whole, with energy flowing according to well-understood laws of physics (Kirchoff's law, etc.). The bandwidth market has no corresponding inherent large-scale feature. The requirement that supply and demand must balance in electricity leads directly to pool-type price discovery mechanisms (markets). In data transmission, routing/switching protocols state the laws for the formation of flows. The choice of protocols and the possibility to combine several complementary technologies allow a much higher degree of flexibility and control over data routing. This degree of control is indeed necessary for the design and operation of efficient data networks and obviates any requirement for centralized pool pricing.

Electrical power networks are a means of distributing energy, with energy being the traded asset in the electricity market, whereas data networks are themselves the underlying commodity of every bandwidth contract.

Supply/demand disassociation. This refers to the fact that the good that is in demand is not usually the one supplied. In electricity the demand is for power at, say, one location but the power is usually supplied from another location. This has been a particular problem in some markets where power was available at one location but, for congestion or legal reasons, could not be transported to where it was needed.

In bandwidth, demand is for end-to-end service but the actual supply of the good is – at a physical level and investment level – on a link-by-link basis. Certainly, collections of links are owned by single entities, so for subsets of the entire network, investment policy will have some coherence. Also new links can be created, but the disassociation remains.

Network externalities. This is the increasing economies of scale argument where the utility of a service grows, the more people – or devices – are attached. In general the utility growth is some power of the number of users, not simply a linear effect [29].

This increasing economies of scale, if it is present at all electricity networks, is a linear or sublinear phenomenon. More users of electricity means a bigger market for electrical devices but there is little interaction. Also the market in the developed world is basically saturated. The openings for new consumption of electricity are very limited. On the contrary, demand for bandwidth is continuously increasing, largely due to network externalities.

Note that we use the term network effects rather than network externalities later to describe the

[2] The PSERC archive contains an excellent collection of research on deregulated electricity markets at www.pserc.wisc.edu.

effect of geographical arbitrage on price development.

Non-storability. Inventories act to smooth variations in supply and demand. When no inventories exist, prices can jump if supply or demand changes suddenly. Prices can also changes suddenly when the perception or expectation of supply or demand status suddenly changes. Bandwidth is non-storable so price jumps and spikes (in both directions) are to be expected.

This is a determining factor in electricity price modeling. Jumps and especially spikes are observed often deriving from weather events (e.g., summer 1998, 1999 in Texas, Australia, California) sometimes in combination with equipment failures. In fact even in commodities where storage is possible, like oil, large-scale political events can still cause jumps and spikes in prices. Oil demonstrated a spike in 1990 during the Gulf War and several jumps, in both directions, depending on changes in OPEC policies from 1973 to the present.

Trading and settlement timescale. In some electricity markets, 30-min and even 10-min blocks are priced, traded and settled. To date, bandwidth contracts have had the character of the regulated electricity industry in that monthly or annual contracts are the norm. With universal trading contracts (i.e., bandwidth deregulation and contact feature standardization) and carrier-neutral pooling points we expect this to change and approximate the electricity market much more closely.

Liquidity. Currently the bandwidth market is less liquid than the electricity or indeed any other commodity market. Trading volumes are increasing, and ongoing deregulation of the industry along with universal trading contracts will contribute to higher levels of liquidity. In any case, we do not expect all traded locations to be equally liquid in the future.

Demand inelasticity. Provided that most consumers do not react to the timescale of market trading and settlement this will be a feature of total point-to-point demand on this timescale. On an individual link basis, demand will be elastic thanks to automation technologies, such as software agents, electronic auctions (e.g., [24]) and least-cost routing, which allow fast switching between substitutes (alternative paths). This elasticity will exploit and contribute to market liquidity. Inelastic demand is a current feature of short-term (days to weeks) electricity markets.

We are being cautious when assuming inelastic point-to-point demand and do so because we examine network effects which involve only demand shifting between substitutes, not changes in total demand. This is why in the analysis that follows we assume that point-to-point demand is conserved, whereas generally demand for a particular substitute is fairly inelastic but not necessarily completely inelastic.

Growth. The Internet, and network bandwidth available, have had periods of 100% growth every 3–4 months. In the past few years this has slowed to only 100% per year. In addition enormous amounts of dark fiber are being laid to take advantage of and push further growth. [3] The energy industry is growing at a much slower rate than the Internet, or indeed, bandwidth. In fact energy growth in the First World is barely 5% per year, if that, partly because of increasing efficiency.

Deregulation. In the bandwidth market there is currently a transition between mostly closed proprietary networks (such as those owned by Quest, MCI Worldcom, Level3, AT&T, Enron, etc.) and interoperability. This interoperability takes the form of a common infrastructure layer (e.g., IP), open public points of interconnection and universal trading contracts (UTC) with QoS guarantees.

These UTCs will have effects that mirror those of deregulation in the conventional utilities markets. These markets have been deregulated over a period of years and in a highly heterogeneous fashion. We may expect to see this same patchy development in bandwidth. However instead of geographically restricted patches there will be increasingly linked interconnection points.

Technological development. This is a dominant factor in bandwidth development encompassing two related areas: bandwidth and switching/routers. Bandwidth increases with both the increase in packing down a single wavelength and with the

[3] See Phone + magazine 12/18/2000.

increase in the number of wavelengths that can go down a single fiber (Dense Wavelength Division Multiplexing, DWDM). Switches are inherently parallel devices so they also exhibit a combined improvement: they improve as chips improve, and they improve with packing multiple switching units. A transition to all optical switching is expected within the next 5–10 years. The combined effect of technological development and competition is a continuous drop in the cost of transporting a Megabit per second per mile.

Technological development is today almost irrelevant in electricity markets. The only significant new factor has been the development of small gas turbines which have short response times (seconds to minutes). However these are expensive and have been built primarily as peaking units.

Capacity expansion. It takes a relatively long time to build a new fiber network and many months to put a new long-distance conduit in place. However once the conduit is in place new fibers can be pulled through (added) quite quickly, say, in a few weeks. Also conduits can house dark fiber (cable without the equipment necessary for transmission). In addition, with multimode fiber and DWDM, more wavelengths can be added to those already present in a lit fiber. If new equipment is required there is the added constraint of manufacturing schedules. Thus different amounts of capacity can be added over a range of timescales.

This is different from electricity where the fastest that significant new capacity can be added is approximately one year for a small gas-fired turbine (we ignore micro-turbines as having too low a capacity to be relevant) to several years for a thermal plant on a brown-field site to a decade for a new green-field development.

Supply elasticity. Suppliers of bandwidth may have significant flexibility in assigning network resources towards the fulfillment of different end-to-end contracts. This is due to routing and bandwidth management tools which allow for a number of different allocations, depending on which contracts the supplier wishes to offer. The amount of flexibility depends on the design of the underlying network. Generally, a network with more switching points will provide more flexibility (at the expense of QoS, because more contention/failure points are added).

In a similar manner, a power plant's resources can be assigned to different markets, but the allocation problem is different since it occurs only at one place, the plant, not on the distribution network.

Supply/demand balance. Utilization studies have produced widely varying results for different parts of the Internet and for private networks with results ranging from a few percent for some Internet backbones to observations of congestion on some links or switches on different days or at different times of day. Some next-generation research networks and especially backbones offering higher service quality are underutilized, but there is little, if any, evidence that the same holds for other large parts of the global infrastructure, such as corporate intranets.

This appears to be similar to electricity in general with the observation of congestion, i.e., blackouts, in electricity markets, but the driving factors are different. In electricity the biggest factors driving congestion are weather and equipment failure. The latter occur rather frequently in data networks.

Term structure of volatility. In commodity markets in general one of the stylized facts is that there is more volatility in the near-term forward market than in the long-term forward market. It is also usual that for non-investment commodities this volatility does not asymptote to zero. In the bandwidth market, whilst in some very long term there may be stability, we consider it more likely that within any reasonable planning horizon, say up to 5 years, there are increasing levels of uncertainty. So far prices are continuously falling (with rare exceptions), but even if we were to assume that this trend will continue, there is significant uncertainty concerning the rate of decline (both on a global and local level).

Positive prices. We assume that prices for bandwidth are always positive. In energy, zero real prices are observed. This is caused by large thermal or nuclear plants that cannot ramp down as quickly as demand can drop. Thus when demand is very low and drops suddenly these plants may give away power free as this is their only means of disposal. That is, in electricity, there is no free disposal.

4. Price development

In the previous section we described factors that are important for the development of bandwidth prices in general. In this section we will consider how these factors combine to give specific models for traded commodities. The basic method is shown in Fig. 1. The inputs to price development are: the network graph of traded contracts G; the initial link (contract) prices G_{init}; the network function f, which expresses how arbitrage opportunities are removed by market forces, and the time constant τ for their removal, i.e., the market liquidity (see Section 4.2); and the stochastic process models for each link (see Section 4.1). These are finally combined in Section 4.4.

The price development that Fig. 1 describes is that observed prices (the Market Observed Outcome, G_{arb}) may contain arbitrage opportunities. Market forces act both to remove the existence of arbitrage and to disturb the prices via normal and unusual information and changes in supply and demand. These two processes in general combine to produce the next set of observed prices. The

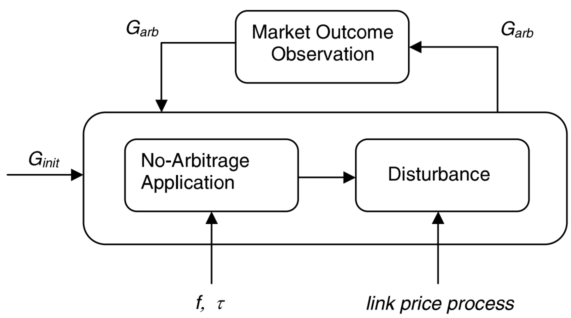

Fig. 1. Telecom commodity price development. Inputs are an initial network graph G_{init} of traded contracts and their prices, the network function f describing geographical arbitrage removal by load balancing, the speed of geographical arbitrage removal (market liquidity) τ, and the stochastic processes for individual links (contracts). Arbitrage removal may be combined with individual link price process. The prices observed in the market G_{arb} may contain geographical arbitrage opportunities. See Section 4 for details.

arbitrage removal occurs relative to the observed state of contract prices. However, as stated above, other market forces are still acting and this occurs at the same time. Thus the new state may not be arbitrage free, even with high liquidity, because there is always the potential for new arbitrage opportunities to be created. In practice we may model the two forces (geographical arbitrage removal and "the rest") separately or in combination. For simplicity we will first describe them separately (Section 4.1 for link prices and Section 4.2 for geographical arbitrage and liquidity) and then give a combined formulation (Section 4.4).

When considering bandwidth prices it is vital to be precise with respect to defining the underlying traded commodities. We model price development at the link level. We consider links to be indivisible contracts offered between pooling points. A pooling point is a facility for the exchange of traffic at a particular geographical point among trading partners in the commodity market. Any party may combine link contracts to form end-to-end contracts. Equally any party buying such an end-to-end contract may be able to split it according to the pooling points along the route to create new (link or multilink) contracts. Thus prices between any pair of pooling points may be observed on the market but these are formed from link prices. We do not model multilink price processes directly.

We model link price development as a combination of three factors: link price changes, geographical arbitrage, and liquidity. If the market is completely illiquid and if geographical arbitrage appears, there will be no action by market participants affecting prices to move the market in a direction to remove these opportunities. That is, traffic will not take the cheapest available route if the market is illiquid. On the other hand in a completely competitive, i.e., liquid, market, arbitrage opportunities will only last until the next trade (at most).

4.1. Independent link price models

The independent link price process models the evolution of prices on a link as if this link were isolated. We describe below the introduction of different forms of dependence. The realized prices in the market for the link are formed from these price proposals together with geographical arbitrage arguments and liquidity considerations. For the moment we are just interested in the link price process. This represents, as far as possible, the evolution of supply and demand of link capacity that providers make available in the form of indivisible contracts between the link endpoints. We first describe the generating equations and then their rationale.

4.1.1. Independent link prices: SRR model

Prices are generated for each link based on an Orstein–Uhlenbeck process with the addition of a process for the long-term mean, spike and jump terms. We also incorporate limited regime switching induced by the spike terms on the effective mean. We term this process a Shock-Regime-Reverting or SRR process.

We make the hypothesis that if there were no network then each link price S would develop as follows:

$$X = \log(S),$$
$$\mathrm{d}X = \eta(\overline{X} + GU - X)\,\mathrm{d}t + \sigma\,\mathrm{d}W + G\,\mathrm{d}U + H\,\mathrm{d}V,$$
$$\mathrm{d}\overline{X} = -\nu\,\mathrm{d}t + \rho\,\mathrm{d}Z,$$

where U is a two-state $\{0, 1\}$ semi-Markov process, and we make an identification between the states and the numbers zero and unity with rate parameters λ_U, μ_U and

$$G = \begin{cases} \mathrm{Gamma}(g_U, \alpha_U) & \text{if } \mathrm{d}U = +1, \\ G & \text{otherwise.} \end{cases}$$

Clearly Gamma stands for a Gamma distribution with scale parameter g and shape parameter α, thus $\mathrm{Gamma}(g, \alpha)$ has mean $g\alpha$ and variance $g^2\alpha$. When U jumps from state $\{0\}$ to state $\{1\}$ then X increases by G and the mean to which the process is reverting also increases by G. When U jumps from state $\{1\}$ to state $\{0\}$ then X decreases by the same amount it previously increased and the extra term in the mean is dropped. Thus a price spike is created. The price will stay in its current state for an exponentially distributed amount of time as given by the rate parameters.

V is a Poisson process with rate parameter λ_V and

$$H = \begin{cases} \mathrm{Gamma}(g_{\mathrm{up}}, \alpha_{\mathrm{up}}) & \text{if jump up,} \\ -\mathrm{Gamma}(g_{\mathrm{down}}, \alpha_{\mathrm{down}}) & \text{if jump down.} \end{cases}$$

Jumps may be equally probable in both directions or not as determined by the probability of an up jump p_{up}. These jumps act additively on the logarithm of the spot price and thus represent a multiplication of the current spot price.

Here η is the speed of price reversion to the average price $\overline{X}$; σ is the scale of the driving Brownian motion of price change increments $\mathrm{d}W$; ν is the (positive) instantaneous rate of average price decrease, and there is an uncertainty about this rate of size ρ (recall that $X = \log S$ so we are hypothesizing log-normal changes in S); $\mathrm{d}Z$ is a Brownian motion uncorrelated with $\mathrm{d}W$. Note that $\overline{X}$ is merely called an average price in that it is the price towards which S reverts. It is not actually an arithmetic average in that sense.

Note that as there is no storage there is no requirement that the process be a Martingale under any particular risk-neutral measure, although this could be arranged.

4.1.2. Link prices: rationale

We will now describe the price process in more detail. First note that the process is a semi-Markov jump diffusion with regime switching, thus the present determines the future. We assume, as in the original paper by Merton ([19] and at more length in [20]), that ordinary market news moves the price continuously and is responsible for the basic driving Brownian motion $\mathrm{d}W$.

Commodity markets often show reversion to some long-run mean, and we expect telecom to be no exception. Hence the use of an Orstein–Uhlenbeck process. However, in a significant departure from non-manufactured commodities, e.g. oil, wheat, etc., there are clear – non-zero – expectations on the speed of telecom capacity development. We expect the long-run mean, $\overline{X}$, to mimic the technological development of communications capacity with its exponential improvement (ν). The degree of the exponential improvement is not known but we may use an estimate, for example from the IBM General Technology Outlook. Whereas single-mode fiber capacity has shown exponential growth, the development of multi-mode (DWDM) transmission has radically improved capacity. Other such disruptive improvements are possible, e.g. long-distance transmission with no repeaters, all optical switching, etc. We model this uncertainty with the scale, ρ, of the driving Brownian motion dZ.

Spikes in prices have been observed in electricity prices and are the result of demand being very close to available supply followed by, say, some equipment failure. Congestion is observed on telecom networks, and equipment failures leading to outages have been observed. Thus we include spikes as a feature of the link price proposal processes. We define a spike to be a sudden increase in price followed quickly by a similar decrease in price. During the spike the mean for reversion is altered to include the magnitude of the spike. This change in regime is reversed when the spike ends. The change in mean during the spike is because the underlying cause, where this is equipment failure or something similar, has changed the underlying network. Thus the mean to which the price reverts should reflect the underlying change. When the underlying cause is fixed, the price and the mean reverse their previous changes.

Spike sizes are modeled with a Gamma distribution. This particular form is not important – what is important is that spikes generate reversible step changes in price. These discontinuous changes in price fit into the unusual information or Type II events in [20].

Price jumps are observed in oil prices, largely as a result of the perceived status of OPEC. Given that the owners of long-distance networks also form an oligopoly, there is the potential for price jumps. These may be local to a single link or more general. We only deal with the simplest case of independent effects on different links. Jump occurrences are modeled with a Poisson process with a given rate that describes how many jumps are expected per unit time. These jumps may be positive or negative, and again are modeled with Gamma distributions.

4.1.3. Dependent link prices

The simplest form of dependence between different link prices is to introduce correlations in the driving Brownian motions for the long- and short-term price variations, i.e., dZ and dW, between different links. We do not go into this in-depth here because we want to highlight a much more network-specific form of dependence: geographical arbitrage. Note however that the introduction of a correlation structure across a node allows us to trivially model the introduction of a pooling point on a previously undivided arc. If the new pooling point is actually redundant and all demand and supply actually crosses it, we can use perfect correlations in the driving processes on both sides of it. This still permits separation of rare events on either side of the pooling point.

4.2. Geographical arbitrage and liquidity

Geographical arbitrage is the term that has been used to describe the existence of at least two different end-to-end prices between two pooling points joined by a single link at a given end-to-end QoS. These two end-to-end prices may each be formed from one or more links but will both provide a required QoS level. Clearly in a liquid market this situation will not persist, if all other factors are equal. More precisely we offer the following definition.

Definition 1. A *simple geographical arbitrage* opportunity exists when multiple links can be substituted for a single link and when the total price of the substituted links is less than that of the single link. We assume of course that the QoS is equivalent between the single link and the end-to-end QoS of the substituting links.

Recall that a link represents an indivisible contract. Not all contracts offered on the market may be indivisible in general, thus we have the following specification of geographical arbitrage.

Definition 2. A *geographical arbitrage* opportunity exists when multiple contracts can be substituted for a single contract and when the total price of the substituted contracts is less than that of the single contract. We assume of course that the QoS is equivalent between the single contract and the end-to-end QoS of the substituting contracts.

Simple geographical arbitrage is important because it provides an immediate downward pressure on the price of the single link. How fast this pressure acts depends on two factors: how easy the substitute is to identify, and how liquid the market is. A decrease in the price of a single link implies that some part of total end-to-end demand has shifted from this link to the cheaper alternative path. The increase in demand in the alternative path should result in a price increase on all links of this path. The question arising here is how to quantify the effect the movement of demand from a link to an alternative path has on the price level.

Now let us put all these ideas together. Let (a,b) be a given link, i.e., an arc on which indivisible contracts are available in the market. Let Λ_{abq} be the set of all paths between a and b that provide at least a QoS q. Now we set

$$p_\Lambda = \{p'_k \mid k \in \Lambda_{abq}\},$$

where p' are formed from the observed prices for links. With respect to geographical arbitrage the next price observed in the market for the link (a,b) at a QoS q is p_{abq} and is thus given by (we will combine this with the stochastic process for the link in Section 4.4):

$$p_{abq} = (1 + \mathrm{e}^{-\tau\Delta t} f(p_\Lambda, p'_{abq}))p'_{abq},$$

where p'_{abq} was the previous observed price on the link (a,b) at QoS q. Here $\mathrm{e}^{-\tau}$ describes how fast the no-arbitrage correction $f(p_\Lambda, p'_{abq})$ takes effect. The relaxation constant τ is the quantification of the system liquidity. The function $f(p_\Lambda, p'_{abq})$ encapsulates the degree of the arbitrage opportunities available relative to the observed price for the link (a,b) and the appropriate correction to the link price. The no-arbitrage correction function f also embodies the speed and extent to which applications and electronic agents can re-balance flow in the network on the timescale of network price development. Note that the prices on the alternative paths will also be affected.

If the observed link price p'_{abq} is the cheapest alternative out of p_Λ then there is no arbitrage opportunity and no correction takes place. We call f the *network function*.

Let us now examine the case where there is just one cheaper alternative path (i.e., one simple geographical arbitrage opportunity). The correction on the direct link (labeled d here) from p_d to p'_d is the effect of a left shift of the demand curve by x for this link. Remember that total end-to-end demand is inelastic in the timescale we examine (short, i.e., here, one day), so this same demand is directed to the alternative path and added to the demand of each link in this path, resulting in shifts to the right by the same amount x, as shown in Fig. 2.

Demand for bandwidth can be modeled with a constant-elasticity function $q = A(p)^{-E}$, where q is quantity demanded at a price p, E the price elasticity of demand, and A a scaling factor. Supporting evidence for the use of such a curve for bandwidth is given in [18], but for the timescale of our experiments a linear approximation $q = l_d p + m_d$ is sufficient. This is because price corrections resulting from geographical arbitrage are small relative to the price level. So for each link we compute l_d as the slope of the tangent of the constant-elasticity function at the current price.

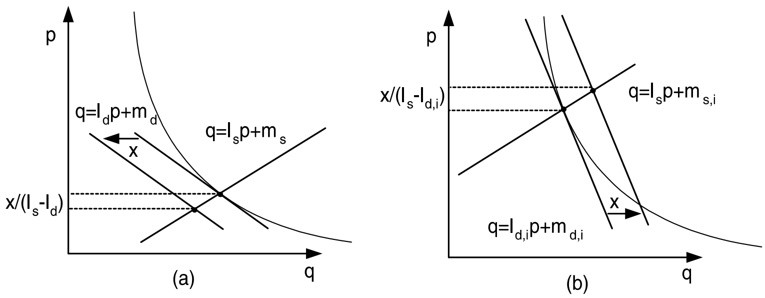

Fig. 2. Load balancing.

Elasticity is uniform across links and can be estimated from real-world data. Estimates from various sources (e.g., Lucent Technologies and France Telecom) regarding demand elasticity are given in [18], with values ranging from 1 to 1.6 for a long-term planning horizon (about five years). Elasticity of demand is also discussed in [3], where hypothetical values of 1 and 1.4 are used for a 5-year horizon, the source of the analysis being Telegeography. Our timescale is much smaller and elasticity of demand is generally inversely proportional to the timescale. Values of $E \geqslant 2$ are plausible, including the case of completely inelastic demand where E goes to infinity. The scaling factor A can be estimated from market data (price and transaction volume), but in the absence of such data we assume A to be constant and uniform across links, with $A = 1$.

We generally assume constant elasticity of supply, i.e., the market normally operates far from the point of drying out, which is expressed as $q = l_s p + m_s$. On a short timescale, expanding a network is out of the question, so the only relevant factor here is the speed at which capacity supplied on a set of ports at a pooling point location can be altered. This depends on the network management technology employed at the pooling points and in providers' networks. Given that pooling points will employ the same or similar technologies for bandwidth management and that total supply on each link should be dominated by the same set of long-distance providers, we assume l_s to be uniform across all links. Capacity offered at the pooling points can be altered at the speed of bandwidth management operations (very fast) and the incremental cost of offering one extra unit of bandwidth on our timescale is close to zero. Therefore supply will be very elastic and thus greater than l_d.

For both supply and demand elasticities more accurate estimates can only be obtained from market data, of which today there is not a sufficient amount (hardly any for spot trades). There is more information on long-term behavior of the bandwidth market, which we utilize in the long-run decrease of mean prices due to technology advances.

We are now ready to resolve a simple arbitrage case by treating demand as a network flow which is conserved point-to-point, while allowing to shift from the direct link to alternative paths so as to achieve a load-balancing effect (see Fig. 3). The result of the correction should be a state in which

$$p_d - a_d x = \sum_{i=1}^{n} p_i + \sum_{i=1}^{n} a_i x,$$

where $a = 1/(l_s - l_d)$ is the change in price resulting from a unitary change in quantity demanded. In this state there is no more arbitrage, hence no more correction takes place. Solving for x yields

$$x = \frac{p_d - \sum_{i=1}^{n} p_i}{a_d + \sum_{i=1}^{n} a_i}.$$

Knowing x, the price of the direct link is corrected to $p'_d = p_d - a_d x$ and that for every link on the alternative path to $p'_i = p_i + a_i x$. The effect of this update mechanism, given our assumptions on supply and demand, is a larger (absolute) correction for the more expensive links in a path than for the less expensive ones.

Generally in geographical arbitrage situations, there will be $m > 0$ paths connecting points a and b that are cheaper than p_{ab}. Let x_k be the amount of network flow shifting from the direct link to all links of path $k \in \{\Lambda_{abq}/(a,b)\}$. Note that if we only permit x_k to be positive this implies that we allow substitution of alternate paths for the direct link, but that we do not allow the alternate paths to act as substitutes for each other. There may be no end-to-end flow on a substitute path even though each link may have flow. Thus in general end-to-end flow cannot move from a multilink substitute path to anywhere, which suggests this

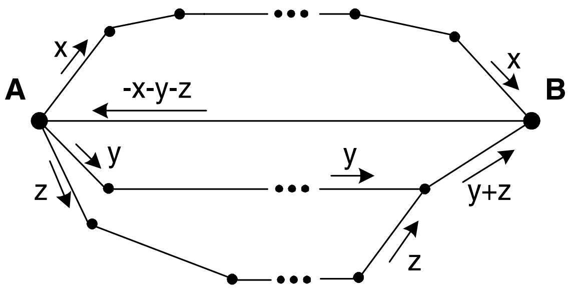

Fig. 3. Shifts in demand described as network flows.

restriction. Then the no-arbitrage state for the direct link and k can be written as

$$p_d - a_d \sum_{i=1}^{m} x_i = \sum_{i=1}^{n} p_i + \sum_{i=1}^{n} a_i x_k + \sum_{i=1,\, i\neq k}^{m} a_{ki} x_i,$$

where

$$a_{ki} = a_{ik} = \sum_{j\in k\cap i} a_j$$

is the sum of a's for all links that paths k and i have in common. The added complexity in the above expression is necessary because paths will generally not be disjoint. Also,

$$x_i \geqslant 0.$$

Let Δ_k be the *arbitrage size* (absolute difference of proposed prices) between (a, b) and path k. The no-arbitrage state can be re-written as follows:

$$\left(a_d + \sum_{i=1}^{n} a_i\right) x_k + \sum_{i=1,\, i\neq k}^{m} (a_d + a_{ki}) x_i - \Delta_k = 0$$

or

$$\sum_{i=1}^{m} (a_d + a_{ki}) x_i - \Delta_k = 0, \quad a_{kk} = \sum_{i=1}^{n} a_i.$$

We extend the previous expression for path k to the following (m by m) linear system of equations, for all alternative paths, using matrices:

$$\mathbf{Ax} - \mathbf{b} = 0,$$

where

$$\mathbf{x} = (x_1, x_2, \ldots, x_m)^{\mathrm{T}}, \quad x_i \geqslant 0,$$
$$\mathbf{A} = a_d \mathbf{1}_m + (a_{ij}), \quad i, j \in \{1 \ldots m\},$$
$$\mathbf{1}_m = (u_{ij}), \quad u_{ij} = 1 \quad \forall i, j \in \{1 \ldots m\}$$

and

$$\mathbf{b} = (\Delta_1, \Delta_2, \ldots, \Delta_m)^{\mathrm{T}}.$$

This is a linearly constrained linear optimization problem. The form $Ax = b$, $x \geqslant 0$ is the form of a standard Linear Programming (optimization) problem, where Programming is meant in the sense of method *not* computer program. A standard reference is [2]. Linear Programming problems are typically solved by either interior point (polynomial worst case) or simplex (exponential worst case but very good in practice) methods. There is a practical problem, however, with the construction of A because this matrix requires all simple paths having a lesser length (price) than the direct link p_{ab} to be found. (Simple paths are those with no repeated edges.) Efficient algorithms for the K-shortest simple path problem, in the sense of having polynomial complexity in the worst case, exist with the most efficient requiring K known in advance (e.g., [16,17]). However, there is a better alternative that allows us to avoid constructing A before we start the optimization. In fact using the iterative scheme described below we avoid the constraint $x \geqslant 0$ because it always holds implicitly from our construction so our load balancing just uses matrix inversion at each step.

Algorithm 1 (*Iterative No-Arbitrage*). Given: a graph G of link prices; the direct link price p_{ab} for the direct link d. Put the expensive path set $E = d$ and the cost of all (any) path in E is $\mathscr{C}(E) = p_{ab}$.

do

1. find the cheapest path from node a to node b. Suppose that this path cost is p_{next} and the path is c_{next}
2. if $\mathscr{C}(E) \leqslant p_{\text{next}}$ then finish because no arbitrage possibility remains
3. load balance between E and c_{next} updating prices on all links in paths in E and on path c_{next}
4. $E = E \cup c_{\text{next}}$

enddo

We now need to demonstrate that the algorithm terminates and is correct. To demonstrate termination note that at each iteration one alternate path that was cheaper that any path in the set E enters E and does not leave. G is finite so the algorithm must terminate. At termination all the paths in the set E have the same cost and there is no cheaper alternative. Thus no arbitrage has been achieved and the algorithm is correct. We still need to state how the load balancing and price updating are done. This is achieved by applying the preceding equations, so we solve $\mathbf{Ax} = \mathbf{b}$ at each step. However the entering path is load balanced with the paths in E at each step rather than load balancing the direct link with all shorter paths. This iterative procedure implicitly ensures that the final solution has $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \geqslant \mathbf{0}$ without having to impose this explicitly.

4.3. Feedback or truncation?

We have stated that we expect geographical arbitrage and liquidity to alter the observed prices on individual links. What we have not yet done is to specify how this alteration will affect the price process dynamics on individual links. There are two basic alternatives: feedback of the altered price into the price process, and simple truncation of the actual process to produce the observed process. These two mechanisms will produce different observed prices, and there are important distinctions with respect to how price spikes and jumps are dealt with.

Let us consider the specific case when a link price process S is given by Geometric Brownian Motion (GBM), i.e.,

$$dS = S\mu \, dt + S\sigma \, dW.$$

Note that in telecom commodities we expect μ to be negative. Now let us consider the Euler discretization, namely

$$S_{i+1} = S_i + S_i\mu\Delta t + S_i\sigma\Delta W_i.$$

(For our purposes we do not require a Milstein or higher-order discretization scheme.) The last term ΔW_i is basically a standard normally distributed random variable.

Suppose that we calculate the observed price to be S'. Now if we use truncation there will be no change in the above equation. On the other hand if we feed back the observed price into the equation, i.e., back into the representation of the market dynamics we obtain

$$S_{i+1} = S'_i + S'_i\mu\Delta t + S'_i\sigma\Delta W_i$$

for the price at the next time step following truncation. In this case the form of the equation, the market dynamics, is unchanged but uses a different starting point for the generation of the next price proposal, namely, the next price is based on the arbitrage-corrected previous price rather than on the previous observed price.

We find the argument for feedback convincing and use it hereafter in our normal price calculations. This leads to the combined link price process in Section 4.4. Spikes and jumps of the link price represent unusual situations or information in the market for the link. Jumps are the simpler situation and if we use feedback then they are simply included in the previous framework. Spikes on the other hand are more complex because they incorporate the jump down from the unusual state as well as the jump to the unusual state. Also, spikes are of short duration even though this duration is uncertain.

An example of how a spike happens is useful at this point. Consider a single link which has a current 50% utilization. Now suppose that half of the traffic is price elastic, for example it may be bought on the spot market, and half is price inelastic, such as longer-term commitments of the providers. Now suppose there is an equipment failure and the link capacity is reduced to 20%. The price elastic traffic will consider the alternative route prices and react accordingly, increasing traffic on alternative (QoS-equivalent) routes as determined by demand elasticity. However there is an inelastic demand that is unmet by the remaining capacity on the link. This will, for certain, take the cheapest alternative route. Hence the effective price for traffic on the single link is the cost of providing service on the next-most-expensive alternative. Thus, in this example, for the duration of the spike the price for traffic on the single link will move exactly as the price for the cheapest alternative route. The realization of a price spike is a truncated process rather than one with feedback. There is truncation because the alternative prices do not affect the underlying cause of the price spike – an unusual event (in the sense of Merton's analyses [19,20]).

We can also model a spike in two different ways. The first way we have already discussed as a change in the price of a single link that results in a truncated observed price. There is a second type of spike that may be of interest as a way of modeling a more general network failure. Suppose that the spike moves the proposal price on the link to S. Now consider all the alternative paths with prices less than S, and set their prices to S. This could model the spreading effect of a network outage. There are several alternatives as to how to allocate the new price between the links on alternative paths. The simplest is perhaps to simply increase all of them in proportion until the sum equals S. This is not necessarily the end of the effect though. We may consider the effect of these new link prices in a similar way as to how we dealt with the original spike on the first link. There would be a limit to the propagation of the network outage at some point, but the effect of a single outage could spread widely. This type of spreading has been observed in some network disasters, but we take this as beyond the scope of the current work.

4.4. Combined price process

We have proposed two sources of price changes for a link (contract) in the network contract graph: an SRR stochastic process, and load balancing following exploitation of geographical arbitrage. We may combine both approaches to obtain the following stochastic differential equation:

$$
\begin{aligned}
X &= \log(S),\\
\mathrm{d}X &= \eta(\overline{X} + GU - X)\,\mathrm{d}t + \sigma\,\mathrm{d}W + G\,\mathrm{d}U\\
&\quad + H\,\mathrm{d}V + \alpha\,\mathrm{d}A,\\
d\overline{X} &= -v\,\mathrm{d}t + \rho\,\mathrm{d}Z.
\end{aligned}
$$

These are the same equations as in Section 4.1 except that we have added a term $\alpha\,\mathrm{d}A$ to describe the geographical (no-)arbitrage effects. This term embodies the network function f described in Section 4.2, and α plays the role of the previous liquidity term τ in the same section. $\mathrm{d}A$ describes a stochastic process that is zero whenever no geographical arbitrage effects are present. When geographical arbitrage effects are present, $\mathrm{d}A$ provides the appropriate correction to remove arbitrage from the previously observed prices in the contract network. The above equations express the fact that market forces are acting at the same time as arbitrageurs to remove arbitrage opportunities (by profiting from them). Indeed they are part of the market, merely modeled explicitly in this case.

In general arbitrageurs could take into account expected movements of the market (as embodied in $E[\eta(\overline{X} + GU - X)]$) in the size of their actions in $\mathrm{d}A$. The expectation here would be with respect to the real measure as the underlying good is non-storable. This would become more important with low liquidity because then the timescale of their actions would increase. We have not taken this into account in our simulations in Section 5 because there we consider a high-liquidity market although the simulation explicitly uses the equations given in this section.

5. Price development for a simple network

The combination of link price processes with Geographical Arbitrage and Liquidity (GAL hereafter) is new so we present a study of a simple network to help build up intuition. We consider an undirected triangular network in which two of the sides start at the same price and the starting price of the third side is varied over various simulations. This is the most basic market setup where geographical arbitrage can occur, and gives an isosceles triangle with respect to prices. We vary the ratio of one side to the other two between 10% and 190% so that the total price to go around the triangle (at the start) runs from 2.1 to 3.9 in some arbitrary price units. We also set the QoS offered on each side to unity and the allowed QoS to two. Thus there are always two alternative paths between any pair of distinct nodes. The parameters used for these price simulations are given in Table 1. In this simulation we consider a highly liquid market so arbitrage opportunities only last for the time step on which they are observed. New geographical arbitrage opportunities may arise at each time step but load balancing (demand shifts) acts to fully remove them on each subsequent time step together with the usual price drivers embodied in the stochastic processes for the links. A triangular network may appear too simple to observe anything of interest but we show here that the network demonstrates a rich set of features.

Perhaps the most important parameter is the short-term volatility, which we examine over a range from 10% to 200%. Spot price volatility (day-ahead volatility) in at least seven current markets for other non-storable commodities (i.e.,

Table 1
Parameters for simple network experiments[a]

Parameter	Symbol	Value
Number of simulations per parameter combination		1024
Simulation length	t_{max}	1 year (= 252 trading days)
Simulation granularity	Δt	1 day
Price trend (time to halve)	$-\nu$	1.25 years
Trend uncertainty	ρ	0–40% per year
Short-term price volatility	σ	0–200% per year
Price reversion to trend (time to halve)	η	3 months
Price jumps	G	None
Price spikes	H	None
Liquidity	α	1

[a] Liquidity refers to how long arbitrage opportunities last before market forces (arbitrageurs, etc.) remove them. In this simulation we consider a highly liquid market. Parameters when referring to equations reference Section 4.4. Note that common random numbers were used across different parameter value combinations for variance reduction.

electricity) has been observed at more than 150% and over 200% for five of them (CINERGY INTO, ENTERGY, MAIN, PALO VERDE, PJM, data source *Power Markets Week*, 1999). One-month forwards were also observed at 45–95%. Thus, for a non-storable commodity, the high side of the range we are considering is perhaps more realistic than the low side.

5.1. Arbitrage existence and net present value

Fig. 4 summarizes our results for the existence of arbitrage, i.e., how much it happens in our setup, and the Net Present Value (NPV) of these arbitrage opportunities relative to the NPV of the entire network. For all NPV calculations we use continuous compounding with a constant 5% (per year) discounting.

As expected arbitrage opportunities increase in frequency with increasing volatility (upper panels), be this short-term price volatility or the volatility of the long-term price trend. Short- and long-term volatilities have different effects both on magnitude and on their relation with the shape of the network. This may be because when arbitrage occurs the price that is affected is the observed price S, and this feeds back directly into the process for S. Hence, when interactions with the long-term trend contribute to arbitrage this long-term trend $\overline{X}$ is not directly affected because it is S that is changed by arbitrage. Also in our base configuration, i.e., with zero volatility, there is no arbitrage and hence the long-term trend tries to keep it this way independent of any arbitrage effects. There is no connection between the long-term-trend price processes so, if there was no arbitrage for them to begin with, they can be regarded as attempting to keep it that way. In short, the independent long-term price trends on the different links act against arbitrage.

The NPV of arbitrage opportunities (lower panels) also increases with volatility to almost 2.5% of the NPV of the total network. This is true even with the highly liquid market we are considering. Long-term price trend uncertainty – alone – is insignificant in creating valuable arbitrage opportunities, i.e., less than 0.1% of network NPV. This is probably because the scale of arbitrage opportunities (i.e., their value) is mainly determined by changes in $\sigma\, dW$, and changes in $\overline{X}$ are scaled down by the speed of mean reversion η.

5.2. GAL effects on observed prices

Fig. 5 shows the percentage change in the mean spot price after one year (252 trading days) and the standard deviation of the observed changes in prices after one year as a percentage of mean prices. We are considering a network in the form of an isosceles triangle so we need to examine the effect on prices of the two sides of the same length (which we term the fixed sides, although their prices change at each time step in each simulation) and of the other side (which we term the variable

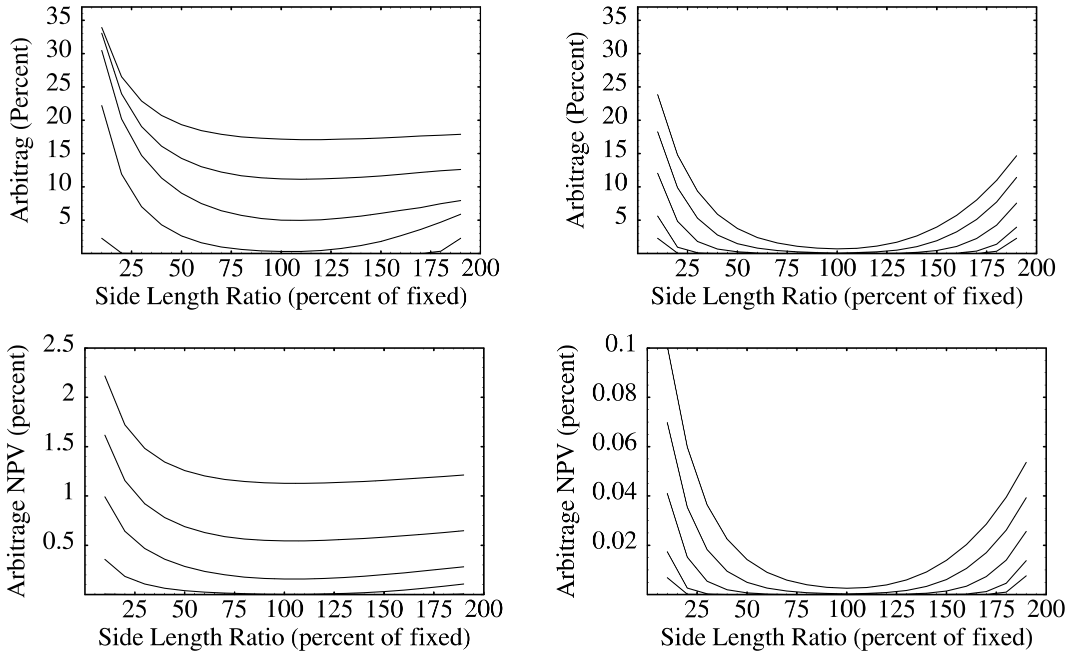

Fig. 4. Arbitrage existence and value. *Left panels*: Here lines indicate increasing short-term price volatility (10%, 50%, 100%, 150%, and 200%) at 0% long-term trend uncertainty. Note that the NPV of arbitrage opportunities (lower left panel) for a short-term price volatility of 10% is indistinguishable from zero. *Right panels*: These display increasing long-term trend uncertainty (0%, 10%, 20%, 30%, and 40%) at 10% short-term price volatility. Note especially the different scales for the lower two panels. In all panels larger effects are seen for higher volatilities.

side in that we vary the starting price ratio from 1:10:10 through 19:10:10, variable:fixed:fixed).

Note first that there is a change in the mean price observed after a year even without network effects (i.e., arbitrage induced load balancing). This is a common phenomenon in stochastic calculus for processes involving logarithms where volatility affects the mean, and is even seen in GBM. It is important here because it means that changes due to arbitrage effects must be compared to a baseline that changes with short-term volatility. This baseline is indicated in all panels with light diagonal crosses.

Comparison of the left (fixed side) and right (variable side) panels shows that the prices of these two side categories behave differently, and indeed even in opposite ways. There are two things going on here, firstly there is the simple difference in side lengths and, more subtly, there is the fact that two of the sides have a common length. The importance of this changes with the ratio of the side lengths. Clearly for very acute triangles either of the fixed sides can easily cause arbitrage whereas for very obtuse triangles only one side (the variable one) can easily generate arbitrage opportunities. Thus we should always expect to see most change relative to the baseline with acute triangular networks. This is indeed what we observe in Fig. 5. Generally the short side (heavy dots, right top panel) becomes longer and the longer sides (heavy dots, left top panel) become shorter. The scale of these changes in the mean prices is very different. In the most acute case the short side may increase in price by 250% rather than decreasing by the expected (baseline) case of around 5%, whereas the long side simply decreases in price by 30% rather than the expected (baseline) case of around 5%. Note that the baseline case will always be the same for the two different sides because when they are isolated they behave exactly the same apart from a constant term which, of course, drops out when percentage changes are considered. Thus in the

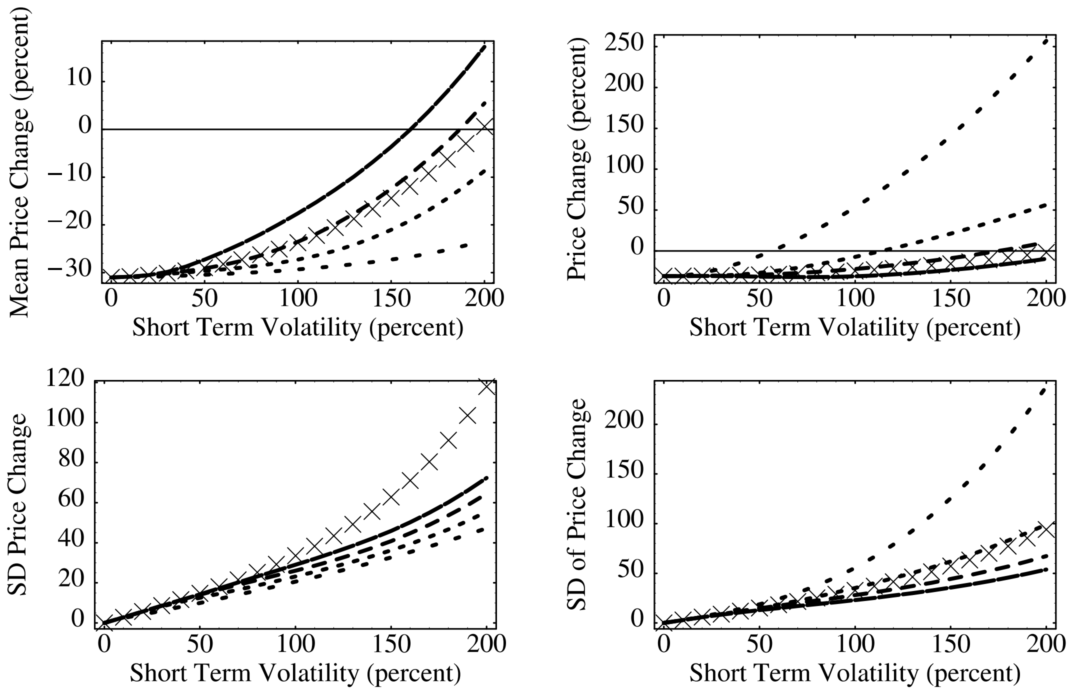

Fig. 5. Network effects on price development versus short-term price volatility (σ) for the "fixed" sides (*left panels*) and for the "variable" side (*right panels*). In all panels diagonal crosses indicate a network with no geographical arbitrage removal, i.e., isolated links. Heavy dotted to full lines indicate increasingly obtuse triangular networks from {1 *dots* – 5 – 10 – 15 *full line*}:10:10. Note that the vertical scales are all different. *Upper panels:* mean percentage change in observed prices. *Lower panels:* standard deviation of changes in observed prices as a percentage of the observed prices.

most acute case, 1:10:10, the arbitrage effects push the triangle towards 2.5:7:7 after one year. In the most obtuse case, 15:10:10, after one year we see a push towards 14:10.5:10.5. We note that, as could be expected from symmetry, 10:10:10 appears to be a stable configuration in that its observed mean price changes little relative to the baseline. Recall that while arbitrage effects may push towards a symmetric situation the long-term price process will act to maintain the original side ratio. Thus the stable limiting configuration for non-equilateral triangles will not be a symmetric 10:10:10 configuration but some balance of the two opposing forces.

Effects of price changes on standard deviation are also mixed. There is a clear progression in effect with increasing side ratio. For the variable side the standard deviation (SD) of the observed price changes decreases in going from acute to obtuse. The opposite is observed for the fixed side, which approaches the baseline for increasing obtuseness. The variable-side SD crosses the baseline and continues to move away from it. With a very acute triangle the (short) variable side is continually pulled by both of the other sides through arbitrage effects but this quickly decreases with increasing obtuseness.

Clearly knowing the topology of alternate routes is vital for understanding price dynamics in a network, and we have here a quantification for a particular limited situation. These results suggest that network effects of geographical arbitrage on price – and on price variations – can be at least as important as the effects of the stochastic processes that affect each individual link.

5.3. GAL effects on network NPV

Fig. 6 shows the effect of geographical arbitrage liquidity on the NPV of a triangular network. This

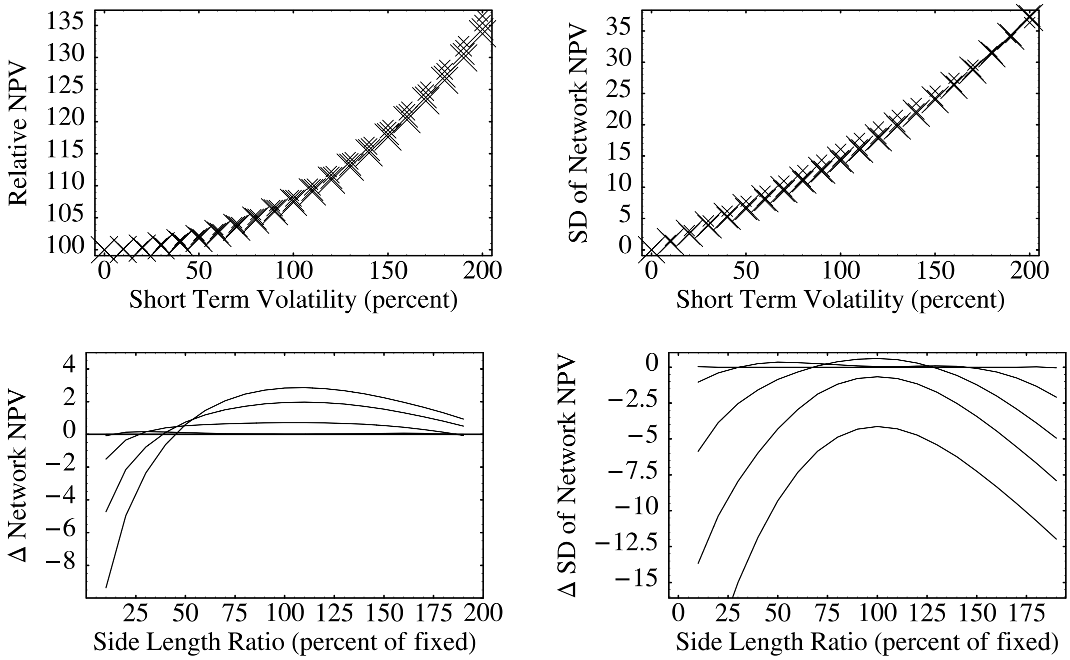

Fig. 6. Network effects on network NPV. *Top panels* show the mean NPV and the standard deviation of NPV (SD) for baseline, isolated link triangle. Mean NPV is shown as a percentage of mean NPV with zero volatility. SD of NPV is shown as a percentage of the mean NPV that corresponds to the same short-term volatility level. Increasing sized crosses show increasingly obtuse triangle networks from {1 *smallest* – 5 – 10 – 15 – 19 *biggest*}:10:10. *Lower panels* show the additional effect of network effects (geographical arbitrage and load balancing) on network mean NPV and SD of network NPV, i.e., the changes ("Δ"'s). The most acute-angled triangles show the most extreme effects (different from zero).

NPV is the sum of the NPV for each edge of the triangle. As before for price changes on individual links, the baseline, isolated link, triangle NPV changes (increases) with increasing short-term volatility (right top panel). The difference between different networks with different triangle ratios is less significant than the change in NPV relative to zero volatility, which is nearly 140% of zero volatility NPV. The range of changes in network NPV across triangle shapes is less than 10%. There is even less difference with respect to triangle shape for the SD of the network NPV in terms of the mean NPV for the same short-term volatility.

The lower panels of Fig. 6 show the additional effect of geographical arbitrage on network NPV. For acute-triangle networks, there is a decrease in mean network NPV up to 10% (left lower panel). This gradually changes with increasing obtuseness to an increase of more than 2% of mean network NPV, maximizing at around equilateral configurations and then gradually decreasing. Changes only become significant for short-term volatilities of 50% and higher. There is always a decrease in the SD of the network NPV (right lower panel). This decrease starts at around a 20% decrease for acute triangles at 200% short-term volatility, then becomes less pronounced at a side ratio of 10:10:10 before becoming more pronounced again to a decrease of 12% for very obtuse triangles.

Thus network effects affect the total NPV around the triangle much less than the prices of the individual links. The difference in the percent mean price change for an individual link is up to 250% both up and down, depending on the side you consider, whereas for the total this effect is always less than 10%. In some sense the triangular network is, at least in the mean, a self-hedging instrument whereas individual links certainly are not. The decrease in the SD of the network NPV is easy to understand because whenever arbitrage is observed prices are moved closer together, depending to some extent on liquidity. In these experiments we used a sufficiently high liquidity to eliminate arbitrage on the next time step after it was observed. There is clearly a trade-off between the value of the arbitrage opportunities and the extent to which no-arbitrage affects network price development. A lower liquidity would have increased the value of the arbitrage opportunities by having them last longer but would have reduced the other effects.

6. Price development on network market

We used a realistic international topology (see Fig. 7) to analyze the applicability of our results. Specifically, we quantified the potential presence of triangles of link contracts and their side ratio distribution.

6.1. Contract network construction

We will now construct a real-world version of the graph of indivisible bandwidth contracts. As today the market is at a stage of expansion and transformation, there is no single data source at present, certainly nothing like a topology of the underlying link contracts. We are, however, in a position where we can construct a plausible future topology using information on commercial optical fiber backbone maps, pooling point operator deployment plans, and trades on the OTC bandwidth marketplaces. The result is a map of what *could* be offered by providers in the bandwidth market as *indivisible* contracts.

We constructed the contract map by overlaying several carrier backbone maps as starting point. We then create an edge (or link) between two pooling point locations A and B in our map if there is at least one backbone that could connect A and B without crossing any other pooling point. If all possible network-level paths cross other pooling point locations, A-to-B cannot be offered as an indivisible contract. An example of this is Dallas–London. All network paths would cross the New York or Washington DC pooling points, therefore Dallas–London is not an edge in our graph. The backbone maps we included in the set of potential providers are

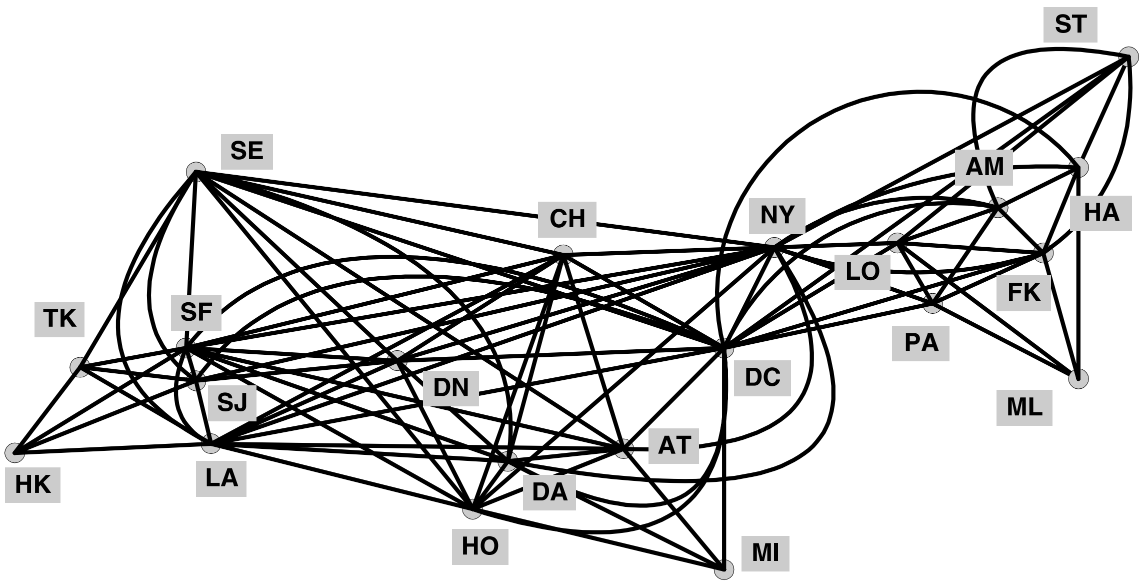

Fig. 7. Combined contract network topology for Asia, the USA and Europe. Two-letter labels give city names, e.g., NY = New York, HK = Hong-Kong. For construction see Section 6.1.

- Williams Communications Global Network Services Interactive Map, http://www.williamscommunications.com
- UUNET Internet Network Maps, http://www.uunet.com
- AT&T United States IP Backbone Network Map – 4Q2000 View, http://www.att.com
- Global Crossing Network Maps, http://www.globalcrossing.com
- The Level 3 Network – United States and International Route Maps, http://www.level3.com
- Enron Boardband Services United States Network Map, http://www.enron.net
- Genuity United States Fiber Optic Network, http://www.genuity.com
- Qwest Americas IP Backbone Map, http://www.qwest.com
- Cable & Wireless Global N^3 IP Backbone (3Q 2000) and European Ipergy Network Maps, http://www.cw.com
- TeliaNet European IP Backbone Network – Interactive Map, http://www.telia.net
- AboveNet Global IP Network Map, http://www.abovenet.com
- Onyx Networks Global IP Infrastructure Map, http://www.onyx.net
- KPNQwest EuroRings Map, http://www.kpnqwest.com

The set of 21 pooling points is a subset of currently operating and planned locations, communicated to the authors by pooling point developers and operators (Enron, Lightrade) combined with destinations that have been popular in the OTC market ([6]). This yields Tokyo, Hong-Kong, Seattle, San Francisco, San Jose, Los Angeles, Denver, Dallas, Houston, Chicago, Atlanta, New York, Washington DC, Miami, London, Paris, Milan, Frankfurt, Hamburg, Amsterdam, and Stockholm. Even though the map would look different for subsets or supersets of the pooling points chosen, the current version provides a useful view of the topology of the underlying good. As a rough approximation of an edge's length we compute the geographical distance between two locations using the Haversine formula [4]. This method will generally underestimate the actual length of a network connection.

Some simple observations we can make are that the USA are almost fully meshed, Europe to a lesser degree, and USA–Europe even less. What is not immediately visible, but relevant to our analysis – as we show later – is the fact that many pooling points are clustered in relatively small, densely connected regions (Asia, US West Coast, US East Coast, Texas, Europe), with a few really long edges from one local cluster to another.

6.2. Contract network topology

Using the potential contract map developed in the preceding section, we analyzed the topology for the presence of triangular subnetworks. Recall that we found network effects to be most pronounced with acute-angled triangles where the side ratio was 1:10:10. We found 187 triangular subnetworks. We found that if we allowed the two "fixed" sides to differ by only 10% 86 triangles remained (46% of the total). Relaxing this requirement to 20% included 129 triangles (69%), and to 30% included 175 triangles (94%). We used 10% bins to construct frequency histograms of the triangle side ratio distribution for these three cases, shown in Fig. 8. These ratios were constructed from geographical distance ratios rather than from observed prices. However, at least some network costs are roughly linear with distance, and these will dominate the mean trend (v) for long distance routes.

The most immediate observation from this subnetwork classification is that there is an overwhelming number of acute-angled triangles (1–5:10:10) and that this does not change with a relaxing of the requirement to be included. As this requirement is relaxed there is the suggestion of a second peak at very obtuse angles, say around 17:10:10. There are very few triangles with a side ratio between 5:10:10 and 15:10:10. The reason for such a high proportion of acute-angled triangles in the potential contract network is simple. There are concentrations of high-tech and population on the East and West coasts of the USA with a few mid-Western additions. Thus you get short sides within such concentrations and much longer ones between different concentrations. The Atlantic and Pacific Oceans only serve to reinforce this trend,

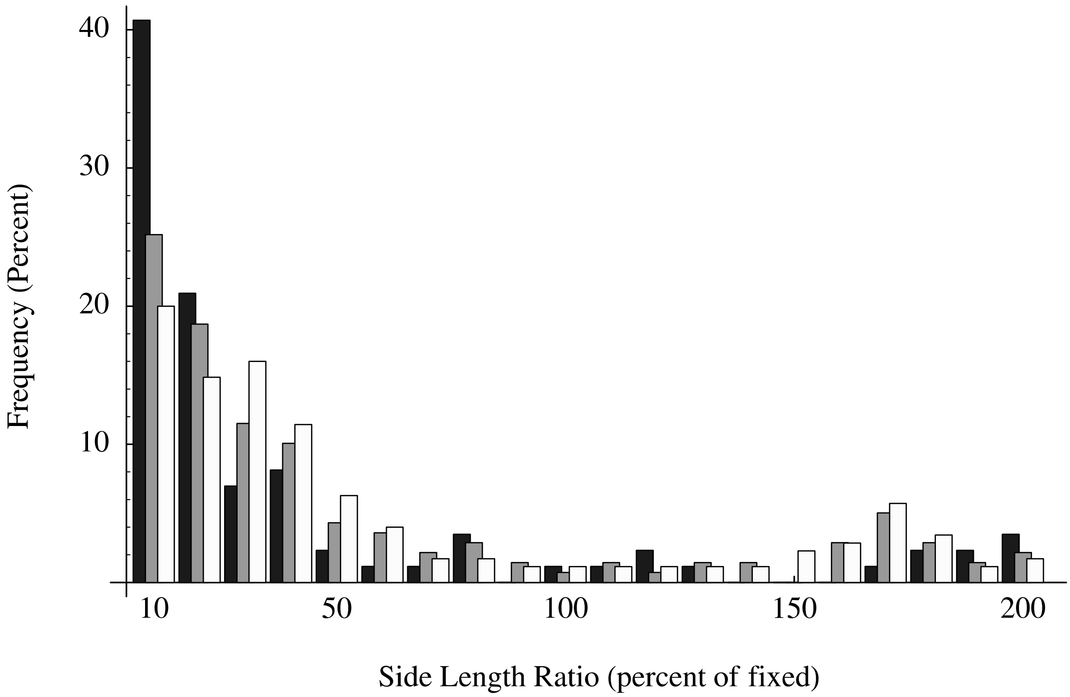

Fig. 8. Triangle distribution in international combined carrier network with respect to ratio of "variable" side of isosceles triangle to "fixed" sides. Darkest bars are triangles in which the two "fixed" side lengths are within 10%, lighter bars are for 20% and 30%.

regardless of the difference in cost between land and undersea transmission cables.

We thus conclude that the presence of a high proportion of acute-angled triangles in the network is a simple result of geography, and expect this to be a very robust result. This implies that the network effects of geographical arbitrage, which are present whatever the ratio, will be very pronounced because these are most extreme for such topologies.

7. Discussion and conclusions

We have introduced a method for constructing telecom commodity spot price processes that reflects their unique characteristics. The commodity envisaged is point-to-point bandwidth between pooling points. This model is applicable to price forecasting in a network context, valuation of network infrastructure applicable to investment or swapping, and as an input for pricing derivative contracts.

Three factors combine to produce the prices observed: link price stochastic processes, geographical arbitrage, and liquidity. Geographical arbitrage affects observed prices by shifting demand to cheaper alternative routes at equivalent QoS. We examined a simulated price development over one year (252 trading days) on a simple triangular network with varying side length (price) ratios using realistic parameter estimates. By constructing a projection of the future market topology we demonstrated that acute angled triangular subnetworks (1–5:10:10) are an overwhelming majority. This, in fact, is a direct consequence of geographical concentrations of high-population and high-tech areas.

We showed that geographical arbitrage would be a common feature of even highly liquid markets occurring up to 35% of trading days. The effect is most pronounced for highly acute-angled triangles. Short- and long-term volatilities lead to different quantitative and qualitative network effects. We used short-term volatilities up to 200% of the individual link processes, which were mean reverting with negative drift, so the actual observed short-term volatilities, without network effects, only went up to 80%. Arbitrage NPV was up to 2.2% of network value with high short-term volatility, but is insignificant (less than 0.1%) for long-term volatility (ρ) up to 40%.

In the simulation, prices for network links decreased on average to reflect decreasing technology cost (ν). Changes in mean prices for network links were altered by network effects (geographical arbitrage) and the difference increased with increasing short-term volatility. The price change could be in either direction depending on the triangle side ratio and which side was being examined. The volatility (standard deviation) of the observed prices was almost always decreased by network effects and this decrease could be up to 50%.

Total network value (NPV) over one year saw a decrease up to 10% with acute triangles and an increase up to 3% for equilateral triangles owing to network effects. Volatility of network value showed a consistent decrease up to 30% with the least decrease for mildly acute triangles (5:10:10).

Link price processes are inspired by jump diffusions used in oil and electricity. The major difference is the explicit inclusion of geographical (no-)arbitrage terms that express effects of network topology on price dynamics. No previous work includes network effects explicitly in price dynamics. We consider short- and long-term dynamics in a similar way as to [28], who had short- and long-term variations (or alternatively a stochastic convenience yield [26]). This is also similar to the Pilipović model (described in [30] and used in [8]). A difference to previous work is that we take the mean drift of the long-term dynamics explicitly negative. Because of technology advances, prices on average decrease rather than increase. The other additional new factor relative to these sources is the addition of jump and spike terms similar to [13] and especially regime switching models [1,10–12]. We expect spikes to be much more important than price jumps. In the oil price, jumps are basically the result of changes in the status of OPEC and different states of this status may be prolonged. In bandwidth it would be legally difficult for the oligopolistic providers of the resource to create or sustain cartel pricing. In electricity, and possibly in bandwidth, spikes are usually the result of a lack of overprovision for peak demand of the resource combined with either equipment failure and/or some correlated increase in demand (weather for electricity and probably seasonal, e.g., Christmas, for bandwidth).

Geographical arbitrage is a new factor for price dynamics not seen in other commodities. We use load balancing with economic arguments to move out of arbitrage situations, and thus have a network effect on link prices.

Many applications of this work are possible, some of which we have already illustrated in our results: valuation of network infrastructure, design of network asset portfolios, decision support tools for network expansion or new pooling point location, and price scenario modeling. For example, recall that the value (NPV) of all three sides of a triangle is almost self-hedging with respect to network effects. Without a method to generate price scenarios for network infrastructure, no quantitative optimization is possible and that is also an application of the current work.

Clearly modeling spot prices is only a beginning for network commodities. A vital next step is to extend this work to forward markets as has been done for electricity, e.g., [11,21]. The extension is non-trivial to the extent that a commodity is non-storable. Instead of deriving the forward curve behavior from the spot price it is necessary to model the forward curve development explicitly and the spot price is just the shortest-term forward. Forward markets are expected to be larger and more active than spot markets, as is true for most commodities, and indeed already exist to a limited extent for bandwidth. Another extension to this work is going into more depth on effects of rare events linked with QoS constraints.

In conclusion, we have proposed a method to model telecom commodity prices taking into account network effects inherent to bandwidth markets. These network effects produce highly significant changes in price development and network value. These changes depend on network topology for both sign and magnitude. Comparison with a plausible future network topology derived from more than 10 major carriers (e.g. AT&T) and new entrants (e.g., Level3, Enron) showed that extreme topologies, from a network-effect point of view, will be common. This model is applicable to price forecasting in a network context, valuation of network infrastructure applicable to investment or swapping, and as an input for pricing derivative contracts.

Acknowledgements

We gratefully acknowledge the assistance of Mike Moore of Amerex Bandwidth, James Hawthorne of Enron Broadband Services, and Michael Smith of Lightrade for providing market information. We are also grateful to an anonymous referee for help in sharpening the focus of this work.

References

[1] N.P.B. Bollen, S.F. Gray, R.E. Whaley, Regime-switching in foreign exchange rates: evidence from currency option prices, J. Derivatives 6 (1998) 38–49.

[2] D. Bertsimas, N. Tsitsiklis, Introduction to Linear Optimization, Athena Scientific, Belmont, MA, 1997.

[3] V. Thompson, Boom or Bust, New Carrier, Emap Business Publications, London, UK, Novermber/December 2000.

[4] B. Chamberlain, What is the best way to calculate the distance between two points?, Online US Census Bureau GIS FAQ, http://www.census.gov/cgi-bin/geo/gisfaq?Q5.1.

[5] G. Cheliotis, A market-based model of bandwidth and a new approach to end-to-end path computation with quality guarantees, Presented at the MIT/Tufts Workshop on Internet Service Quality Economics (ISQE), Cambridge, MA, December 2–3, 1999; also IBM Research Report RZ 3294.

[6] G. Cheliotis, Bandwidth trading in the real world: findings and implications for commodities brokerage, Presented at Third Berlin Internet Economics Workshop, Berlin, Germany, May 26–27, 2000; also IBM Research Report RZ 3244.

[7] S.S. Chiu, J.P. Crametz, Suprising pricing. Risk, Bandwidth Special Report, June 2000, pp. 15–17.

[8] S.S. Chiu, J.P. Crametz, B. Null, Commodity pricing fundamentals and telecommodity applications, in: The Telecoms Trading Revolution, Risk Books, London, 2000, pp. 91–107.

[9] D.D. Clark, A model for cost allocation and pricing in the internet, in: L.W. McKnight, J.P. Bailey (Eds.), Internet Economics, MIT Press, Cambridge, MA, 1997.

[10] S. Deng, Stochastic models of energy commodity prices and their applications: mean-reversion with jumps and spikes, PSERC Working Paper 98-28, 1998, available at http://www.pserc.wisc.edu/.

[11] S. Deng, B. Johnson, A. Sogomonian, Exotic electricity options and the valuation of electricity generation and transmission assets, PSERC Working Paper 98-13, 1998, available at http://www.pserc.wisc.edu/.

[12] J.-C. Duan, I. Popova, P. Ritchken, Option pricing under regime switching, Working Paper, 1999, available at http://www.mgmt.purdue.edu/faculty/ipopova/.

[13] M.A.G. Dias, K.M.C. Rocha, Petroleum concessions with extendible options using mean reversion with jumps to model oil prices, in: 3rd Real Options Conference, 1999, pp. 1–27, available at http://www.puc-rio.br/marco.ind/main.html.

[14] N. Economides, The economics of networks, Int. J. Industrial Organization 14 (1996) 673–699.

[15] T. Gros, The development of a global bandwidth commodity market, in: The Telecoms Trading Revolution, Risk Books, London, 2000, pp. 1–10.

[16] E. Hadijconstantinou, N. Christofides, An efficient implementation of an algorithm for finding K shortest simple paths, Networks 34 (1999) 88–101.

[17] N. Katoh, T. Ibaraki, H. Mine, An efficient algorithm for K shortest simple paths, Networks 12 (1982) 411–427.

[18] G. Lanning, D. Mitra, Q. Wang, M.H. Wright, Optimal planning for optical transport networks, Philos. Trans. Royal Soc. Lond. A 358 (1773) (2000) 2183–2196.

[19] R.C. Merton, Option pricing when underlying stock returns are discontinuous, J. Financial Economics 3 (1976) 125–144.

[20] R.C. Merton, On the mathematics and economics assumptions of continous-time models, in: Financial Economics: Essays in Honor of Paul Cootner, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[21] M. Manoliu, S. Tompaidis, Energy futures prices: term structure models with kalman filter estimation, University of Texas at Austin, Center for Computational Finance Preprint, February 1998.

[22] D. Pilipović, Energy Risk: Valuing and Managing Energy Derivatives, McGraw-Hill, New York, 1998.

[23] I.C. Paschalidis, J.N. Tsitsiklis, Congestion dependent pricing of network services, IEEE/ACM Trans. Networking 8 (2000) 171–184.

[24] B. Rupp, V. Grimm, A proposal for an automated bandwidth exchange, Presented at the 3rd Berlin Internet Economics Workshop, Berlin, Germany, May 26–27, 1999.

[25] D.J. Songhurst (Ed.), Charging Communications Networks: From Theory to Practice, Elsevier, The Netherlands, 1999.

[26] E. Schwartz, The stochastic behaviour of commodity prices: implications for valuing and hedging, J. Finance 52 (1997) 923–973.

[27] S. Shenker, D. Clark, D. Estrin, S. Herzog, Pricing in computer networks: reshaping the research agenda, ACM Comput. Commun. Rev. 19 (1996) 244–268.

[28] E. Schwartz, J.E. Smith, Short-term variations and long-term dynamics in commodity prices, Manage. Sci. 46 (7) (2000) 893–911.

[29] H.R. Varian, Market structure in the network age, Presented at Understanding the Digital Economy, Department of Commerce, Washington, DC, May 25–27, 1999.

[30] P. Wilmott, Paul Wilmott on Quantitative Finance, Wiley, Chichester, 2000 (Chapter 62, Energy Derivatives).

Chris Kenyon is a Research Staff Member at the IBM Zürich Research Laboratory and has a Ph.D. in Applied Mathematics from Cambridge University. His current research interests include telecommunications and e-commerce infrastructure economics, mathematical finance and real options. He is a member of INFORMS.

Giorgos Cheliotis received the M.Sc. degree in Electrical Engineering and Computer Science from the National Technical University of Athens (NTUA) in 1997. He is currently working toward the Ph.D. degree in Electrical Engineering and Computer Science from NTUA and is with the IBM Zurich Research Lab since 1998. His current research interests include network design, electronic commerce and internet economics. He has been actively involved in research regarding bandwidth commodity markets since 1998.