Wireless Network Capacity Management: A Real Options Approach

Y. d'Halluin[*], P.A. Forsyth[†], and K.R. Vetzal[‡]

May 10, 2004

Abstract

This paper applies financial option valuation methods to new wireless network capacity investment decision timing. In particular, we consider the case of network capacity for cellular telephone service. Given a cluster of base stations (with a certain traffic capacity per base station), we determine when it is optimal to increase capacity for each of the base stations contained in the cluster. We express this in terms of the fraction of total cluster capacity in use, i.e. we calculate the optimal time to upgrade in terms of the ratio of observed usage to existing capacity. We study the optimal decision problem of adding new capacity in the presence of stochastic wireless demand for services. A four factor algorithm is developed, based on a real options formulation. Numerical examples are provided to illustrate various aspects of the model.

Keywords: real options, optimal network planning decisions, network capacity, market price of risk, investment timing option

1 Introduction

Wireless networks are now regarded as essential tools, dramatically impacting how people approach personal and business communications.[1] As new network infrastructure is built and competition between wireless carriers increases, subscribers are becoming ever more demanding of the service and voice quality they receive from network providers. Wireless operators need to provide a high quality of service to customers while maximizing profit. Network managers are faced with the prospect of decreasing revenue per user (on a per minute basis) and increasing demands on networks from new features in wireless equipment [13].

In this paper, we apply a real options framework to the problem of optimal investment in new wireless capacity for cellular telephone service. The advantage of the real options paradigm over traditional capital budgeting methods such as net present value (NPV) lies in its ability to incorporate additional managerial flexibility. Perhaps the most common example is the timing of an investment decision. Basic NPV analysis simply involves comparing the discounted future cash

[*]Y. d'Halluin, School of Computer Science, University of Waterloo, Waterloo ON, Canada N2L 3G1 (e-mail: ydhallui@elora.uwaterloo.ca).

[†]P.A. Forsyth, School of Computer Science, University of Waterloo, Waterloo ON, Canada N2L 3G1 (e-mail: paforsyt@elora.uwaterloo.ca).

[‡]K.R. Vetzal, Centre for Advanced Studies in Finance, University of Waterloo, Waterloo ON, Canada N2L 3G1 (e-mail: kvetzal@watarts.uwaterloo.ca).

[1]Note that our focus throughout this paper is solely on networks for cellular telephones, not other types of wireless networks.

flows generated by an investment with its cost today. However, this ignores management's ability to select when to make the investment. For example, it may be better to wait until more information is known about product market demand. It is well known that this type of situation bears a strong resemblance to an American-style financial call option. Such a contract allows its holder the right to purchase some underlying asset at a specified exercise price during some period of time. To make the analogy with real options, consider the project being evaluated as the underlying asset and its cost as the exercise price. The real options approach has been used to analyze a wide variety of corporate investment decisions in the past couple of decades. General accounts of the theory and some detailed examples of applications can be found in texts such as [6, 16].

This paper is specifically concerned with the question of when to expand capacity, in the particular context of a wireless communications network. Previous papers dealing with the general question of optimal capacity levels which use a real options approach include [12, 5, 2, 3]. A general overview is provided in Chapter 12 of [6]. In [12], a firm facing a stochastically evolving demand chooses an optimal initial level of capacity to install. It can subsequently expand capacity, but it can never reduce it. A quite similar problem is explored in [5], but where the firm's technology features increasing returns to scale at low capacity levels and decreasing returns at high levels. In [2], the case considered is where the firm makes a once and for all choice of optimal capacity—it has no opportunity to add extra capacity subsequently. Some general features of this previous literature include the following. First, capacity additions are incremental, so that the firm gradually adds capacity a bit at a time, rather than jumping discretely to a new higher level (though this can be optimal in [5] over certain capacity ranges with increasing returns to scale). Second, the firm's investment horizon is infinitely long. Third, these studies generally assume that there is no lead time required for capacity expansions. In other words, once a decision is made to expand, it can be implemented immediately.

In contrast, a study of when to add capacity in the context of bandwidth for Internet traffic is provided in [3]. In the situation considered, technological factors dictate that extra capacity must be added in discrete levels, the firm has a finite investment horizon (due to the assumption that in an environment with rapidly evolving technology, new products will make investments in current technology rapidly obsolete), and there is a significant lead time for capacity to be installed once expansion decisions are made. In this paper, we consider an application similar to that in [3], except in the case of wireless network capacity for cellular telephone service. A significant constraint on the algorithm provided in [3] is that the lead time required for an upgrade determines the upgrade decision interval. For example, if it takes three months for new capacity to be installed, the firm can only evaluate whether to upgrade capacity every three months. The present paper relaxes this restriction, so that the upgrade decision interval is independent of lead time. This comes at a signficant computational cost in that an extra dimension is added to the optimization problem. However, this does provide an additional benefit in that payments for capacity upgrades can be spread out over time, rather than occurring all at once when the upgrade decision is made. Another important difference between the present paper and [3] is that in the context of this work, there is no traded asset available to facilitate estimation of a parameter known as the market price of risk. Consequently, alternative methods are required, as discussed in detail below.

The remainder of this paper is organized as follows. Section 2 describes the modeling framework, while Section 3 presents the mathematical model. The computational algorithm is described in detail in Section 4. Section 5 provides the estimated model parameters, and Section 6 contains various simulated results. Conclusions are given in Section 7.

2 Background

The revenue received by the owner of a wireless network is determined by the prevailing price per minute and the amount of traffic. The average price per minute of usage paid by wireless users has been decreasing, with relatively little uncertainty (see Figure 1). On the other hand, an examination of wireless network traffic reveals some interesting features. Although traffic has obviously been increasing, it has done so in an uneven way, with apparent randomness. This is shown for an illustrative switch in Figure 2.[2] In Appendix A, we show that not all network traffic movements can be attributed to deterministic drift and cyclical patterns. When cycles, trends, anomalous drops and statutory holidays are removed, large volatility in network traffic remains. This leads to the following observation. Since price is relatively smooth, and demand fluctuates considerably, even after factoring out predictable cycles, the primary risk factor for revenue is the level of demand. The model we develop below is based on this reasoning.

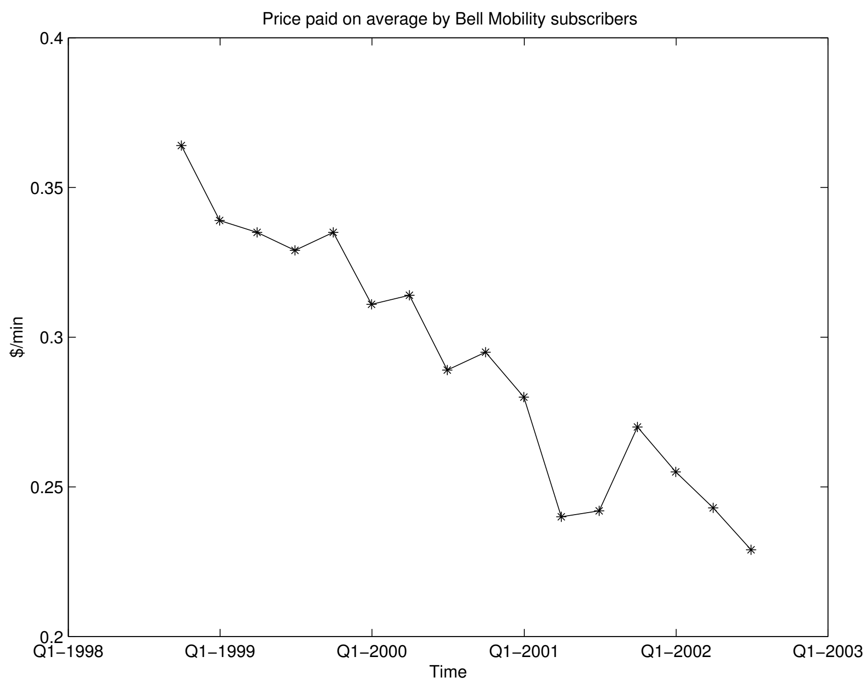


Figure 1: *Average price paid by wireless network subscribers in $ per minute. The price is obtained by dividing the average revenue per user by the usage per subscriber. The price decline can be fit reasonably well by the function* $\mathcal{P}(t) = \mathcal{P}_0 \exp(-\mu_p t)$*, where* $\mu_p = -.08/year$*. These data were obtained from Bell Canada quarterly financial reports.*

We now provide a description of some aspects of the current environment in the industry. For a given set of base stations (i.e. a *cluster*) and a desired *grade of service/blocking probability*, traffic engineers predict the amount of capacity (or the number of carriers) necessary to satisfy the given demand while maximizing revenue. When there is high traffic, the base stations experience very high *blocking* and new equipment must be installed at the base station level (i.e. *carriers*). Figure 3 provides a representation of a simplified cluster of base stations/cell sites. A typical

[2] In telecommunications, a switch is a network device that selects a path or circuit for sending a unit of data to its next destination.

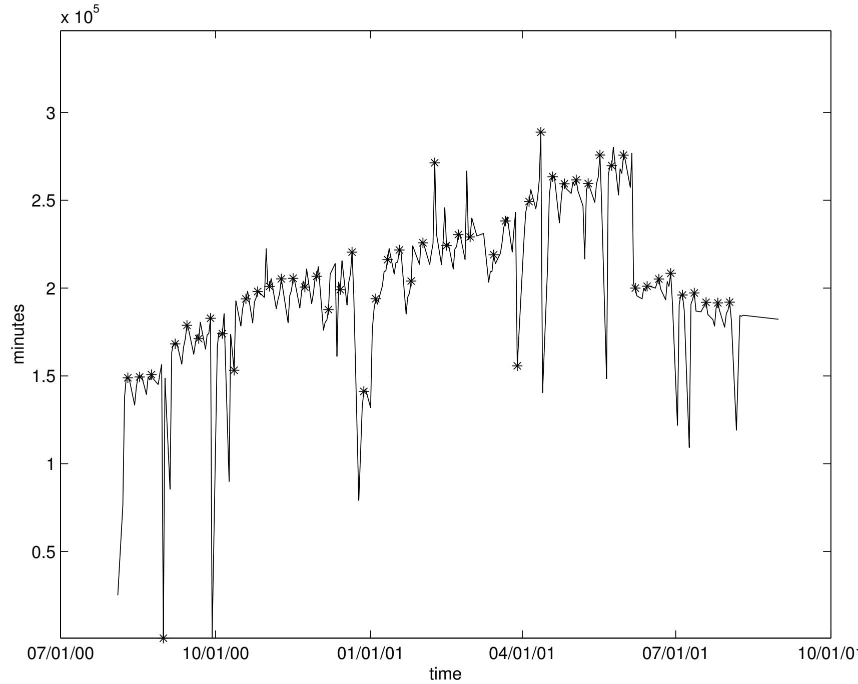


Figure 2: *Daily bouncing busy hour data traffic on a representative switch [11]. To remove weekly cycles, the weekday with the most traffic on average in a year is selected (Thursday in our sample). Asterisks are used to indicate this day. Bouncing busy hours are a measure of peak demand during the day.*

cluster contains at least 20 cell sites. Each cell site has a certain coverage area that is divided into sectors. An easy solution to blocking would be to conduct a traffic study at the cell site level and increase the capacity of the cells that are experiencing too much blocking. However, due to the Code Division Multiplexing Access (CDMA) [9] technology that is currently used in leading edge wireless networks, it is not possible to only add carriers to the cell sites where high blocking occurs. A user on a network using CDMA technology may talk simultaneously to many base stations since a principle called *soft hand-off* is used [9]. As such, when there is too much blocking on a particular cell site, all the cell sites within the cluster must have a new carrier added to maintain homogeneity.

With this in mind, we can consider traffic and grade of service/blocking probability. Traffic is measured in *Erlang* units [10]. An Erlang is defined as the average number of simultaneous calls, or equivalently, the total usage during a time interval divided by the length of that interval. Most network management systems measure usage during a one hour interval. The *blocking probability* is the probability that a call is blocked when there is no channel available. The blocking probability is evaluated for the load during peak daily traffic periods (known as *bouncing busy hours*). In this work, a blocking probability of 2% is used, consistent with current practice in the industry [11]. The calculation used to determine the bouncing busy hour is as follows:

1. For each sector of every cell site, determine which hour had the most traffic for a given day. This is called the *busiest bouncing hour*. This hour will most likely be different for each sector. Note that the hour begins at the top of the hour, i.e. the busiest hour is not the 60 consecutive minutes where traffic is the highest.

2. For a given carrier, add up the number of minutes during the busiest hour for every sector

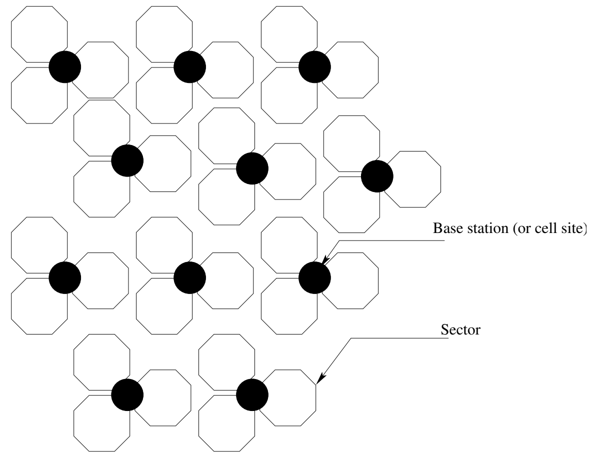


FIGURE 3: *An example of a cluster of cell sites. A typical cluster contains at least 20 base stations. The coverage area or sector is generally not octagonal, but this diagram gives an idea of how cell sites are deployed.*

of every cell site. This number will be the bouncing busy hour traffic for that switch on that particular day.

Based on Erlang-B or Erlang-C mathematical models, the relationship between blocking probability and demand can be established. Details are provided in [10].

Initially, we assume that we have an existing cluster where each base station has a base level of capacity (i.e. a single carrier per base station, or approximately 20 users can handled). Using Erlang tables, this means 13.2 Erlang of traffic can be handled at 2% blocking. If a second carrier is added, 40 users can be accommodated, meaning 31 Erlang of traffic can be handled at 2% blocking. The carried load at 2% blocking more than doubles when a new carrier is added. Table 1 provides a synopsis of an Erlang table.

Number of carriers	Users	Erlang	Minutes of traffic in the hour
1	20	13.2	2376
2	40	31.0	5580
3	60	49.6	8928

TABLE 1: *Correspondence table between users, Erlang and minutes of traffic in the hour that can be approximately handled by a carrier at 2% blocking. Carriers usually cover three sectors.*

3 Mathematical Model

As noted above, the level of network traffic should be viewed as the main risk factor for the revenue stream of a wireless network operator. We adopt the following simple model for network traffic. Let Q represent bouncing busy hour network traffic (in units of minutes per bouncing busy hour). During a short time interval dt, the change in Q is given by

$$dQ = \mu Q dt + \sigma Q dz, \tag{3.1}$$

where μ is the *drift* or *growth rate*, σ is the *volatility*, and dz is the increment of a Wiener process. Readers unfamiliar with Wiener processes can think of dz as being a normally distributed random variable with mean zero and variance dt. Equation (3.1) is a stochastic process known as geometric Brownian motion. It basically says that the expected change in Q during dt is $\mu Q dt$, but there is some uncertainty or noise involved due to the $\sigma Q dz$ term. An implication of equation (3.1) is that Q is lognormally distributed. Note that if $\sigma = 0$, the model gives exponential growth at the rate μ, i.e. $Q = Q_0 \exp(\mu t)$, where Q_0 is some constant. Because of differences in their scaling with time, the drift and volatility terms have different effects. For short time periods, the volatility term has a dominant influence, while over longer periods the drift term becomes more important.

Our ultimate objective is to determine the optimal time for an equipment upgrade. This is found as a byproduct of calculating the value of the network operator's investment in equipment. We use a dynamic programming approach to maximize the value of this investment, and this in turn will tell us the optimal upgrade strategy.

For expositional ease we begin by ignoring the upgrade decision entirely and simply describe how to calculate the value of an investment for a fixed given level of capacity. Let V denote the value of the investment and let $\overline{Q}$ be the capacity level. We consider an investment horizon of T years. At this date, the value of the investment is assumed to be given by

$$V(Q, \overline{Q}, T) = f(Q).$$

Our methods can be used with any suitable choice of $f(Q)$, but for simplicity we assume that the value of all capital investment at T is equal to its salvage value, and we further assume that this salvage value is zero (i.e. $f(Q) = 0$). Implicitly, we are assuming that new technology renders all existing equipment obsolete and worthless at T. Our dynamic programming approach solves backward from this investment horizon date to today. We let t be time running in the forward direction, and $\tau = T - t$ be time running backward. We suppose that revenues accrue continuously, and are given by $R(Q, \overline{Q}, t)$, measured in dollars per year.

Based on standard hedging arguments in the financial valuation literature (e.g., [7, 17, 16], and provided for completeness in this article in Appendix B), a partial differential equation for the value of the investment V is found to be

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 Q^2 \frac{\partial^2 V}{\partial Q^2} + (\mu - \kappa\sigma)\, Q \frac{\partial V}{\partial Q} - rV + R(Q, \overline{Q}, t) = 0, \tag{3.2}$$

where r is the instantaneous risk free rate of interest and κ is the "market price of risk". Informally, the tradeoff between the risk of an investment which has a value dependent on Q and its anticipated return is captured by κ. Since we are solving backward from the investment horizon T to today, it is convenient to transform the forward equation (3.2) into the backward equation

$$\frac{\partial V}{\partial \tau} = \frac{1}{2}\sigma^2 Q^2 \frac{\partial^2 V}{\partial Q^2} + (\mu - \kappa\sigma)\, Q \frac{\partial V}{\partial Q} - rV + R(Q, \overline{Q}, \tau). \tag{3.3}$$

Equation (3.3) is the valuation equation for a given capacity level $\overline{Q}$, where the network operator receives continuous revenues. The next step is to add in maintenance costs. We assume that these are constant over time and are paid at discrete time intervals τ_{maint} (e.g. monthly). With time running backward, let $\tau^-_{\text{maint}} = \tau_{\text{maint}} - \epsilon$ and $\tau^+_{\text{maint}} = \tau_{\text{maint}} + \epsilon$, where $\epsilon > 0$, and $\epsilon \ll 1$. Equation (3.3) is then augmented by the condition

$$V(Q, \overline{Q}, \tau^+_{\text{maint}}) = V(Q, \overline{Q}, \tau^-_{\text{maint}}) - M\Delta\tau_{\text{maint}}, \tag{3.4}$$

where M is the maintenance cost in dollars per year and $\Delta\tau_{\text{maint}}$ is the time interval between maintenance payments (expressed as a fraction of a year).

With this in mind, we now consider the possibility of upgrading to a higher level of capacity. This adds considerable complexity. We need to keep track of the maximum capacity of each cluster. This will be a discrete variable $\overline{Q}_j$, where there are $j_{\max}$ possible capacity levels, indexed by $j = 0, 1, \ldots, j_{\max} - 1$. We allow upgrades to any higher level of capacity (e.g. we could upgrade to the highest possible level, skipping all intermediate levels). In other words, we can move from $\overline{Q}_j$ up to $\overline{Q}_u$, where $u \in \{j + 1, \ldots, j_{\max} - 1\}$.

The decision of whether or not to upgrade (and to which level) is assumed to be made periodically, at a discrete set of observation times $t_{\text{obs}} = \{0, \Delta t_{\text{obs}}, 2\Delta t_{\text{obs}}, \ldots\}$. Typically, we will let $\Delta t_{\text{obs}} = 1$ month. If a decision is made to upgrade at some time $t_{\text{up}} \in t_{\text{obs}}$, then the actual upgrade is completed at $t_{\text{up}} + \gamma$. We assume that $\gamma/\Delta t_{\text{obs}}$ is an integer. Note that γ is the amount of time necessary to order and set up the equipment.

Due to the fact that the dynamic programming approach works backward in time, at any time in t_{obs} we cannot know when (or if) an upgrade decision was made, and to which capacity level. Hence, we have to solve for all possible times at which an upgrade could occur, and all possible capacity levels for an upgrade. In addition to the variable u described above which indicates the capacity level of a potential upgrade, we need an additional discrete counter variable l to track the times at which upgrades might occur. There are $l_{\max} + 1$ possible values for this variable, these being $0, 1, \ldots, l_{\max}$, where $l_{\max} = \gamma/\Delta t_{\text{obs}}$. The value of $l = 0$ is used to store the value of the investment at the existing level of capacity. Values of $l > 0$ correspond to the amount of time elapsed (working backward and measured in terms of $\Delta\tau_{\text{obs}} = \Delta t_{\text{obs}}$) since a potential upgrade was completed. For example, suppose that $\gamma = 3$ months, $\Delta\tau_{\text{obs}} = 1$ month and $l_{\max} = 3$. If $l = 1$, then the equipment came online 1 month ago, and the upgrade decision will be made 2 months from now.[3] Similarly, if $l = 2$, the equipment was ready 2 months ago, but the actual decision of whether or not to upgrade will occur 1 month from now. If $l = 3$, the decision to upgrade will be made immediately, and the equipment was operational 3 months ago. Note that at each observation time $\tau_{\text{obs}} = T - t_{\text{obs}}$, the value of l changes. Let τ^- (τ^+) denote backward time right before (after) an observation time, and let l^- and l^+ be the value of the counter variable at τ^- and τ^+. Then we have

$$l^+ = l^- + 1 \text{ if } l^- < l_{\max} \tag{3.5}$$

This is because the elapsed time will be incremented by one if we have not yet reached the time when the upgrade decision occurs.

Our valuation function now depends on five variables: the level of demand Q, the existing capacity level $\overline{Q}_j$, the upgrade decision level indicator u, the discrete time counter variable l, and backward time τ: $V = V(Q, \overline{Q}_j, u, l, \tau)$ (as well as the periodic maintenance costs). As mentioned above, $l = 0$ is used to indicate the value of the investment at the existing level of capacity. Note that this also requires $u = j$.

The only remaining factor to include is the cost of upgrading. Let $C_{j\to u}(\tau)_{x\to y}$ denote some (possibly time-dependent) designated partial payment from state x to state y of the total cost of upgrading from a cluster with maximum capacity $\overline{Q}_j$ to one with maximum capacity $\overline{Q}_u$.[4] Note

[3] Recall again that time is running backward in the dynamic programming approach being described here. In terms of forward time, the upgrade decision has been made in the past and the equipment will be installed in the future; but in backward time the upgrade decision will be made in the future and the equipment installation has occurred in the past.

[4] The designated partial payment can be specified in a variety of ways. For instance, with four months lead time, one-quarter of the cost could be paid in each of the four months. Alternatively, all of the cost could be paid up front.

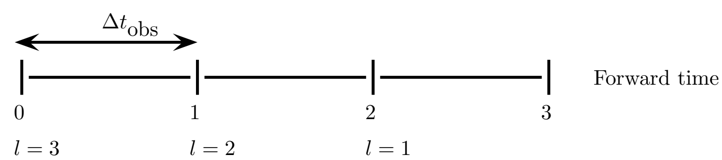


FIGURE 4: *This figure illustrates a possible sequence of payments for upgrade costs. Suppose that $\gamma = 3$ months, $\Delta\tau_{obs} = 1$ month and $l_{\max} = 3$. Partial upgrade costs are paid at times $0, 1, 2$ going forward in time. Going backward in time equation (3.6) represents partial payments for $l = 3, 2$ and equation (3.7) indicates what occurs at time $l = 1$.*

that it is assumed below that partial upgrade costs are paid at the beginning of the month (in forward time), which causes some notational complexity here. We can think of $x = l$ and $y = l-1$, but only for $l = 2, \ldots, l_{\max}$. There is no partial upgrade cost payment from $l = 1$ to $l = 0$. However, there is such a payment from $l = 0$ to $l = l_{\max}$, so we have to allow for this, leading to the notation $C_{j\to u}(\tau)_{x\to y}$ (in this case $x = 0$ and $y = l_{\max}$).

We assume that these upgrade costs are paid at the observation times τ_{obs}, implying that

$$V(Q, \overline{Q}_j, u, l, \tau^+_{\text{obs}}) = V(Q, \overline{Q}_j, u, l-1, \tau^-_{\text{obs}}) - C_{j\to u}(\tau^-)_{l\to l-1}, \quad l = 2, \ldots, l_{\max} - 1 \tag{3.6}$$

where $u = j+1, \ldots, j_{\max} - 1$, $\tau^+_{\text{obs}} = \tau_{\text{obs}} + \epsilon$, and $\tau^-_{\text{obs}} = \tau_{\text{obs}} - \epsilon$, where $\epsilon > 0$, and $\epsilon \ll 1$. At $l = 1$, we have

$$V(Q, \overline{Q}_j, u, 1, \tau^+_{\text{obs}}) = V(Q, \overline{Q}_u, u, 0, \tau^-_{\text{obs}}) \tag{3.7}$$

Consistent with the discussion above, equations (3.7) and (3.6) indicate that partial upgrade costs are assumed to be paid at the beginning of the month (going forward in time) so that $C_{j\to u}(\tau)_{1\to 0} = 0$, and it is possible to upgrade from $\overline{Q}_j$ to $\overline{Q}_u$, but the new equipment will not be ready for some time. For example, suppose that $\gamma = 3$ months, $\Delta\tau_{\text{obs}} = 1$ month and $l_{\max} = 3$, then upgrade costs are paid at times $0, 1, 2$ going forward in time. Thus going backward in time equation (3.6) represents partial payments for when $l = 3, 2$ (corresponding to forward times $0, 1$ and equation (3.7) indicates what occurs at when $l = 1$ (forward time 2). This is depicted in Figure 4.

Furthermore, when working backward in time, upgrade costs are paid before upgrade decisions are made, and as a consequence we need to store $V(Q, \overline{Q}_j, u, l_{\max}, \tau^-_{\text{obs}})$ into a temporary variable $W(Q, \overline{Q}_j, u, l_{\max}, \tau^-_{\text{obs}})$ before equation (3.6) is applied. This temporary variable is then used in the decision of whether or not to upgrade. In particular, at each upgrade decision date τ_{up}, we maximize the value of the investment V by comparing the value of the investment in the current level of capacity with that of all possible completed upgrades to higher capacities. This implies

$$V(Q, \overline{Q}_j, j, 0, \tau^+_{\text{up}}) = \max\left[V(Q, \overline{Q}_j, j, 0, \tau^-_{\text{up}}), W(Q, \overline{Q}_j, u, l_{\max}, \tau^-_{\text{up}}) - C_{j\to u}(\tau^-)_{0\to l_{\max}}\right], \tag{3.8}$$

for $u = j+1, \ldots, j_{\max} - 1$, $\tau^+_{\text{up}} = \tau_{\text{up}} + \epsilon$, and $\tau^-_{\text{up}} = \tau_{\text{up}} - \epsilon$, where $\epsilon > 0$, and $\epsilon \ll 1$. Equation (3.8) indicates that a decision to add capacity will only be taken if the value of the investment with the extra capacity exceeds the value of the investment without it. An important assumption implicit in equation (3.8) is that while new equipment is ordered, installed, and tested, the current stream of revenue is not interrupted. In other words, while there is lead time, there is no down time.[5]

Following the financial option valuation literature, the implication of the extra state variables j, l, and u is that we have to solve a set of partial differential equations of the form of equation (3.3), one for each upgrade possibility. Let $V(Q,\tau)_{j,u,l}$ denote the value of an investment given the continuous variable Q, and the discrete variables j, u, and l. Then we have to solve a collection of problems

$$\frac{\partial V_{j,u,l}}{\partial \tau} = \frac{1}{2}\sigma^2 Q^2 \frac{\partial^2 V_{j,u,l}}{\partial Q^2} + (\mu - \kappa\sigma)\, Q \frac{\partial V_{j,u,l}}{\partial Q} - rV_{j,u,l} + R(Q, \overline{Q}_j, \tau), \tag{3.9}$$

where the revenue function $R(\cdot)$ now explicitly depends on the capacity level via j, and updating rules analogous to equations (3.4) for maintenance costs[6], (3.6) for upgrade costs, and (3.8) for upgrade decisions are also applied.

4 Details of the Computational Algorithm

In this section, we provide a more complete description of the algorithm that we use. For simplicity, we assume that $\Delta\tau_{\text{obs}} = \Delta\tau_{\text{maint}} = \Delta\tau_{\text{up}}$, so that $\tau_{\text{obs}} = \tau_{\text{maint}} = \tau_{\text{up}}$. Note that $\Delta\tau_{\text{up}}$ corresponds to the time interval between upgrade decisions (e.g. one month, six months, one year). At observation times, then any of the following events can take place:

- maintenance costs are paid;
- partial or complete payments are made for capital expenditure;
- decisions about possible upgrades are made; and
- upgrades come on line.

The exact sequence of these observation time events will be made clear in the discussion below.

We now give some more details about the revenue function. As noted above, it is assumed that the owner of the cluster receives continuous payments. For the maximum capacity $\overline{Q}$ (measured in minutes per bouncing busy hour), we have

$$R(Q, \overline{Q}_j, \tau) = \min(Q, \overline{Q}_j) P(\tau). \tag{4.1}$$

The price function $P(\tau)$ is given by

$$P(\tau) = P_0 \exp\left(-\alpha(T - \tau)\right), \tag{4.2}$$

where $P_0 Q$ has units of dollars per year and α is a decay parameter. The payment received can be no larger than the maximum capacity of the cluster multiplied by the price. Consistent with our earlier observations about price paid by users per minute (see Figure 1), we assume that this is a known decreasing function of time.

Turning to upgrade costs, recall that $C_{j\to u}(\tau)_{x\to y}$ is a designated fraction of the total upgrade cost from a cluster of maximum capacity $\overline{Q}_j$ to one with maximum capacity $\overline{Q}_u$ (during the transition from state x to state y). These upgrade costs are assumed to decay exponentially over time with the same parameter α as the price term above.[7]

[5]It would be a straightforward extension to incorporate down time into the model.

[6]Note that the maintenance costs are assumed to depend on the capacity level, i.e. $M = M_j$.

[7]More precisely, only the total cost of an upgrade is assumed to decline over time. If the upgrade costs are paid at observation dates throughout the lead time interval, then the individual payments for a particular upgrade do not decay exponentially.

Next we provide a more complete summary of our algorithm. As noted above, a variety of events can happen at observation times, but for simplicity we denote the time (working backward) before any of these events as τ^- and the time following the event as τ^+. The length of time between observation times is $\Delta\tau = 1$ month. For illustrative purposes, we assume here that the lead time required for the upgrade is 3 months. We can think of the algorithm as having the following steps:

- **Step 1:** Impose the condition $V_{j,u,l}(Q, 0) = f(Q)$ at the investment horizon date $\tau = 0$.
- **Step 2:** Solve the collection of PDEs (3.9) to the first observation time.
- **Step 3:** Consider the observation time as a time when maintenance costs are paid. Update the solution by subtracting these costs, as described in the pseudo-code given in Algorithm 1.
- **Step 4:** Consider the observation time as partially completed upgrades move one step closer to completion, upgrade costs are paid and completed upgrades to higher capacity levels become available. The solution updating is shown graphically in Figure 5. Pseudo-code is given in Algorithm 2.
- **Step 5:** Consider the observation time as an upgrade decision date and apply equation (3.8). This is sketched in the pseudo-code provided in Algorithm 3, and depicted graphically in Figure 6.
- **Step 6:** If $\tau = T$, terminate. Otherwise, solve the collection of PDEs to the next observation date and repeat Steps 3-6.

Algorithm 1 At each maintenance cost payment date, V is reduced by the amount of these costs.

```
for j = 0, ..., j_max − 1 do
  V_{j,j,0}(Q, τ⁺) = V_{j,j,0}(Q, τ⁻) − M_j Δτ
  // loop over the different clusters
  for u = j + 1, ..., j_max − 1 do
    // update the solution with the appropriate maintenance cost
    for l = 1, ..., l_max do
      V_{j,u,l}(Q, τ⁺) = V_{j,u,l}(Q, τ⁻) − M_j Δτ
    end for
  end for
end for
```

As noted above, the implementation of our algorithm requires solving a set of partial differential equations (3.9). This is accomplished using a general numerical partial differential equation solver such as described in [18, 19]. The approach involves a finite volume discretization of equation (3.9) along the Q-axis, representing traffic demand. As it is beyond the scope of this paper, we will not present the details of the scheme here. Interested readers should consult references such as [8, 18, 19, 4].

5 Parameter Values

This section describes the estimation of parameter values. Details regarding estimates of the growth rate μ and the volatility σ can be found in Appendix A. By averaging estimates for these two

Algorithm 2 At each observation date, update the solution of the potential upgrades from the cluster with maximum capacity $\overline{Q}_j$ to that with maximum capacity $\overline{Q}_u$.

```
for j = 0, ..., j_max − 2 do
  // loop over the different clusters
  for u = j + 1, ..., j_max − 1 do
    //copy solution of l_max into temporary variable
    W_{j,u,l_max}(Q, τ−) = V_{j,u,l_max}(Q, τ−)
    // loop over the upgrade possibilities for cluster j
    for l = l_max, ..., 2 do
      // loop over the elapsed time since the upgrade was completed
      V_{j,u,l}(Q, τ+) = V_{j,u,l−1}(Q, τ−) − C_{j→u}(τ−)_{l→l−1}
    end for
    //at l = 1 (upgrade costs are paid at the beginning of the month)
    V_{j,u,1}(Q, τ+) = V_{u,u,0}(Q, τ−)
  end for
end for
```

Algorithm 3 At each upgrade decision date, compare the value of the investment if a decision is taken to upgrade to a higher level of capacity with the value of the existing level of capacity.

```
for j = 0, ..., j_max − 2 do
  // loop over the cluster with capacity j
  for  u = j + 1, ..., j_max − 1  do
    // loop over the upgrade possibilities for cluster j
    V_{j,j,0}(Q, τ+) = max [V_{j,j,0}(Q, τ−), W_{j,u,l_max}(Q, τ−) − C_{j→u}(τ−)_{0→l_max}]
  end for
end for
```

parameters from data obtained for three different switches, we find values of $\mu \approx .30$ per year and $\sigma \approx .65$ per year$^{\frac{1}{2}}$.

We next consider the market price of risk κ. In Appendices C and D, two approaches are presented to estimate this parameter. Using these different methods, and based on data for different switches, we estimate values of $\kappa \approx .17$, $\kappa \approx .10$ and $\kappa \approx .03$. While these three estimates may be seen as quite different, it turns out that our results (in terms of when it is optimal to upgrade capacity) are not very sensitive to values of κ across the range of these estimates.

In terms of upgrade costs and time, our estimates were obtained from industry practitioners [11] and are as follows. The total hardware cost of adding one carrier is approximatively \$150,000. Consequently, adding one carrier for each base station of the cluster would cost around \$150,000 $\times$ 20 = \$3,000,000. Once an order has been placed to upgrade a cell site with an additional carrier, it takes approximately two months before the hardware is delivered, one month to install the hardware, and one month to set up and optimize the new carrier. Hence, it takes about four months from order placement for the equipment to be available on line. Finally, note that we assume that upgrade costs decline over time. The rate of decrease is assumed to be the same as that for average revenue per user, as described below.

The monthly maintenance fee of a cluster is the cost of a cell site technician and a T1 cable connection. A cell site technician maintains approximately 20 cell sites. The cost of a cell site technician is assumed to \$150,000 per year. Cell site leasing costs are approximately \$1,500 per

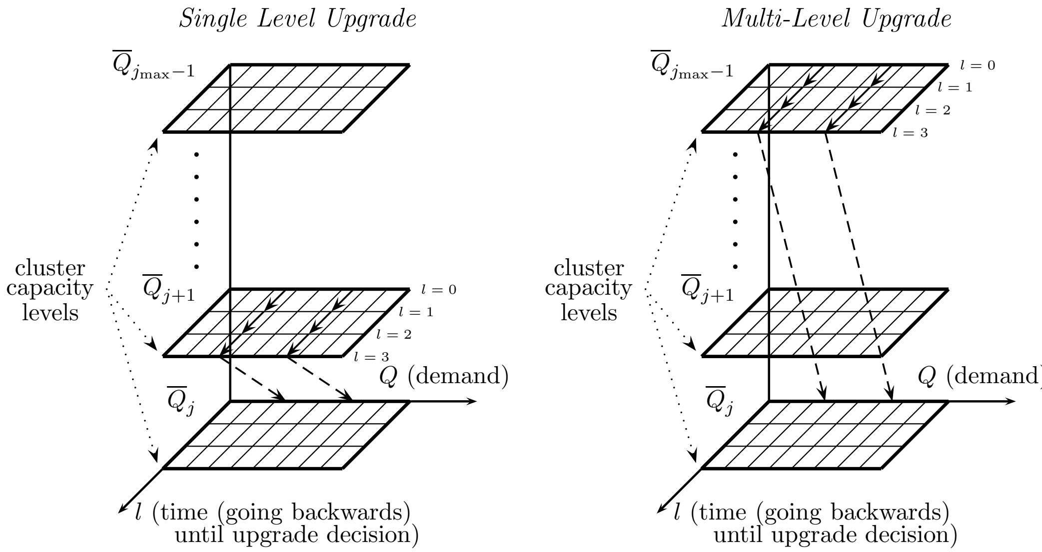


FIGURE 5: *The flow of information at observation dates. In the left panel, a possible upgrade decision from level $\overline{Q}_j$ to the next higher level, $\overline{Q}_{j+1}$ is being evaluated. In the right panel, the potential upgrade is to the highest possible level, $\overline{Q}_{j_{\max}-1}$. Time is running in the backward direction, and it is assumed that it takes three months for a capacity upgrade to be implemented once the decision has been taken. Consider, for example, the left panel. The value of a completed investment into the higher capacity level $\overline{Q}_{j+1}$ is shown along the line along the Q-axis (i.e the $l = 0$ line) in the $\overline{Q}_{j+1}$ plane. At the next monthly observation date, the value is moved one step in to the $l = 1$ line. Note that on this line, the capacity in use is actually $\overline{Q}_j$, because the upgrade is incomplete. Similar updating occurs at the succeeding observation dates, until the $l = 3$ line is reached. At this point, the decision of whether or not to upgrade to this capacity level is made by comparing the value of the investment along this line with that along the Q-axis for the existing $\overline{Q}_j$ capacity level. If a decision is made to upgrade to capacity level $\overline{Q}_{j+1}$, the value from the $\overline{Q}_{j+1}$ level along the $l = 3$ line migrates down to the $l = 0$ line in the $\overline{Q}_j$ plane, as indicated by the dashed line. The evaluation of the possible multi-level upgrade in the right panel is similar.*

month and T1 back-haul costs are approximately \$500 per month per T1. Since there is one T1 connection per carrier, a three carrier site would have three T1 connections provisioned. Electricity and warranty costs are respectively \$250 and \$200 per cell site. Thus the total annual maintenance cost for cluster with n carriers per cell site is about $\$150{,}000 + \$1{,}500 \times 20 \times 12 + n \times \$500 \times 20 \times 12 + (\$250 + \$200) \times 20 \times 12$. Table 2 shows the overall maintenance costs when there are one, two, and three carriers per cell site.

In order to determine revenues, we need to determine the price function $P(\tau)$ (see equations (4.1) and (4.2)). The most recent data shown in Figure 1 shows that the average revenue per user (ARPU) is approximately \$0.229/min. However, this value corresponds to the total revenue received based on daily traffic, not net revenue based on bouncing busy hour traffic. Consequently, to estimate P_0 we must adjust the average revenue per user (ARPU) appropriately. The total revenue producing minutes in a day is approximately ten times the bouncing busy hour minutes (according to [11]) and net revenue per user is approximately 70% of the average revenue per user.

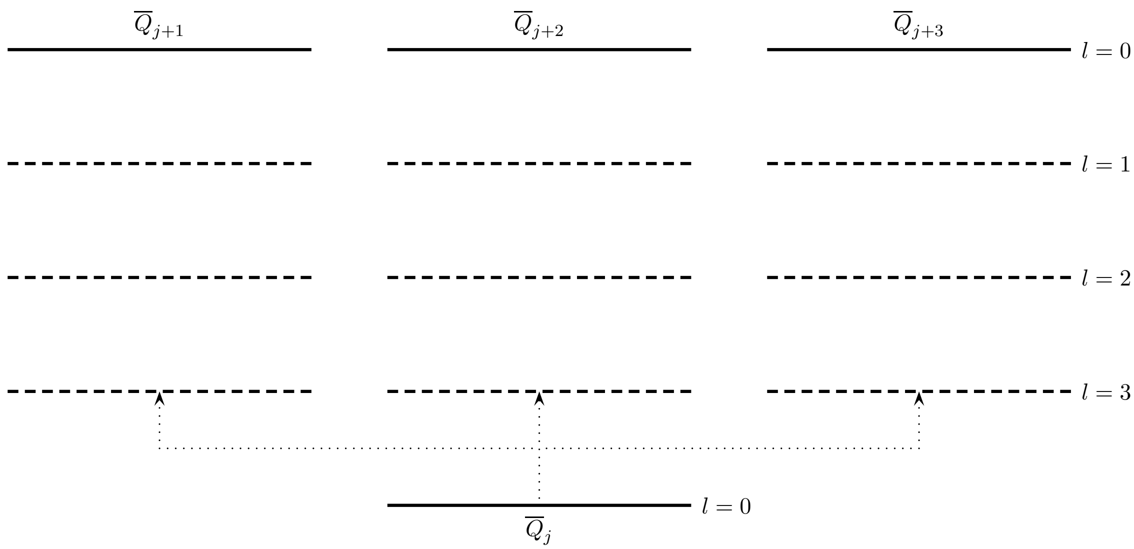


Figure 6: *Illustration of the upgrade decision. At each upgrade decision date, the value of the investment at the current level of capacity $\overline{Q}_j$ and the current level of demand, $V(Q,\tau)_{j,j,0}$, is compared with the values of possible completed investments to various higher levels of capacity, $V(Q,\tau)_{j,u,3}$. In this case, we are considering upgrades to 1, 2, or 3 higher levels (i.e. $u \in j+1, j+2, j+3$), and assuming that it takes 3 months to complete the investment. In other words, working backward in time, we start 3 months before the upgrade decision date with the values of completed investments corresponding to the different higher capacity levels. The counter variable l is incremented at each observation date (and appropriate adjustments are made to the value V, arising from items such as capital costs being paid). When 3 months have elapsed, we have reached the upgrade decision date, so we compare the values of the various investments into higher levels of capacity with the current level, and choose the highest value at this time.*

We assume that revenues are zero over weekends. Thus,

$$P_0 = \text{ARPU} \times .70 \times 10 \times 250, \tag{5.1}$$

where 250 is the approximate number of days in the year excluding weekends. Using the ARPU value of \$0.229 per minute gives $P_0 = 400.75$ dollars per minute per bouncing busy hour per year.

The maximum total annual revenue for a given capacity level can then be calculated as follows. Suppose we are considering a cluster with a single carrier per cell site. From Table 1, 2,376 minutes of traffic can be handled per bouncing busy hour per cell site. Multiplying by 20 (the number of cell sites in the cluster) gives a total of 47,520 minutes of traffic per bouncing busy hour for the cluster. Multiplying this by $P_0 = 400.75$ gives annual revenues of \$19,043,640. The maximum total revenues for one, two, and three carriers cell sites are presented in Table 3.

To complete the determination of the $P(\tau)$ function, we need to determine the decay parameter α. For voice traffic, the price per carrier is decreasing every year, but not by a significant amount. Vendors offer features to increase the traffic handling capability of each carrier every couple of years. Such enhancements can be used to justify keeping the dollars per carrier rate relatively stable. We will assume a decay factor of 5% per year for average revenue per user. Note that this decay factor is not the same as in Figure 1 (i.e. 8%). Figure 1 was constructed using Bell Canada quarterly financial reports. The decay factor and volatility were estimated using the methods described in Appendix A. From our discussions with network operators [11], we believe that a decay factor of $\alpha = 5\%$ is more representative for the expected future decrease in ARPU. Note that, as mentioned above, this decay parameter is also applied to upgrade costs.

We end this section by listing the parameters needed for equation (3.9) and summarizing our assumptions about the characteristics of a cluster. The model parameters for equation (3.9) are presented in Table 4. Our assumptions about the features of a cluster are as follows:

- A cluster is composed of 20 cell sites.
- Each cell site starts with one carrier and can be upgraded to at most three carriers.
- Traffic is homogeneous throughout the cluster.
- It takes four months after order placement until the upgraded cluster is completely operational. There is no loss of revenue from the existing equipment during this lead time.
- Each cell site of the cluster has three sectors.
- The maximum number of minutes per hour that can be handled at 2% blocking for a carrier of a cell site is given in Table 1.

6 Results

We now provide some illustrative results. Before beginning, the following two observations should be made. First, we are concentrating on the upgrade decision, not the value of the investment. This is equivalent to identifying the location of a free boundary. This boundary will move over time. In fact, it will generally increase as time moves forward, as shown in the bandwidth capacity management context evaluated in [3]. The intuition for this is straightforward. As we get closer to the investment horizon T, there is less time available for the higher revenues associated with higher traffic and expanded capacity to cover the costs of upgrading. Of course, things are more complicated due to factors such as upgrade costs decreasing over time and the possibility of upgrading to more than a single higher capacity levels, but this is the basic idea. Note, however, that if we get close enough to T (within the lead time required to install the extra capacity), we will never upgrade, no matter how high demand is, simply because the new capacity will never provide any revenue to offset the expansion costs. Keeping this in mind, we will present results only for the upgrade decision at $t = 0$.[8] Second, we will not concentrate here on the effects of the parameters of the traffic time

[8] In fact, the optimal upgrade decision as a percentage of existing capacity in use does not change too much for $0 \leq t \leq 2$, provided we are considering fairly frequent evaluations of the upgrade decision. If we suppose that upgrade decisions are made only on annual basis, then we find more significant differences.

Number of carriers per cell site	Cluster annual maintenance cost
1	$738,000
2	$858,000
3	$978,000

TABLE 2: *Estimated annual maintenance costs for a cluster.*

Number of carriers per cell site	Cluster maximum annual revenues
1	\$19,043,640
2	\$44,723,700
3	\$71,557,920

Table 3: *Estimated maximum annual revenues for a cluster.*

Parameter	Value
Investment horizon (T)	5 years
Decay in price (α)	.05/year
Growth rate (μ)	.3/year
Volatility (σ)	.65/year$^{\frac{1}{2}}$
Risk free interest rate (r)	.04/year
Market price of risk (κ)	[.03;.10;.17]/year$^{\frac{1}{2}}$

Table 4: *Model parameters for equation* (3.9).

series, μ and σ. The effects of these two parameters are qualitatively similar to those reported in [3] in the bandwidth context. Faster rates of the growth parameter μ obviously give rise to faster upgrades to capture the extra revenue earlier. Higher levels of the volatility parameter σ work in the opposite direction—with increased uncertainty, it is better to wait longer before upgrading. This is because the chances of traffic demand decreasing substantially are higher with more uncertainty.

Since we are solving the set of partial differential equations (3.9) on a discrete grid of points and over a discrete set of timesteps, we begin by exploring how sensitive the upgrade decision is to the level of discretization. Table 5 shows the upgrade decision for various levels of grid refinements and timesteps. The upper part of the table is for the case where we consider adding a single carrier to each cell site of the cluster, while the lower part shows results where we add two carriers to each cell site. Table 5 reveals that the estimated percentage of existing maximum capacity at which it is optimal to upgrade is not overly sensitive to the grid used, changing by at most around 3% as the grid is refined for a given upgrade decision interval. The table also shows that the estimated percentages are changing only slightly as we go from two levels of grid refinement to three, indicating that the algorithm has converged to a reasonable level of accuracy. Subsequent results reported below use three levels of grid refinement.

Table 6 contains results for three different values of the market price of risk κ, for a variety of upgrade decision intervals, ranging from monthly to annually. As with Table 5, we assume that all of the cluster capacity (at 2% blocking) is available and the stream of revenues is capped when the maximum capacity is reached. The percentage in use of the maximum cluster capacity at which it is optimal to upgrade rises with κ. This result is to be expected because as κ increases, the drift term $(\mu - \kappa\sigma)$ of equation (3.9) decreases, and the upgrade should therefore occur later. The magnitude of this effect is not large, however: the difference between the lowest and highest values of κ, in terms of the optimal upgrade percentage, is only about 10%. This indicates that the upgrade decision is not very sensitive to our estimate of κ.

In Table 6, we find that at present it would be optimal to add a new carrier to each cell site of the cluster if about 94% (a little lower for $\kappa = .03$, slightly higher for $\kappa = .17$) of its maximum

Refinement Level	Upgrade Decision Interval Monthly	Quarterly	Semi-annually	Annually
	Add One Carrier per Cell Site			
0	90%	80%	75%	60%
1	90%	77.5%	72.5%	60%
2	90%	77.5%	72.5%	60%
3	90%	77.5%	72%	60%
	Add Two Carriers per Cell Site			
0	102%	88%	84%	70%
1	101.5%	88%	82%	70%
2	101.25%	87.5%	81%	67.5%
3	101%	87.25%	81%	67.5%

TABLE 5: *The effect of grid and timestep refinement on today's upgrade decision. The numbers indicate the percentage in use of the maximum capacity of the cluster at 2% blocking at which it is currently optimal to upgrade. We allow 100% usage of the total cluster capacity. Above 100%, the revenue stream is capped by the cluster maximum capacity. Parameters are* $r = .04$, $\kappa = .03$, $\mu = .30$, $\sigma = .65$, *and* $T = 5$ *years. It takes four months between the time the equipment is ordered and the time it is on line. The initial grid (refinement level 0) has 70 points. The timestep size for this grid is 1/250 (i.e. there are 250 timesteps per year). For each succeeding refinement, the number of grid points is doubled by inserting new points halfway between previously existing ones, while the timestep size is halved.*

capacity is reached for a monthly upgrade decision interval. As the upgrade decision interval is increased from monthly to annually, the upgrade percentage decreases from around 94% to about 65%. The comparable numbers when considering adding two new carriers to each cell site are about 106% for monthly upgrade decisions, dropping to around 72% when the decisions are made annually. Intuitively, this simply reflects the fact that with less frequent decisions it is better to upgrade earlier, since there are fewer opportunities to make decisions. For example, suppose traffic is currently high. If upgrade decisions are made annually, there is a large potential cost to not upgrading today in terms of foregone revenue. This is mitigated substantially if the decision can be re-evaluated in a month.

In Table 6, we observe that in some cases it is optimal to wait before upgrading the cluster until traffic demand is above its maximum capacity. While these levels may appear somewhat surprising and unrealistic, we are in effect simply saying that customers experience a lot of blocking before it is optimal to upgrade the cluster. In practice, since there is a financial penalty with poor quality of service, these levels of upgrades are never reached. As a final comment about Table 6, we observe that in some circumstances it can be optimal to add two new carriers per cell site, rather than one. This is the case if traffic demand is quite high, and if upgrade decisions are made relatively infrequently.

A concern which could be raised regarding this analysis is that the assumption that 100% usage (at 2% blocking) is quite aggressive. However, our framework allows us to take into account criteria such as quality of service. For instance, it is conceivable that engineers prefer a safety buffer between the maximum capacity (at 2% blocking) and the capacity available to customers. Once this threshold is reached, the quality of service deteriorates. To compensate, customers may

Upgrade Decision Interval	Add One Carrier $\kappa = .03$	$\kappa = .10$	$\kappa = .17$	Add Two Carriers $\kappa = .03$	$\kappa = .10$	$\kappa = .17$
Monthly	90%	94.5%	99.25%	101.25%	106.5 %	112%
Quarterly	77.5%	82%	86.5%	87.5%	92.25%	97.5%
Semi-annually	72.5%	76.25%	81%	81%	86%	91%
Annually	60%	65%	70%	67.5%	72.5%	77.5%

TABLE 6: *The effect of the market price of risk κ on today's upgrade decision. The numbers indicate the percentage in use of the maximum capacity of the cluster at 2% blocking at which it is currently optimal to upgrade. We allow 100% usage of the total cluster capacity. Above 100%, the revenue stream is capped by the cluster maximum capacity. Parameters are $r = .04$, $\mu = .30$, $\sigma = .65$, and $T = 5$ years. It takes four months between the time the equipment is ordered and the time it is on line.*

receive rebates or free calls. Of course, customers may also seek other vendors.

We make the following simple change to our model to investigate how quality of service can affect the upgrade decision. Let ϕ be the safety factor, representing a percentage of the maximum capacity of the cluster. Replace the revenue equation (4.1) by

$$R(Q, \overline{Q}_j, \tau) = \begin{cases} \mathcal{P}(\tau) \times Q & \text{if } Q \leq \phi\overline{Q_j} \\ \mathcal{P}(\tau) \times \overline{Q_j} \times \max\left(1 - \frac{Q - \phi\overline{Q_j}}{\phi\overline{Q_j}}, 0\right) & \text{otherwise.} \end{cases} \tag{6.1}$$

For example, $\phi = 90\%$ implies that full revenues are received only when traffic is less than or equal to 90% of the maximum capacity of any given cluster (at 2% blocking). Above this threshold, revenues eventually drop to zero as demand keeps increasing. Effectively, we are adding a financial penalty as the quality of service deteriorates. Note that this penalty applies even if $\phi = 100\%$. Whereas the earlier revenue equation (4.1) simply caps the revenue at the capacity level for demand above capacity, the modified equation (6.1) reduces the revenue received when demand exceeds capacity. It might also be possible to develop penalty functions based on the effect of quality of service on customer "churn rates" (i.e. the loss of customers to other vendors as a result of poor service), or other criteria. This differs from the cases considered above in Tables 5 and 6 where revenue was simply capped once capacity was reached.

In Table 7, we present the results for different values of the safety level ϕ, ranging from 100% to 80%. The upgrade decision is considered every six months and $\kappa = .10$. The table considers the case where we add a single carrier to each cell site. Similar results are obtained when upgrading involves adding two carriers per cell site. The first row in the table, where the safety factor is indicated as "none", corresponds to the situation in the previous tables where revenue is simply capped when demand is higher than available capacity. In Table 7, we observe that as the safety factor decreases, the upgrade occurs at lower values of the percentage of the total capacity of the cluster. Similar behaviour is found when upgrade decisions are made for other time intervals (e.g. quarterly, or monthly). This phenomenon makes intuitive sense: a financial penalty for poor quality of service provides an enhanced incentive to upgrade.

Safety Factor	Add One Carrier
None	76.25%
100%	65 %
90%	62.5%
80%	60%

Table 7: *The effect of the safety factor ϕ on today's upgrade decision. The numbers indicate the percentage in use of the maximum capacity of the cluster at 2% blocking at which it is currently optimal to upgrade. Parameters are $r = .04$, $\kappa = .10$, $\mu = .30$, $\sigma = .65$, and $T = 5$ years. It takes four months between the time the equipment is ordered and the time it is on line. Upgrade decisions are made every six months.*

7 Conclusion

In this paper, we have explored the management of wireless network capacity using a real options formulation. While the method presented here is similar to that described in [3], the limitations embedded in the algorithm in that work are alleviated by developing a four dimensional model. This enables us to consider different upgrade decision intervals independent of the time period before the new equipment becomes operational.

In practice, current upgrade decisions often appear to be based purely on quality of service criteria. We believe that it is important to also consider financial criteria in terms of maximizing net revenues. By developing appropriate penalty functions which assign a cost to poor quality of service, we can combine both financial and quality of service criteria. This approach will require managers to assign a cost to quality of service issues. Penalty functions could be real financial incentives provided to users (e.g. during high blocking periods, all calls are free), or they could be based on customer churn rates.

Acknowledgment

This work was funded by Bell University Laboratories at the University of Waterloo. We would like to thank Brian O'Shaugnessy and Darlene Farrow for many useful discussions. In particular, we would also like to thank Derek MacAvoy for providing us with invaluable technical information on numerous occasions.

References

[1] C. Chatfield. *The Analysis of Time Series: An Introduction*. Chapman and Hall, London, Fourth edition, 1989.

[2] T. Dangl. Investment and capacity choice under uncertain demand. *European Journal of Operational Research*, 117:415–428, 1999.

[3] Y. d'Halluin, P. A. Forsyth, and K. R. Vetzal. Managing capacity for telecommunications networks under uncertainty. *IEEE/ACM Transactions on Networking*, 10:579–588, 2002.

[4] Y. d'Halluin, P. A. Forsyth, K. R. Vetzal, and G. Labahn. A numerical PDE approach for pricing callable bonds. *Applied Mathematical Finance*, 8:49–77, 2001.

[5] A. K. Dixit. Irreversible investment with uncertainty and scale economies. *Journal of Economic Dynamics and Control*, 19:327–350, 1995.

[6] A. K. Dixit and R. S. Pindyck. *Investment Under Uncertainty*. Princeton University Press, 1994.

[7] J. C. Hull. *Options, Futures, and Other Derivatives*. Prentice Hall, Inc., Upper Saddle River, NJ, Fifth edition, 2002.

[8] D. Kröner. *Numerical Schemes for Conservation Laws*. Chichester, New York, 1997.

[9] J. S. Lee and L. E. Miller. *CDMA Systems Engineering Handbook*. Artech House, Boston, 1998.

[10] WestBay Engineers Limited. What is an Erlang? `http://www.erlang.com/whatis.html`.

[11] D. MacAvoy, 2002. Bell Mobility, Associate Director Access Technology Planning. Private communication.

[12] R. S. Pindyck. Irreversible investment, capacity choice, and the value of the firm. *American Economic Review*, 78:969–985, 1988.

[13] S. Romero. Success of cellphone industry hurts service. *The New York Times*, November 18, 2002.

[14] S. A. Ross, R. W. Westerfield, J. F. Jaffe, and G. S. Roberts. *Corporate Finance*. McGraw-Hill Ryerson Limited, Toronto, Third Canadian edition edition, 2003.

[15] E. S. Schwartz and M. Moon. Rational pricing of internet companies revisited. *Financial Review*, 36:7–26, 2001.

[16] L. Trigeorgis. *Real Options: Managerial Flexibility and Strategy in Resource Allocation*. MIT Press, 1996.

[17] P. Wilmott. *Derivatives: The Theory and Practice of Financial Engineering*. Wiley, Chichester, 1998.

[18] R. Zvan, P. A. Forsyth, and K. R. Vetzal. Robust numerical methods for PDE models of Asian options. *Journal of Computational Finance*, 1:39–78, Winter 1998.

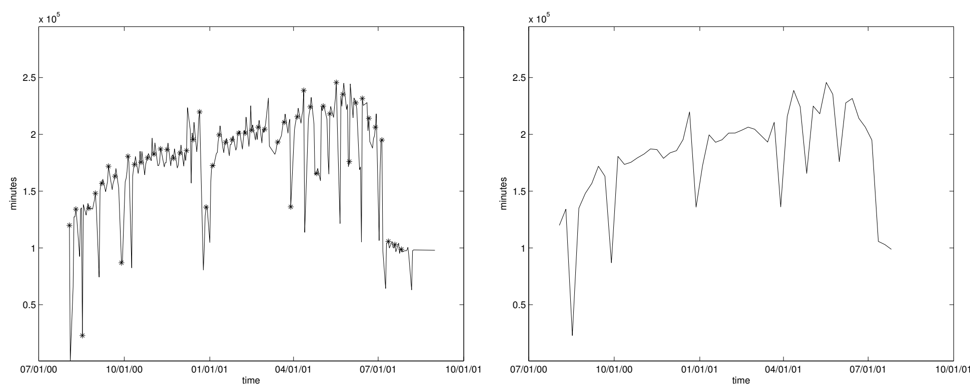
[19] R. Zvan, P. A. Forsyth, and K. R. Vetzal. A finite volume approach for contingent claims valuation. *IMA Journal of Numerical Analysis*, 21:703–731, 2001.

Appendices

A Estimation of Growth Rate and Volatility Parameters

This appendix describes how we estimate the growth rate parameter μ and the volatility parameter σ of equation (3.1). We obtain daily bouncing busy hour traffic data for one year from [11] for three different switches. In this appendix (and also in Appendices C and D dealing with the market price of risk), we provide parameter estimates based on all three switches to get some idea of the range of possible values. We then average these values across the switches to obtain our estimated parameter values.

An initial investigation of the data for the switches showed strong autocorrelation of the time series within each week, as depicted in Figure A.1(a) for a representative case. This is not surprising, since we expect that there will be repetitive patterns within each week. To filter out these effects, we average the yearly network traffic for each day of the week separately, and choose the day with the highest average bouncing busy hour traffic (see Figure A.1(b)). This turns out to be Thursday. We then use this same day each week to estimate week to week effects.



(a) *One year total daily bouncing busy hour traffic. The stars indicate the weekly highest traffic days.*

(b) *One year total daily bouncing busy hour traffic. The plot shows data for Thursdays, the day of the week with the highest average level of traffic.*

FIGURE 7: *Total daily bouncing busy hour traffic for a representative switch.*

After applying this simple filter to remove weekly effects, Figure A.1(b) reveals several suspicious large drops in traffic. Basing our volatility estimate on this data would produce a very high value. Some of the short term declines are simply due to statutory holidays or scheduled maintenance. As these are known events for low network traffic, we should not take them into account when estimating volatility. Consequently, the dates corresponding to statutory holidays and suspicious changes are smoothed out using interpolation. The month of December is also ignored since it is a known low traffic period. Figure 8 presents the time series once the holidays and the large drops have been removed and smoothed out.

We begin by estimating the drift term μ. An implication of equation (3.1) is that traffic is distributed lognormally. Thus we use logarithmic weekly changes, and estimate the growth rate using a least squares method. A plot of the data with the estimated trend removed for a representative switch is provided in Figure 9.

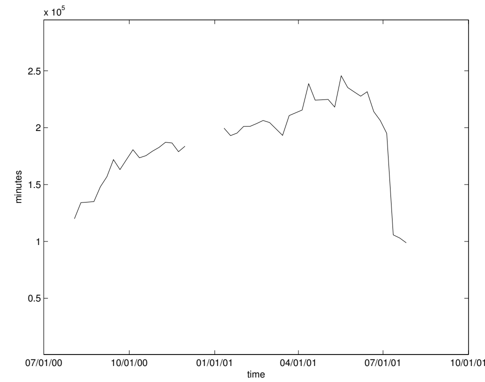


FIGURE 8: *One year total daily bouncing busy hour traffic without known low traffic periods such as major holidays, suspicious events, and the month of December. The plot shows data for Thursdays, the day of the week with the highest average level of traffic.*

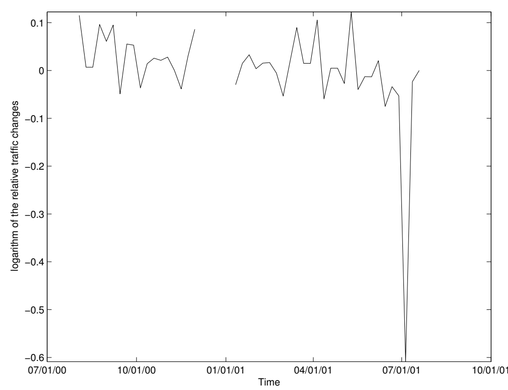


FIGURE 9: *One year logarithmic daily bouncing busy hour relative traffic changes. Known holidays and suspicious changes have been replaced using interpolation. The month of December is ignored and the trend of the time series has been removed. The plot shows data for Thursdays, the day of the week with the highest average traffic.*

If equation (3.1) is a reasonable model for our data, then the detrended logarithmic relative traffic change time series should contain only random errors. To assess this, we use the Ljung-Box Q-statistic [1]. For each switch, we perform the test for the period before and after December for 4 lags. The p-values for the null hypothesis of zero autocorrelation are presented in the two right hand columns of Table 8. They indicate that the hypothesis of no autocorrelation is not rejected in either time period for any of the three switches.

We next estimate the volatility parameter σ by calculating the annualized standard deviation of the detrended time series for each switch. Table 8 shows the estimated values of the drift term μ and the volatility σ for each switch in the two left hand columns. Clearly, there is substantial variation in these parameters across the three switches. For the application presented in this paper, we will simply average the estimates for each parameter across the different switches, resulting in $\mu \approx .30$ and $\sigma \approx .65$. It should be noted that network traffic is highly volatile, even after the various filters described above have been applied. The estimated volatilities for the switches range from about .30 to .90. These can be compared to volatilities in the range of .15 to .30 for stock market indexes.

	Drift	Volatility	Ljung-Box test p-values	
	μ	σ	(before Dec.)	(after Dec.)
Switch A	-.24	.90	.83	.92
Switch B	.41	.74	.29	.61
Switch C	.73	.32	.42	.59

TABLE 8: *Parameter estimates and Ljung-Box test results.*

B Derivation of the Mathematical Model

This appendix presents a brief derivation of equation (3.2) using standard financial arguments [7, 17]. Recall that the level of network traffic Q is assumed to follow geometric Brownian motion (equation (3.1)). Let $V(Q,t)$ be the value of an investment dependent only on Q and time t. Using Itô's lemma, the process followed by $V(Q,t)$ is

$$dV = (\psi V + R)dt - \theta V dz, \tag{B.1}$$

where R represents revenue in dollars per year,

$$\psi V = \frac{\partial V}{\partial t} + \mu Q \frac{\partial V}{\partial Q} + \frac{1}{2}\sigma^2 Q^2 \frac{\partial^2 V}{\partial Q^2}, \tag{B.2}$$

and

$$\theta V = -\sigma Q \frac{\partial V}{\partial Q}. \tag{B.3}$$

Let us pick two investments V_1 and V_2 expiring at some future time. From equation (B.1), we have

$$dV_1 = (\psi_1 V_1 + R_1)dt - \theta_1 V_1 dz,$$
$$dV_2 = (\psi_2 V_2 + R_2)dt - \theta_2 V_2 dz.$$

Both V_1 and V_2 have the same factor of uncertainty dz. We can thus construct a portfolio Π composed of V_1 and V_2 such that the return of this portfolio Π is non-stochastic. Let x_1 be the fraction of the amount invested in V_1 and x_2 be the fraction of the amount invested in V_2. Note that $x_1 + x_2 = 1$. The return on the portfolio is given by

$$\begin{aligned} d\Pi &= x_1 \frac{dV_1}{V_1} + x_2 \frac{dV_2}{V_2} \\ &= \left(x_1\psi_1 + x_1 \frac{R_1}{V_1} + x_2\psi_2 + x_2 \frac{R_2}{V_2} \right) dt - (x_1\theta_1 + x_2\theta_2)\, dz. \end{aligned} \tag{B.4}$$

Choosing $x_1 = -\theta_2/(\theta_1 - \theta_2)$ and $x_2 = \theta_1/(\theta_1 - \theta_2)$, we have $x_1\theta_1 + x_2\theta_2 = 0$. Thus we have

$$\begin{aligned} d\Pi &= \left(x_1\psi_1 + x_1 \frac{R_1}{V_1} + x_2\psi_2 + x_2 \frac{R_2}{V_2} \right) dt \\ &= r\Pi dt, \end{aligned} \tag{B.5}$$

where the second equality comes from the fact that as Π is riskless, it must earn the risk free rate of return r. It follows that

$$\frac{\psi_1 + \frac{R_1}{V_1} - r}{\theta_1} = \frac{\psi_2 + \frac{R_2}{V_2} - r}{\theta_2}. \tag{B.6}$$

Define κ as the value of each side of equation (B.6), i.e.

$$\frac{\psi_1 + \frac{R_1}{V_1} - r}{\theta_1} = \frac{\psi_2 + \frac{R_2}{V_2} - r}{\theta_2} = \kappa.$$

Dropping the subscripts, we have shown that if V is an investment dependent on Q and t, such that

$$dV = (\psi V + R)dt + \theta V dz,$$

then

$$\psi - r + \frac{R}{V} = \theta\kappa. \tag{B.7}$$

Substituting ψ from equation (B.2) and θ from equation (B.3) into equation (B.7), we find

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 Q^2 \frac{\partial^2 V}{\partial Q^2} + (\mu - \kappa\sigma)\frac{\partial V}{\partial Q} - rV + R = 0. \tag{B.8}$$

C Estimation of the Market Price of Risk

This section describes one method of estimating the market price of risk κ. We follow the approach described in [15], which uses stock market data. However, Bell Mobility is not a publicly traded company, so we need to find public Canadian companies which are in the wireless communications business. Two such firms are Rogers Wireless Communications Inc. and Microcell Telecommunications Inc.

For each company, we must first compute the systematic risk exposure using the standard capital asset pricing model.[9] Because each firm has significant amounts of debt outstanding, we will initially use the "levered" equity beta β as the measure of the systematic risk. To estimate β, we run linear regressions of returns for each stock versus the return on the market. We use the Toronto Stock Exchange (TSE) 300 index as a proxy for the return on the market. We obtain daily total return data for the TSE 300 index and each firm from the Canadian Financial Markets Research Centre database. The sample for Microcell runs from October 9, 1997 to December 29, 2000. The sample for Rogers runs from August 9, 1991 through December 29, 2000. Figure 10 provides plots of the return for the market versus the return for Microcell and Rogers.

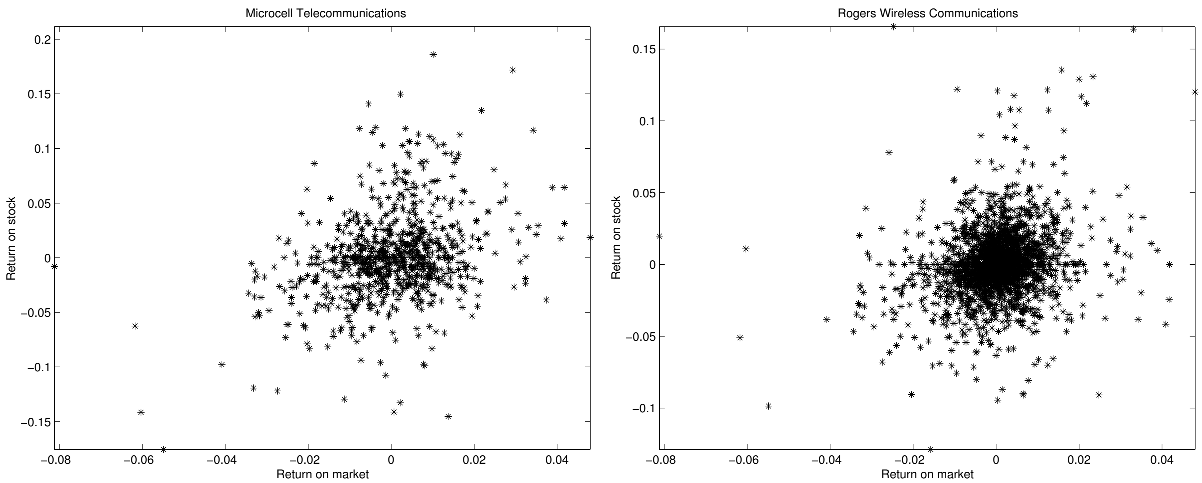
Figure 11 shows the line of best fit superimposed on each point representing pairs of daily return data. For Microcell, we find $\beta = 1.0455$, while $\beta = .5603$ for Rogers.

Having estimated the levered betas for both firms, we next undo the effects of leverage by calculating the beta for a hypothetical unlevered (i.e. no debt) version of each firm.[10] To compute the unlevered firm's beta from the levered equity beta, the market value of the firm's debt and equity must be estimated, along with its corporate tax rate. The unlevered firm's beta is then given by

$$\beta^{\text{unlevered}} = \frac{E}{E + (1 - T_c)D}\beta, \tag{C.1}$$

[9] Readers unfamiliar with these concepts should consult a corporate finance text such as [14].

[10] Note that the calculations to follow involve some simplifying approximations commonly used in corporate finance (such as the assumption that each firm's debt is perpetual).



(a) *Microcell.* (b) *Rogers.*

FIGURE 10: *Daily return on TSE 300 index (horizontal axis) versus daily return on wireless telecommunications firm (vertical axis).*

where

$$E = \text{market value of equity (total number of shares times share price)},$$
$$D = \text{market value of long term debt},$$
$$T_c = \text{corporate tax rate}.$$

As an initial approximation, for each company we will assume a tax rate of 40%. Tables 9 and 10 contain the information we use to calculate the unlevered beta for each company.

Long term debt	$1,887,048,000
Corporate tax rate	40%
Stock price (on February 13, 2002)	$2.68
Number of shares	240,000,000

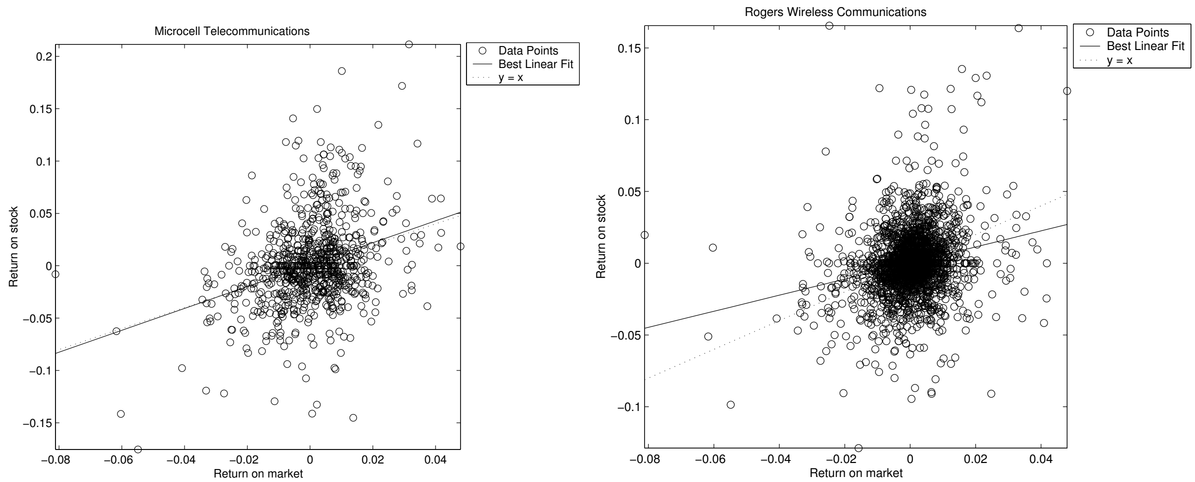
TABLE 9: *Microcell Telecommunications Inc. corporate data. Dollar figures are in Canadian funds. The data was obtained from* `http://www.globeinvestor.com/` *on February 13, 2002 and Microcell's quarterly financial reports.*

Using equation (C.1), we find that

$$\beta^{\text{unlevered}}_{\text{Microcell}} = \frac{2.68 \times 240}{2.68 \times 240 + (1. - .4) \times 1887.048} \times 1.0455$$
$$= .37876,$$

and

$$\beta^{\text{unlevered}}_{\text{Rogers}} = \frac{17.78 \times 144.4}{17.78 \times 144.4 + (1. - .4) \times 2305.638} \times .5603$$
$$= .36411.$$



(a) *Microcell. The slope of the regression line $\beta = 1.0455$. The regression $R^2 = .3301$.*

(b) *Rogers. The slope of the regression line $\beta = 0.5603$. The regression $R^2 = .2150$.*

FIGURE 11: *Regressions of daily returns for wireless communications firms on TSE 300 index.*

Long term debt	\$2,305,638,000
Corporate tax rate	40%
Stock price (on April 18, 2002)	\$17.78
Number of shares	144,400,000

TABLE 10: *Rogers Wireless Telecommunications Inc. corporate data. Dollar figures are in Canadian funds. The data was obtained from* `http://www.globeinvestor.com/` *on April 18, 2002 and Rogers' quarterly financial reports.*

These two values are remarkably close. Averaging them, we estimate that the unlevered Bell Mobility β is given by $\beta^{\text{unlevered}}_{\text{Bell Mobility}} = .3714$. Note that our use of unlevered betas means that our real option valuation is biased low for an investment project financed with debt, as interest tax shields have not been accounted for. The value of these tax shields could be added later, if desired. However, it might be argued that, given the current financial situation in the telecommunications sector, new debt financing is unlikely to be available at present.

Now that we have estimated the unlevered beta for Bell Mobility, we can compute the market price of risk κ. As mentioned, we follow the methodology described in [15]. For readers unfamiliar with this approach, we provide a short and somewhat simplified description here.

Suppose there is a single stochastic factor X (in our context this is Q), which follows the risk-adjusted process:

$$dX = (\mu - \lambda)X dt + \sigma X dz, \tag{C.2}$$

where μ is the real world drift, σ is the volatility, λ is the risk premium, and dz is the increment of a Wiener process. Note that the risk premium $\lambda = \kappa\sigma$ (i.e. the market price of risk multiplied by the volatility). Let the firm's stock price be S. From Itô's lemma, we have:

$$\frac{dS}{S} = \frac{\left[\frac{1}{2}\sigma^2 X^2 S_{XX} + (\mu - \lambda) X S_X + S_t\right]}{S} dt + \frac{\sigma X S_X}{S} dz, \tag{C.3}$$

where the risk premium is:

$$\frac{\lambda X S_X}{S}. \tag{C.4}$$

The intertemporal capital asset pricing model (ICAPM) is then applied in the following way. The firm's stock β, denoted by β_S, depends on the covariance between returns on the market portfolio M and returns on the stock. This can be written as a function of the "β" of the stochastic factor X:

$$\beta_S = \frac{\sigma_{SM}}{\sigma_M^2} = \frac{XS_X}{S}\frac{\sigma_{XM}}{\sigma_M^2} = \frac{XS_X}{S}\beta_X, \tag{C.5}$$

where σ_{SM} is the covariance between changes in S and M and similarly for σ_{XM}. In the ICAPM, the expected return on the stock is:

$$r_S = r_f + \beta_S(r_M - r_f) = r_f + \frac{XS_X}{S}\beta_X(r_M - r_f) \tag{C.6}$$

where r_f denotes the risk free rate of interest and r_M is the expected return on the market portfolio. Equating the risk premium from (C.4) with that implied in (C.6) gives:

$$\begin{aligned}\frac{\lambda XS_X}{S} &= \frac{XS_X}{S}\beta_X(r_M - r_f) \\ &\Rightarrow \lambda = \beta_X(r_M - r_f).\end{aligned} \tag{C.7}$$

Using (C.5), we have:

$$\lambda = \frac{S\beta_S}{XS_X}(r_M - r_f), \tag{C.8}$$

i.e. the risk premium is a function of the expected excess market return, the firm's current stock price, the β of the firm's stock price, the current level of the stochastic factor X, and S_X.

Returning to our context, we then have:

$$\lambda = \kappa\sigma = \frac{S\beta^{unlevered}_{Bell\,Mobility}}{QS_Q}(r_M - r_f), \tag{C.9}$$

where S is Bell Canada Enterprises (BCE)'s current stock price, Q is the current level of traffic, and S_Q is the first derivative of the stock price with respect to the level of traffic.

All the parameters from equation (C.9) are known (or have been estimated previously) except S_Q, r_M and r_f. For the risk free rate r_f, we assume a value of $r_f = .04$. We assume that the expected market return r_M is 6% higher than the risk free rate (consistent with the average level for the past 50 years of Canadian data). Thus we have $r_M = .10$. S_Q is a more challenging parameter since there is no direct data from which we can determine it. Consequently, to estimate S_Q, we run linear regressions of BCE's stock price on the traffic data for the various switches. Our estimates of S_Q are the slope coefficients of these regressions. We obtain $S_Q = 1.8656 \times 10^{-4}$ for switch A, 9.0057×10^{-5} for switch B, and 8.4996×10^{-5} for switch C.

On August 2, 2002, BCE's stock price was \$39.08. Combining this information with our estimated values of S_Q and using $Q = 1.9197 \times 10^5$ (the level of bouncing busy hour traffic on August 2, 2002) and equation (C.9) gives

$$\lambda = \begin{cases} \frac{39.08 \times .3714}{1.9197\,10^5 \times 1.8656\,10^{-4}} \times (.1 - .04) = .0243 & \text{(Switch A)} \\ \frac{39.08 \times .3714}{1.9197\,10^5 \times 9.0057\,10^{-5}} \times (.1 - .04) = .0504 & \text{(Switch B)} \\ \frac{39.08 \times .3714}{1.9197\,10^5 \times 8.4996\,10^{-5}} \times (.1 - .04) = .0534 & \text{(Switch C)} \end{cases}$$

Hence, using the estimates of σ for each switch from Table 8 and the values of λ calculated above, equation (C.9) gives estimates of $\kappa \approx .05$ (based on switch A), $\approx .07$ (based on switch B), and $\approx .17$ (based on switch C). Averaging these gives a value of around .10.

One of the key assumptions above was that both Microcell Telecommunications and Rogers Wireless Communications face corporate tax rates of 40%. This is potentially problematic, especially for Microcell, which has been in financial difficulty and may not be likely to be in a tax paying position. Repeating our calculations, but assuming that each firm has a tax rate of 0%, we obtain $\kappa \approx .04$ (based on switch A), $\approx .05$ (based on switch B), and $\approx .13$ (based on switch C), with an average of about .075. Any other combination of tax rates for the two firms lying between 0% and the statutory rate of 40% gives rise to estimates of κ lying between the values calculated when both firms have rates of 0% and 40%.

D An Alternative Approach to Estimate the Market Price of Risk

In this appendix we present a more traditional approach to estimate the market price of risk. This is described in standard texts such as [7]. It is based on the correlation between market returns (i.e. TSE 300 index) and changes in the bouncing busy hour traffic levels. Using this approach, it is possible to estimate the market price of risk as

$$\kappa = \rho \kappa_M, \tag{D.1}$$

where ρ is the correlation between the bouncing busy hour changes and market returns and κ_M is the market price of risk for the stock market. Using the capital asset pricing model, the market price of risk is $\kappa_M = (r_M - r_f)/\sigma_M$, where r_M is the market return, r_f is the risk free rate and σ_M is the market volatility. We assume (as in Appendix C that $r_M - r_f = .06$. We calculate $\sigma_M = .2383$, using historical market data over the same time period as our network traffic data series. Over this period, we find estimates of ρ of .1367 for Switch A, .1171 for Switch B, and .1188 for Switch C. Using equation (D.1) we obtain estimates for κ of .0344 for Switch A, .0295 for Switch B, and .0299 for Switch C. Consequently, our estimates of κ range from a high value of $\approx .17$ (based on one particular switch in Appendix C) to a low of around .03 (based on all switches in Appendix D).