Wireless Communications*

by

Joshua S. Gans, Stephen P. King *and* Julian Wright**

Handbook of Telecommunications Economics, Volume 2 (forthcoming)

* We would like to thank Tomaso Duso, Natalie Lippey, Lars-Hendrik Röller, Aaron Schiff, and Tommaso Valletti for providing useful comments, and the Center for Research in Network Economics and Communications for funding research assistance. We also would like to than Aaron Schiff, Richard Hayes and Ryan Lampe for outstanding research assistance.

** University of Melbourne (Gans and King) and National University of Singapore (Wright). All correspondence to j.gans@unimelb.edu.au.

1 Introduction

In 1895, Guglielmo Marconi opened the way for modern wireless communications by transmitting the three-dot Morse code for the letter 'S' over a distance of three kilometers using electromagnetic waves. From this beginning, wireless communications has developed into a key element of modern society. From satellite transmission, radio and television broadcasting to the now ubiquitous mobile telephone, wireless communications has revolutionized the way societies function.

This chapter surveys the economics literature on wireless communications. Wireless communications and the economic goods and services that utilise it have some special characteristics that have motivated specialised studies. First, wireless communications relies on a scarce resource – namely, radio spectrum – the property rights for which were traditionally vested with the state. In order to foster the development of wireless communications (including telephony and broadcasting) those assets were privatised. Second, use of spectrum for wireless communications required the development of key complementary technologies; especially those that allowed higher frequencies to be utilised more efficiently. Finally, because of its special nature, the efficient use of spectrum required the coordinated development of standards. Those standards in turn played a critical role in the diffusion of technologies that relied on spectrum use.

In large part our chapter focuses on wireless telephony rather than broadcasting and other uses of spectrum (e.g., telemetry and biomedical services). Specifically, the economics literature on that industry has focused on factors driving the diffusion of wireless telecommunication technologies and on the nature of network pricing regulation and competition in the industry. By focusing on the economic literature, this chapter complements other surveys in this Handbook. Hausman (2002) focuses on technological and policy developments in mobile telephony rather than economic research per se. Cramton (2002) provides a survey of the theory and practice of spectrum auctions used for privatisation. Armstrong (2002a) and Noam (2002) consider general issues regarding network interconnection and access pricing while Woroch (2002) investigates the potential for wireless technologies as a substitute for local fixed line telephony. Finally, Liebowitz and Margolis (2002) provide a general survey of the economics literature on network effects. In contrast, we focus here solely on the economic literature on the mobile telephony industry.

The outline for this chapter is as follows. The next section provides background information regarding the adoption of wireless communication technologies. Section 3 then considers the economic issues associated with mobile telephony including spectrum allocation and standards. Section 4 surveys recent economic studies of the diffusion of mobile telephony. Finally, section 5 reviews issues of regulation and competition; in particular, the need for and principles behind access pricing for mobile phone networks.

2 Background

Marconi's pioneering work quickly led to variety of commercial and government (particularly military) developments and innovations. In the early 1900s, voice and then music was transmitted and modern radio was born. By 1920, commercial radio had been established with Detroit station WWJ and KDKA in Pittsburgh. Wireless telegraphy was first used by the British military in South Africa in 1900 during the Anglo-Boer war. The British navy used equipment supplied by Marconi to communicate between ships in Delagoa Bay. Shipping was a major early client for wireless telegraphy and wireless was standard for shipping by the time the Titanic issued its radio distress calls in 1912.[1]

Early on, it was quickly recognized that international coordination was required for wireless communication to be effective. This coordination involved two features. First, the potential for interference in radio transmissions meant that at least local coordination was needed to avoid the transmission of conflicting signals. Secondly, with spectrum to be used for international communications and areas such as maritime safety and navigation, coordination was necessary between countries to guarantee consistency in approach to these services. This drove government intervention to ensure the coordinated allocation of radio spectrum.

2.1 Spectrum Allocation

Radio transmission involves the use of part of the electromagnetic spectrum. Electromagnetic energy is transmitted in different frequencies and the properties of the energy depend on the frequency. For example, visible light has a frequency between 4×10^{14} and 7.5×10^{14} Hz.[2] Ultra violet radiation, X-rays and gamma rays have higher frequencies (or equivalently a shorter wave length) while infrared radiation, microwaves and radio waves have lower frequencies (longer wavelengths). The radio frequency spectrum involves electromagnetic radiation with frequencies between 3000 Hz and 300

---

[1] Numerous references provide a history of wireless communications. For a succinct overview, see Schiller (2000).

[2] One Hertz (Hz) equals one cycle per second.

GHz.[3]

Even within the radio spectrum, different frequencies have different properties. As Cave (2001) notes, the higher the frequency, the shorter the distance the signal will travel, but the greater the capacity of the signal to carry data. The tasks of internationally coordinating the use of radio spectrum, managing interference and setting global standards are undertaken by the International Telecommunication Union (ITU). The ITU was created by the International Telecommunications Convention in 1947 but has predecessors dating back to approximately 1865.[4] It is a specialist agency of the United Nations with over 180 members.

The Radiocommunication Sector of the ITU coordinates global spectrum use through the Radio Regulations. These regulations were first put in place at the 1906 Berlin International Radiotelegraph Conference. Allocation of the radio spectrum occurs along three dimensions – the frequency, the geographic location and the priority of the user with regards to interference. The radio spectrum is broken into eight frequency bands, ranging from Very Low Frequency (3 to 30 kHz) up to Extremely High Frequency (30 to 300 GHz). Geographically, the world is also divided into three regions. The ITU then allocates certain frequencies for specific uses on either a worldwide or a regional basis. Individual countries may then further allocate frequencies within the ITU international allocation. For example, in the United States, the Federal Communications Commission's (FCC's) table of frequency allocations is derived from both the international table of allocations and U.S. allocations. Users are broken in to primary and

---

[3] One GHz equals one billion cycles per second.

[4] See Productivity Commission (2002) and the ITU website at www.itu.int.

secondary services, with primary users protected from interference from secondary users but not vice versa.

As an example, in 2003, the band below 9 kHz was not allocated in the international or the U.S. table. 9 to 14 kHz was allocated to radio navigation in both tables and all international regions while 14 to 70 kHz is allocated with both maritime communications and fixed wireless communications as primary users. There is also an international time signal at 20kHz. But the U.S. table also adds an additional time frequency at 60 kHz. International regional distinctions begin to appear in the 70 to 90 kHz range with differences in use and priority between radio navigation, fixed, radiolocation and maritime mobile uses. These allocations continue right up to 300GHz, with frequencies above 300 GHz not allocated in the United States and those above 275 GHz not allocated in the international table.[5]

The ITU deals with interference by requiring member countries to follow notification and registration procedures whenever they plan to assign frequency to a particular use, such as a radio station or a new satellite.

2.2 The range of wireless services

Radio spectrum is used for a wide range of services. These can be broken into the following broad classes:

- *Broadcasting services*: including short wave, AM and FM radio as well as terrestrial television;

[5] For full details see *The FFC's on-line table of frequency allocations*, Federal Communications Commission Office of Engineering and Technology, revised as of January 17, 2003, http://www.fcc.gov/oet/spectrum/table/fcctable.pdf

- *Mobile communications of voice and data*: including maritime and aeronautical mobile for communications between ships, airplanes and land; land mobile for communications between a fixed base station and moving sites such as a taxi fleet and paging services, and mobile communications either between mobile users and a fixed network or between mobile users, such as mobile telephone services;
- *Fixed Services*: either point to point or point to multipoint services;
- *Satellite*: used for broadcasting, telecommunications and internet, particularly over long distances;
- *Amateur radio*; and
- *Other Uses*: including military, radio astronomy, meteorological and scientific uses.[6]

The amount of spectrum allocated to these different uses differs by country and frequency band. For example, in the U.K., 40% of the 88MHz to 1GHz band of frequencies are used for TV broadcasting, 22% for defense, 10% for GSM mobile and 1% for maritime communications. In contrast, none of the 1GHz to 3 GHz frequency range is used for television, 19% is allocated to GSM and third-generation mobile phones, 17% to defense and 23% for aeronautical radar.[7]

The number of different devices using wireless communications is rising rapidly. Sensors and embedded wireless controllers are increasingly used in a variety of appliances and applications. Personal digital assistants (PDAs) and mobile computers are regularly connected to e-mail and internet services through wireless communications, and wireless local area networks for computers are becoming common in public areas like airport lounges. However, by far the most important and dramatic change in the use of

[6] Adapted from Productivity Commission (2002)

[7] See Annex A from Cave (2001)

wireless communications in the past twenty years has been the rise of the mobile telephone.

2.3 The rise and rise of mobile telephony

The history of mobile telephones can be broken into four periods. The first (pre-cellular) period involved mobile telephones that exclusively used a frequency band in a particular area. These telephones had severe problems with congestion and call completion. If one customer was using a particular frequency in a geographic area, no other customer could make a call on that same frequency. Further, the number of frequencies allocated by the FCC in the U.S. to mobile telephone services was small, limiting the number of simultaneous calls. Similar systems, known as A-Netz and B-Netz were developed in Germany.

The introduction of cellular technology greatly expanded the efficiency of frequency use of mobile phones. Rather than exclusively allocating a band of frequency to one telephone call in a large geographic area, a cell telephone breaks down a geographic area into small areas or cells. Different users in different (non-adjacent) cells are able to use the same frequency for a call without interference.

First generation cellular mobile telephones developed around the world using different, incompatible analogue technologies. For example, in the 1980s in the U.S. there was the Advanced Mobile Phone System (AMPS), the U.K. had the Total Access Communications System (TACS), Germany developed C-Netz, while Scandinavia developed the Nordic Mobile Telephone (NMT) system. The result was a wide range of largely incompatible systems, particularly in Europe, although the single AMPS system was used throughout the U.S.

Second generation (2G) mobile telephones used digital technology. The adoption of second generation technology differed substantially between the United States and Europe and reverses the earlier analogue mobile experience. In Europe, a common standard was adopted, partly due to government intervention.

Groupe Speciale Mobile (GSM) was first developed in the 1980s and was the first 2G system. But it was only in 1990 that GSM was standardized (with the new name of Global System for Mobile communication) under the auspices of the European Technical Standards Institute. The standardized GSM could allow full international roaming, automatic location services, common encryption and relatively high quality audio. GSM is now the most widely used 2G system worldwide, in more than 130 countries, using the 900 MHz frequency range.

In contrast, a variety of incompatible 2G standards developed in the United States. These include TDMA, a close relative of GSM, and CDMA, referring to Time and Code Division Multiple Access respectively. These technologies differ in how they break down calls to allow for more efficient use of spectrum within a single cell. While there is some argument as to the ‘better’ system, the failure of the U.S. to adopt a common 2G standard, with the associated benefits in terms of roaming and switching of handsets, meant the first generation AMPS system remained the most popular mobile technology in the U.S. throughout the 1990s.

The final stage in the development of mobile telephones is the move to third generation (3G) technology. These systems will allow for significantly increased speeds of transmission and are particularly useful for data services. For example, 3G phones can more efficiently be used for e-mail services, and downloading content (such as music and videos) from the internet. They can also allow more rapid transmission of images, for example from camera phones.

An attempt to establish an international standard for 3G mobile is being moderated through the ITU, under the auspices of its IMT-2000 program. IMT-2000 determined that 3G technology should be based on CDMA systems but there are (at least two) alternative competing systems and IMT-2000 did not choose a single system but rather a suite of approaches. At the ITU's World Radiocommunication Conference in 2000, frequencies for IMT-2000 systems were allocated on a worldwide basis. By 2002, the only 3G system in operation was in Japan, although numerous companies have plans to roll out 3G systems in the next few years.

The growth in use of mobile telephones has been spectacular. From almost a zero base in the early 1980s, mobile penetration worldwide in 2002 is estimated at 15.57 mobile phones per 100 people worldwide. Of course, the level of penetration differs greatly between countries. In the United States, there were 44.2 mobile telephones per 100 inhabitants, with penetration rates of 60.53 in France, 68.29 in Germany, 77.84 in Finland and 78.28 in the United Kingdom. Thus, in general mobile penetration is lower in the U.S. than in the wealthier European countries. Outside Europe and the U.S., the penetration rate in Australia is 57.75, 62.13 in New Zealand, and 58.76 in Japan. Unsurprisingly, penetration rates depend on the level of economic development, so that India had only 0.63 mobile telephones per 100 inhabitants in 2002, with 1.60 for Kenya, 11.17 for China, and 29.95 for Malaysia. The number of mobile phones now exceeds the number of fixed-wire telephone lines in a variety of countries including Germany, France, the United Kingdom, Greece, Italy and Belgium. However, the reverse holds, with fixed-lines outnumbering mobiles in the United States, Canada, and Argentina. Penetration rates were close to equal in Japan in 2001, but in all countries, mobile penetration is rising much faster than fixed lines.[8]

The price for mobile phone services are difficult to compare between countries. In part this reflects exchange rate variations, but more importantly pricing packages and the form of pricing differs significantly between countries. Most obviously, different countries have different charging mechanisms, with ‘calling party pays’ dominating outside the United States. But in the United States and Canada ‘receiving party pays’ pricing often applies for calls to mobile telephones. Different packages and bundling of equipment and call charges also make comparisons difficult. A major innovation in mobile telephone pricing in the late 1990s was the use of pre-paid cards. This system, where customers pay in advance for mobile calls rather than being billed at a later date, has proved popular in many countries. For example, in Sweden, pre-paid cards gained 25% of the mobile market within two years of their introduction (OECD, 2000, p.11).

Despite the changing patterns of pricing, the OECD estimates that there was a 25% fall in the cost of a representative ‘bundle’ of mobile services over its member countries between 1992 and 1998 (OECD, 2000, p.22).

---

[8] All statistics from the ITU-D Country Database.

3 Economic Issues in Wireless Communications

3.1 Spectrum as a scarce resource

Radio spectrum is a natural resource, but one with rather unusual properties. As noted above, it is non-homogeneous, with different parts of the spectrum being best used for different purposes. It is finite in the sense that only part of the electromagnetic spectrum is suitable for wireless communications, although both the available frequencies and the carrying capacity of any transmission system depend on technology. The radio spectrum is non-depletable; using spectrum today does not reduce the amount available for use in the future. But it is non-storable.

Under ITU guidance, spectrum has been allocated to specific uses and then assigned to particular users given the relevant use. Traditionally, user assignment was by government fiat. Not infrequently, the user was government owned.

Privatizations in the 1980s and 1990s, and the success of (at least limited) mobile telephone competition in some countries, resulted in a more arms-length process of spectrum allocation developing in the 1990s. Users of radio spectrum, and particularly users of 2G and 3G mobile telephone spectrum, have generally been chosen by one of two broad approaches since the early 1990s – a 'beauty contest' or an auction.

A 'beauty contest' involves potential users submitting business plans to the government (or its appointed committee). The winners are then chosen from those firms submitting plans. There may be some payment to the government by the winners, although the potential user most willing to pay for the spectrum need not be among the winners. For example, the U.K. used a beauty contest approach to assign 2G mobile telephone licenses in the 1990s. Sweden and Spain have used beauty contests to assign

3G licenses.

France used a beauty contest to assign four 3G licenses. The national telecommunications regulator required firms to submit applications by the end of January 2001. These applications were then evaluated according to preset criteria and given a mark out of 500. Criteria included employment (worth up to 25 points), service offerings (up to 50 points) and speed of deployment (up to 100 points). Winning applicants faced a relatively high license fee set by the government. As a result, there were only two applicants. These firms received their licenses in June 2001, with the remaining two licenses unallocated (Penard, 2002).

The concept of using a market mechanism to assign property rights over spectrum and to deal with issues such as interference goes back to at least the 1950s when it was canvassed by Herzel (1951) and then by Coase (1959). But it was more than thirty years before spectrum auctions became common. New Zealand altered its laws to allow spectrum auctions in 1989 and in the early 1990s auctions were used to assign blocks of spectrum relating to mobile telephones, television, radio broadcasting and other smaller services to private management (Crandall, 1998). In August 1993, U.S. law was modified to allow the FCC to use auctions to assign radio spectrum licenses and by July 1996 the FCC had conducted seven auctions and assigned over 2,100 licenses (Moreton and Spiller, 1998). This included the assignment of two new 2G mobile telephone licenses in each region of the U.S. through two auctions.[9] In 2000, the U.K. auctioned off five 3G

[9] Formally, the FCC spectrum auctions were for personal communications services (PCS). In addition to 2G mobile telephones, relevant services included two-way paging, portable fax machines and wireless computer networks. See McAfee and McMillan (1996) for an early appraisal of these auctions.

licenses for a total payment of approximately $34b.[10]

Auctions have involved a variety of formats including ‘second price sealed bid’ in New Zealand, modified ascending bid in the U.S. and a mixed ascending bid and Dutch auction format in the U.K.[11] Bidders may have to satisfy certain criteria, such as service guarantees and participation deposits, before they can participate in the auctions. Limits may also be placed on the number of licenses a single firm can win in a particular geographic area, so that the auction does not create a monopoly supplier. From an economic perspective, using an auction to assign spectrum helps ensure that the spectrum goes to the highest value user.

While auctions have been used to assign spectrum to different users, they still involve a prior centralized allocation of bands of spectrum to particular uses. Economically, this can lead to an inefficient use of spectrum. A user of a particular frequency band (e.g. for 3G services) might have a much higher willingness-to-pay for neighboring spectrum than the current user of that neighboring spectrum (e.g. a broadcaster or the military). But the prior allocation of frequency bands means that these parties are unable to benefit from mutually advantageous trade. It would violate the existing license conditions to move spectrum allocated to one use into another use even if this is mutually advantageous. Building on the work of Coase (1959), Valletti (2001) proposes a system of tradable spectrum rights, using the market to both allocate spectrum to uses and simultaneously assign it to users. Interference can be dealt with through the

---

[10] See Binmore and Klemperer (2002) for a review of the U.K. 3G auction. For a survey of auctions in telecommunications see Cramton (2002).

[11] McMillan (1994) notes problems that arose in early spectrum auctions including the 1990 spectrum auctions in New Zealand and the 1993 satellite-television auctions in Australia.

assignment of property rights and negotiation between owners of neighboring spectrum. Valletti notes that both competition issues and issues of mandated standards would need to be addressed in a market for spectrum rights. We deal with the issue of standards later in this section while competition issues are considered in section 5 below.

Noam (1997) takes the concept of tradable spectrum assignment one stage further. Technological advancements, such as the ability for a signal to be broken into numerous separate digital packets for the purposes of transmission and then reassembled on reception, means that the concept of permanent spectrum assignment may become redundant in the near future. As technology advances, Noam argues, spot and forward markets can be used to assign use within designated bands of spectrum. The price of spectrum use would then alter to reflect congestion of use. DeVany (1998) also discusses market-based spectrum policies, including the potential for a future "open, commoditized, unbundled spectrum market system." (p.641)

Conflicts in the allocation of spectrum allocation arose in the FCC auctions in the U.S. The 1850-1910 MHz and 1930-1990MHz bands to be allocated by these auctions already had private fixed point-to-point users. The FCC ruled that existing users had a period of up to three years to negotiate alternative spectrum location and compensation with new users. If negotiations failed, the existing user could be involuntarily relocated. Cramton, Kwerel and Williams (1998) examine a variety of alternative 'property rights' regimes for negotiated reallocation of existing spectrum and conclude that the experience of the U.S. reallocations is roughly consistent with simple bargaining theory.

While economists have generally advocated the assignment of spectrum by auction, auctions are not without their critics. Binmore and Klemperer (2002) argue that a number of the arguments against auctions are misguided. But both Noam (1997) and Gruber (2001b) make the criticism that spectrum auctions automatically create a non-competitive oligopoly environment. Gruber argues that technological change has generally increased the efficiency of spectrum use and increased the viability of competition in wireless services. For example, in terms of spectral efficiency, GSM mobile telephone services are approximately four to thirty times more efficient than earlier analogue systems (Gruber, 2001b, Table 1). An auction of spectrum rights, however, is preceded by an allocation of spectrum. The government usually allocates a fixed band of spectrum to the relevant services. Further, the government usually decides on the number of licenses that it will auction within this band. So the price paid at the auction and the level of *ex post* competition in the relevant wireless services are determined by the amount of spectrum and the number of licenses the government initially allocates to the service. While the auction creates competition for the scarce spectrum, it does not allow the market to determine the optimal form of competition. Noam argues that flexibility of entry needs to be provided by the assignment system in order to overcome the artificial creation of a non-competitive market structure.

3.2 Complementarities in spectrum use

Using spectrum to produce wireless communications services can lead to synergies between services and between geographic regions. In the U.K., 3G spectrum auction, the potential synergies between 2G and 3G mobile telephone infrastructure was noted by Binmore and Klemperer:

> [T]he incumbents who are already operating in the 2G telecom industry enjoy a major advantage over potential new entrants … . Not only are the

> incumbents' 2G businesses complementary to 3G, but the costs of rolling out the infrastructure (radio masts and the like) necessary to operate a 3G industry are very substantially less than those of a new entrant, because they can piggyback on the 2G infrastructure. (2002, p.C80)

Thus, there are synergies in terms of being able to supply new products to an existing customer base using existing brands, and economies of scope between 2G and 3G services.

Geographic synergies are evident from the FCC 2G auctions. Moreton and Spiller (1998) examine the two 1995-96 mobile phone auctions in the U.S. They run a reduced-form regression on the winning bid for each license and a number of factors designed to capture the demographics of the relevant license area, the competitive and regulatory environment, and the effects of any synergies. These were ascending bid auctions so that the winning price is approximately equal to the second-to-last bidder's valuation for the license. As such, the relevant synergies relate to the network of the second-to-last bidder, to capture any effect of this network on the value of that bidder.

To capture the effect of geographic synergies, Moreton and Spiller assume that the expected network associated with any bidder is the same as the actual post-auction network. They categorize geographic synergies as either 'local' or 'global'. Local synergies consider the relationship between value of a license in one area and ownership of 2G licenses in neighboring geographic areas. Global synergies look at the total extent of the second-to-last bidder's national network.

Moreton and Spiller find strong evidence of local synergies. "At the local level, our results indicate that groups of two or more adjacent licenses were worth more to a single bidder than to separate bidders." (p.711)[12] These local synergies appear to fall rapidly as the geographic area covered by adjacent licenses increases and evidence of global synergies is weak.

Local coverage by existing cellular services tended to reduce the price paid for 2G licenses in the Moreton and Spiller study. This appears to run counter to the Binmore and Klemperer argument for economies of scope between different mobile telephone services. Moreton and Spiller argue that the negative relationship may reflect a reduction in competition. Firms are reluctant to bid strongly against existing analogue mobile telephone incumbents and prefer to use their limited resources elsewhere. This argument, however, is weak. In an ascending bid auction, participants will bid up to their own valuations and if there are positive synergies between existing analogue mobile services and 2G services, this should raise the value of the license to the second-to-last bidder regardless of any other parties bids.

As expected, Moreton and Spiller find that the value of a 2G license increases with market population and population growth rate and decreases with the size of the area served. These results are broadly consistent with Ausubel, et.al. (1997) and are intuitive. Population and demand are likely to be positively correlated so that for any given level of competition, increased population will tend to increase expected profits. But increased geographic region tends to raise the roll-out cost of the 2G cellular network for any population size, lowering expected profits.

The Moreton and Spiller study find some evidence that those jurisdictions where

[12] Ausubel et.al., (1997) also find evidence that local synergies were a significant determinant of price in the U.S. mobile telephone spectrum auctions.

regulators require tariff filing for the existing analogue mobile phone networks tend to have higher values for the 2G licenses. This suggests that tariff filing on existing services may have an anti-competitive effect leading to higher prices overall. The potential anti-competitive effects associated with regulatory notification and publication of price information has been shown in other industries (e.g. Albaek, et.al., 1997).

3.3 Standards

The adoption of standards has been a long-standing issue in wireless communications. The International spectrum allocation system is a form of standardization – ensuring that certain frequencies are used for certain purposes on a regional or world-wide basis. Standardization, often with government involvement, has also been a key factor in the success of new wireless technology.

For example, when television was first introduced in the U.S., there were a variety of competing potential technologies. In 1939, NBC began experimental television broadcasts in New York, but the FCC ordered it to cease broadcasting until the FCC had approved a standard. As a result, in 1940, the National Television Standards Committee (NTSC) was formed and their recommended standard was adopted by the FCC. The FCC mandated that television broadcasters must use this standard.[13] Governments in other countries also had a strong role in determining television standards, often to be broadcast by state-owned companies. For example, in the 1950s, Western European countries adopted alternative technologies that were incompatible with NTSC, known as PAL and SECAM.

---

[13] See Rohlfs (2001, ch.12) for background on the adoption of standards in U.S. television.

In contrast, AM radio in the U.S. did not involve government mandated standards. In fact, radio competition evolved before both the 1934 Communications Act and the formation of the FCC. A significant difference in mobile telephony between the U.S. and Europe relates to standards. For analogue cellular systems, Europe had a variety of incompatible systems while the AMPS system was ubiquitous in the U.S. But the reverse holds true for 2G mobile, with GSM being the government imposed standard in Europe but incompatible alternative systems being offered in the U.S.

There has been extensive analysis of the incentives for private firms to adopt standardized products or components. Economides (1989) considers both pricing and industry profits when firms can produce either compatible or incompatible component products. If components are compatible, then customers can assemble a system using different sellers' individual components. If the components are incompatible, however, a consumer can only buy a system by assembling components from the same seller. For a given number of firms, Economides shows that both prices and profits are higher when firms adopt compatible technology even in the absence of any direct consumer network externalities. This is due to the effect of compatibility on the intensity of competition. Given the prices set by other firms, a reduction in a firm's own price of one component tends to raise the demand for just this component when systems are compatible. But the same price decrease raises the demand for the firm's entire system of components when they are incompatible across firms. This leads to more intense price competition when components are incompatible.

For single products, in the absence of network externalities, differentiated consumer tastes mean that product variety is likely to be preferred. If there are direct network externalities, so that each consumer benefits when more consumers buy a product that is compatible with their own choice, standardization can create consumer benefits. Again, incompatibility tends to result in more intense competition, so that prices and profits are higher when a desirable standard is chosen.

As Shy (2001) notes, both the components and the network-externality models of compatibility show that choosing a standard tends to raise prices, lower consumer welfare and to raise social welfare. This does not, however, mean that private and social incentives for standardization are aligned. As Katz and Shapiro (1985) note, firms may have too little or too great an incentive for standardization from a social perspective. Further, even if private firms adopt a standard through market interaction, this need not be the most desirable standard from a social perspective (Economides, 1996).[14]

If competitive markets do not necessarily result in optimal standard choice, then there is a potential role for government. However, it cannot be assumed that government standard setting is necessarily superior to market-based processes. Governments may not set appropriate standards. This, in part, reflects the information limitations government faces, particularly when dealing with rapidly evolving technology. Standards are also often associated with proprietary intellectual property. This means that governments must deal with problems relating to 'essential' intellectual property when choosing a standard. For example, Bekkers, Verspagen and Smits (2002) consider the intellectual property conflicts that arose in the choice of the GSM standard in Europe. Further, being chosen as the appropriate standard can be profitable for the firm that owns and controls that

[14] See Economides (1996) for an extensive survey of the economics of compatibility choice in network industries. For more recent surveys, see Shy (2001) and Leibowitz and Margolis (2002).

standard, leading to the potential for corruption in the process of standard choice.

History shows that government choice of standard is not a panacea for market failure. As Rohlfs notes,

> [T]he history of television illustrates the wide range of possible outcomes from governmental standard setting – from effective action to utter failure. … The history of television illustrates many of the substantial defects that inhere in government decision-making; for example, the role of political influence/corruption and the pursuit of unworthy protectionist goals. Nevertheless, in bandwagon markets, government standard setting is sometimes superior to the alternative inefficient competitive process. (2001, pp.164-165)

While standards have been important in wireless communication, there has been little formal economic analysis focused specifically on wireless. Some recent papers, including Lehenkari and Miettinen (2002) and Haug (2002) focus on the history of technological adoption and the role of government for the NMTS and GSM standards.

Standardization is most important when there are strong network externalities. It can be argued that mobile telephony, unlike say television broadcasting, has few network externalities. Direct network externalities appear limited. Historically, mobile phone calls have tended to be to or from the fixed line network. Even for mobile to mobile calls, there is little need for compatibility as calls from say a CDMA phone can terminate on an AMPS phone through the fixed line network. The main network benefit for mobile phones relates to roaming – where the out-of-area coverage of a particular technological standard becomes important. Thus, while incompatible standards may lead to customer lock-in as moving service providers requires purchase of a new telephone, there are few network externalities that justify a government imposed standard. In such circumstances, it is most likely better to allow competing standards to "fight it out in the market" (Rohlfs, 2001, p.141. See also Shapiro and Varian, 1999, 264-267).

Koski and Kretscher (2002) empirically estimate the effects of standardization through two alternative approaches. For a series of 32 industrialized countries, they consider how a variety of factors, including standardization influence (a) the timing of the initial introduction of digital mobile telephony in a country and (b) the diffusion of digital mobile technology in the country. The initial adoption decision is modeled by considering the probability of adoption by a country after 1991 – the year before retail 2G services were first offered in Finland. In other words, the authors model the probability that a country will adopt digital mobile technology in any year *t* given that it has not previously introduced 2G technology. Diffusion is modeled through a standard S-shaped (logistic growth curve) process.

Koski and Kretscher find that standardization has a positive but insignificant effect on the timing of initial entry of 2G services. Thus, the adoption of a specific national standard for 2G is statistically unrelated to the timing of adoption of 2G by a country. This is not unexpected, even if there are network effects. While Koski and Kretscher argue that standardization should encourage early entry, by reducing technological uncertainty, when the standard is chosen by a government or regulator there is no reason to expect that the choice of standard will be timely.

Network effects are more likely to be important for the diffusion of mobile technology. Koski and Kretscher find that standardization has significantly facilitated the diffusion of 2G mobile technology. This result is broadly in line with Gruber and Harold (2001a) who analyze the diffusion of both analogue and digital mobile technology over 118 countries. Countries that had competing analogue systems had significantly slower rates of diffusion than single standard countries. They also find that standardization is associated with faster diffusion for digital technology, although this effect is both smaller and less precise than for analogue technology. Gruber and Harold (2001a) conclude that "the disadvantages of competing systems (network effects and scale economies) were dominant during the analogue era. During the digital era, the disadvantages may have been partly balanced by the advantages from technological systems competition." (p.1210)

A strong empirical prediction from the theoretical literature on standards is that standardization can lead to higher prices as it dampens competition. This prediction is consistent with Koski and Kretschers' findings. Their regressions show a significant positive relationship between standardization and price, and they conclude that "firms implement less aggressive pricing strategies when competition takes place within a single standard." (2002, p.26)

4 Diffusion and Demand for Mobile Telephony

4.1 Diffusion

While standardization appears to increase the rate of diffusion of mobile technology within a country, what else affects the rate of take up of mobile phones? Gruber and Verboven (2001b) estimate the diffusion of mobile technology over the fifteen states in the European Union using an S-shaped diffusion path.[15] They address a number of specific issues:

[15] For a non-technical overview of the diffusion of mobile telephones in the EU, see Gruber (1999).

(a) *Did the switch from analogue to digital systems increase the rate of diffusion?* Gruber and Verbovens' analysis suggests that the move to 2G systems in Europe led to a rapid increase in the diffusion of mobile technology. This diffusion effect was significantly greater than the acceleration effect built in to an S-shaped diffusion process. However, Gruber and Verboven do not isolate the cause of this acceleration. One explanation is standardization. 2G technology was introduced using the GSM standard across the EU, replacing a variety of incompatible analogue systems. Thus, the network effects associated with roaming may underlie the increased diffusion. Alternatively, as noted above, 2G mobile systems are significantly more efficient in terms of spectrum use than analogue systems. The capacity expansion associated with the introduction of 2G systems could lead to a drop in prices that encourages mobile take-up, regardless of the degree of competition.

(b) *Does increased competition increase the rate of diffusion?* Intuitively, increased competition (say from monopoly to duopoly) should lead to lower prices and increase the rate of diffusion of mobile technology. This intuition is verified by Gruber and Verboven. Moving from monopoly to duopoly increased the rate of diffusion although this effect was more important for analogue technology and was smaller than the effect on diffusion of the move to digital technology.

(c) *Do 'late' adopters catch up with 'early' adopters?* Gruber and Verboven show that convergence is occurring within the EU. Late adopting countries tend to have a rate of diffusion that is faster than early adopters, although the lead by early adopters is predicted to last for more than a decade on their analysis.

These empirical findings are broadly supported by other work. Koski and Kretschmer (2002) find that a wider diffusion of analogue mobile technology has a significant effect on the early take-up of 2G technology. This may reflect that spectrum capacity was being reached in those countries with greater analogue mobile phone use, and is consistent with the reduced capacity constraints of digital systems accelerating the rate of 2G diffusion.

Koski and Kretschmer (2002), Gruber and Verboven (2001a) and Gruber (2001a) all find that increased competition increases the rate of diffusion of 2G technology. Ahn and Lee (1999) estimate the demand for mobile connections using data from 64 countries. Consistent with the 'competition' hypothesis, they find that mobile demand is decreasing in price, although only the effect of monthly charges is statistically significant.

Koski and Kretschmer (2002) do not consider 'catch up' explicitly but do find that countries with a higher GDP per head adopted 2G technology earlier. This is consistent with Ahn and Lee (1999). Gruber and Verboven (2001a) also find that "countries with higher income per capita tend to be more advanced in adopting mobile phones." (p.1204) Further, they consider convergence in terms of the level of a countries development and show that convergence in diffusion of mobile technology is much faster in rich countries than in poor countries. Thus, a 'late starter' among poorer nations will take a lot longer to catch up with an early starter than would a comparable 'late starter' among rich countries.

4.2 The relationship between fixed and mobile telephony

Are fixed-line telephones and mobiles complements or substitutes in demand? Theoretically, the answer is ambiguous. To the extent that mobile telephones offer similar call functions to fixed-line telephones, we would expect there to be substitution in demand (Woroch, 2002). But mobile telephones are often used for short calls that would not be possible on a fixed-line telephone and such calls are often made to or from fixed-line telephones. Thus, the diffusion of mobile technology increases the benefits accruing to a fixed-line subscriber, potentially increasing demand for fixed-line services.

In the U.K., OFTEL concluded on the basis of qualitative survey evidence that fixed-line and mobile telephones, to a significant degree, are complements. "[T]he advent of the mobile has, to a significant degree, expanded the market for making calls, rather than substituting for fixed calls, implying that a large majority of mobile calls are complementary to fixed calls." (OFTEL, 2001, paragraph A1.14)

This conclusion is backed up by Gruber and Verboven (2001a). "[C]ountries with a large fixed network tend to be more advanced in adopting mobile phones." (pp.1024-5) While this effect is diminishing over time, Gruber and Verboven conclude that "the fixed network is largely viewed as a complement to mobile phones." Similarly, Anh and Lee (1999) find that "[t]he number of fixed lines per person … has a positive influence on the probability of mobile telephone subscription." (p.304)

A number of recent studies, however, suggest that mobile and fixed-line telephony are substitutes in demand. Horvath and Maldoom (2002) criticize the OFTEL conclusion as potentially confusing complementarity with individual tastes. If individuals who have a greater propensity to make telephone calls tend to have both mobile and fixed-line telephones then simply noting the correlation between ownership does not give any information about the degree of substitution between mobile and fixed-line services. They use survey data to try and correct for this taste effect, concluding that an individual's spending on fixed-line telephony decreases significantly when the individual also has a mobile telephone.

While Horvath and Maldoom consider call spending, other studies relate mobile and fixed-line penetration rates. Cadima and Barros (2000) find that the ability to access mobile telephony reduces demand for fixed-line services. The availability of mobile services leads to approximately a ten percent decrease in the fixed-wire telephony penetration rate in their study. Sung and Lee (2002) use Korean data and estimate that a 1% increase in the number of mobile telephones results in a reduction of 0.10-0.18% in new fixed-line connections and a 0.14-0.22% increase in fixed line disconnections.

The conflict between these empirical results may partially be explained by differences in countries and the life-cycle of mobile technology. As Gruber (2001a) and Gruber and Verboven (2001a) find increased waiting lists for fixed-line telephones has a positive effect on the diffusion of mobile telephones. Thus, in lesser developed countries we would expect access to a mobile phone to substitute for rationed access to fixed-line connections. In developed countries, initial use of mobile telephones is likely to involve mainly mobile-to-fixed or fixed-to-mobile calls. In such circumstances, call benefits can dominate leading to complementarity between fixed and mobile services. But as mobile penetration rises and mobile-to-mobile calls increase in importance, mobile phones may become substitutes for fixed-line services. This latter effect may be exacerbated as technology advances, both reducing the cost of mobile services and improving mobile functionality.

While there has been significant empirical work on the demand-side relationship between fixed and mobile services, there has been little work considering the supply-side. Koski and Kretschmer (2002) find that those countries with more competition in fixed-line telephone services introduced 2G mobile services earlier. They argue that this reflects substitution in supply between fixed and mobile telephones. Liberalization of fixed-line services tends to raise competition and lower profits in these services, making early entry into mobile look relatively more attractive for telecommunications firms.

4.3 Costs

A small number of papers have attempted to determine the presence of scale economies in mobile telephony. The results from these studies tend to be contradictory. While McKenzie and Small (1997) find that mobile telephony generally exhibits diseconomies of scale, and at best has constant returns to scale, Foreman and Beauvais (1999) find that mobile telephony exhibits modest increasing returns to scale.

The disagreement between such cost studies is unsurprising. In fixed-line local telephony, the presence or absence of increasing returns to scale has been subject to heated debate for much of the last thirty years (e.g. see Shin and Ying, 1992). In local telephony, the debate has immediate regulatory implications regarding the presence or absence of a natural monopoly in the local loop. In contrast, for 2G mobile, robust competition has been shown to be viable in many countries. Even if it is possible to find some degree of increasing returns to scale in mobile telephony, both the practicality and the benefits from mobile competition seem undeniable.

5 Regulation and Competition

This section surveys the literature dealing with possible anticompetitive behaviour in wireless markets and the associated regulation of wireless services. The focus is on issues that are of potential concern to policy makers – including potential barriers to wireless competition (arising from incomplete coverage, roaming, standards, and number portability), the possibility of tacit collusion, discriminatory on- and off-net pricing, and the regulation of access prices.

5.1 Limitations of wireless competition

Given the success of wireless services, at least in terms of the spectacular growth in cellular subscriptions (see section 2.3 for details), it is perhaps surprising that a main focus for the economics literature on wireless has been on the limitations of wireless competition. To some extent this probably reflects the historical structure of the industry, which given the allocation of scarce spectrum often only allowed two networks to exist, for example in the U.S. and U.K.. A natural question to ask is whether two networks are enough to ensure competitive conditions, a question we will address in section 5.1.1 which reviews the literature on tacit collusion in wireless markets.

Section 2 introduced several other properties of the wireless market that might cause policymakers to be concerned about the likely levels of competition and prices. These included sunk costs, high switching costs (lock-in occurs through long-term contracts, a lack of phone portability and a lack of number portability), differentiation in coverage, and inconsistent standards. In section 5.1.2 we review the competition and regulatory aspects of these issues. We also review the literature relating to pricing in wireless markets, dealing with the possibility of discriminatory on- and off-net pricing and the high price of international roaming.

5.1.1 Tacit collusion

A number of studies have suggested tacit collusion was present during the early period of wireless competition, when many wireless markets contained just two operators. For the U.K., Valletti and Cave (1998, pp. 115-116) point to tacit collusion to explain the seven years of stable and similar prices (1985-1991) when the market was a duopoly, followed by the sudden decrease in prices, the emergence of multiple tariff options, and the divergence of prices between operators following the entrance of two new competitors in 1993.

Stoetzer and Tewes (1996, pp. 305-307) suggest tacit collusion characterized the German market during the duopoly period up to 1994. In addition to the fact that there were only two operators during this period, entry was not possible (due to spectrum not being available to competitors), tariffs were relatively simple and easily observable by competitors, and the operators appeared to be quite similar. These factors all tend to make tacit collusion easier to sustain. For Germany, Stoetzer and Tewes point to stable prices but strong competition in the service dimension (such as geographic coverage and network specific value added features) as further evidence the two operators were colluding over prices.

Parker and Röller (1997) and Busse (2000) conduct formal econometric analyses of the tacit collusion hypothesis using U.S. data. In each sub national geographic market, the U.S. Federal Communications Commission (FCC) granted licences to two cellular operators, with one going to the existing local wireline operator. A duopolistic structure was thus established. Both Parker and Röller and Busse use panel data on the different geographic markets over the period from 1984 to 1988 to test the degree to which duopolistic competition is consistent with the observed market outcomes. Both find evidence of non-competitive prices consistent with tacit collusion.

Parker and Röller estimate a standard structural model of competition (see Bresnahan, 1989 for a survey of this type of approach), making use of conjectural variations to imbed different types of competitive behaviour. The relevant conduct parameter θ equals one for a monopoly, zero for perfect competition, one-half for non-cooperative duopoly (Cournot), and is greater than one-half for collusive duopoly. They allow this parameter to depend on a variety of market characteristics that might increase the likelihood of tacit collusion, such as the extent to which operators compete with each other in multiple markets (multimarket contact) and the extent to which competitors have cross-ownership in some other market. They also allow the parameter to depend on the extent of state regulation (regulators sometimes requested or required operators to file tariffs for informational purposes).

An appealing feature of the data on wireless competition for this kind of estimation is that regulators fixed the structure of the industry. After the licenses were awarded, there was an initial monopoly period, followed by duopoly until around 1996 when further entry occurred. The length of this initial monopoly period depended on technical constraints. The first operator to provide a cellular service knew they would face a competitor (given the duopoly policy) and so there was presumably no point trying to price low initially to deter entry. Once a duopoly was established, there was also no point for the firms to try to deter further entry, since such entry was restricted by the lack of additional licenses. Additional licenses were not awarded until 1996 or later, and according to Parker and Röller there was anecdotal evidence that wireless operators did not expect any entry over the period they study. This allows market behaviour under duopoly to be compared to behaviour under the initial monopoly without worrying about the possibility of entry deterrence behaviour in either case. It also allows the initial monopoly period to be used to test the appropriateness of the empirical specification.

Parker and Röller allow their conduct parameter θ to vary across the monopoly period and the duopoly period. They estimate θ to be 1.079 with a standard error of 0.17

in the former period, which means their model cannot reject monopoly behaviour during the period the market was known to be one of monopoly. In comparison, for the duopoly period they estimate θ to be 0.857 with a standard error of 0.042, meaning they can reject perfect competition, non-cooperative behaviour, and cartel (monopoly) behaviour. The estimate is consistent with collusive behaviour somewhere between that of Cournot and a monopoly.

The conduct parameter was found to depend positively and significantly on variables that are thought to make tacit collusion easier; namely, the extent of multimarket contact, the extent of cross-ownership in other markets, and the ability of operators to voluntarily report prices to a regulatory authority[16]. Overall their empirical evidence is consistent with the view that over the period from 1984 to 1988, wireless operators in the U.S. were engaging in tacit collusion to sustain high prices.

Busse (2000) is interested in how firms tacitly collude. She hypothesises that U.S. wireless operators tacitly colluded over the period from 1984 to 1988 by setting identical prices in multiple markets in which they serve. Identical pricing arises when a firm sets the same price schedule across different markets – rather than different firms setting the same price schedules within the same market. In her view, multimarket contact facilitated tacit collusion not only by enhancing the ability of firms to punish deviators, but also by increasing firms' scope for price signalling and coordination. Thus, multimarket contact alone may not be sufficient to enable wireless operators to tacitly collude. They may need a way to communicate their behaviour, which could be established by setting identical price schedules in multiple markets.

[16] Mandatory disclosure of prices did not have a significant effect on the conduct parameter.

Using a probit regression, she finds multimarket contact makes operators significantly more likely to set identical price schedules in the different markets they engage in, controlling for the similarity of the characteristics of the markets, the geographic proximity of the markets, and an indicator variable for each operator (to capture any idiosyncratic tendency of firms to price identically across the markets they serve). Second, using a panel regression of a firm's price on various demand characteristics, firm and time effects, and dummies for the use of identical pricing, she finds that when a wireless operator sets identical prices across markets, its price increases by 6.9% (with a t-statistic of 2.69).[17] Combined, this evidence is suggestive of identical pricing across markets being used as a way to facilitate tacit collusion.

The above analysis suggests it is the combination of multimarket contact and identical pricing that is responsible for higher prices. Given this, it is surprising that Busse does not interact these variables directly. Instead, she adds multimarket contact as an additional explanatory variable in the price regressions and finds it is not important after controlling for the identical pricing dummy variable. The finding that multimarket contact does not lead to higher prices even when the identical pricing dummy variable is dropped is inconsistent with the finding from Parker and Röller that multimarket contact leads to more collusive behaviour. Given both studies consider essentially the same sample, this suggests one of the specifications is misspecified. Interestingly, Busse also finds higher prices are associated with price-matching, in which multiple competitors set the same price schedules in a given market, suggesting an additional avenue for tacit

[17] The endogenous variable, price, is averaged over usage levels for each firm's price schedule. Similar results are obtained using a single point along each firm's price schedule, say corresponding to usage of 500 minutes per month.

collusion.

Put together, the above studies are certainly indicative of tacit collusion during the early period of duopoly in wireless markets. The extent to which this estimated collusion has persisted in the face of more recent entry is something worthy of further empirical investigation. The anecdotal evidence from Valletti and Cave (1998) and Stoetzer and Tewes (1996) is that competition has indeed become stronger following entry in the U.K. and Germany. It would be interesting to use U.S. data to see whether any breakdown of tacit collusion could be linked to specific market features, such as the process of entry, the number of new entrants, de-regulation, or a breakdown in identical pricing. It would also be interesting to compare the estimated conduct parameter (a'la Parker and Röller) in the post-entry period with the pre-entry period.

The finding of tacit collusion raises issues of whether regulation can be used to limit tacit collusion, or whether regulation is in fact a source of tacit collusion. The result of Parker and Röller that tacit collusion is stronger in markets where operators can voluntarily report prices to a regulatory authority but not in markets where operators must report prices is consistent with the view that regulation that allows for voluntary price reporting promotes tacit collusion and higher prices. Hausman (1995) and Hausman (2002) makes the case that cellular prices are higher because of regulation.[18] Hausman uses data from the 26 states in the U.S. that had voluntary or mandatory disclosure of mobile prices (regulated states), and compares it to data from the other 25 states which did not. Based on instrumental variables estimation in which "regulation" is treated as a

---

[18] Shew (1994) also argues that regulation is responsible for higher cellular prices. Ruiz (1995) found that the regulatory variables did not significantly explain prices, and concluded that the analysis did not imply any policy suggestions.

jointly endogenous variable (along with price), he finds that regulated states have prices that are 15 percent higher, holding other variables constant (population, average commuting time, average income). In 1996, after such regulation was eliminated due to an act of Congress, prices in the previously regulated states were not significantly different from the non-regulated states.

Duso (2003) takes a different approach, explicitly accounting for the endogeneity of regulation. He estimates an equation for the determinants of regulation. This allows for reverse causality in which regulation can be less likely when it would be most effective in reducing prices, reflecting the effect of lobbying activity on the regulatory choice. Controlling for this endogeneity of regulation, Duso finds that prices in regulated markets are lower than the prices cellular operators would have set had these markets not been regulated, although this effect is not statistically significant. Duso also finds that regulation would have resulted in a significant decrease in prices had the non-regulated markets been regulated. The latter result reflects the success of lobbying activity against regulation that would have reduced prices a lot.

Another paper that accounts for the endogeneity of regulation (in fact of deregulation) is that of Duso and Röller (2003). They consider the impact that deregulation (that is, entry) has had on productivity in the mobile phone sector across OECD countries. They again test for reverse causality, in this case whether countries that are more productive are more likely to deregulate. They estimate two-equations (a policy equation and a market equation). The first is the effect of deregulation on productivity. The second is the effect of productivity on the decision to deregulate. They find that deregulation does lead to higher productivity, but the impact is about 40% smaller once they control for reverse causality (the second equation).

The unique features of the regulatory environment for wireless markets make it ideal for testing theories of industrial organisation and regulation. We just briefly mention some other recent work that utilizes these unique features. Duso and Jung (2003) provide an empirical investigation of the relationship between firms' lobbying expenditures and product market collusion using data from the U.S. cellular industry between 1984 and 1988. They find a significant negative two-way relationship between the strength of collusion in the product market and firms' lobbying expenditures. Collusive conduct decreases political activities, while higher contributions increase competition in the product market. Reiffen, et.al. (2000), use data from the cellular industry in the U.S. to study the possibility that wireless operators that are also local exchange carriers supplying interconnection to rival cellular operators may discriminate against their rivals.[19] The exogenous geographic differences in the carriers' ability and incentive to discriminate in the U.S. make the industry ideal for such a study. Miravete and Röller (2002) also use historical data on wireless markets in the U.S. to estimate an equilibrium oligopoly model of horizontal product differentiation when firms compete in nonlinear tariffs.

5.1.2 Strategic instruments to limit competition

Aside from the empirical studies of tacit collusion and regulation, most of the literature on competition in wireless markets is based on descriptive accounts of competition, with a focus on the barriers to effective competition. For instance, Valletti and Cave (1998) provide a detailed discussion of competition in the U.K. wireless market

[19] Reiffen and Ward (2002) provide a survey of this line of work.

up to 1997. In addition to the possibility of tacit collusion in the pre-entry period, Valletti and Cave also discuss the various strategic instruments that wireless firms may have used to raise switching costs and enhance network effects, so as bind consumers to their networks. These include a lack of number portability, the use of SIM card locking, incompatible standards, and network-based price discrimination. The possibility of limiting competition through differential coverage is raised by Valletti (1999 and 2002), while roaming policy has been the subject of recent regulatory interest. These themes are discussed here in relation to the available literature.

(a) Number portability

Number portability ensures wireless subscribers can maintain the same phone number when switching to a rival's network. Since consumers generally value maintaining the same phone number, number portability reduces switching costs. What are the implications of reduced switching costs for wireless competition and prices?

Farrell and Klemperer (2002) provide a survey of the switching cost literature, arguing in general that the lock-in associated with switching costs has an ambiguous effect on competition and prices. Generally, it leads to less aggressive *ex-post* competition (once customers are locked-in), but it also leads to more aggressive *ex-ante* competition (to attract market share in the first place). On balance, the present value of current and future prices and profits can be higher or lower as a result. Switching costs suggest a particular pattern of prices, a "bargain-then-rip off" structure. This approach predicts that as wireless ownership matures, the percentage of locked-in consumers will increase and prices should rise. However, this does not necessarily imply wireless firms are earning supranormal profits. Through their *ex-ante* competition for customers (for the market), they may have already competed away these future *ex-post* rents.[20] This is less likely to be true if for the period over which operators were building market share they engaged in tacit collusion.

While there is no general and unambiguous result on the effects of switching costs, there seems to be some consensus that an increase in switching costs increases industry profits, especially where firms try to sustain or artificially create switching costs (Farrell and Klemperer, 2002). If this is true, then the wireless industry should be opposed to the introduction of number portability, which indeed seems to be the case.[21] There is an offsetting efficiency effect that could also explain why operators are opposed to number portability. If number portability is costly, but the operators bearing the cost are unable to fully pass on this cost to the consumers who wish to port their number, then there may be excessive use of number portability (with switching), resulting in higher costs to operators. Number portability could also have differential effects on new entrants versus established incumbents, also explaining why it might be difficult to get industry agreement in support of it.

Aoki and Small (2001) construct a model of number portability in telecommunications between two suppliers of differentiated products, each of which sets a two-part tariff. Number portability is modelled as involving a reduction in switching

---

[20] Although policy intervention to control *ex-post* rents of any incumbent that exploits locked-in consumers amounts to expropriation of the incumbents' *ex-ante* investments in attracting customers, if incumbents expect such a policy the result could be a more efficient time path of prices, avoiding the usual "bargains-then-rip off" pattern.

[21] Perhaps because of this, number portability has been mandated in the telecommunications sector in Australia, Denmark, Hong Kong, the U.K. and the U.S., among other countries. In the EU there is a directive that states that all EU countries have to supply mobile number portability.

costs and an increase in the marginal cost of making calls (as additional routing-related tasks must be performed to establish connections). The model of competition is one in which lower switching costs result in lower fixed fees set by carriers, but in which the increased marginal cost of call production raises usage fees. The implications of number portability for consumer or social surplus in this model are ambiguous, a finding that contrasts with the normal regulatory presumption in favour of number portability. Their modelling suggests empirical work is needed to measure the two opposing effects, and to quantify the consumer and welfare effects of number portability.

In Aoki and Small, the consumer and social benefits depend on the market structure (symmetric or asymmetric) and who (incumbent or entrant) bears the cost of porting consumers. They consider several different market configurations with the most relevant to wireless being a structure in which not all consumers buy initially from the firms (infant industry). In this model there are two periods. In the first period there is a monopoly. In the second period some new consumers enter, and an entrant competes with the incumbent. Unlike the first period customers, the new customers can choose between two firms without incurring a switching cost.[22] Aoki and Small conclude "In this case, under plausible assumptions, reductions in the cost of switching benefit consumers and the entrant, and have no effect on the incumbent. The consumer gains come entirely from expansion of the market."

Gans and King (2001b) also model number portability as a switching cost-reducing device. They focus on the case in which an established incumbent, which

---

[22] However, as with the Gans and King paper discussed below, this is still a static model of competition in that firms only compete in one period. As such, there is no *ex-ante* competition for the market, and so the normal trade-off between *ex-ante* competition for the market and *ex-post* competition in the market does not apply. This explains why both papers have the property that higher switching costs imply higher prices.

already has all of the potential customers, faces a new entrant. Unlike Aoki and Small, they allow the consumer to be charged for porting, which reduces any excessive usage of porting by consumers. However, this (along with the case in which entrant pays for porting) raises the possibility that the incumbent will then inflate the cost of providing number portability, so as to reduce switching. In the face of this effect, a superior regulatory solution may be to provide the customer with the opportunity to own their own phone number, and the carriers the option of buying this back from customers. Provided Coase-type bargains can be reached, this should ensure number portability is chosen only if it is efficient, and that the incumbent provides number portability in the most cost effective way.[23]

The above models are orientated towards addressing number portability for the wireline sector, between an established incumbent and a new entrant, in a mature market. A more interesting benchmark for wireless is one in which multiple (possibly symmetric) carriers are already established, and in which there are new customers to be attracted. In this case, switching costs may actually encourage entry, as a large customer base encourages incumbents to harvest its base relative to winning new customers, thus letting smaller firms catch up (Farrell and Klemperer, 2002). On the other hand, switching costs will tend to deter entry if entrants have to attract existing customers to cover their fixed costs, which could be the case when an expensive block of spectrum has to be purchased to enter in the first place.

Finally, it is worth noting that even if number portability is mandated, this does

---

[23] In practice, as Gans and King (2001b) point out, asymmetric information will likely prevent this first best outcome being achieved. See also the discussion in Gans, King and Woodbridge (2001).

not mean it will be effective. Networks may find ways to limit number porting, through poor advice about the possibility of number porting, or by making porting troublesome for consumers (Oftel, 2001, Section 2.2.3).

(b) SIM card locking

SIM-locking stops mobile handsets obtained through one operator from being used to get a rival operator’s service. For a fee (currently around twenty pounds in the U.K.), handsets can be unlocked to be used on a rival’s network. Such fees are a way to tie together the ownership of a handset to the subscription to a wireless service. As such, their main impact exists when consumers do not already have contracts involving the subsidised purchase of a handset and an ongoing subscription service. Even in such cases, SIM-locking is only likely to have a limited impact, both since the fee is not huge, and since many consumers will wish to update their handsets anyway.

Unlike number portability, the decision to maintain SIM-locking rests with each individual carrier. In the U.K., fees for unlocking SIM cards have remained despite increased competition and despite regulatory investigations into the fees. Oftel (2002a) summarises the arguments that have been put forward for and against the fees.

Justifications for the fees include that without SIM locking there is a risk of a ‘chicken and egg’ situation, whereby providers would not develop services without being sure of having enough available handsets with the functionality to deliver them, but manufacturers would not add this functionality without more certainty that it would be used, and that SIM locking enables operates to provide a handset subsidy which encourages more participation in the market (to the benefit of existing users). On the other hand, Oftel argues SIM locking raises switching costs for consumers,[24] prevents competition for services over the same handset (where consumers have multiple SIM cards), limits consumers choosing separately their preferred handsets and preferred service provider, and prevents suppliers providing just handsets or just service provision without providing the other.

Oftel's current response to these fees is to make consumers more aware of them in the first place. Although this will not reduce the switching costs implied by these fees, it may make consumers more likely to take the fees into account in deciding which network to join in the first place, thereby increasing the demand for operators that set lower fees for unlocking SIM cards.

(c) Standards and competition

The adoption of incompatible standards in wireless telephony increases switching costs, and has an effect similar to SIM locking. For instance, in New Zealand the two wireless networks (Telecom and Vodafone) operate on different technologies so that to switch between them, consumers need to purchase a different handset. To the extent that consumers expect to update their handsets fairly regularly, this may not be a large barrier to switching operators. Moreover, with a well-developed second hand market for handsets, the switching cost implied by incompatible standards is further reduced (to the transaction costs of trading in the second hand market).

[24] This increase in switching costs may be offset by handset subsidies. Oftel (2001, Section 2.3.1) notes, "… it is possible that the current levels of switching are partly supported by handset subsidies because these subsidies lower the cost of switching when handsets are SIM locked."

(d) Network-based price discrimination

Historically, the number of mobile-to-mobile calls has not been very significant, at least relative to the number of mobile-to-fixed calls.[25] As the penetration of wireless subscribers has increased, so has the proportion of mobile-to-mobile calls. This makes the use of discriminatory tariffs between on- and off-net calls more relevant.[26]

In a model with two symmetric but horizontally differentiated telecommunications firms, Laffont et.al. (1998) show that if firms are allowed to price discriminate, they will set their on- and off-net usage prices at cost and recover all profits through their fixed charges (monthly rentals). This implies that the incentive for network-based price discrimination (along the lines above) only arises to the extent the cost of terminating calls on the rival wireless network is higher than the cost of terminating calls on one's own wireless network. The main reason for any difference in the on-net versus off-net cost of terminating calls is if the interconnection charge agreed for mobile-to-mobile calls is set above or below cost. Using the Laffont et.al. (1998) framework, Gans and King (2001a) show that the operators will want to set this interconnection price below cost to soften competition.[27] This implies that wireless firms should charge consumers *less* for calls made to rival networks than to their own network. Clearly, this is not the case in practice. If there is any price discrimination, it tends to be to price off-net

---

[25] According to Valletti (1999), in the U.K. in mid 1997 around 95% of wireless calls terminate on wireline networks.

[26] By 2000/2001 off-net mobile-to-mobile calls in the U.K. accounted for 13% of all revenue generated (Oftel, 2001, Section 1.6).

[27] Mobile-to-mobile interconnection is also discussed in Section 5.2 below.

calls more than on-net calls.[28] This suggests there is some other reason for network-based price discrimination.

Berger (2002) provides a model that can explain why operators may want to set off-net retail prices higher than on-net prices. His main idea is that when consumers care about receiving calls (this would seem to be quite important for wireless), firms will want to make it expensive to call the rival network, so as to lower the demand for subscription to the rival network. This effect makes networks set higher off-net prices than otherwise would be the case, since by doing so they obtain the added benefit of making their rival less attractive to subscribe to. They then compete by setting lower on-net prices (in his model firms set only usage fees). Like Gans and King, in Berger's model, firms may still want to set the interconnection price below cost, although in his model this helps offset the inflated off-net prices and so is generally a good thing.[29]

(e) Differential coverage

An important attribute of wireless is its coverage. Over what geographic area (or percentage of the population) does the mobile phone work? A network that allows consumers to use their phone over a wider area is, other things equal, more valuable. Full coverage can be viewed as the wireless equivalent of universal service, in that it requires more expensive locations to be provided for, although not necessarily at the same price. This could serve some social (or political) objectives, such as ensuring people living in remote areas can contact and be contacted by others on the network, and ensuring that

[28] In Oftel's review of the U.K. wireless sector, it noted the relatively high price of off-net mobile-to-mobile calls as an area of concern (Oftel, 2001, Section 2.6.1).

[29] Section 5.2 provides a different reason why mobile firms may set high off-net prices, which reflects inflated mobile-to-mobile termination charges.

people can always remain in contact even when they travel into more remote areas.

Valletti (1999) emphasizes the role of differential coverage in limiting wireless competition. He argues that operators may vertically differentiate themselves, opting either for full coverage or for minimal coverage (where the minimum allowable coverage is determined by regulation).

In Valletti's model, consumers differ in their willingness to pay for calls (income) but are otherwise homogenous with respect to their need for coverage. Firms play a two-stage competition game, choosing coverage in the first period and price in the second period. As a result of the vertical differentiation structure, in equilibrium firms differentiate themselves in terms of coverage, with one firm opting for maximal coverage and the rival firm opting for less coverage (minimum coverage if the regulated minimum coverage level is binding). The analysis suggests if regulators ensure a high minimum coverage of all operators, they will reduce vertical differentiation, resulting in more intense price competition.[30]

A classic result in the literature on vertical differentiation is that even if incumbent operators earn positive economic profits, no matter how low are investment costs, there may be no further entry (Sutton, 1991). Valletti (2003) applies this result in the wireless context, showing in a model in which consumers move randomly between city and rural areas, and in which wireless operators choose their coverage in the rural area (as well as whether to enter the city), that there will be a 'natural oligopoly' equilibrium. Despite the existence of positive economic profits, there will be no further

[30] Minimum coverage conditions were required by Oftel in the U.K. Note however, minimum coverage conditions may be easily met if a few cells in large population centers enable carriers to say they cover a large proportion of the population.

entry. Operators have a strong incentive to use differing levels of coverage in order to vertically differentiate and relax price competition.

Both of Valletti's models assume away horizontal differentiation. With both horizontal and vertical differentiation, firms may no longer want to differentiate vertically, and instead may tend to set similar coverage levels.[31] This seems to be the norm in practice, with most providers attempting to provide full coverage. Valletti argues that as more spectrum is released, price competition will become stronger, and the incentives to vertically differentiate will become greater. In this regard, evidence from the different geographic regions of the U.S. should be telling.

(f) Roaming

Roaming arises when a network has incomplete coverage. If a geographic area is not covered by a subscriber's own network then the wireless subscriber may still be able to make or receive calls in the area if it is able to 'roam' on another network's infrastructure. It is useful to distinguish 'domestic' roaming from 'foreign' roaming. Domestic roaming happens when an operator wants roaming rights on a *rival's* network. Foreign roaming happens when an operator wants roaming rights on another network that does not (normally) compete for the same customers. In most countries, wireless firms operate at a national level so domestic roaming is equivalent to national roaming and foreign roaming is equivalent to international roaming. In the U.S., these need not be the same, as some firms operate across many regions and others in only specific regions. Likewise, with the development of some multimarket mobile operators, international

[31] See for instance, Dos Santos Ferreira and Thisse (1996). Similarly, if firms tacitly collude then they may not want to vertically differentiate.

roaming may be achieved within the same network (or alliance of networks).

Where firms do not compete against one another, as is frequently the case when operators are in different countries, the networks are complements to each other, and they both can benefit from roaming agreements. Consistent with this, Valletti (2003) notes that international roaming is far more prevalent than national roaming. However, international roaming is often alleged to be excessively expensive (Sutherland, 2001). Surveys conducted in 1999 indicate roaming charges are between two and ten times the price of the same or a similar domestic call. The difference in price between roamed and non-roamed international mobile calls to the same destination within the EU can be up to 500%. Oftel (2001, Section 2.70) and Sutherland suggest a lack of customer information may explain why retail prices for international roaming have remained so high. While plausible, this does not explain high wholesale charges. Typically, the operator providing the roaming service charges a wholesale fee based on its normal retail prices plus a margin. The home operator than adds a mark-up to this, often between 10% and 25%. The result appears to represent a case of double marginalisation, which would be neither in the operators' or the consumers' interests.

One reason such high roaming rates could be in the operators' interests is if the demand for roaming is very price inelastic (compared to regular calls). Given the common costs of a mobile network, Ramsey style pricing could be used to justify setting high rates for providing international roaming services (both wholesale and retail), although not necessarily to the levels chosen by the operators. With high wholesale charges for roaming, one might also expect that each foreign operator will undercut its rival (at the wholesale level) to be the network of choice for the home networks. Perhaps collusion at the wholesale level (through the negotiation of roaming agreements) can prevent such undercutting. High retail mark-ups are then sustained by the lack of customer information.

When operators compete for the same customer base, networks are substitutes and denying competitors domestic roaming is sometimes argued by policymakers as a way to make it more difficult for small entrants to compete. Entrants have to provide full coverage to compete on equal terms with incumbent operators, and it may be difficult to attract customers until they have comprehensive networks. This tends to raise the stakes of entry, although it is not so clear how it can prevent entry. In fact, it seems quite possible that if two partial coverage networks agree on high roaming fees, retail price competition could be weakened.

Roaming is also likely to affect investment incentives. On the face of it, (cheap) roaming would seem to weaken the incentive of firms to invest in infrastructure (full coverage), since using a rival's network may be more profitable. A free-riding problem could result in insufficient investment. However, limiting their own incentive to invest could be good for wireless operators, if one models them as competing in a Cournot fashion. This is the insight Foros et.al. (2002) develop in a model of investment and roaming in a wireless market.

In particular, they explore the implications of roaming policy for the upgrade of networks from 2G to 3G. The framework used is that of a three-stage game with two firms, where in the first stage roaming quality is chosen, in the second stage firms choose how much to invest, and in the final stage they compete in quantities. An extended version of the model also incorporates a third firm, a virtual operator that competes in the third stage of the game.

If wireless firms are not allowed to cooperate at the investment stage, they will choose too much roaming quality in the first stage. To understand this result, note that under roaming, each firm’s investment increases the demand for the other firms’ networks. By choosing a high roaming quality, the effect of each firm’s investment on the demand faced by rival firms is also high. It is assumed this spillover cannot be internalised through a pricing mechanism. With large spillovers under non-cooperative investment, each firm will invest too little. This has the effect of lowering output and raising prices in the third stage.

When a virtual operator is allowed to enter in stage 3, the incentives for roaming change. The two physical operators may now want to choose a high level of roaming between themselves and a low level with the virtual operator, so as to block the third firm from sharing the market. Such blocking of the third firm may be socially desirable, since with a high level of roaming the virtual firm may induce the incumbent firms to invest less at the second stage.

Some countries, such as the U.K., have mandated roaming during the period over which competitors are rolling out their network. The above analysis suggests it could be important to take into account the investment incentives such a policy may have, as well as any direct effects roaming has on retail competition. Mandated roaming may result in entrant networks choosing limited coverage which tends to perpetuate the regulation of roaming and delay full-blown infrastructure competition. Hausman (2002) also emphasizes the negative implications for investment of giving competitors a “free option” to roam existing network.

5.2 Access pricing and caller vs. receiver pays

Broadly speaking, there are three different access prices that are of interest in wireless markets.

1. A call from a mobile operator to a fixed-line network (a mobile-to-fixed call) involves the mobile operator collecting all the revenue for the call, but the fixed-line network providing the termination of the call. As a result, the fixed-line network usually charges for the termination of mobile-to-fixed calls. Typically, the termination of mobile-to-fixed calls is treated in the same way as the termination of other calls to fixed-line networks, being regulated on a cost-basis.

2. A call from a fixed-line network to a mobile operator (a fixed-to-mobile call) often leads to an access price being set by the mobile operator. In the majority of markets, the termination of fixed-to-mobile calls has been priced many times higher than the termination of mobile-to-fixed calls, sometimes resulting in the regulation of fixed-to-mobile termination prices.[32] An exception is the U.S., where this access price is the same as the mobile-to-fixed access price due to the symmetry requirements under the 1996 *Telecommunications Act*. Other exceptions are Canada, Hong Kong and Singapore, where there are no fixed-to-mobile termination charges. Instead, like the U.S., mobile operators recover the cost of terminating fixed-to-mobile calls from the mobile callers directly (receiver pays).

3. A call from a mobile operator to another mobile operator (a mobile-to-mobile call) may involve mobile operators charging each other for termination. Such access prices are often negotiated privately between mobile operators.

In this section we will review the recent literature that has developed on fixed-to-mobile termination, dealing with the related issues of caller versus receiver pays and mobile-to-mobile interconnection as we go.

The first paper to model the optimal pricing of fixed-to-mobile termination was that of Armstrong (1997). Armstrong noted the high price of fixed-to-mobile calls in the

---

[32] Mobile termination charges are regulated in the U.K using a price-cap approach, with the cap reducing towards estimates of long-run incremental cost over four years. In Australia the regulator ties year-on-year changes in mobile termination charges to changes in retail prices of mobile services.

U.K. In July 1996 the average per-minute retail charge for fixed-to-mobile calls was around 25 pence. In his view, this high price reflected two distortions – an inflated fixed-to-mobile termination charge (16.5 pence per-minute), and a margin that fixed-line operators put on fixed-to-mobile calls over and above the high termination charge (8.5 pence per-minute).[33] Armstrong provides a simple model of fixed-to-mobile calls which deals with the optimal pricing of the two components.

Armstrong assumes mobile operators are perfectly competitive and symmetric. There are a fixed number of subscribers that the mobile operators compete for. Each mobile operator is assumed to charge the same amount for terminating fixed-to-mobile calls, and earns a margin equal to this termination charge less its costs of terminating calls.[34] Competition drives mobile operators to return this margin to its subscribers, which takes the form of a subsidy to consumers to connect to the network. Armstrong argues that high fixed-to-mobile termination charges can explain the handset subsidies so often observed in the wireless market.

Assuming fixed-line operators obtain no margin on fixed-to-mobile calls, Armstrong shows the socially optimal fixed-to-mobile termination charge is equal to cost. This ensures fixed-to-mobile calls are priced at cost, which is the only source of a possible inefficiency in the model. Such cost-based termination of fixed-to-mobile calls eliminates any subsidy to mobile subscribers.

Armstrong considers how optimal termination charges change when three

---

[33] Even back in 1997, these issues had already been dealt with by regulation in the U.S. The 1996 *Telecommunications Act* already required that fixed-to-mobile termination be priced at the same level as mobile-to-fixed termination, and fixed-to-mobile retail prices be regulated on the same basis as other outgoing fixed-line calls.

[34] In his model, mobile-to-mobile calls and mobile-to-fixed calls do not play any role in the analysis.

important assumptions are relaxed. The optimal termination charge is lower if mobile subscribers get utility from receiving calls, since then mobile subscribers benefit from lower fixed-to-mobile prices.[35] The optimal termination charge is also lower if the price of fixed-to-mobile calls involves a margin above the termination cost, as might be expected with a dominant fixed-line network that is free to set the fixed-to-mobile price. A lower termination charge can help offset the mark-up put on fixed-to-mobile calls by a dominant fixed-line operator. Note this "subsidising a monopolist" approach is also applicable to mobile carriers to the extent competition between them is imperfect, suggesting it could also be used to justify higher termination charges. Finally, Armstrong notes that the optimal termination charge is higher to the extent a subsidy to mobile subscribers encourages more consumers to subscribe. An increase in mobile subscribers provides a benefit to fixed-line consumers who now can make more fixed-to-mobile calls, which increases the social benefits of above-cost termination charges.

An implicit assumption in Armstrong's model is that fixed-line callers alone determine the number and length of calls to a mobile subscriber. If in practice the number and length of fixed-to-mobile calls is jointly determined by both caller and receiver, then above cost termination charges should lead competing mobile carriers to set *negative* reception fees. By paying consumers to receive calls, mobile carriers will more directly promote increased fixed-to-mobile termination revenue than an equivalent subscription subsidy. This can explain why consumers are not charged for receiving calls in countries with high (above-cost) termination charges. However, negative reception fees are not

---

[35] It is worth noting that the same logic could also be applied to other types of calls, such as mobile-to-fixed calls, where the fixed-line subscriber obtains benefits from receiving calls from mobile subscribers. Traditionally the benefits of call receivers have been ignored.

observed in practice. This could reflect the fact that with high fixed-to-mobile prices and no reception fees, the number and length of calls will be predominately determined by the fixed-line caller and not the mobile receiver. That is, fixed-line callers will usually be the first to want to hang up for price reasons. In this case, there will be little or no benefits to mobile carriers of further subsidising reception fees rather than subscription fees.[36]

Doyle and Smith (1998) develop a model of fixed-to-mobile calls that allows caller pays and receiver pays to be compared. Their model assumes two competing mobile operators and a single unregulated fixed-line network. They assume the demand for outgoing mobile calls determines the demand for fixed-to-mobile calls. They show that each mobile network will set its fixed-to-mobile termination charge above cost, and on top of this, the fixed-line network will impose its own bottleneck mark-up, resulting in a double marginalisation problem. They then modify their model to consider the receiver pays principle. They consider a set-up in which the fixed-to-mobile caller pays a price equal to the normal price of a fixed-line call. The mobile subscriber receiving the call makes up the difference between the posted price of a fixed-to-mobile call, set by the mobile operator, and the lower charge paid by the fixed-line caller. They argue under this receiver pays principle, double marginalisation is eliminated and prices are lower. In some cases termination charges set by the mobile network will be lower than under the caller pays principle.

Their analysis can be interpreted in an alternative way. Their set-up is equivalent to regulating the price of a fixed-to-mobile call at the price of a regular fixed-line call,

---

[36] There may be other problems with negative reception fees. In this regard, Tommaso Valletti pointed out in private correspondence that some time ago in Italy, mobile operators set negative reception fees only to find that people were calling their mobile phones from office lines so as to obtain these rebates.

and imposing zero termination charges for mobile carriers. With no termination revenues from the fixed-line network, mobile operators will be forced to charge mobile callers directly to recover their costs of terminating calls. With positive (and potentially high) reception charges, call receivers will have a strong interest in the number and length of calls to their phones. A lower reception charge will encourage mobile subscribers to accept more calls (or allow longer conversations). Both reception and subscription charges will be uniquely determined. In this case, competition between mobile carriers will drive their reception charges towards the cost of terminating calls. Under this interpretation, it is not that the receiver pays principle helps solve the problem of inflated termination charges, but rather that the regulation of low (or zero) termination charges leads carriers to charge receivers for incoming calls.

Gans and King (2000) and Wright (1999) model the incentives of competing mobile operators in setting termination charges. They assume two mobile operators that compete in a Hotelling framework, and that fixed-line subscribers choose the number of fixed-to-mobile calls.[37] Gans and King consider the case in which the fixed line network can set differential fixed-to-mobile prices depending on which mobile operator the calls terminate on, but where consumers are ignorant about which mobile network they are calling. Wright makes a formally equivalent assumption, that a uniform fixed-to-mobile price is set. In both cases, with a fixed number of mobile subscribers in the market, competing mobile operators set their respective termination charges to the point where they completely choke off all demand for fixed-to-mobile calls.

[37] As noted in the discussion of reception charges, the fact that fixed and mobile subscribers do not jointly determine the number of fixed-to-mobile calls is broadly consistent with the fact that in equilibrium it is the fixed-line caller that pays for the call.

Using a more general model of competition, Wright (2002) clarifies the source of this market failure. When a mobile carrier increases its termination charge, there will be two types of effects on its profits. First, the increase in its termination charge will increase the fixed-to-mobile price, and so decrease the number of fixed-to-mobile calls. This will change the carrier's termination profit obtained per subscriber, and so have a direct impact on its profits. In addition, any change in termination profit per subscriber will influence the strength of competition for mobile subscribers, having an indirect impact on the carrier's profits (and other carriers' profits).

When the fixed-line networks set differential fixed-to-mobile prices depending on which mobile network calls terminate on (and assuming callers know which operator they are calling), competition will lead mobile operators to set termination charges to the same level as would be set by a single monopoly mobile operator. Competing cellular firms will want to maximize termination profits so as to subsidize their mobile subscribers as much as possible.

With uniform fixed-to-mobile pricing (or customer ignorance), any increase in termination charges and decrease in fixed-to-mobile demand is shared across all mobile carriers. Essentially, fixed-to-mobile demand becomes more elastic. This effect implies each mobile operator will want to set a higher (monopoly) termination charge. However, as termination charges are increased above this monopoly level, there are two additional effects. On the one hand, termination profits per subscriber will decrease, which will decrease the mobile operator's profits. On the other hand, the increase in termination charges will increase the uniform fixed-to-mobile price, which will decrease the termination profits per subscriber that rivals can capture. This makes rivals compete less aggressively for mobile subscribers. When both operators set their termination charges at the monopoly level, the first effect will vanish while the second effect remains. This implies that each operator will want to set a termination charge above the monopoly level. In the special case of the Hotelling model considered by Gans and King (2000) and Wright (1999), the second effect continues to dominate no matter how high the two carriers termination charges are set, and equilibrium termination charges in fact eliminate fixed-to-mobile calls altogether. Each operator will want to set termination charges a few cents above its rival.

Relative to this rather extreme outcome, regulation of lower termination charges generally benefits all parties, including mobile operators and mobile subscribers.[38] It is unambiguously desirable. Relative to the monopoly termination charges that result when discriminatory fixed-to-mobile pricing is effective, regulation of lower termination charges benefits the fixed-line network and fixed-to-mobile callers, but generally makes mobile operators and mobile subscribers worse off. However, starting from the monopoly termination charge, a small decrease in the termination charge will have only a second-order effect on mobile operators and subscribers, but will have a first-order effect on fixed-to-mobile prices. Welfare will be higher when termination charges are regulated below those set by competing firms.

Wright (1999) calculates welfare maximizing termination charges when consumers vary in their subscription benefits of a mobile phone. When consumers

---

[38] In the Hotelling model of competition used by Gans and King (2000) in which there are a fixed number of mobile subscribers, mobile operators are indifferent over the common fixed-to-mobile termination charge. However, in a more general model of competition, including when demand by mobile subscribers is elastic, mobile operators will be better off avoiding termination charges above the monopoly level. See Wright (2002).

already have access to the fixed-line network, but many do not have access to the mobile network, a subsidy to mobile subscribers from high termination charges has two beneficial effects. First, both fixed-line subscribers and existing cellular subscribers benefit from being able to call (and be called by) new mobile subscribers who join because of subsidised subscription charges. Efficiency is also enhanced by above-cost termination charges when retail competition between mobile carriers is limited. This "monopoly subsidy" justification for a high termination charge is neutralised to the extent the fixed-to-mobile price also needs to be subsidised to offset monopoly markups. Wright takes into account these various effects and finds using a calibrated version of the model that the socially optimal termination charge can be several times cost when half of all consumers have mobile subscriptions (in equilibrium). As mobile subscription tends towards saturation point, the need to tax fixed-to-mobile calls to subsidise mobile subscription diminishes. Given an increased tax-base (an increased number of fixed-to-mobile calls), the efficient mark-up of termination charges above cost decreases.

Armstrong (2002a) analyses socially optimal termination charges, assuming away complications arising from mobile-to-mobile calls and other non-linear network externalities. In addition to the effects discussed previously, he shows that when mobile subscribers internalise the benefits of those calling them (for example, they might care about the price friends and family pay to call them), mobile carriers will have less incentive to set high termination charges. Mobile subscribers directly care about how much fixed-to-mobile callers pay and so are less inclined to join a network that offers cheap subscription only because it is expensive for others to call the network.[39]

The models of fixed-to-mobile termination of Armstrong, Gans and King, and Wright all assume away any role for the fixed and common costs of operating fixed-line and mobile networks. Implicitly it is assumed that retail mark-ups on these services are sufficient for fixed and common costs to be fully recovered. In the face of fixed and common costs, solving for the Ramsey optimal termination charges taking into account the various externalities described above would be required to determine the optimal solution.[40] However, given Ramsey considerations are at least as important for the fixed-line operator as they are for mobile operators, the Ramsey argument alone does not provide an automatic justification for high fixed-to-mobile termination charges. If fixed-to-mobile demand is inelastic, then it may be socially optimal for fixed-line networks to exploit this by setting high fixed-to-mobile retail mark-ups, thereby enabling other fixed-line prices to be lowered. High fixed-to-mobile mark-ups would then imply socially optimal fixed-to-mobile termination charges are lower.

Armstrong calls the situation in which firms compete for customers but in which access to these customers is required by third parties one of “competitive bottlenecks.” Mobile operators have a bottleneck because when someone wants to call a subscriber to their network, the calling party has no choice but to call their network. However, unlike a typical bottleneck situation, mobile firms compete to attract the subscribers in the first

[39] Such subscribers may still join a cheaper network funded by high fixed-to-mobile termination charges since they can always take advantage of cost-based mobile-to-fixed prices to call back fixed-to-mobile callers. This tends to make fixed-to-mobile demand more elastic, and the monopoly fixed-to-mobile termination charge lower.

[40] The Competition Commission in the U.K. has argued against the use of Ramsey pricing to set fixed-to-mobile termination charges, arguments summarized and rebutted in Oftel (2002b). See also Competition Commission (2003) for a further lengthy discussion of why it rejects Ramsey pricing.

place. Competition transfers the rents from the access bottleneck to the mobile subscribers.

Despite the bottleneck for mobile termination, mobile operators do not necessarily have market power. The market is a two-sided one, with services being provided to both fixed-to-mobile callers and mobile subscribers.[41] It is wrong to isolate the effects on just one side of the market, and ignore the effects on the other. After all, without mobile subscribers, there can be no fixed-to-mobile calls. The competitive bottleneck does not necessarily lead to higher overall prices of fixed-to-mobile calls and mobile services. Rather, with a competitive mobile sector, it leads to a distortion in the *structure* of prices between fixed-to-mobile calls and mobile services. Regulation may be required to achieve a more efficient structure of prices in this market. However, this does not indicate a problem of market power. As Gans (1999) argues, "the *less concentrated* the mobile network market, the higher will be the level of fixed to mobile call charges." With perfect competition between mobile operators, the distortion in prices between the two sides of the market may be at its most severe (for instance, if there is uniformity of fixed-to-mobile prices, or if there is some customer ignorance over which network they are calling). Moreover, the need to regulate termination charges of small carriers may be

[41] A recent literature on two sided markets has emerged that can be applied to these issues. Two-sided markets have the property that there are two types of users that wish to make use of a common platform, each of which obtains benefits that depend on the number of users of the opposite type. As Armstrong (2002b) notes, fixed and mobile telephony is a type of two-sided market. One type of user is the mobile subscriber who values receiving calls, and the other is the fixed-to-mobile caller who values being able to call mobile subscribers. The two-sided markets literature suggests similar issues to those discussed here arise in many different industries including shopping malls, yellow pages, software/hardware, payment systems, and matchmaking services to name a few. See also Caillaud and Jullien (2001), Guthrie and Wright (2003), Parker and Van Alstyne (2000), Rochet and Tirole. (2001) and Schiff (2003).

even stronger, since small carriers will tend to set even higher termination charges.[42]

A more general framework of fixed and mobile networks would have the number of fixed-to-mobile calls be jointly determined by both fixed and mobile callers. Taking into account that both callers and receivers obtain some utility from calls, and allowing for the possibility that networks can charge consumers directly for making and receiving calls, in general the socially optimal termination charge could be negative, zero, below cost, at cost, or above cost.[43] Seen in this light, the termination charge is just the instrument that the fixed-line and mobile networks can use to get closer to their desired structure of retail fees (caller and reception fees). Bill and keep (zero termination charges) might have some appeal in this context, as suggested by DeGraba (2002), but depending on various asymmetries in the two sides of the market, so might various other approaches. It would be especially interesting to consider the case in which users get random utility from calls, and so can choose to be callers, or wait to be receivers. In this case, the question is whether there are any particular advantages of having callers pay (given originators of calls are more likely to be the ones with higher willingness to pay) and therefore whether there is a general argument for positive fixed-to-mobile termination charges. Empirically, it would also be fascinating to try and address whether receiver pays, as opposed to caller pays, has restrained growth in mobile services.[44]

Finally, a word on mobile-to-mobile calls. Traditionally, mobile-to-mobile

[42] To the extent that fixed-to-mobile callers act as though there is a uniform price of calling mobile subscribers, a small mobile carrier will face very inelastic fixed-to-mobile demand.

[43] Hermalin and Katz (2001) and Jeon *et al*. (2004) provide frameworks in which callers and receivers may obtain benefits from calls, and be charged for calls.

[44] The faster growth of mobile subscriptions in (caller pays) Europe compared to (receiver pays) North America, the rapid growth of subscriptions in Mexico and Peru following their shift to caller pays from receiver pays, and the lackluster performance of mobile subscription growth in India under receiver pays (Srivastava and Sinha, 2001) are all noteworthy in this regard.

termination has been thought about in the same way as other two-way interconnection problems. In this case, the literature predicts that operators will agree to below cost termination (see Gans 1999, Gans and King, 2000, and Berger, 2002). However, with high mobile penetration, mobile-to-mobile calls are an important substitute for fixed-to-mobile calls. If consumers are charged too much for making fixed-to-mobile calls, they will simply make mobile-to-mobile calls instead. This substitution possibility should help constrain the ability of the fixed-line operator to mark up fixed-to-mobile calls. To the extent that mobile-to-mobile calls are cost based, this will also undermine the ability of mobile carriers to raise their fixed-to-mobile termination charges much above cost. Fixed-to-mobile demand will become highly elastic. To overcome this customer arbitrage, mobile carriers may want to set mobile-to-mobile termination charges at similar (high) levels as fixed-to-mobile termination charges. This could explain the emergence of off-net price discrimination involving higher retail prices for mobile-to-mobile calls that terminate on a rival's network versus calls that terminate on the same mobile network.[45] This also suggests the analysis of mobile-to-mobile termination charges may be quite closely tied to that of fixed-to-mobile termination. Future research would be usefully directed towards a deeper analysis of mobile-to-mobile calls.

---

[45] Off-net calls averaged over four times as much as on-net calls in the U.K. in 2000/01. See Competition Commission (2003, 2.122). More surprisingly, they were significantly higher than on-net calls, even allowing for differences in termination charges. See Competition Commission (2003, footnote 1). As discussed in section 5.1.2(d) above, Berger (2002) provides one explanation for this phenomenon when consumers value incoming calls. An alternative explanation for why a large mobile network may want to set high charges for calls to rival mobile networks is to generate an artificial network effect, so that customers will be drawn to its own (large) network to avoid higher off-net prices.

References

Ahn, H. and M-H. Lee (1999) "An econometric analysis of the demand for access to mobile telephone networks", *Information Economics and Policy*, 11, 297-305.

Albaek, S., P. Mollgaard and P. Overgaard (1997) "Government assisted oligopoly coordination? A concrete case," *Journal of Industrial Economics*, 45, 429-443.

Armstrong, M. (1997) "Mobile telephony in the U.K." Regulation Initiative Discussion Paper No. 15, London Business School.

Armstrong, M. (2002a) "The theory of access pricing and interconnection," in M. Cave, S. Majumdar, and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, North-Holland.

Armstrong, M. (2002b) "Competition in Two-Sided Markets," *mimeo*, Nuffield College, University of Oxford.

Ausubel, L., P. Cramton, P McAfee and J. McMillan (1997) "Synergies in wireless telephony: Evidence from the MTA auction," *Journal of Economics and Management Strategy*, 6, 497-527.

Bekkers, R., B. Verspagen and J. Smits (2002) "Intellectual property rights and standardization: the case of GSM", *Telecommunications Policy*, 26, 171-188.

Berger, U. (2002) "Network competition and the collusive role of the access charge." *mimeo*, Vienna University.

Binmore, K. and Klemperer, P. (2002) "The biggest auction ever: The sale of the British 3G telecom licences", *The Economic Journal*, 112, C74-C96.

Bresnahan, T. (1989) "Empirical studies in industries with market power" in R. Schmalensee and R. Willig (eds.), *Handbook of Industrial Organization*, North-Holland.

Busse, M. (2000) "Multimarket contact and price coordination in the cellular telephone industry." *Journal of Economics and Management Strategy*, 9 (3): 287 – 320.

Cadima, N. and P. Barros (2000) "The impact of mobile phone diffusion on the fixed-link network", CEPR Discussion Paper 2598, London.

Caillaud, B. and B. Jullien (2001) "Competing Cybermediaries," *European Economic Review*, 45, 797-808.

Cave, M. (2001) *Independent Review of Radio Spectrum Management*, Consultation Paper, HM Treasury and Department of Trade and Industry, London.

Coase, R. (1959) "The Federal Communciations Commission", *The Journal of Law and Economics*, 2, 1-40.

Competition Commission (2003) "Vodafone, O_2, Orange, and T-Mobile" U.K, 18 February 2003.

Cramton, P. (2002), "Spectrum Auctions," in M. Cave, S. Majumdar, and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, North-Holland.

Cramton, P., E. Kwerel and J. Williams (1998) "Efficient relocation of spectrum incumbents", *Journal of Law and Economics*, 41, 647-675.

Crandall, R. (1998) "New Zealand spectrum policy: A model for the United States?", *Journal of Law and Economics*, 41, 821-840.

DeGraba, P. (2002). "Bill and keep as the efficient interconnection regime?: A Reply." *Review of Network Economics*, 1: 61-65.

DeVany, A. (1998) "Implementing a market-based spectrum policy", *Journal of Law and Economics*, 41, 627-646.

Doyle, C. and J. Smith (1998) "Market structure in mobile telecoms: Qualified indirect access and the receiver pays principle." *Information Economics and Policy*, 10: 471 – 488.

Duso, T. (2003) "Lobbying and regulation in a political economy: Evidence from the U.S. cellular industry." *Public Choice*, forthcoming.

Duso, T. and A. Jung (2003) "Lobbying and collusion in a regulated industry: The U.S. mobile telecommunications market." mimeo, Humboldt University Berlin.

Duso, T. and L.-H. Röller (2003) "Endogenous Deregulation: Evidence from OECD Countries" mimeo, Wissenschaftszentrum Berlin (WZB).

Economides, N. (1996) "The Economics of networks", *International Journal of Industrial Organization*, 14, 673-700.

Economides, N. (1989) "Desirability of compatibility in the absence of network externalities", *American Economic Review*, 79, 1165-1181.

Farrell, J. and P. Klemperer (2002) "Coordination and lock-in: competition with switching costs and network effects" forthcoming chapter in the Handbook of Industrial Organization, Vol 3.

Dos Santos Ferreira, R. and J.-F. Thisse (1996) "Horizontal and Vertical Differentiation: The Launhardt Model" *International Journal of Industrial Organization*, 14: 485-506.

Foreman, R.D. and E. Beauvais (1999) "Scale economies in cellular telephony: Size matters", *Journal of Regulatory Economics*, 16, 297-306.

Foros, O., B. Hansen, and J-Y. Sand (2002) "Demand-side spillovers and semi-collusion in the mobile communications market" *Journal of Industry, Competition and Trade*.

Gans, J.S. (1999) "An evaluation of regulatory pricing options for mobile termination services," *mimeo*., University of Melbourne, December.

Gans, J.S. and S.P. King (2000) "Mobile network competition, customer ignorance and fixed-to-mobile call prices," *Information Economics and Policy*, 12: 301 – 327.

Gans, J.S. and S.P. King (2001a) "Using 'Bill and Keep' Interconnect Arrangements to Soften Network Competition", *Economic Letters*, 71 (3), pp.413-420.

Gans, J.S. and S.P. King (2001b) "Regulating endogenous customer switching costs" *Contributions to Theoretical Economics*, 1: 1-29.

Gans, J.S., S.P King and G. Woodbridge (2001), "Numbers to the People: Regulation, Ownership and Local Number Portability," *Information Economics and Policy*, 13 (2), pp.167-180.

Gruber, H. (2001a) "Competition and innovation: The diffusion of mobile telecommunications in Central and Eastern Europe", *Information Economics and Policy*, 13, 19-34.

Gruber, H. (2001b) "Spectrum limits and competition in mobile markets: the role of license fees", *Telecommunications Policy*, 25, 59-70.

Gruber, H. (1999) "An investment view of mobile telecommunications in the European Union", *Telecommunications Policy*, 23, 521-538.

Gruber, H., and F. Verboven (2001a) "The evolution of markets under entry standards and regulation – the case of global mobile telecommunications", *International Journal of Industrial Organization*, 19, 1189-1212.

Gruber, H., and F. Verboven (2001b) "The diffusion of mobile telecommunications services in the European Union", *European Economic Review*, 45, 577-588.

Guthrie, G. and J. Wright (2003) "Competing Payment Schemes," University of Auckland *Working Paper* No. 245.

Haug, T. (2002) "A commentary on standardization practices: lessons from the NMT and GSM mobile telephone histories", *Telecommunications Policy*, 26, 101-107.

Hausman, J. (1995) "The cost of cellular telephone regulation," *mimeo*, MIT.

Hausman, J. (2002) "Mobile telephony" in Cave, M., S. Majumdar, and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, North-Holland.

Herzel, L. (1951) "Public interest and the market in color television regulation", *University of Chicago Law review*, 18, 802-816.

Horvath, R., and D. Maldoom (2002) "Fixed-mobile substitution: a simultaneous equation model with qualitative and limited dependent variables", DotEcon Discussion Paper, DP 02/02, www.dotecon.com, London.

Jeon, D-S., J.-J. Laffont, and J. Tirole (2004), "On the receiver pays principle*." Rand Journal of Economics*, 35(1), pp.85-110.

Katz, M. and C. Shapiro (1985) "Network externalities, competition and compatibility", *American Economic Review*, 73, 424-440.

Koski, H., and T. Kretschmer (2002) "Entry, standards and competition: firm strategies and the diffusion of mobile technology", *mimeo*, LSE, London.

Laffont, J-J., P. Rey and J. Tirole (1998a), "Network Competition I: Overview and Nondiscriminatory Pricing," *RAND Journal of Economics*, 29 (1), pp.1-37.

Laffont, J-J., P. Rey and J. Tirole (1998b), "Network Competition II: Price Discrimination," *RAND Journal of Economics*, 29 (1), pp.38-56.

Lehenkari, J., and R. Miettinen (2002) "Standardization in the construction of a large technological system – the case of the Nordic mobile telephone system", *Telecommunications Policy*, 26, 109-127.

Leibowitz, S.J. and S.E. Margolis (2002), "Network Effects," in M. Cave, S. Majumdar, and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, North-Holland.

McAfee, R.P. and J. McMillan (1996) "Analyzing the airwaves auction", *Journal of Economic Perspectives*, 10, 159-175.

McKenzie, D. and J. Small (1997) "Econometric cost structure estimates for cellular telephony in the United States", *Journal of Regulatory Economics*, 12, 147-157.

McMillan, J. (1994) "Selling spectrum rights", *Journal of Economic Perspectives*, 8, 145-162.

Miravete, E. and L-H. Röller (2002) "Competitive nonlinear pricing in duopoly equilibrium: The early U.S. cellular telephone industry" mimeo, Department of Economics, University of Pennsylvania, August 4.

Moreton, P. and P. Spiller (1998) "What's in the air: Interlicense synergies in the Federal Communication Commission's broadband personal communications service spectrum auctions", *Journal of Law and Economics*, 41, 677-716.

Noam, E.M. (1997) "Beyond spectrum auctions – taking the next step to open spectrum access", *Telecommunciations Policy*, 21, 461-475.

Noam, E.M. (2002), "Interconnection Practices," in M. Cave, S. Majumdar, and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, North-Holland.

OECD (2000), *Cellular mobile pricing structures and trends*, May, Paris.

Oftel (2001) "Effective competition review: mobile" Office of Telecommunications, U.K, 26 September.

Oftel (2002a) "Review of SIM-locking policy" Office of Telecommunications, U.K, 26 November.

Oftel (2002b) "Ramsey pricing – Oftel's response to letter of 4 July" U.K. 12 July.

Parker, G. and M. Van Alstyne (2000) "Information Complements, Substitutes, and Strategic Product Design," *Working Paper* No. 299, March. Tulane University and University of Michigan.

Parker, P. and L.-H. Röller (1997) "Collusive conduct in duopolies: Multimarket contact and cross-ownership in the mobile telephone industry." *RAND Journal of Economics*, 28 (2): 304 – 322.

Penard, T. (2002) "Competition and strategy on the mobile telephony market: a look at the GSM business model in France", *Communications and Strategies*, 45, 49-79.

Productivity Commission (2002) *Radiocommunications*, Report no. 22, AusInfo, Canberra.

Reiffen, D., L. Schumann, and M. Ward (2000) "Discriminatory dealing with downstream competitors: Evidence from the cellular industry." *Journal of Industrial Economics*, XLVIII: 253 – 286.

Reiffen, D. and M. Ward (2002) "Recent empirical evidence on discrimination by regulated firms." *Review of Network Economics*, 1: 39-53.

Rochet, J. and J. Tirole (2001) "Platform Competition in Two-Sided Markets," *mimeo*, Toulouse University.

Ruiz, L.K. (1995) "Pricing strategies and regulatory effects in the U.S. cellular telecommunications duopolies." In G.W. Brock (ed.) *Towards a Competitive Telecommunications Industry* Lawrence Erlbaum Associated, Mahwah, N.J., 13-54.

Schiff, A. (2003) "Open and Closed Systems of Two Sided Networks," *Information, Economics and Policy*, 15(4), 425-442.

Schiller, J. (2000), *Mobile Communications*, Addison-Wesley.

Shapiro, C. and H. Varian (1999) *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, Boston.

Shew, W.B. (1994) "Regulation, competition and prices in the U.S. cellular telephone industry." Working Paper, American Enterprise Institute.

Shin, R. and J. Ying (1992) "Unnatural monopolies in local telephone", *RAND Journal of Economics*, 23, 171-183.

Shy, O (2001) *The Economics of Network Industries*, CUP, Cambridge, U.K.

Srivastava, L. and S. Sinhab (2001) "TP case study: fixed–mobile interconnection in India" *Telecommunications Policy*, 25: 21-38.

Stoetzer, M. and D. Tewes (1996) "Competition in the German cellular market?" *Telecommunications Policy*, 20: 303 – 310.

Sung, N. and Y-H. Lee (2002) "Substitution between mobile and fixed telephones in Korea", *Review of Industrial Organization*, 20, 367-374.

Sutherland, E. (2001) "International roaming charges: over-charging and competition law" *Telecommunications Policy*, 25: 5-20.

Sutton, J. (1991) "Sunk costs and market structure." MIT Press, Cambridge, MA.

Valletti, T. (1999) "A model of competition in mobile communications." *Information Economics and Policy*, 11: 61 – 72.

Valletti, T. (2001) "Spectrum trading", *Telecommunications Policy*, 25, 655-670.

Valletti, T. (2003) "Is mobile telephony a natural oligopoly?" *Review of Industrial Organization*, 22, 47-65.

Valletti, T. and M. Cave (1998) "Competition in U.K. mobile communications," *Telecommunications Policy*, 22: 109 – 131.

Woroch, G.A. (2002), "Local Network Competition," in M. Cave, S. Majumdar, and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, North-Holland.

Wright, J. (1999), "Competition and Termination in Cellular Networks," *Working Paper*, Centre for Research in Network Economics and Communications, University of Auckland.

Wright, J. (2002) "Access pricing under competition: An application to cellular networks." *Journal of Industrial Economics*, L: 289 – 315.