Aalto-yliopisto
Kauppakorkeakoulu

Efficient use of radio spectrum: the Administrative Incentive Pricing (AIP) approach

Kansantaloustiede

Maisterin tutkinnon tutkielma

Salla Emilia Taskinen

2015

Taloustieteen laitos
Aalto-yliopisto
Kauppakorkeakoulu

Aalto University School of Economics Abstract

Master's Thesis 30.12.2014
Emilia Taskinen

Efficient use of radio spectrum: the Administrative Incentive Pricing (AIP) approach

Purpose of the study

The purpose of this study is to offer an overview on spectrum management methods and to provide a detailed analysis of one of them, namely the Administrative Incentive Pricing (AIP). This thesis discusses the background, purpose and determination of AIP by first shortly introducing the alternatives for AIP and then critically assessing the alternative ways to construct such payments.

The main economic models discussed include the opportunity cost based theory by Smith and NERA (1996) as well the subsequent, more refined model by Levine and Rickman (2007), which extends the Smith-NERA methodology to account for market structure and interference constraints. These models and the alternative ways of assigning frequencies are critically viewed against their core assumptions, regulator's objectives for spectrum management as well as from the point of view of economic efficiency.

Methodology

The study is conducted as a literature review combining literature and publications from scientific research regarding spectrum management and formation of AIP payments, consultative research commissioned by the telecommunications authorities as well as publicly available information on realized market transactions for spectrum.

Results of the study

Taking into account the growing demand and utilization of spectrum resources and given its scarcity, spectrum management and efficiency of use is crucial. Traditional administrative allocation and assignment methods need to be replaced by market based methods or complemented by market mimicking methods such as the AIP in order to fulfill efficiency.

AIP payments are currently constructed through three different kinds of methods basing the payment either on the concept of opportunity cost, realized market prices or treating the formation as an optimization problem where the regulator maximizes overall welfare with respect to the spectrum fee given the interference and resource constraints. The latter method (by Levine and Rickman) combines economic modelling and information theory arriving at a group of equations determining optimal AIP. The key conclusion is that in a setting where interference, market structure and overall welfare including consumer and producer surplus as well as the revenue impact for the government from imposing AIP are accounted for, the optimal AIP should be higher whenever spectrum sharing is possible. The AIP then act as Ramsey tax across sectors of the economy being inversely related to the elasticity of demand.

The method by Levine and Rickman explicitly accounts for most of the crucial elements of spectrum and spectrum markets, but further studies are needed especially to account for dynamic efficiency and how AIP payments may be used to promote it.

Key words: Economics, frequencies**,** efficiency, spectrum pricing, spectrum management, Administrative Incentive pricing, AIP

Table of Contents

1. Introduction

Rapid development and increasing demand of new types of technological products and services utilizing radio spectrum[1], such as smartphones and tablets using mobile services, combined with the scarcity of spectrum resources[2] attest to the need of efficient frequency use. The aspiration for efficiency can most clearly be seen in the increased interest of telecom regulators to find new spectrum management methods to complement or even substitute traditional methods such as beauty contest and auctions in order to help ensure that frequencies are used as efficiently as possible. There are for example numerous studies conducted for communications regulators on spectrum management and pricing in particular during the last five years or so, some of which are also discussed in this thesis.

From an economic point of view efficient use of spectrum is by no means a trivial issue; according to the Economist (2004), at the beginning of the 21st century as much as half of the total value of frequencies was still wasted on uneconomic uses causing social losses to the society. Presumably the situation has improved since then as efficient market-based mechanisms (especially auctions) to divide frequencies have become more popular, but the magnitude of significance of the issue in economic terms remains great. In addition to not being in their socially optimal use, some frequencies are not used at all or are used only a part of the time. These features have been claimed to be consequences of generally making spectrum a de facto private good through licensing as well as consequences of the lack of incentives to use spectrum efficiently. This in turn further attests to the importance of efficient use of spectrum.

Despite the growing significance of spectrum in the society and the vast economic benefits it may bring when used optimally (or losses caused when used inefficiently), there have been very few new approaches to spectrum management during the last couple of decades. Academic research is still dominated by traditional division and pricing methods such as beauty contests and auctions. This thesis concentrates on the new approaches describing and discussing their determination,

---

[1] Spectrum is an umbrella term which describes a band of electro-magnetic frequencies. In this thesis it specifically refers to radio spectrum, commonly understood to cover frequencies from approximately 10kHz to 300GHz which are usable for communications purposes (see e.g. Cave, Martin & Webb 2007).

[2] There has been quite extensive discussion on whether spectrum actually is a finite and thus a scarce resource since the development of more efficient technologies allows higher capacity utilization (see e.g. Staple and Werbach 2004). However, the mainstream view is that spectrum resources currently are limited and thus need to be regulated e.g. with the help of licensing. These issues are discussed in more detail in chapter 2 of the thesis.

development and implications for future spectrum management. The focus is specifically on Administrative Incentive Pricing (AIP), a frequency pricing method used together with traditional administrative methods for dividing spectrum, which do not attach a price for the spectrum resources. Other new approaches, which are significant in order to provide a holistic view on spectrum management and to describe the possible role and significance of AIP, as well as traditional spectrum management methods in the future, include unlicensed spectrum and the ideas of abundant spectrum (as opposed to the traditional view of scarcity) and spectrum as a natural resource.

The basic idea of AIP is to improve the efficiency of resource use by shifting spectrum from lower value use to higher value use. The shift is incentivized by imposing a fee on spectrum users to encourage them to give up under- or unused spectrum. In other words, AIP is a cost on hoarding and it presupposes some form of property rights for spectrum in order for the cost to be imposed correctly. The most common property rights regime is the one with exclusive licenses, which is the primary tool for spectrum management in most countries (Doyle 2006, p. 2). While AIP imposes a price for spectrum the actual division of frequencies (i.e. frequency licenses) is executed through administrative means such as beauty contests or lotteries.

Setting the AIP fee naturally requires thorough consideration as fees set too high can cause under-utilization of the resource whereas fees that are substantially lower than the value of spectrum to the user(s) encourage hoarding, which in turn may cause congestion (Cave, Doyle, Webb, 2007, p.167). The fee is usually based on the opportunity cost of spectrum use, since it is regarded to be a clear, relatively easily attainable measure in line with efficient outcomes, at least under perfect competition assumptions. The core method, called the Smith-NERA method is based on this principle. However, as will be shown in chapter 4 there are also two alternative ways proposed to calculate the optimal AIP price; a method by Levine and Rickman (2007) which extends the Smith-NERA methodology to account for market structure and interference constraints, and a method which bases AIP payments on prices realized in market transactions.

AIP should not be confused with pure cost recovery fees set to license holders. The sole purpose of cost recovery fees is to cover the frequency governance costs of the regulator, such as costs of dividing the frequencies between uses and users, whereas AIP payments are intended to reflect the value of the frequency and aim to incentivize its efficient use. In Europe these two frequency fees (termed as administrative charges and fees for rights of use) are actually required to be kept separate by law (EU Authorization Directive, 2002, articles 12 and 13). Also in the U.S. the responsible regulator, the Federal Communications Commission (FCC) separates these charges by terming them regulatory fees and application fees (Doyle 2007, p.1). Thus, spectrum license holders may be faced with two kinds of payments; cost recovery payments utilized by basically all regulators around the world (Cave et al. 2007, p.167) and incentive pricing fees utilized only by some regulators.

Relative to alternative spectrum management methods, which currently include auctions, frequency trading, beauty contests, first-come-first-served methods and lotteries, AIP is a new method of promoting efficient use of spectrum. In its earliest forms as an "AIP like" incentive fee aimed at improved efficiency of spectrum use the method has been utilized in the United Kingdom since 1998 (Ofcom) and in the New Zealand since 2009 (ACMA). In Ireland, Canada and Spain the regulators have applied spectrum fees exceeding the cost recovery fees, but without an explicit aim at doing it according to clear economic principles (ComReg, CRTC, CMT). Even though AIP is not currently in use in Finland the Ministry of Transport and Communications (MINTC), which is the authority responsible for spectrum management in Finland, has investigated this opportunity (see e.g. MINTC 2009). In addition, frequency assignment and pricing are current themes in Finland; the second auctioning of frequencies in the nation's history (the 800MHz frequency band) was just held in 2013 and discussion around its successfulness has thrived in the Finnish media.

Due to the novelty of the subject AIP has not been studied substantially. As was previously pointed out, academic research is centered around the traditional spectrum management methods and research around auctions is especially vast. Thus, there is a need for a concise representation of current literature and research around AIP and how it relates to spectrum management objectives and other spectrum management methods. The basic concepts and ideas go far back to the principles of microeconomics which have been applied to construct the framework for implementing AIP.

The key materials on AIP discussed in this thesis include the book "Essentials of modern spectrum management" by Cave, Doyle and Webb (2007), the subsequent work of Doyle in "The pricing of radio spectrum: using incentives mechanisms to achieve efficiency" (2007) and "The need for a conservative approach to the pricing of radio spectrum and the renewal of radio spectrum licenses" (2010). In addition, reports prepared by consultants in cooperation with economics professionals for regulators are discussed; namely the "Study into the use of Spectrum Pricing" by the National Economic Research Associates Economic Consultants (NERA) and Smith System Engineering Ltd

(1996), “An economic study to review spectrum pricing” by Indepen Consulting Ltd, Aegis Systems Ltd and Warwick Business School (2004) and “Administrative Incentive Pricing for Radiofrequency Spectrum” by Aegis Systems Ltd and Plum Consulting Ltd (2008). These reports present the current way of applying the AIP, which is utilized in the U.K. and New Zealand. Levine and Rickman (2007) have further developed the AIP approach from Smith-NERA (1996) and Indepen (2007) by constructing a more rigorous mathematical way of calculating the optimal AIP in their paper “Optimal Administered Incentive Pricing” which is discussed in section 4.2.

As stated above, this thesis concentrates on Administrative Incentive Pricing in theory and practice. The thesis describes AIP and explains how it is expected to promote efficient use of spectrum resources. It also discusses the different situations in which AIP can and should be utilized and demonstrates the current methodologies used in determining AIP. In addition, the thesis discusses the potential role of AIP (as well as other spectrum management methods) in the future by introducing the issue of unlicensed spectrum and the ideas of abundant spectrum and spectrum as a natural resource. The thesis is mainly executed as a literature review combining theory on spectrum economics, opportunity cost based pricing and a practical application regarding spectrum management that is the AIP. The thesis presents the currently used methods of determining AIP, but refrains from more specific calculation of incentive prices for different spectrum band uses (mobile and fixed services, broadcasting etc.).

An important conceptual issue regarding spectrum division is the difference between dividing spectrum to uses such as broadcasting or radio services and further assigning the frequencies (i.e. licenses) between users such as broadcasting operators and radio service providers. The former is often referred to as allocation and the latter as assignment of frequencies (European Commission, 2012). Thus, allocation defines the license, i.e. the frequency band, the geographic area, the time period, and the restrictions on use whereas assignment defines the licensee (Cramton 2003, p. 28). Allocation between uses has been traditionally determined through a process of negotiations between national and international regulators of spectrum (see e.g. Doyle 2007 and subsection 3.1 of the thesis) and changes in it are usually extremely slow or even unfeasible. This can be due to international harmonization agreements restricting the use or some technical constraints and the costs related to overcoming them. An example of the latter are the costs related to renewing or substituting equipment which has been built to operate with certain frequencies. In other words, there is a dependency of infrastructure and equipment on certain technologies utilizing particular frequencies (and being unable to utilize others) which leads to restrictions in frequency allocation.

This in turn leads to inefficiencies in spectrum use when initial allocations are not optimal. Alternatively one can conclude that the system of allocating and assigning frequencies (through licensing) has lead to the current technologies being developed as opposed to what might have been if different kinds of allocation and assignment mechanisms, such as freely tradable spectrum rights or more unlicensed spectrum, had been used. As opposed to allocation between uses, the assignment of spectrum between different users is a process conducted and managed by the national authorities. It is executed with the help of various alternative spectrum management methods discussed in chapter 3.

AIP payments can influence both assignment and allocation, since the license renounced as a result of the payment ends up with a new user, which may operate in a different market than the previous owner, i.e. have a different use for the resource. However, in practice there is likely to be much greater discretion over the assignment of rights in the short to medium term, as international agreements and technical restrictions restrict the changes in allocations. The evaluation of both allocative and assignment impacts is however of importance, since concentrating only on assignment may neglect significant opportunities to enhance efficient use of spectrum. This is easy to understand knowing that efficient use of spectrum resources requires allocating them to the most socially profitable uses; if this cannot be done due to restrictions imposed then some of the potential social benefits are lost. As a matter of fact, it has even been suggested that improvements in allocation impose significantly larger efficiency gains than improvements in assignment (Cramton 2003, p.28). The described definitions of allocation and assignment are followed throughout this thesis.

The structure of the thesis is as follows. First the economics of radio spectrum such as demand, supply and efficiency determination as well as the concept of spectrum value are introduced in chapter 2 to build a basis for further discussion. Then alternative frequency assignment methods are described and compared in chapter 3, especially with respect to their fulfillment of economic efficiency, as it is the main objective of AIP utilization and the spectrum management in general. The current and suggested methods for calculating AIP are introduced and compared in chapter 4, again mainly in relation to the concept of economic efficiency. Chapter 5 summarizes the thesis findings and makes suggestions for further studies.

2. Economics of radio spectrum

This chapter goes through the fundamental economic characteristics of radio spectrum, i.e. the part of the electro-magnetic frequencies which covers frequencies from approximately 10kHz to 300GHz and can be used mainly for communications purposes. These characteristics are essential in understanding the ways in which radio spectrum is managed; including the motivation behind Administrative Incentive Pricing payments. The characteristics discussed include supply and demand, efficiency in the context of spectrum, as well as the concept of interference as the main motivation behind spectrum management. In addition, it is quite natural (and yet very rare) to draw parallels between spectrum and traditional natural resources (such as land, fisheries or fossil fuels) since the resemblance in many characteristics is evident. Thus, the viewpoint of spectrum as a natural resource is discussed in order to better understand the possible alternative ways of viewing spectrum resources. The chapter is concluded by a discussion on one of the key concepts of this thesis, namely spectrum value and its determinants.

It is noteworthy that even these basic characteristics are not all unambiguous; they may not be mutually agreed upon or have not been studied extensively. A good example of a disagreement is viewing spectrum supply as limited, an approach changing due to technological development. Another example of a new approach to economics of spectrum yet to be extensively studied is the previously described viewpoint of spectrum as a natural resource. Since both examples can have significant effects on the way spectrum can and should be managed (see chapter 3 next for this discussion) they are also brought up in this chapter discussing the fundamental economic properties of spectrum.

2.1 Demand

Frequencies can be regarded as inputs of production; they have no intrinsic value, but their value is constructed through utilizing them to produce different kinds of products and services. No other resources are required to produce spectrum, but some factors of production such as labor and capital are needed to make use of it. Thus, the total demand for spectrum can be derived from the demand of the end products and services produced by using spectrum as an input (Indepen, Aegis and Warwick, 2004, p.17). These products and services are multifold and include both commercial and public services. Examples of the former include TV and radio broadcasting as well as mobile phone services, GPS devices and wireless consumer electronics such as microwaves and automatic garage door openers. The latter include services of the military and other public safety promoters, the use of emergency frequencies being one example. In addition, radio spectrum is used to provide products and services utilized by both private end consumers (citizens, firms) as well as public entities: examples such as navigation and aviation applications are the most common ones.

Frequencies differ in terms of their physical properties, which the technologies utilizing them have to account for.[3] Simply put the range and penetration power are higher with lower frequencies and thus less infrastructure (cell sites including masts, towers and related equipment such as transmitters) is needed to cover a larger area. In contrast higher frequencies have a larger bandwidth capacity, i.e. they allow the signal to carry more data, which is why they are often utilized for example in the urban areas with many users. Due to the different characteristics different frequency bands are not equally useful for all purposes. This clearly implies that the demand for spectrum is not homogenous either; some frequency bands have significant excess demand while others remain relatively unused or specialized to certain applications. Good examples of the former are the 3G and 4G frequencies[4] which face a huge excess demand due to their commercial value especially for the telecommunications operators. The strong demand is reflected in the realized auction prices for these frequencies, which are throughout significantly higher than prices paid for other frequencies with different usages (see e.g. FCC and spectrum auctions). Examples of frequencies with a lower demand are frequencies of specialized usage such as the ones used for radio astronomy, which can utilize even the extremely high frequencies (EHF's) at about 30-300GHz. The demand is low simply because the utilization of these frequencies requires special equipment and technology.

The regulator can affect spectrum demand through incentive pricing, e.g. by setting AIP payments. The size of the effect of a price change depends naturally on the elasticity of demand of spectrum (marked by ϵ in the analysis in section 4.2). As spectrum is an input to production the relevant term to discuss is the elasticity of derived demand for spectrum. This in turn is dependent on (see Marshall 1920, Hicks 1932 and the Hicks-Marshall Law of Derived Demand)[5]:

[3] For a detailed description of spectrum's physical properties see e.g. Electromagnetic waves (InTech, 2011), available at: http://www.intechopen.com/books/electromagnetic-waves. In addition, the key properties affecting spectrum value are discussed in detail in subsection 2.6.3.
[4] 3G and 4G technologies operate mainly at 700MHz-2,6GHz frequency bands, for more information see e.g. International Telecommunication Union (ITU) at www.itu.int
[5] Elasticity of demand of an input (i.e. elasticity of derived demand) and variables affecting it were first discussed by Marshall (1920) and Hicks (1963) resulting in the so called Hicks-Marshall Law of Derived Demand. The law was originally constructed in the context of labor demand. Here it is applied in the context of spectrum demand

- The price elasticity of demand for all the services and end products produced using spectrum as an input
- The availability and price elasticity of supply of alternative inputs – and thus the production technologies available
- Spectrum's share of total cost of production

The price elasticity of demand for the services produced using spectrum as an input differs greatly between the services produced. The price elasticity for many commercial applications is very high (e.g. consider customers’ tendency to compare and switch their mobile operators), whereas the opposite is often true for non-commercial/public services (such as emergency services). As for alternative inputs for spectrum, substitutes do exist. Firstly, lack of spectrum can to some extent be substituted by increasing capital (i.e. building infrastructure). As an example one can think of a situation where a mobile communications firm lacks lower frequencies which require less infrastructure since their range is longer, but has a sufficient amount of higher frequencies, which may be as suitable for the technology utilized / service provided as the lower ones but due to their shorter range require more infrastructure. Thus, the lack of lower frequencies can to some extent be made up for by utilizing the higher frequencies by increasing capital K. In addition, as this example demonstrates a lack of one type of spectrum input (e.g. lower frequency bands) can be substituted by other frequencies (e.g. higher ones). Thus, since different frequencies are not identical, spectrum itself is a substitute for spectrum. The substitutability naturally depends on the technologies available and their suitability for different spectrum bands.

Traditionally the cost of spectrum relative to total costs of production has been nonexistent, since spectrum has been assigned practically for free, except for the relatively small cost recovery fees imposed by the regulators to cover their spectrum management costs (see the next chapter 3 discussing alternative spectrum management methods for more detailed information).

The demand for spectrum is also affected by innovation: development of more efficient technologies may result in demand decreases as fewer spectrum resources are needed to produce the same amount of output.

2.2 Supply

By nature spectrum is a common access resource since it is available to anyone. It is an intangible resource which, as previously mentioned, is not produced or refined from anything; it just exists as a part of the electromagnetic spectrum. However, in most countries practicing spectrum management most of the spectrum is owned by the state and leased (using licenses) under various terms of use. Thus, spectrum frequencies are effectively made de facto private goods through regulation. An exception is the so-called unlicensed spectrum, for which exclusive usage rights are not imposed. Unlicensed spectrum is discussed in more detail in subsection 3.5.3.

Since spectrum cannot be used up (although it can in theory be in full utilization), it is also a non-exhaustible resource. Despite its non-exhaustible nature the supply of spectrum is fixed in a sense that there exists a certain, finite amount of spectrum – the relevant spectrum bands under discussion here being radio spectrum bands, which cover frequencies from approximately 10kHz to 300GHz and are usable for communications purposes (Cave, Doyle & Webb 2007, p.4). However, as was previously discussed in relation to spectrum demand the heterogeneity (of spectrum use and thus demand) implies that even though the overall spectrum resources are regarded as scarce, there exist also spectrum bands with excess supply.

In addition, the development of more efficient technologies, such as cognitive radios and ultra-wideband[6], enables a higher capacity utilization rate of frequencies which lessens the scarcity of frequencies. As Staple and Werbach (2004, p.50) state: "…the extent to which there appears to be a spectrum shortage largely depends not on how many frequencies are available but on the technologies that can be deployed". This in turn implies that if more efficient technologies could develop rapidly and with relatively low costs, spectrum as a resource would evolve from scarce to abundant. This would have profound effects on the way spectrum could and should be managed. The current mainstream view however is, that spectrum resources are limited since at least for now technologies efficient enough to challenge this do not exist, they are too costly or complex to be employed commercially or can only be utilized with respect to some of the frequencies (Cave et al. 2007, p. 11-23). This fact is significant since the idea of scarcity is key in justification of spectrum regulation in general and thus the use of AIP, as will be discussed next.

[6] Cognitive radios refer to a technology which allows radios to move across the frequency band seeking for free spectrum capacity which can then be utilized. Ultra-wideband in turn refers to a technology, which can be used at a low energy level for short-range, high-bandwidth communications using a large portion of the radio spectrum.

2.3 Interference, spectrum re-use and the need for regulation

Scarcity of spectrum resources combined with high demand causes congestion which in turn often results in interference between different spectrum users. In other words, when users of limited spectrum resources "transmit at the same time, on the same frequency and sufficiently close to each other they will typically cause interference which might render both of their system unusable. Even if users transmit on neighboring frequencies, they can still interfere since with practical transmitters signals transmitted on one channel "leak" into adjacent channels, and with practical receivers signals in adjacent channels cannot be completely removed from the wanted signal" (Cave et al. 2007, p.3). The first type of interference is called co-channel interference and the leakages into adjacent channels adjacent-channel interference. Thus, interference is a negative externality imposed by one user of spectrum on other users and it is driven by congestion.

From the early days of spectrum management interference has been the main reason justifying the need for spectrum regulation by the governments and their agencies. As Melody (1980, p.393) summarizes: "Cooperation among all users is essential if the spectrum is to be used effectively by anyone." Another important reason behind regulatory needs is that sufficient access to spectrum has to be ensured for applications of social or public value. Examples of these kinds of applications are the emergency services.

A national regulator is a natural entity to execute the required supervision in coordination with other national regulators as well as international organs. Even though one cannot rely solely upon the market mechanism in achieving efficiency regulators can utilize market based mechanisms such as auctions or market outcome mimicking incentives such as AIP to achieve superior outcomes relative to alternatives such as pure administrative assignment of spectrum (Cave et. al 2007, p.171). The form of regulation and the regulative authorities are discussed in more detail in chapter 3.

Since frequencies differ in their demand the constraints imposed by interference are not equal for all parts of the radio spectrum. For frequencies with excess supply interference poses no problem whereas for high demand (the most valuable) frequencies careful management is needed to ensure that as many users as possible are able to utilize the resource without unnecessary interference. One way of accounting for the differing interference constraints when defining the AIP payments is by combining basic graph theory with economic theory as will be shown in subsection 4.2 along with the optimal AIP calculations by Levin and Rickman (2007).

Interference requires the use of different frequencies for some communications, but the possibility of sharing frequencies exists for others. Frequency reuse means using the same radio frequencies on radio transmitter sites within a geographic area, provided that they are separated by sufficient distance in order to minimize interference (Althos - GSM tutorial). In other words, to avoid harmful interference adjacent areas (sometimes called cells) use different frequencies, but geographic areas which are sufficiently far away from each other may use the same spectrum resources. In practice this is seen for example in the fact that in different countries the same frequencies are (re-)used; e.g. operators in Finland and Sweden may both use the same frequencies to provide mobile services.

2.4 Efficiency and spectrum

The key purpose of spectrum management is to maximize the value of frequency use to society by encouraging *efficient use of spectrum* and thus allowing as many efficient users as possible while also ensuring manageable interference between users (Cave et al. 2007, p.3). In order to create spectrum management tools to try to fulfill this objective one must have a clear understanding on what is meant by efficiency in spectrum policy. This in turn is not as unambiguous as one might think; various practices in determining efficiency in spectrum policy exist (concepts such as allocative, productive, technical, dynamic and functional efficiency are often mentioned) and they are not always correctly understood by regulators designing the policy tools. In some cases, these efficiency measures can even be conflicted and achieving one does not guarantee that other efficiency indicators are successfully fulfilled. An example of this kind of a situation is an auction which manages to allocate spectrum to the bidders who value them the most and can use them most efficiently in the long term. Still, in the short term these acquired resources may be left fallow, which indicates that even though the so called allocative efficiency (see the definition below) is fulfilled, technical efficiency requiring constant full utilization of the resource is left unfulfilled (for a more detailed discussion see e.g. Freyens & Yerokhin, 2011).

In general efficiency is concerned with the society utilizing scarce resources such as spectrum in order to satisfy differing needs of various agents, which include for example consumers, firms and the regulator as a social planner. To be more specific, this thesis follows the approach given by

Indepen et al. (2004) and followed by Cave et al. (2007) as well as Doyle (2007), which suggests that efficiency in the context of spectrum management should be understood as economic efficiency. Economic efficiency accounts for both static (efficiency existing at a point in time) and dynamic efficiency (efficiency over time) by having three dimensions: Allocative, productive and dynamic efficiency.

Allocative efficiency has to do with the economy producing the most desired types of goods and services in a way that Pareto optimality is satisfied (Indepen et al. p.20). In other words, spectrum should be allocated across different uses in a way that the mix of goods and services produced is optimal; no other mix can increase the well-being of one economic agent without harming the well-being of another agent[7]. Allocative efficiency can thus be improved by encouraging the utilization of spectrum as an input in the production of products and services most valued by the consumers.

Productive efficiency refers to producing the goods and services at the lowest possible cost where cost is measured in terms of inputs such as capital, labor and spectrum (Doyle 2007, p.2). Thus, being productively efficient implies producing on the production possibility frontier. As productive efficiency requires that no inputs are "wasted" it is closely related to technical efficiency, a concept frequently emphasized in spectrum management. Technical efficiency is concerned about the utilization rate of spectrum, i.e. that the maximum output is produced with a minimal amount of inputs, and it is often seen as an integral part of productive efficiency (see e.g. Doyle 2007). It is important to note the additive nature of the different efficiency measures. For example technical efficiency is a part of the overall economic efficiency, but it can as well occur while the overall economic efficiency is left unfulfilled: the spectrum resources may be fully utilized, but by non-optimal users.

Dynamic efficiency in turn refers to frequencies being allocated and used in a way that encourages (an optimal amount of) innovation and R&D (Cave et al. p.170). It can also be interpreted as allocating the inputs to production over time in a way which maintains productive and allocative efficiency in response to changes in technology and consumer preferences (Ofcom, 2006, p.54). Thus, through dynamic efficiency it is possible to further improve efficiency over time; for example by investing in the development of new technologies a spectrum (license) owner can increase the

[7] It is noteworthy that a Pareto efficient solution is not necessarily socially optimal; there may be a way to increase one economic agent's welfare more than another agent's welfare decreases which increases social welfare.

utilization rate of frequency and in economic terms shift his/her production possibilities frontier outwards producing more with the same resources.

When all the conditions necessary for economic efficiency are achieved, the economy satisfies the requirements of perfect competition thus enabling the attainment of socially optimal solutions as well. Markets utilizing spectrum naturally do not satisfy these strict requirements: there exist externalities, the most important of which is the interference discussed above. The consequent need for regulation as well as various transactional, administrative as well as political constraints restrict the flexibility to allocate and assign spectrum to the most socially valuable users and uses. Thus, the regulators must compromise between different objectives while accounting for interference. These issues are discussed in more detail in chapter 3 in connection with spectrum management and its methods.

2.5 Natural resource properties of spectrum

Spectrum as a resource with significant value to the society and established need for regulation is in many fundamental ways similar to the traditional natural resources such as land, forests or fossil fuels. Actually, many governmental actors as well as researchers seem to be agreeing on the issue that spectrum is, at least to some extent, a natural resource, but have yet to fulfill their implicit promise to treat it as such (Ryan 2005). This subchapter shortly discusses the connection between natural and spectrum resources. The issue is of relevance, since it may have impacts on AIP's (as well as other current pricing and assignment methods') use and validity in the future. On one hand, relating spectrum to natural resources might mean that current economic models for spectrum regulation must be adapted to accommodate the special features of natural resources. On the other hand, this might open up new possibilities of regulating and pricing spectrum.

Spectrum possesses a variety of features that have traditionally been regarded as properties of natural resources. It is similar to air or sunshine in a sense that they are all ubiquitous, i.e. can be found and exist everywhere. In principle spectrum is also non-excludable and non-exhaustible (even though it can be fully utilized). The scarcity and possibilities for externalities have however resulted in a situation where spectrum has become a common property resource, collectively managed by governments (nationally and through international cooperation) and leased under various terms;

much in the way in which for example land tenures function. Another way in which spectrum resources resemble land resources is the heterogeneous nature of both.

However, spectrum resources also differ from natural resources in some elementary ways. In addition to being non-exhaustible, spectrum resources are also instantly renewable; whenever a certain application stops using a frequency the same frequency becomes usable for any other application. In any case an interesting question arises: what could we learn from all the existing research and policies applied to natural resources?

2.6 Spectrum value

Considerations of spectrum value are essential to administrative pricing decisions, since the basic idea of AIP is to encourage efficient use of spectrum resources through a fee which reflects the value of spectrum in its optimal, feasible use. This fee then gives spectrum users incentives to re-consider their current use and need of spectrum. As a result spectrum resources can be re-allocated or re-assigned from lower value use to higher value use implying better overall efficiency.

However, the problem a regulator faces while reflecting on spectrum pricing is the complex nature of spectrum value. This is well explained by ITU (2012, p.1) which states that the building blocks of spectrum value are as much political and socioeconomic as they are purely financial. Financial value refers to the value derived through market sales of spectrum (e.g. auctions), which actually reflects the private value of spectrum to its users, i.e. what they are willing to pay for the resource. Due to market imperfections this is usually not consistent with the value of spectrum to the society (or social value), which also accounts for an array of objectives of the regulator (the social planner), such as market structure (see section 3.1 for a detailed description of spectrum management objectives). These objectives may not always be commercially viable, but are seen as socially preferable; such as the rollout of services into rural areas. Valuing spectrum also raises the question of whether frequencies used to produce public services such as defense and emergency services should be priced and thus be subject to AIP in the first place.

This section 2.6 discusses the possible inconsistency between private and social value of spectrum along with key spectrum drivers. Since spectrum prices reflect (private) spectrum value and are in a market environment determined by demand and supply this section is strongly linked to those of demand (2.1) and supply (2.2.). Thus all previous discussion naturally applies here.

2.6.1 Private value

It seems plausible that a rational firm values access to spectrum, or in the presence of a license regime the license, based on expected (discounted) future returns provided by the access. Aegis and Plum (2008) offer a useful framework to examine private spectrum value for the spectrum utilizing firms. According to them the total spectrum value (TV) for the firms consists of two elements: the expected net present value of future returns (NPV) and the option value (OV). The expected net present value of returns include returns from spectrum use i.e. from enhancement of existing services or creation of new ones (termed as the project based value, PV) and defensive or strategic value (DV) from gaining additional profits by utilizing some level of market power. The defensive value is thus concerned with acquiring spectrum resources to protect ones market share e.g. by restricting entry of new players or raising competitors' costs. Defensive value is assumed to be nonnegative based on the presence of imperfections in the market.

The concept of option value becomes relevant when there exists significant uncertainty over future applications and their value and there are sunk or irreversible costs associated with investments (Aegis & Plum 2008, p.9). It refers to the value of flexibility spectrum offers even if left partly or totally unutilized; it offers the spectrum (license) holder better abilities to respond to changing circumstances by keeping the spectrum "on hold". Positive option values would imply clearly positive prices even for spectrum resources for which there exists excess supply. The determination of spectrum value as defined by Aegis and Plum (2008) can be summarized as:

$$TV = PV + DV + OV, \text{ where } DV, OV \geq 0 \text{ and } PV + DV = NPV \text{ of returns} \qquad (1)$$

However, the application of regulatory control as well as the competition law implies that at least in theory the significance of defensive value should be small and it can be largely ignored when assessing spectrum values. In addition, spectrum licenses often carry with them different kinds of coverage and rollout conditions, or even straightforward "use it or lose it" –type of terms. This in turn implies that the spectrum license owner rarely gets to keep the acquired spectrum unutilized. Thus, the role of option value is also assumed to be quite insignificant. So, the main component of spectrum value from the users' point of view is thus the spectrum's ability to generate revenues/profit for the firm.

The current methods of determining AIP payments are mainly based on the idea of the opportunity cost of spectrum use (see chapter 4 for more details) and thus reflect the private value of spectrum. The same is true for spectrum resources that are auctioned (or traded in secondary markets), since the bidders in the auction naturally bid according to their own private valuations of the spectrum resource. This emphasis of private value may however differ from the value of spectrum to society, especially when the markets are not perfect. The social value aspect is discussed next.

2.6.2 Social value

According to basic economic theory, in the absence of externalities the private optimum level of production equals the social optimum. Thus, in such a market the social value of spectrum equals private value, i.e. the valuation of the most efficient firm. This valuation in the absence of market distortions was depicted in the previous subsection by the gains from enhancement of existing services or creation of new ones, i.e. the project based value (PV). As was already acknowledged however, in the market for spectrum the externality of interference exists implying that social and private valuations of spectrum differ.

Furthermore, if the firm acquires significant market power upon obtaining the license, social and private values diverge (McMillan 1995, p.193). Obtaining market power would be indicated by a positive defensive value parameter (DV) in the previous subsection 2.6.1. This also seems to often be the case for spectrum resources since the current license assignment methods have in many countries led to highly concentrated market structures (Milgrom, Levin & Eilat 2011, p.12). Therefore, the optimality of using private valuation information as a basis of pricing spectrum is further challenged. However, private spectrum valuations may be the only feasible valuations available or at least most easily attainable, since e.g. information on realized auction prices is available. In addition, they may be close enough to the optimum given that the distortions (such as externalities and market power impacts) are relatively small. Aegis and Indepen (2005, p.5)[8] also acknowledge that opportunity cost estimates used as a basis for AIP fees may not need to be adjusted to account for the social value, because the opportunity cost estimates are calculated in the presence of policies, such as coverage requirements, that are designed to promote the social aspects.

[8] This article focuses specifically on pricing frequencies that are used to provide broadcasting services, but the same conclusion can be drawn for all spectrum resources

Thus, in addition to taking into account the gains from enhancement of existing services or creation of new ones social value includes considerations about factors such as market structure and investment regulation – for example ensuring investments and thus the existence of services in rural areas. The mechanisms used to calculate AIP payments, introduced in chapter 4, differ with respect to their considerations on private versus social value.

2.6.3 Key value drivers

Valuation differences between spectrum bands result from various spectrum properties, physical as well as other, a part of which were already discussed in subchapter 2.1. Trying to provide a complete description of the properties would be an onerous task providing very little benefit for the further analysis. However, it is useful to identify the key value drivers in order to be able to discuss the AIP payment formation and justification in the forthcoming chapters. For this purpose a classification of the key value drivers is formed based on Smith & NERA (1996) and Aegis and Plum (2008).

While considering the value of a certain spectrum asset, the following aspects should be taken into account:

A. Frequency amount
   - Bandwidth
   - Area sterilized

B. Frequency properties
   - Propagation characteristics
   - Location of use – e.g. urban vs. rural
   - Possible international harmonization

C. Existence of alternatives
   - Utilization possibilities for different applications
   - Re-use opportunities (frequency sharing)
   - Congestion level

D. Other qualitative factors
    - Convenience of use
    - Ease of equipment availability
    - Maintenance or quality of transmissions

From the point of view of AIP, which is most often determined based on opportunity cost, the especially interesting aspects relate to the existence of alternatives (point C above). Naturally the more applications can use the band the more demand there is and thus the value of the spectrum band is increased. The same value increase through demand explains why congested bands are more valuable than uncongested ones. Possibilities for frequency sharing naturally reduce congestion and thus can be expected to have an opposite effect on the value of the frequency.

Chapter 2 discussed the fundamental economic characteristics of spectrum thus building a basis for further discussion of spectrum allocation, assignment and pricing as well as the motivation behind applying AIP. As spectrum is an input of production its demand was shown to be derived from the demand of the end products and services produced using frequencies. The heterogeneity of demand was also justified and main reasons behind varying demands across different spectrum bands were explained to be caused by the heterogeneous nature and the vast amount of different applications for spectrum resources. The scarcity of spectrum supply was shown to be a controversial issue; however the current understanding being that the supply is limited. The concept of interference was established as the main motivation for spectrum management and efficiency in the spectrum context was defined as economic efficiency, which was shown to include the concepts of allocative, productive and dynamic efficiency. As a curiosity spectrum was paralleled with natural resources, since they are similar in many ways and this approach may arouse new ways of managing and pricing spectrum. Finally, one key element of AIP, spectrum value was discussed from the private and social point of views. Private spectrum value was broken down into project and defensive values (forming the total returns from spectrum usage) and the option value. It was also shown that this private value is likely to divert from the social value of spectrum whenever market distortions are present. Finally, the key value drivers were identified as the amount of frequency (bandwidth), its key properties (propagation characteristics, location of use and harmonization), the existence of alternative uses and possibilities for sharing.

Next the different spectrum management practices, i.e. alternative ways of assigning and pricing frequencies, are introduced and compared. The comparisons are made especially with respect to the fulfillment of economic efficiency. The main objective of chapter 3 is to give the reader a thorough understanding on alternatives for the AIP method and advantages as well as downsides of AIP relative to other spectrum management methods.

3. Spectrum management – description and comparison of the alternative methods

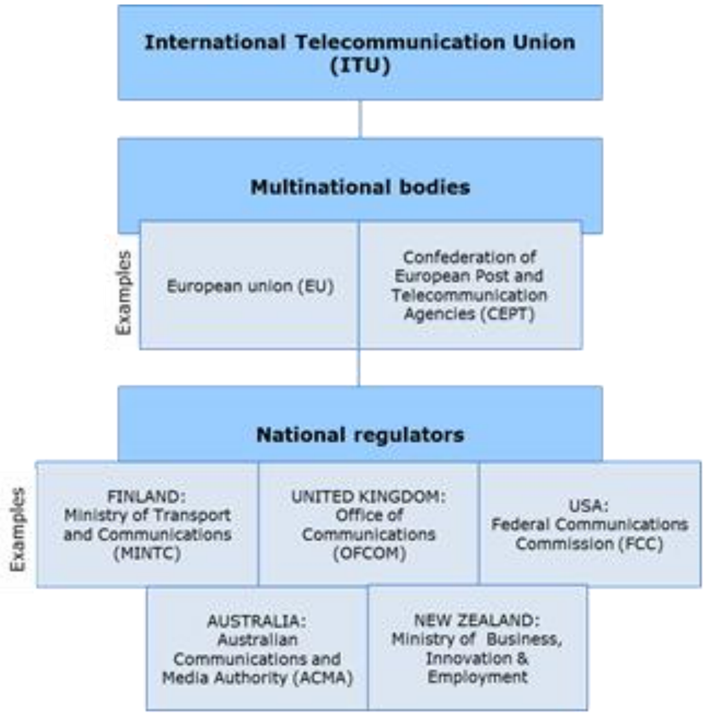
This chapter offers a description of spectrum management, its objectives as well as key methods. The different methods are also compared with respect to the attainment of the objectives of the regulator ("social planner") and especially relative to economic efficiency (i.e. allocative, productive and dynamic efficiency), since that is the main goal to be fulfilled by the use of AIP. The main focus is naturally on AIP and the comparisons are made in orders to understand when and why AIP is preferred as a pricing method relative to other alternatives and how it is used to complement alternative spectrum management methods.

3.1 Spectrum management and its objectives

The main objective of spectrum management is naturally to ensure that the value to the society from scarce spectrum resources is maximized. This is done by allowing as many efficient users as possible while keeping interference at an acceptable level. To fulfill this task spectrum is allocated to different uses and further assigned to users.

As was discussed in 2.3 the need for regulation is justified by the existence of interference. Due to the fact that interference can extend beyond national geographical boundaries and since there exists also inherently international uses of spectrum, such as aviation, spectrum management needs to operate at international as well as national level (Cave et al. 2007, p.5). The figure 1 below, constructed based on Cave et al. (2007) illustrates the international spectrum management framework. The examples given for each level of regulation (light blue boxes) consist of the ones discussed in this thesis and are not meant to be exhaustive.

Figure 1. The spectrum management framework



Source: Cave et al. 2007 (modified)

The International Telecommunication Union (ITU) is an agency of the United Nations responsible for information and communication technologies (ITU 2012). It accounts for the highest level of spectrum management allocating the global radio spectrum to different uses, which vary from prescriptive, such as satellite, to uses allowing significant interpretation and variation, such as fixed or mobile (Cave et al. 2007, p.5). Under ITU there are the multi-national bodies further coordinating and harmonizing spectrum management across regions, such as the European Union (EU) and Confederation of European Post and Telecommunication Agencies (CEPT). National regulators, such as the Ministry of Transport and Communications in Finland (MINTC) then operate within the guidelines provided by the international regulating bodies. It is important to note that the guidelines provided are non-binding, but deviations by individual countries are expected not to cause interference on other countries (Cave et al. 2007, p.6).

Since the use of a certain frequency band is often defined through international coordination and harmonization in the way described above (although the allocation might not be binding as such) the national regulator’s ability to allocate spectrum to different uses is restricted. Thus, spectrum management at the national level is usually concerned with assignment decisions within predefined uses rather than allocation between uses. However, the assignment methods introduced can be applied also to allocation of spectrum, where the national regulator assesses the possible benefits of different uses and allocates spectrum accordingly possibly deviating from the international guidelines. Consequently the spectrum management methods presented in this chapter are often referred to as allocation methods as opposed to assignment methods.

As was previously stated, in most countries the primary tool for spectrum management is a licensing system, which is a form of property rights. A spectrum license gives its holder an exclusive right to transmit at a given frequency. License conditions, which define the contents of the property right, i.e. what the license entitles the license holder to do and on the other hand what the holder is required to do, are multifold. They can for example be defined to restrict the particular technology that can be used such as GSM, or a particular use, such as mobile (Cave et al. 2007, p.105). However, many regulators nowadays express willingness to grant more service and technology neutral licenses imposing fewer restrictions on the use of spectrum (see e.g. Ministry of Transport and Communications in Finland 2012). Naturally this kind of deregulation can be claimed to increase economic efficiency through allowing spectrum to be used to produce the services most valued in the society (increasing allocative efficiency) with a technology that is regarded the most effective (increasing technical and thus productive efficiency). It also allows for experimenting with new technologies which may increase the dynamic efficiency through innovation. Thus, in theory the license conditions should be as unrestricted as possible implying that the national regulators would also be in charge of allocation of spectrum in addition to its assignment between users. Yet in practice the problem which arises is again the scarcity imposed interference; the more non-harmonized the usage terms the more likely interference is to occur between users. In other words, the existence of externalities (i.e. interference) requires deviations from socially optimal solutions.

As the social planner the regulator has a set of objectives to fulfill in order to maximize the spectrum's value to the society. The main goals of the regulator usually include the following (ITU 2012):

- Efficient usage and assignment of the spectrum resources
- Rapid and effective introduction of a new wireless technology (i.e. broadband wireless access or BWA)
- Reduction of the digital divide, through the development of wireless service in remote, rural or generally low population density areas

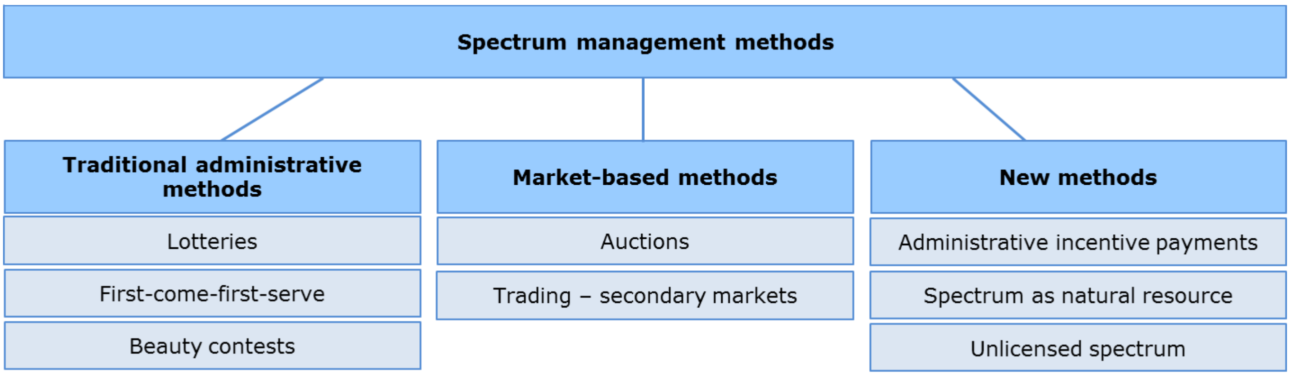
- Protection or promotion of social welfare and/or public service
- Minimization of potential interference and coexistence issues
- Government revenue generation

Some of the objectives, such as efficient usage of spectrum resources and promotion of social welfare are clearly complementary goals both increasing overall welfare. Instead goals such as increasing rural rollout do not necessarily increase the overall social welfare but are however imposed and desirable due to equity reasons. This is true for the rollout objective since such investments are costly with little revenue gained, and thus not economically viable for firms to make. The social losses made by heavy investments are unlikely to be recompensed by the increase in consumer surplus since very few customers are located in these sparsely populated rural areas. This example illustrates the contradicted nature of some of the objectives of the regulator and emphasizes the need for prioritization.

3.2 Spectrum management methods

As for the assignment and/or pricing methods by which key spectrum management objectives can be reached, regulators have three sets of methods in their use. Firstly there are traditional administrative methods, which do not include any marked-based processes but are, as their name suggests, purely administrative giving the regulator a lot of power over the assignment or pricing of the spectrum resources. These methods include lotteries, first-come-first-serve methods and beauty contests. Secondly, regulators may use market-based methods, which include auctions and secondary markets for spectrum, i.e. spectrum trading. These methods impose fewer information requirements for the regulators, since the regulators do not have to administratively set the price, but it is set by a market process. However, even the use of these methods requires careful planning of the framework to ensure optimal outcomes: aspects to consider include the terms with which participants are allowed to participate (e.g. financial credibility), possible bidding limitations (e.g. bidding caps) as well as the technical execution of the process (e.g. programs used). Thirdly, the regulators have in their use a set of newer pricing and assignment methods, which are here termed as “new methods”. There have not been many new approaches introduced into spectrum management during the last few decades, but the three most significant ones, AIP (combined with administrative methods), the viewpoint of frequencies as natural resources and the special case of unlicensed spectrum are discussed under the title “new methods”. Figure 2 summarizes the available spectrum management methods.

Figure 2. Spectrum management methods



Source: Modified from various spectrum regulator sources (ITU, Ofcom, MINTC, FCC)

When different assignment/pricing methods are compared, the essential setting which the regulator faces is that of the principal-agent setting familiar from economic theory. The task of spectrum assignment and pricing can be thus thought of as a "principal-agent" type of game between the regulator ("principal") and the users of spectrum ("agents"). The problem the regulator faces while deciding the optimal mechanism for dividing frequencies as well as governing their use is that of asymmetric information. The task would be easy if the regulator possessed sufficient information about the individual valuation (and thus the cost structures as well as technological solutions available) that spectrum users have for spectrum. In that case the choice of the assignment method would be insignificant, since the regulator would always be able to offer Pareto efficient solutions, i.e. optimal prices for the different spectrum licenses. In the real world however, the problem of adverse selection exists as the regulator is imperfectly informed about the characteristics of the spectrum users. The existence of this problem makes the different allocation methods unequal from the economic efficiency point of view or at least imposes many requirements on the information the regulator should have in order to price spectrum administratively in a way that promotes efficiency.

Market-based mechanisms which are designed to reveal the valuation without the regulator needing to determine it "based on an enlightened guess" are preferred in this respect. However, there may be other reasons related to e.g. political pressure or some social value such as equity which favor the use of administrative methods. The AIP in turn is a method "in the middle" in a sense that it is combined with administrative methods to incorporate market-based incentives to them. The alternative methods as well as benefits and weaknesses of each alternative are discussed next so that each of the alternative approaches constitutes its own subchapter. The main focus is on the attainment of economic efficiency, i.e. how well the alternative methods satisfy allocative, productive and dynamic efficiency goals. This is due to the reason that achieving economic efficiency is the main goal of AIP utilization (see e.g. Ofcom 2010).

3.3 Traditional administrative methods

First we concentrate on the traditional assignment methods, which as such impose no price on frequencies (although it is common for the regulators to impose small cost recovery fees), but can be combined with pricing schemes such as the AIP. These include lotteries, first-come-first-serve methods and beauty contests, sometimes also called hearings. Administrative licensing methods, especially beauty contests, are still widely used even though market-based mechanisms have increased their popularity during the last couple of decades.

3.3.1 Lotteries

Lotteries are random selection processes whereby the licensees are selected by chance. Due to their obvious lack of any kind of systematic aspiration towards economic efficiency, or for any of the other key objectives of spectrum management depicted in 3.1, they are nowadays rarely used in spectrum management (FCC 2012). The key motivation behind their use, mainly in the 1980's, was that they succeeded in assigning the licenses quickly. However, as demand started to grow, this became evermore challenging. An illustrative example of this is the appliance of lotteries in the U.S. in 1982, when beauty contest awards lacked severely behind causing costs to the applicants, the government and ultimately to the public as forgone services (McMillan 1994, p.4). Consequently the government tried replacing the beauty contests with lotteries, but the prospect of windfall gains attracted nearly 400 000 applications, some of which were submitted by users not being technically competent to develop and operate the subsequent spectrum utilizing services.

3.3.2 First-come-first-serve approach

The first-come-first-serve (FCFS) approach, according to its name, is an administrative decision assigning the license(s) to the first credible applicant(s). It is typically used for low-valued frequencies with weak demand, since for those frequencies there will be no necessity to resolve mutually exclusive or competing requests (ITU 2010, p.17). The FCFS method was especially popular before the 21st century as there was enough spectrum in almost every band to accommodate most or all users and permit adequate separation among potentially incompatible uses (Neto &

Wellenius, 2005, p.2). An example of frequencies assigned using the FCFS method are link frequencies in Finland.

The credibility of applicants requires that in order to be granted the license the applicant must adhere to certain technical standards and regulations. This aspiration towards selecting a credible applicant is what essentially separates the FCFS method from lotteries implying slightly more efficient assignments of the resource. However, savings in administrative costs of allocation/assignment relative to more complex management methods seems to be the only significant benefit the FCFS method offers. In addition, this benefit is likely to be revoked by the probability of not assigning spectrum to the user valuing it the most, i.e. the probability that the most efficient user of the resource is not the quickest to respond to the offer.

3.3.3 Beauty contests

In a beauty contest the awarding authority (regulator) releases an invitation to bid for the spectrum licenses in question. The invitation contains a set of criteria, such as population to be served (i.e. coverage), speed of deployment, project viability, spectrum efficiency and ability to stimulate competition, based on which responses are evaluated. The selection criteria can be weighted depending on the objectives of the regulator. After the responses to the invitation have been submitted the awarding authority judges the quality of applicants' responses against the criteria and assigns spectrum licenses accordingly. (Ofcom 2012, FCC 2012) Since a beauty contest includes actually assessing the benefits the applicants would bring to society instead of assignment based on draws or FCFS principles it is the most interesting administrative spectrum management method as far as economic efficiency is concerned.

Beauty contests are still widely used, processes well established and understood by the regulators (tenderers) as well as the applicants. For example in Finland, licenses even to the most valuable radio frequencies have traditionally been assigned using beauty contests. Market based mechanisms were first tried as the 2,6GHz licenses were auctioned to mobile communications operators in 2009. The second frequency auction, for one of the most valued spectrum bands in the 800MHz area is scheduled to be held during 2013 (MINTC 2012).

The main issues with beauty contests relate to the low benefit-cost ratio: such comparative processes can be very time-consuming and resource-intensive (especially related to more straightforward traditional methods) and yet not able to assign spectrum to the agents valuing them the most, thus leading to non-optimal solutions from the efficiency point of view. Administrative decisions and their impartiality are easy to call into question and whereas the regulator may be able promote national/societal goals more easily through tendering terms, comparative processes are in practice often decided on the basis of minor differences among applicants (ITU 2010, p.18). The award usually favors established companies (e.g. incumbents), since they are able to cite a track record to support their case (Ofcom 2012), which may impair dynamic efficiency by restricting the market entry of innovative, new companies.

In general, as the traditional administrative methods do not impose any price on frequencies, the issue of economic rents or windfall profits arises. In addition, users of spectrum have few incentives to give up underused spectrum or on the other hand invest in spectrally efficient technologies or services. Next we will move from discussing pure administrative methods to market-based methods, which address and correct some of these issues.

3.4 Market-based methods

Instead of relying on a regulator to perform spectrum allocation or assignment, market-based methods are based on the assumption that market mechanisms, while properly monitored and supported, are the most effective way of complementing the task. Market-based assignment methods include auctions and spectrum trading or a secondary market for spectrum licenses.

3.4.1 Auctions

Auctions represent a market based pricing and assignment[9] mechanism whereby the price and the licensee of the frequency are determined in a bidding process. Aside from pre-determined requirements and conditions for the tendering process, only price matters. The idea of auctioning airwaves rather than assigning them through administrative licensing methods was first proposed by Leo Herzel in his 1951 article “'Public Interest' and the Market in Color Television Regulation.” followed by Ronald Coase in his 1959 article “The Federal Communications Commission”. However, it took approximately forty years before the administrative licensing methods described previously started to be replaced by market based pricing of frequencies. New Zealand was the first

[9] As was discussed previously, the methods discussed may in theory also be used to allocate spectrum between uses, but in practice this is very rare

country in the world to legislate spectrum auctions in 1989 and it also executed the first auction only a year later in 1990 (McMillan, 1994, p.147-148). Another pioneer has been the United States which held its first spectrum auction in 1994 and has since executed nearly a hundred auctions (FCC, 2012). In Europe the first auctions were held at the beginning of the 21st century as 3G licenses were assigned. Currently auctions are the dominant assignment method for spectrum and they are carried out especially when there exists strong competition for scarce spectrum with a high commercial value (Milgrom et al., 2011, p. 22).

Spectrum auctions are generally considered to be the most efficient spectrum management tools in achieving allocative efficiency. This is due to the fact that in a competitive bidding situation where only the price paid matters (in determining the winner(s)), the licenses are obtained by those who value them the most and are thus best equipped to utilize them effectively. This naturally presupposes that the firms are unbiased in their estimation of the future profits generated by the use of the resource, i.e. in determining the total value (TV) as described previously in 2.6.1.

The framework of value creation is also useful for identifying the key issues and possible shortcomings of auctions as a spectrum management method. As private firms evaluate the value of spectrum resources not only based on returns on spectrum use, but also based on the possible defensive or strategic value (termed as DV in 2.6.1) the spectrum resource entails, an auction could end up in a situation where a firm would be willing to bid on spectrum just to interfere with competitors. This also illustrates the problematic relationship between private and social value and the fact that these two often do not meet. Regulators have however in their disposal measures to prevent such a situation from arising. Auction conditions such as “use it or lose it” are examples of this. In addition, anti-competitive outcomes such as large operators acquiring an undue concentration of the available spectrum can be restricted by limiting the amount of spectrum one applicant may bid on (i.e. establishing bidding caps).

Another key pitfall related to auctions is the phenomenon of winner’s curse. When there is incomplete information, the winner of the tender process tends to overvalue the resource tendered and thus overpay (for empirical proof see e.g. Bajari and Hortacsu, 2003). This in turn may lead to lower investment level and thus hinder development and efficiency especially from the point of view of dynamic efficiency.

Even though auctions are considered the preferred assignment method to ensure initial efficient distribution of the spectrum, the question related to dynamic efficiency remains; how to ensure that spectrum continues to be used in an economically efficient manner in the future**?** As with other resources, economists recommend that spectrum users be allowed to transfer their spectrum rights (ITU 2010, p.18). The emergence of secondary markets for spectrum will be discussed next in subchapter 3.4.2.

3.4.2 Trading – secondary markets

In order to ensure that spectrum resources continue to be used efficiently, secondary markets may be needed. The core idea is very simple and equivalent to that of the basic principle of AIP payments: moving spectrum from lower-value to higher-value uses and users until the value of any marginal unit of bandwidth is equal for all, or until the cost of spectrum to any buyer equals its value to some next-best user.

The main constraint for creating any form of free trading in spectrum is the externality of interference. Most adversaries of free trade state that preventing interference among technically different services would require extremely complex engineering analysis and could lead to litigation among spectrum users. Other counterarguments include for example not satisfying socially desirable requirements and not being able to restrict anti-competitive outcomes. (ITU 2010, p. 17).

A totally free market spectrum approach has not been implemented by any country (ITU 2012, p. 32). However, spectrum trading is to some extent allowed in Australia, Guatemala, New Zealand and the United Kingdom (Doyle 2006, p.1).

Administrative Incentive Payments can be thought of as a gradual step towards trading frequencies as they enable and encourage users to give up frequencies that are either not utilized or are of more value to someone else. AIP will be discussed in short in the next subchapter (3.5.1) under "new approaches" for spectrum management as well as more extensively during the rest of this thesis (chapters 4-5).

All in all, market based mechanisms are preferred over pure administrative ones whenever there exists a sufficient amount of actors in the market (i.e. demand exceeds supply) and there is genuine competition between the players.

3.5 New approaches

As was stated at the beginning of this thesis there have been very few new ways of thinking about spectrum resources, their management and pricing during the last couple of decades. The most significant ones, discussed in this subchapter, include the Administrative Incentive Pricing (AIP), spectrum as a natural resource and the concept of unlicensed spectrum. All of these methods pursue to either find fundamental ways of treating and regulating spectrum differently than before (spectrum as a natural resource and unlicensed spectrum) or to complement more traditional methods in order to enhance efficiency of the resource use (AIP).

3.5.1 The Administrative Incentive Pricing (AIP) approach

As was previously explained, Administrative Incentive Pricing is a frequency pricing method, utilization of which presupposes some form of property rights in order to impose a fee on spectrum usage. Most commonly the property right regime consists of licenses offering exclusive usage rights to spectrum for a specified time and possibly with other usage requirements (e.g. regarding coverage). AIP is combined with the traditional administrative assignment methods, which themselves impose no explicit price for spectrum, to better mimic the market based outcomes. Thus, the AIP approach is not an assignment method, but a pricing method combined with administrative assignment methods.

AIP aims at imposing a market price for frequencies, which encourage spectrum users to give up spectrum that is either unused or otherwise valued less than the charged AIP payment. If AIP truly reflects the highest market valuation or highest value of alternative use, this encourages the transference of spectrum resources to agents valuing them the most.

The use of incentive pricing of spectrum as opposed to or as a complement for the pure administrative allocation and assignment methods was first proposed by Levin as early as 1970 in his paper “Spectrum allocation without market”. Levin’s approach to spectrum management envisaged an incremental path towards efficient pricing, with revealed and stated preference methods being used to reveal opportunity costs. Another economist promoting the issue was Melody who (1980, p.396) identified the substantial possibilities for economic rents or windfall profits to be gained by the firms utilizing spectrum with the contemporary spectrum management mechanisms.

AIP or an AIP like pricing (i.e. prices clearly above pure administrative fees) for frequencies is currently utilized in the United Kingdom, New Zealand, Ireland, Canada and Spain. In addition, its application has been considered in Finland. Many of the current applications are however still implemented without an explicit aim at doing it according to clear economic principles (ComReg, CRTC). The telecom regulator in the UK, Ofcom, defines the role of AIP as follows: “AIP should continue to be used in combination with other spectrum management tools, in both the commercial and the public sectors, with the objective of securing optimal use of the radio spectrum in the long term. AIP’s role in securing optimal use is in providing long-term signals of the value of spectrum” (Ofcom 2010).

Instead of lump sum payments (think of auction payments), AIP payments are charged as an annual fee from the spectrum licensees. Theoretically the sum of the net present values of AIP payments should be equal to the auction price paid for the same frequency. Due to its annual nature, it could be claimed that AIP lessens the need for the spectrum license holders to predict their revenue and profits streams far into the future, thus improving flexibility of the players to operate. At any point in time, when a spectrum user regards the value of spectrum to it less than the AIP payment, it is incentivized to give it up resulting in new assignment (or in some cases allocation) of the resource. In auctions on the other hand, the paid price as a whole is regarded as sunk cost.

AIP may also be used in connection with administrative methods such as beauty contests in order to account for both monetary and non-monetary objectives. In addition, it can also be used on frequencies dedicated to public use, which often are “favored” in spectrum assignments sue to their societal purpose. There are some frequencies for which neither AIP, combined with an administrative assignment method, nor auctioning can be used. Examples include the so-called unlicensed frequencies, which are discussed in subsection 3.5.3.

On the other hand, AIP payments assume that a regulator is able to set a payment that reflects the real value of spectrum to its users. In practice this is a challenging task due to asymmetric information, the many different uses that exist for specific spectrum resources (which one is the one bringing highest value?) and conflicts between private and social value already discussed in section 2.6 of this thesis.

Three different ways of determining the size of AIP payments have been proposed by economists and regulators. The starting point is usually the opportunity cost of spectrum use. The core method, called the Smith-NERA method is based on this principle. In addition to this method, there exist two alternative ways to calculate the optimal price; a method by Levine and Rickman (2007), which extends the Smith-NERA methodology to account for market structure and interference constraints, and a method which bases AIP payments on prices realized in market transactions. All of these methods and especially the most refined method of Levine and Rickman are discussed in detail in chapters 4 and 5.

3.5.2 Spectrum as a natural resource

As previously discussed, spectrum resources resemble natural resources in many respects. Thus, it would be only logical to explore the possibilities that existing natural resource regulation, pricing and trading schemes could bring to the discussion of managing scarce spectrum resources.

The idea of spectrum as a natural resource is discussed by many scientists, recently for example by Ryan (2004 and 2005) in his articles "Application of the Public-Trust Doctrine and Principles of Natural Resource Management to Electromagnetic Spectrum" and "Treating the Wireless Spectrum as a Natural Resource". Ryan concludes that there seems to exist a consensus on electromagnetic spectrum being, to some extent, a natural resource. However, it is not treated as such even though the current ways of regulation and spectrum management may be unsuitable given this fact.

As the focus is on the AIP model this thesis refrains from more detailed discussion around the topic, but brings the issue up as a potential and interesting topic for further studies related to spectrum management.

3.5.3 Unlicensed spectrum

As was previously discussed, the use of AIP presupposes a licensing regime with exclusive usage rights for the license holders. Thus, one cannot thoroughly cover the concept of efficient use of spectrum or meaningfully talk about the future applicability of AIP without also discussing the other alternative to exclusive licenses i.e. the so-called unlicensed (or license-exempt) spectrum. This subchapter provides a concise representation of unlicensed spectrum, its connection with economic efficiency and implications on the use of AIP.

Unlicensed spectrum simply refers to those frequency bands in which users can operate without a license. In other words, different users share the same spectrum resources (aptly often called spectrum sharing). Thus, these parts of spectrum are treated as non-excludable, but still rivalrous (due to interference) common-pool resources as opposed to the private goods approach of licensed spectrum. As with any commons the problem of overuse, congestion and thus interference is likely to occur, when spectrum users do not account for the externality of interference they impose on others (also called the tragedy of the commons in economic literature). In order to avoid excess interference the users must therefore use certified radio equipment and must comply with the technical requirements (e.g. power limits) set by the regulator. The regulator's, whether it is a national regulator or an international governing body, challenge to allocate spectrum resources to specific uses as described previously, includes the decision of how much spectrum to allocate to unlicensed uses.

The idea of unlicensed spectrum as such is not novel; there have always been unlicensed spectrum bands and before the discovery of the value of radio spectrum (due to technological development) and thus the strict regulation of spectrum, spectrum resources were inherently unlicensed. The reason why unlicensed spectrum deserves to be introduced under the heading of "new approaches" is its increased significance since the late 1990s and the current heated discussion around it.

The cause of discussion has mostly to do with dynamic efficiency, i.e. the ability of unlicensed spectrum to encourage innovation. This is due to the fact that many valuable innovations including spectrum, such as the development of Wi-Fi in the 2,4GHz band, have taken place on spectrum bands that are unlicensed (Milgrom, Levin & Eilat 2011, p. 1). This seems natural, since innovation is often best encouraged in an open environment; just think about all the technical applications developed in open-source environments without restricting property rights aspects, an example being the operating system Linux. Consequently, this has aroused the question of whether more spectrum should be allocated to unlicensed uses to encourage innovation. The development of interference restricting technologies has further intensified the debate since it has the possibility to overcome interference-related problems of unlicensed spectrum.

Since unlicensed spectrum lays aside any barriers of entry it is especially efficient in encouraging third-party innovation, i.e. innovation by parties who do not necessarily own any licensed spectrum. This is mainly because the innovators no longer have to seek and pay for the approval of current license holders to let them develop and test their ideas in the spectrum bands that they have no usage right to. The subsequent innovations can be substitutes for technologies utilized by licensed spectrum, thus increasing competition and efficiency of spectrum use; in which case the permission to develop them in the current licensees' spectrum is likely to be revoked. Alternatively they can be complementary technologies, which increase the total demand for spectrum related services thus increasing the value of licensed spectrum. An example of the latter is Wi-Fi, a technology allowing an electronic device such as a smartphone or a tablet to exchange data wirelessly (using unlicensed frequencies) over a computer network. The availability of Wi-Fi increases the demand for electronic devices that are able to utilize it, at the same time increasing demand for e.g. 3G mobile services (using licensed frequencies), since the devices are forced to use 3G outside Wi-Fi hotspots. The value increasing effects that the unlicensed spectrum has on licensed spectrum may even revoke the effect of revenues lost by the society, when instead of licensing spectrum and selling the licenses, spectrum is allocated to unlicensed free use. (Milgrom et al. 2011)

However, as any form of property rights, licenses protect the usage of the resource and increase predictability over future events - or alternatively decrease the risk of disturbances by e.g. competitors. Thus, licensing encourages the licensees to make related investments, such as building the infrastructure. As many uses of licensed spectrum, such as 3G and 4G wireless mobile technologies and radio as well as TV broadcasting, require large infrastructure investments licensing is a preferred method to ensure that these investment are made.

Another concern with licensed spectrum has been the technical and thus productive efficiency of spectrum use (Milgrom et al. 2011, p. 11). This is because exclusive licenses provide the licensees with a right to use the spectrum in question at all times and possibly all over the nation (i.e. national licenses) even if the resources are only needed at certain times a day or in certain geographical areas. For example, just as people tend to consume more electricity during the day than at night many spectrum utilizing services (think about e.g. mobile phone usage within Finland) are consumed during the day rather than during nights. This indicates that at certain times (or in certain areas) the spectrum resources are severely underused. Several studies confirm this underutilization of part of the licensed spectrum (see e.g. Santivanez et al. 2006, Cave et al. 2007, Calabrese 2009). As a solution to the underutilization problem the introduction of unlicensed spectrum or spectrum sharing has been suggested, but the interference constraints have discouraged rapid and substantial changes so far. One example of the first steps towards more sharing can however be seen in the U.S. where FCC in cooperation with the National Telecommunications and Information Administration (NTIA), responsible for managing the spectrum used for federal purposes, announced that it plans to take spectrum resources in the 3.5GHz area (specifically frequencies between 3550-3650MHz), which are currently used for radar and share it with wireless carriers (Arstechnica, 2012).

All in all, having both unlicensed and licensed spectrum is attractive due to the diverse advantages that they offer. In other words, the existence and benefits from unlicensed spectrum do not mitigate the applicability or necessity of other spectrum management mechanisms, such as AIP, but complement them. In addition, one might envision possible hybrid solutions combining elements from unlicensed and licensed spectrum. An example would be a mechanism where spectrum is unlicensed, but users pay an access fee depending on the level of congestion of the band (Cave et al. 2007, p.203). In this case AIP payments (at least in some form) might be applied also to unlicensed spectrum contradicting the prerequisite of a license regime.

In this chapter spectrum management and its key objectives as well as the alternative spectrum management methods were introduced. The alternative methods were also compared to each other primarily against their fulfillment of economic efficiency consisting of productive, allocative and dynamic efficiency. Spectrum management methods were shown to include the administrative methods, i.e. lotteries, first-come-first-serve methods and beauty contests, market-based methods, i.e. auctions and secondary markets for spectrum, as well as the newer approaches of AIP, the viewpoint of frequencies as natural resources and the special case of unlicensed spectrum.

With respect to economic efficiency the market-based methods were shown to be preferred. However, there was also shown to be situations where these cannot be utilized or where other objectives than economic efficiency necessitate the use of other methods; the most important case being unlicensed spectrum, which existence is claimed to encourage innovation increasing dynamic efficiency. Of the methods that are either market based (auctions, spectrum trading) or try to mimic market outcomes (AIP), auctioning was shown to be preferred for high demand and value frequencies under competitive settings whereas AIP can also be utilized in connection with administrative assignment methods and for pricing public service frequencies.

Due to increasing demand and multifold applications for spectrum resources, new ways of thinking about spectrum management, assignment and pricing where shown to be crucial. Regarding spectrum as a natural resource and applying management methods and pricing used in connection with traditional natural resources was identified as a key opportunity for further studies.

Spectrum management was shown to include the international level of coordination through organizations such as ITU, which oversees spectrum management and the allocation of spectrum to uses, multinational level coordination through organs such as the EU and a national level governed by the national communications regulators. The decision-making power of national regulators in practice determining the AIP payments was shown to be restricted by international coordination in a way that it mostly includes the assignment, but not the allocation decisions. The multiplicity and sometimes even contradictory nature of different objectives of regulators was also pointed out.

In the next chapter the existing alternative ways to calculate AIP payments are introduced, compared and analyzed with respect to the fulfillment of the requirement of economic efficiency. Chapter 5 then concludes the thesis.

4. Different methods of defining AIP payments

After the need for administrative pricing of spectrum is established the question of how to determine the fee arises. So far there have been three alternative ways suggested for calculating AIP payments.

Firstly, there is the method based on opportunity cost of spectrum use, which is currently utilized (with minor differences in calculation practices) in the United Kingdom and in New Zealand. This method is based on the fundamental economic understanding that a price based on opportunity cost guarantees that spectrum users cost spectrum resources as any other inputs in their production (Doyle 2007, p.1). This in turn implies productive efficiency. An observation of opportunity cost based pricing with spectrum resources originally dates back to Levin (1970) who stated that although a system of freely-transferable rights that works would be by far the best from a strictly economic viewpoint, it may be impossible to conduct (at least for all spectrum bands). Thus, we should be able to determine shadow prices that are derived from maximum sums that current spectrum users and systems designers would be willing to pay rather than do without some small amount of spectrum (these sums naturally referring to opportunity costs) to ensure efficient use of resources that cannot be priced by the market. This methodology was later on proposed in touch with spectrum pricing by Smith and NERA (1996) and further elaborated by Indepen, Aegis and Warwick Business School (2004). The elaboration mainly consisted of taking into account the possibility of re-allocation of spectrum between uses, while Smith and NERA initially considered only changes in assignments between users as a result of imposing AIP payments. According to the first developers this method will be referred to as the Smith-NERA method.

Secondly, Levine and Rickman (2007) have proposed a more rigorous method which builds on the Smith-NERA methodology. More specifically, Levine and Rickman have developed an optimal pricing scheme that allows for consumer surplus, interference constraints and their implications for productive efficiency, revenue implications and market structure. This method shall be referred to as the Levine-Rickman method from here on. Thirdly, it has been also suggested that AIP payments could be directly derived form observed market prices of spectrum generated in auctions or on the secondary markets for spectrum.

This chapter presents these three alternative ways developed and utilized up to date to estimate AIP payments. The focus is on the methodologies, not on the absolute values given per spectrum band, since the values strongly depend on the characteristics of technologies and uses of spectrum as was discussed in subsection 2.6.3 (the key value drivers). In other words, using these methodologies the AIP payments for different spectrum bands can be calculated, but some adjustments have to be made in order to take into account the individual characteristics of different uses and technologies utilizing the spectrum bands[10]. While discussing the different methods for constructing AIP payments I provide comments and critique related to the key assumptions. Separate sections 4.1.3, 4.2.6 and 4.3 summarize these discussions per model.

4.1 Opportunity cost based AIP payments – the Smith-NERA methodology

In their paper "Study into the use of Spectrum Pricing" prepared for the Radiocommunications Agency Smith and NERA (1996) construct a simple framework to examine AIP payments based on opportunity costs. Their primary focus is on assignment decisions whereas Indepen et. al (2004) extend this framework to cover also allocation changes caused by imposed AIP payments. Thus, the Smith-NERA methodology presented next accounts for both efficiency gains from shifting spectrum resources from inefficient users to efficient ones as well as efficiency gains from altering the allocation, i.e. the use of spectrum. However, the same constraints and hindrances for allocation changes that were discussed in the previous chapters apply.

4.1.1 Striving for productive efficiency

Introduction to the Smith-NERA approach is initiated by introducing the First Welfare Theorem stating that a competitive equilibrium (i.e. a Walrasian equilibrium) is a Pareto optimum. Thus, when perfectly competitive markets prevail in equilibrium the price mechanism establishes relative prices such that the cost to society of producing X in terms of Y reflects consumers' willingness to pay for such a transformation, i.e. the opportunity cost (Indepen et al. 2004, p.21). This theorem attests to the desirability of competitive markets. The Second Fundamental Theorem of Welfare Economics then states that out of all possible Pareto-efficient outcomes, one can achieve any particular one by enacting a lump-sum wealth redistribution and then letting the market take over.

[10] For a more detailed description of the adjustments needed in the AIP payment by use/technology see e.g. Ofocm 2010.

In other words, in a perfectly competitive economy where policy instruments are non-distorting, the 'first-best' welfare maximizing outcome can be achieved.

As far as spectrum management is concerned, the fact that equilibrium prices in a perfectly competitive market are in accordance with efficient outcomes brings us to the subsequent conclusion that prices equating supply and demand for spectrum are likely to promote efficiency. As economies in practice are not perfectly competitive, and appliance of lum-sum wealth redistributions is basically impossible, only second best outcomes are possible. However, according to the general theory of the second best[11] it is not desirable to set prices at 'first-best' levels when distortions persist elsewhere in the economy. In addition, according theory on optimal taxation[12], it is not recommended to tax the use of inputs when pursuing welfare maximizing outcomes in a second-best setting. This would suggest that the use of inputs in a competitive economy should satisfy conditions necessary for productive efficiency. (Indepen et al. 2004, p.21)

Indepen et al. (2004, p.22) conclude that when competitive markets exist government policy should be directed towards the promotion of competition where possible and desirable, and tax instruments should be used mainly on final goods and services to achieve second-best welfare maximizing outcomes. Given this, the use of spectrum should satisfy conditions needed for productive efficiency. If this holds, policy as a whole ought to be consistent with a second-best welfare maximum. As a corollary, setting spectrum prices that promote productive efficiency is desirable for efficiency.

4.1.2 A hypothetical example

After it has been argued that the use of inputs should satisfy productive efficiency, the relationship between spectrum usage, pricing and productive efficiency can be studied. Indepen et al. (2004) as well as Doyle (2007) provide simple, complementary examples utilizing basic microeconomic theory to discuss the link between efficiency and spectrum pricing. These examples allow for identifying the necessary conditions for productively efficient spectrum use and thus act as a

[11] Smith –NERA refer to Lipsey and Lancaster (1956) discussing the theory of the second best; R.G. Lipsey and K. Lancaster (1956) "The general theory of the second best", Review of Economic Studies, vol. 24, pp. 11-32.
[12]Smith –NERA refer toDiamond and Mirrlees (1971) discussing optimal taxation; Peter Diamond and James Mirrlees (1971) "Optimal taxation and public production 1: Production efficiency and 2: Tax rules", American Economic Review, vol. 61, pp. 8-27 and 261-78.

cornerstone for developing incentive spectrum pricing. These examples are presented and discussed next.[13]

Assume that the available spectrum resources lie on a unit interval [0,1] and they are used by two sectors, 1 and 2. In other words, there are two differing uses for the spectrum resources in the economy, for example broadcasting and telephony. Alternatively, one might think of a certain frequency band being assigned between two users; for example a band of frequencies allocated to TV broadcasting use assigned between two TV broadcasting companies. The two sectors utilize two types of inputs, spectrum and labor in order to produce the respective final outputs (broadcasting and telephony services). Thus, labor and spectrum are regarded as substitute goods. Note that the other input in the production in addition to spectrum might as well be any other input, e.g. base stations, so that lack of spectrum resources could be replaced by investing in infrastructure. In other words, substitutes for spectrum exist.

There are many other sectors beside these two in the economy, but they do not utilize spectrum. However, they do make use of other inputs such as capital and they also demand labor. Thus, any amount of labor unused in the two sectors utilizing spectrum is valued in the other sectors of the economy. The total amount of labor in the economy equals L and the wage rate $w>0$ is determined on a competitive market. The labor resources used by sector 1 equal l_1 and by sector 2 l_2, the amount of labor utilized in sectors 1 and 2 equaling $l_1 + l_2 \leq L$. In addition, the prices of all final outputs produced in the economy are determined in a competitive market and firms take the prices as given. Prices for the outputs produced by sectors 1 and 2 are denoted with p_1 and p_2 respectively. However, there exists a market imperfection, which is the lack of a market for spectrum. In other words, spectrum is allocated to the sectors using administrative proceedings such as lotteries, beauty contests or first-come-first serve methods instead of with the help of market mechanisms. Note that this still is the case in many countries. For simplicity assume that frequencies as such (excluding cost recovery) carry a zero price, as was previously shown to usually be the case in administrative allocation. The costs of the regulator from spectrum management are covered through general taxation.

[13] I have intertwined these two examples into one uniform example. In order to do this I have made some minor modifications to the notation of the example presented by Indepen et. al (2004) in order to allow for a sufficiently theoretical presentation - note that the original paper was intended for regulatory use and was thus intentionally expressed in layman's terms rather than using a notation in line with the economic practice. The basic story and results naturally stay unaffected by the notation modifications.

From the available amount of spectrum (=1) let $0 < s < 1$ be the amount of spectrum originally allocated to sector 1 and 1-s allocated to sector 2. The firms in each sector are profit maximizers[14] choosing inputs and thus the output in order to achieve this goal. The profit is maximized with respect to possible spectrum resource constraints reflecting the scarcity of spectrum; the usage of spectrum cannot exceed supply. Other types of scarcity constraints, the interference constraints discussed previously, are imposed by the fact that spectrum can be re-used in some sectors, but not in others due to interference. A more mathematical presentation of the optimization problem is given in subsection XX, but here the basic principle of the opportunity cost based pricing can be shown with this more general example. The total output produced in sector 1 is denoted by $Q_1(s,l_1)$ and in sector 2 by $Q_2(1-s,l_2)$. In this example the production function is assumed to be concave with diminishing returns i.e. $Q_i' > 0$ and $Q_i'' < 0$, with i=1,2, but the logic also applies to production functions of other forms[15].

From the spectrum demand point of view there are naturally three scenarios which may occur: demand for spectrum in equal to supply in each sector, demand is below supply in each sector or demand for one or both sectors exceeds the supply. In the absence of excess demand (the first and second scenarios) it can be argued that the economy is at an efficient point, since no further profit can be gained by substituting costly labor or other inputs with free spectrum resources. If it were possible the profit maximizing and thus cost minimizing firms would have done it ultimately causing excess demand for spectrum.

From the regulators (and efficiency's) point of view the interesting scenario is the last one, where excess demand for spectrum exists. This is the case with many spectrum resources such as the ones used for mobile services (especially the 3G and 4G frequencies mentioned previously). The existence of excess demand raises the question of whether the current resources could be re-allocated to achieve efficiency gains. The efficiency gains could be achieved if a re-allocation of spectrum freed up some of the labor (the other input) resources without necessitating a reduction in the overall output produced in each sector. The freed labor resources could be used in other sectors

---

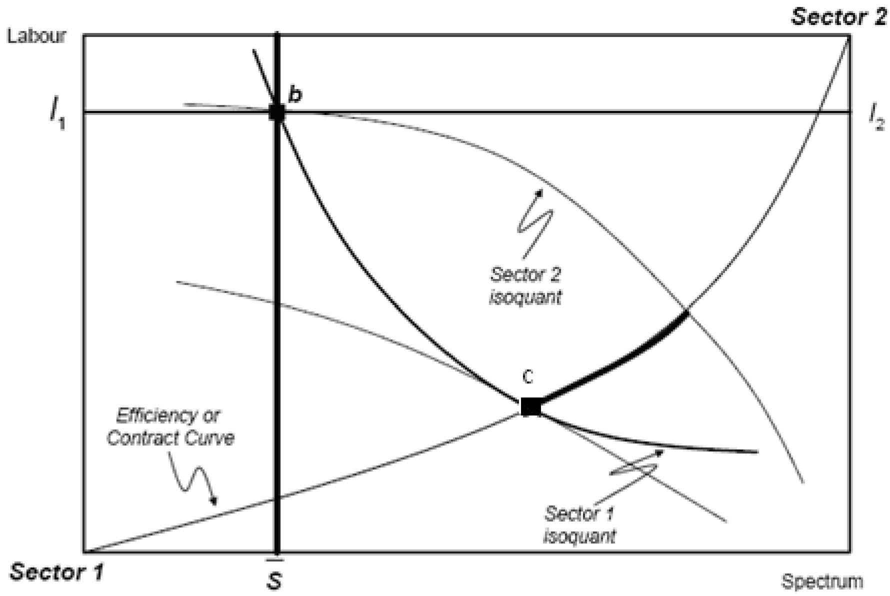
[14] It is important to note one important shortcoming of the illustrative example: there are also firms utililizing spectrum that are not profit-maximizers such as many public sector users of spectrum. The example can be however be extended to cover them by allowing for cost minimization rather than profit maximization and the implications on spectrum pricing prevail. The necessity of imposing AIP on public spectrum users is also discussed in more detail in chapter 5.

[15] Different forms of production functions for spectrum utilizing services and their plausibility, as well as effects on optimal pricing of spectrum, are discussed next in section 4.2.

to increase the overall output of the economy. Thus, the initial spectrum allocation would be productively inefficient.

In order to examine whether a re-allocation of spectrum resources could induce efficiency gains the effect of a marginal change in spectrum allocation is considered. The re-allocation is assumed to be such that outputs Q_1 and Q_2 remain constant, i.e. the use of the other input (labor) is adjusted. Keeping the output constant allows the focus to be on productive efficiency. The re-allocation from an inefficient point to an efficient one can be illustrated in the Edgeworth box in figure 3, where point b illustrates the original inefficient point.

Figure 3. Efficiency illustrated in the Edgeworth box



Source: Indepen et al. 2004

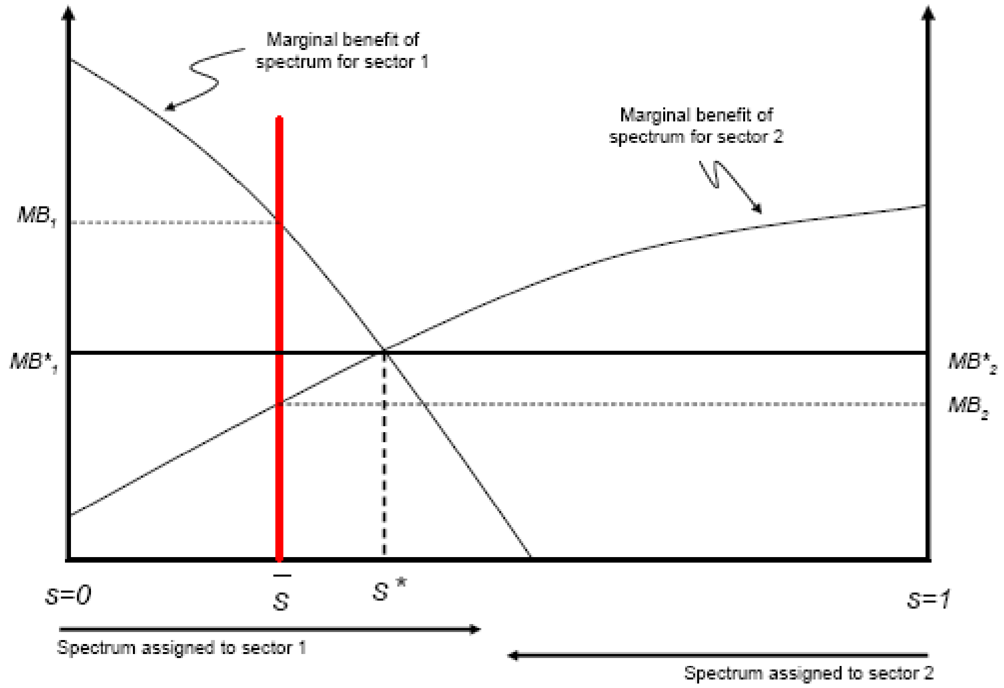
In figure 3 the spectrum inputs are in full utilization (s_{1+} s_2 = s), i.e. technical efficiency is achieved. The isoquants, which portray the different combinations of inputs with which the output remains constant, correspond to the previously determined concave production functions with diminishing returns to scale. Isoquants for sector 1 are depicted as convex to the origin and isoquants for sector 2 in the opposite corner with increasing output towards the origin. As can be seen from the figure current market outcome at point b is inefficient as a re-allocation in spectrum (and consequently in labor resources) brings forth an improvement in overall quantity of output without impairing the quantity produced by the other sector. When a re-allocation cannot benefit the other sector without harming the other, the allocation satisfies productive efficiency and the solution lies on the contract curve of Pareto efficient outcomes i.e. on the curve in which the isoquants of the two sectors are tangential. This is the bolded curve in figure 3. For example, the regulator may strive to the efficient point c in the Edgeworth box. This happens by allocating more spectrum resources to sector 1 and consequently decreasing its use of labor.

We can assess the extent of inefficiency in the initial allocation (point b) in terms of the other input used by the spectrum utilizing sectors, i.e. labor. Suppose that after a marginal change in the spectrum resources, for example an increase (decrease) in s, denoted by Δs, the same output in sector 1 Q_1 can be produced by using Δl_1 units less (more) labor. It is now possible to infer value of spectrum Δs in terms of labor, i.e. $w\Delta l_1$; the value of input resources that would be saved by allocating Δs to sector 1 instead of sector 2. The same holds naturally for sector 2 where the value of Δs is $w\Delta l_2$. These values are the marginal benefits (MB) of spectrum and represent estimates of the marginal opportunity cost of spectrum since by definition opportunity cost is the cost of an alternative that must be forgone in order to pursue a certain action[16]. These values allow AIP payments to be calculated correctly, but require an understanding of close substitutes for spectrum and their relationship with spectrum resources. Thus, the scarcity necessitates trade-offs and trade-offs result in opportunity costs. When prices are set equal to opportunity cost, the firms treat spectrum as any other input in production, choose the inputs to minimize these costs thus achieving productive efficiency.

The marginal benefit curves for the two sectors of the example are illustrated in figure 4. The decreasing returns to scale can be seen in the downward sloping shape of the marginal benefit curves. The initial allocation of spectrum s (amount s for sector 1 and 1-s for sector 2) with the corresponding marginal benefits of the sectors being MB_1 and MB_2 can be seen to be inefficient, since re-allocating spectrum from sector 2 to sector 1 with a higher marginal benefit would improve efficiency.

[16] In fact since the Smith-NERA method looks at opportunity costs calculated at the margin by viewing how small incremental changes in spectrum affect input substitutability, the approach actually studies marginal rate of technical substitution and values attained can be viewed as marginal technical opportunity costs derived from the production functions (Indepen &Aedis 2007, p.13).

Figure 4. Marginal benefit functions of spectrum



Source: Indepen et al. 2004

Efficiency is in turn satisfied at s* where the marginal benefits across sectors are equal and thus no further improvements can be made. In practice, the regulator however does not have the possibility of getting accurate information about the shape of the marginal benefit functions. Nevertheless, this is not necessary since the regulator may use estimated marginal benefits at current assignments and allocations, i.e. the opportunity costs of spectrum at the current situation. Doyle (2007, p.7-9) continues by illustrating this with a following hypothetical example.

First assume that spectrum resources consist of three non-overlapping spectrum bands a,b and c in the interval [0,1]. These frequencies have been allocated to three different and competing uses I, II and III. The current allocations are depicted in table 1 below with the highlighted cells and the numbers stand for marginal benefits across different uses. In addition, a substitute input is depicted in the right-most column.

Table 1. Marginal benefits of spectrum

	Frequency bands			Alternative non-spectrum input
Uses	*a*	*b*	*c*	
I	100	75	0	0
II	35	60	30	0
III	10	10	15	5

Source: Doyle 2007

Thus, use I utilizes frequency band a, use II frequency band b and use III frequency band c. The highlighted cells, i.e. the current allocations depict the estimated opportunity cost values that a regulator can estimate most easily. The way of calculating them in practice is explained in detail in the next subsection 4.1.3.

Again, the inefficiency of the initial allocation can be seen, since the marginal benefits for spectrum bands across uses (the colums in the table) are not equalized. The frequencies in band a allocated to use I are in their most efficient use, since the marginal benefit of the two alternative uses II and III are lower than the marginal benefit at current use. However, frquencies in band b have a higher marginal benefit at use I (MB=75) than in their current use II (MB=60). Thus, by allocating more spectrum from use II to use I increases efficiency. This re-allocation naturally affects the marginal benefits of the uses that are affected by the change. The changes are depicted in table 2 below. Specifically, when more spectrum b is allocated to use I the marginal benefit in use I of frequency band b will fall below 75 due to the decreasing returns to scale. Similarly the marginal benefit in use I of frequency band a falls below 100, since there are now more spectrum resources allocated to use I. In addition, an opposite effect is seen in use II of frequency bands a, b and c.

Table 2. Marginal benefits of spectrum after the first re-allocation

	Frequency bands			Alternative non-spectrum input
Uses	*a*	*b*	*c*	
I	90	70	0	0
II	38	70	32	0
III	10	10	15	5

Source: Doyle 2007

Thus, now the marginal benefits of frequency bands a and b in uses I and II have changed as well as the MB of frequency band c in use II. Now the MB's of frequency band b between uses I and II are equalized, which implies that for frequency band b the allocation is efficient. Again the highlighted cells are the ones that the regulator can more easily calculate in practice, since they correspond to the current allocations and assignments.

We can see that there still is scope for further efficiency gains by re-allocating spectrum in band c form use III (MB=15) to use II (MB=32). Following the same logic as above this re-allocation will decrease the MBs of frequency bands a, b and c in use II as the total amount of spectrum in that use increases. In addition, the change will increase the MBs of frequency bands a, b and c in use III, which now has overall less spectrum. As a result the MBs of spectrum band b between uses I and II will no longer be equal and yet a further re-allocation of frequency band b is needed. After the requisite re-allocations an efficient solution illustrated by table 3 is achieved.

Table 3. Marginal benefits of spectrum at the efficient solution

	Frequency bands			Alternative non-spectrum input
Uses	*a*	*b*	*c*	
I	87	68	0	0
II	36	68	25	0
III	12	12	25	4

Source: Doyle 2007

In table 3 above there is equality of marginal benefits across uses in the two highest values. No further re-allocations would yield better outcomes and thus the solution is efficient.

Thus, arriving at an efficient outcome, i.e. achieving allocative efficiency, is an iterative process where one (or possibly several) re-allocations or assignments are made at a time and the changed marginal benefits then calculated again to show the possible the need for further re-allocations or re-assignments within uses. It can also be seen that in order to achieve the efficient outcome the regulator needs to know about the MBs of frequency bands in neighboring uses. These values are proposed to be evaluated based on the costs of alternatives.

In other words, this method basically suggests that the regulator needs to be able to first identify all frequency bands and associated uses and then determine the marginal benefits for each of them utilizing a least cost alternative, i.e. determining, what is the spectrum user's alternative (substitute) way of offering the service if a marginal amount of spectrum is taken from him. Using these estimations for MB's the regulator identifies the direction of needed change in spectrum re-allocation and sets the prices for spectrum accordingly to encourage these changes. The actual price set depends on whether the maximum marginal benefit is offered by the current use or by a use that currently does not utilize the spectrum resource. In the former case the price is set at the value of the current use whereas in the latter case the price should lie on the interval between MB* and the current use marginal benefit (for a more detailed discussion, see Doyle 2007, p 9-11). This procedure starts an iterative process towards the efficient allocation or assignment, which may take up to five years or so (Doyle 2007, p. 10).

4.1.3 Criticisms of the Smith-NERA method

While the Smith-NERA methodology is simple and thus can be easily communicated to regulators setting the policies, its straightforwardness comes at a price. In particular, the methodology assumes perfect markets and thus refrains from discussing any issues related to market structure. This is a key weakness of the approach, since many of the spectrum-utilizing markets are highly concentrated, which is also reflected as high end prices of spectrum utilizing products and services for the customers. For example, in Germany, the United States, Spain and Greece, where there exist no challenger operators, the mobile data rates per gigabyte are 30-100 times higher than in the highly competed markets in Finland, Denmark and the UK (Taloussanomat, 25.3.2014).

Another key limitation of the model is that it does not explicitly account for interference or interference constraints, which were previously shown to be the major driver behind the need for spectrum management in the first place. This is mainly due to the fact that the Smith-NERA model is based on simplistic economic assumptions of a non-existing state of the market (perfect competition). As will be shown in the next subchapter 4.2 this is a key issue corrected in the approach by Levine and Rickman, achieved by supplementing the basic economic theory with graph theory, which allows for constructing interference constraints to the model.

In addition, the approach does take into account productive efficiency, but it ignores other effects such as the consumers' willingness to pay and revenue raised by the government (Levine &

Rickman 2007, p.2). The dynamic efficiency as such is also left unexplained and depends on the idea that a firm that values the spectrum resources most today is also the best innovator in the long run. If we think about the private value build-up discussed in 2.6.1 we see that this often may not be the case: for example a possibility to stall or damage competitors, indicated by a nonnegative defensive value (DV) parameter, might increase the valuation a spectrum user has for the spectrum resources without having anything to do with the actual willingness and capabilities of the user to guarantee sufficient investments to achieve dynamic efficiency. Finally, the fact that the smallest decrements in spectrum used for evaluating the opportunity cost may actually be very large for some services implies that the assumption of the output remaining constant may be unrealistic (Aegis & Plum 2008, p.19).

Many of these issues are addressed by the extended model for AIP determination by Levine and Rickman, which is discussed next.

4.2 Optimal Administrative Incentive Pricing of spectrum by Levine and Rickman (2007)

As was discussed in the previous chapter the AIP calculation approach developed by Smith and NERA and further enhanced by Indepen et al. has been criticized for its simplistic assumptions as well as omitting certain structures of the actual economies in which the firms operate. To correct some of these flaws Levine and Rickman (2007) have constructed a mathematical framework combining models from information technology and economics to explain the structure and attainment of optimal AIP payments. In particular they take into account interference and market structures (other than perfect competition proposed by Smith-NERA) and allow for revenue implications to the regulator or government from the collection of AIP payments. AIP determination is regarded as an optimization problem where the regulator effectively maximizes overall welfare with respect to the spectrum fee given resource and interference constraints. Their work combines both approaches of productive and allocative efficiency and is introduced and discussed in this subchapter.

4.2.1 Formulation of the spectrum assignment problem

The assignment of spectrum to users can also be called channel assignment since a channel is simply a specified frequency range.[17] The spectrum assignment problem is in this section termed as

[17] For example in Finland the national (tv) *channel 5* operates on frequencies from 174MHz to 181MHz (Ficora 2012).

the channel assignment problem in order to follow the well-known terminology used throughout the literature discussing this issue. The channel assignment problem can be seen as a mathematical problem of dividing scarce spectrum resources (i.e. channels) between competing, though predetermined (since allocation has been conducted) set of demands while taking into account the constraints imposed by interference (Levine & Rickman 2007, p. 3). In other words, the problem specification consists of information on requirements (demand) for spectrum across the system, the constraints imposed to limit interference and the specification of the objective to be fulfilled while satisfying the spectrum requirements and the interference constraints.

The spectrum requirements are introduced by specifying the amount of distinct channels each transmitter site reguires. For n different transmitter sites T_1, T_2, T_3…Tn there exist corresponding demands of m_1, m_2, m_3…m_n channels, where site T_i requires m_i distinct channels. We have a set of constraints each relating to a single transmitter site T_i, known as co-site constraints or to a pair of transmitter sites (T_i, T_j), known as inter-site constraints. For simplicity, the different channels are labeled with integer numbers, which correspond to the location of the channels in the spectrum band. When $f_1^{(i)}$ and $f_2^{(i)}$ are channels both assigned to a transmitter site T_i the co-site constraint requires that,

$$\left| f_1^{(i)} - f_2^{(i)} \right| \geq \kappa_i \qquad (1)$$

where κ_i is a specified minimum channel separation, i.e. distance in the spectrum between two distinct channels, which ensures that interference between the channels is kept tolerable. Then naturally for channels $f^{(i)}$ and $f^{(j)}$ assigned to different transmitter sites T_i and T_j the inter-site constraint requires that,

$$\left| f^{(i)} - f^{(j)} \right| \geq \kappa_{ij} \qquad (2)$$

where similarly to (1) κ_{ij} is a specified minimum channel separation. The interference limiting co-site and inter-site constraints are thus specified by κ_i and κ_{ij} constructing the constraint matrix, where κ_i form the diagonal entries and κ_{ij} the non-diagonal entries. These constraints reflect the use of protection ratio (i.e. the signal-to-interference ratio) in the radio community.

The objective of the channel assignment problem can be specified as either minimizing the span required (i.e. the difference between the highest and lowest channel used) subject to the constraints, or as a fixed spectrum problem where given the maximum span (i.e. the amount of spectrum available) the channels are assigned to as many spectrum requirements as possible. The latter approach is used by Levine and Rickman. This approach implies that the transmitter network and power are fixed and thus effectively taken into account by the constraint matrix. An alternative would be to have these as extra variables in the model to be optimized with the channel assignment, but such theoretical work is scarce and thus the more simplistic approach is taken. (Levine & Rickman 2007, p.5)

The channel assignment problem constructed above has been studied with the help of basic graph theory and specifically the graph-coloring problem.[18] A graph is a collection of abstract 'nodes', of which some are joined by 'edges'. The coloring problem attaches a color to each of the nodes in a way that no adjacent nodes share the same color and the overall amount/number of colors is minimized. This minimum number of colors is called the chromatic number of the graph. This problem relates directly to channel assignment when the nodes are thought of as transmitter sites and the colors as channels. For example, if we determine that m_i equals 1 i.e. each transmitter requires only one channel, and κ_{ij} equals 1 if the nodes T_i and T_j are joined and 0 otherwise we end up with the minimum span channel assignment problem discussed above (since each site requires one channel the values for co-site constraints κ_i are immaterial). In physical terms, we model co-channel (instead of adjacent channel) interference and the edges represent the rough location of potential coverage blackspots. (p.5)

The next step is to relate the channel assignment problem to an economic model explaining spectrum demand in terms of market conditions and costs. This is done by assuming that each node or transmitter site incorporates a local market with an oligopolistic market structure. This is in accordance with many of the actual product and service markets utilizing spectrum in real life (as was previously discussed in 4.1.3) and brings a clear correction to the Smith-NERA model which assumes perfect competition. There should be no restrictions in interpreting the local markets as national markets (e.g. the Finnish mobile communications market) or alternatively as local markets within national markets (e.g. mobile communications market within the Eastern Finland). The firms

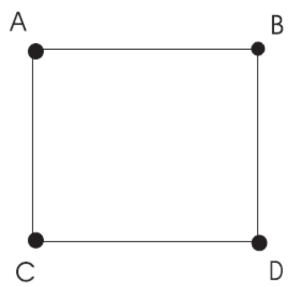
[18] For a more thorough explanation of graph theory and the graph coloring problem see e.g. "Graph Theory" by Diestel Reinhard (Springer 2006).

operating in these local markets are assumed to provide homogeneous products. Note that products across markets can still differ. A spatial interpretation of the transmitter sites is to regard them as cells i.e. specific regions of service. Thus, a transmitter site consists of all the transmitters used by the firms in the local market and they may share some of the transmitters, perhaps against a fee.

In each cell a local oligopoly offers a local service, produced with spectrum resources (channels) as inputs. The firms purchase a license (which is equivalent to charging an AIP from their use) from the regulator to use a certain amount of channels depending on the level of output. Within a cell the firms are so close to each other that no spectrum re-use or sharing can occur between them, and we assume that there exists only co-channel interference, i.e. κ_i=1 in (1) and κ_{ij}=0 or 1 in (2). This assumption by Levine and Rickman is in accordance with reality, as a certain geographical distance is required in order to be able to share frequencies (see chapter 2.3 for more details). The demand of spectrum is defined by the sum of demands of the individual firms, to be modeled in 4.2.2.

Each cell is given a color and a shared color indicates that spectrum sharing is possible between the regions. Figures 5 and 6 illustrate this in a world with four local markets (i.e. four nodes).

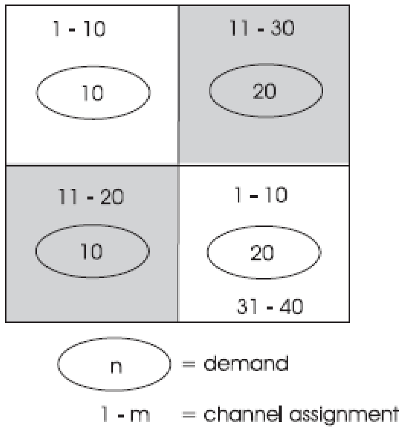
Figure 5. A four node graph of four markets



Source: Levine and Rickman 2007

In figure 5 the nodes A-D represent the transmitter sites around which the local markets are formed.

Figure 6. A coloured map of four markets



Source: Levine and Rickman 2007

In Figure 6 the graph has been colored with two colors in the previously determined way where no adjacent nodes receive the same color. The four local markets are formed around the four nodes. The numbers inside circles represent the demand for frequencies (or channels) in each market and shared colors across diagonal markets indicate the possibility for these markets to share frequencies. Due to this possibility of sharing the total demand for frequencies, which equals 60 channels (10+20+10+20) can be satisfied with a minimum of 40 distinct channels. An exemplary division of frequencies is as follows. Market A is assigned with channels 1-10, of which market D can re-use all ten channels and in addition requires ten channels more, say 31-40. Market B is assigned with channels 11-30, of which market C can re-use channels 11-20.

4.2.2 The core economic model

This subsection constructs the economic model, which is then combined with the spectrum assignment problem depicted in the previous section. First a single local market with N competing firms and a homogenous service offered at price P is considered (sectors and sector-crossing sales are included later on). The total output Q in the local market equals the individual firms' outputs $Q = \sum_{k=1}^{N} q_k$, where k=1,2,…N. The output can be thought of as minutes of a service requiring frequencies as inputs. The demand curve is given by $Q = D(P)$ with the usual property $D'(P) < 0$ stating that demand diminishes as price increases. It is also assumed that $\lim_{P\to\infty} PD(P) = 0$. The inverse demand is denoted by $P = D^{-1}(Q) = P(Q)$. The output of each firm is produced using labor (L), capital (K) and frequencies (Z) as inputs according to a general CES production function. Later on Levine and Rickman specify this, but first the basic form is assumed (without the firm subscript):

$$q = T\left[\gamma_1 L^{\eta} + (1-\gamma_1)\left[\gamma_2 Z^{\varepsilon} + (1-\gamma_2)K^{\varepsilon}\right]^{\eta/\varepsilon}\right]^{1/\eta}, \tag{3}$$

where T measures the total factor productivity and γ_1 and γ_2 are the share parameters and ε and η parameters determining the elasticity of substitution between inputs. Capital and spectrum have been grouped together, giving them an elasticity of substitution of 1/1+ε and the elasticity of substitution between labor and the grouped input (Z,K) being 1/1+η. In other words, if $\varepsilon \in (0,1)$ and $\eta < 0$, spectrum and capital are considered substitutes whereas labor is a complement for these inputs. In reality this seems like a rational assumption since we know that lack of spectrum can to some extent be made up for by building a more extensive infrastructure (requiring capital)[19], but increasing the amount of labor does not increase output if spectrum resources are lacking. On the other hand labor and capital usually can be regarded as at least partial substitutes. Thus, without the main conclusions being compromised, we can alternatively have a production function of the form:

$$q = T\left[\gamma_1 Z^{\eta} + (1-\gamma_1)\left[\gamma_2 L^{\varepsilon} + (1-\gamma_2)K^{\varepsilon}\right]^{\eta/\varepsilon}\right]^{1/\eta}, \tag{4}$$

where spectrum is a complement for capital and labor which in turn are substitutes to each other. In the limit as η and ε tend to 0, both (3) and (4) approach the production function of the Cobb-Douglas form:

$$q = TL^{\theta 1} Z^{\theta 2} K^{\theta 3}; \sum_{i=1}^{3} \theta_i = 1 \tag{5}$$

Given one of these forms (3)-(5) and factor prices (w, r, a) it is possible to formulate the cost function $c(w, r, a)$ as:

$$c_i(w, r, a) = wL_i(w, r, a) + rK_i(w, r, a) + aZ_i(w, r, a); i = 1,2 \tag{6}$$

where $L(w, r, a)$, $K(w, r, a)$ and $Z(w, r, a)$ are the associated factor demands per unit of output which by convention satisfy $\frac{\partial L}{\partial w}, \frac{\partial K}{\partial r}, \frac{\partial Z}{\partial a} < 0$. The firms are assumed to be price takers in the factor

[19] As an example one can think of a situation where a mobile communications firm lacks lower frequencies which require less infra since their range is longer, but has a sufficient amount of higher frequencies which may be as suitable for the technology utilized/service provided as the lower ones but due to their shorter range require more infra. Thus, the lack of lower frequencies can to some extent be made up for by utilizing the higher frequencies by increasing capital K.

markets including the spectrum license market which incorporates all the local markets. In addition, it is assumed that the price elasticity of demand $\in(Q) = -{PdQ}/{QdP}$, is constant with respect to output Q and ϵ >1, which will be justified later on (see function 13).

First the regulator sets the spectrum license price a after which the firms compete in the market given the factor prices. This leads to a subgame perfect equilibrium which can be found by backward induction. Levine and Rickman assume a Cournot-Nash equilibrium at the second stage of the game. Thus, the regulator acts as a Stackelberg leader in the first phase by setting the price. The game is solved for both the case with an exogenous number of firms as well as the case with an endogenous number of firms dictated by a condition of free entry and thus zero profits.

Levine and Rickman (2007, p.8-10) first consider the equilibrium with an exogenous number of firms. Given the core economic model presented above, profits Π for a firm (indexed by k) are:

$$\prod_k = \prod_k (q_k, w, r, a) = [P - c_k(w, r, a)]q_k - F, \tag{7}$$

where F stands for fixed costs, which are independent of output. We can write the total output as $Q = q_k + \overline{q_k}$ where $\overline{q_k}$ stands for the outputs of all but the kth firm. In a market clearing this naturally equals the total demand $D(P)$. In a Cournot-Nash equilibrium the kth firm takes $\overline{q_k}$ as well as the inverse demand $P = D^{-1}(Q) = D^{-1}(q_k + \overline{q_k})$ and spectrum price a as given while choosing its output level as a strategic response. Since at the second stage of the game (when the regulator has set the spectrum price) the firm is a follower in the leader-follower game, and thus a price-taker, there does not exist any strategic bidding for licenses. This is a clear difference compared to traditional auction theory which also discusses and allows for strategic bidding and can be argued to be a defection of the model.

The firm thus maximizes (7 above) with respect to its output q_k given $\overline{q_k}$, a and the market-clearing condition $P = D^{-1}(Q) = D^{-1}(q_k + \overline{q_k})$. The first order condition is given by:

$$P' q_k + P - c_k(w, r, a) = 0 \tag{8}$$

and the second order condition by

$$2P'+q_k P''<0 \tag{9}$$

Rearranging the FOC (8) reveals the familiar mark-up pricing formula for an oligopolist:

$$P=\frac{c_k(w,r,a)}{1-\frac{q_k}{\in Q}} \tag{10}$$

where $\in=\in(Q)=-{PdQ}/{QdP}$ represents the elasticity of demand. As the number of firms increases the price tends to marginal cost $c_k(w,r,a)$. Since it was assumed that the firms are identical (offering homogenous services) it must be that $Q=Nq_k=Nq$, thus each firm produces the same amount of output. Thus, the oligopolist's pricing result (10) becomes

$$P=\frac{c(w,r,a)}{1-\frac{1}{\in N}}=P(w,r,a,N)=P(a,N) \tag{11}$$

where wages w and interest rates r have been omitted from the right-most expression since they are exogenously determined by the general equilibrium in which the market model is embedded. A key interpretation of equation (11) is that the price for the end products and services is dependent on the number of firms i.e. the competitive situation in the market and the price set by the regulator for spectrum. Effectively this simple conclusion further attests to to the need of setting optimal, not excessively high AIP's as, in addition to leaving spectrum unused, a high AIP is reflected as high end prices, which diminishes consumer surplus. The condition ϵN>ϵ>1 ensures that the price is always positive and it is also the second-order condition for the profit maximization. In order to see this we can write a constant elasticity demand curve as $Q=AP^{-\in}$. Differentiating twice with respect to Q we get:

$$P''=\frac{(\in+1)(P')^2}{P} \tag{12}$$

Substituting this for P'' in the second order condition (99), noting that $q_k=q=Q/N$ and using $P'<0$ as well as the stated elasticity of demand $\in=-{PdQ}/{QdP}$ it is relatively easy to see that

$$\epsilon > \frac{1}{2N-1} \tag{13}$$

must hold. And thus, since 2N-1≤ 1for N≥1, clearly ϵ>1 is sufficient for (XX ylempi) to hold. The reasoning for this is that given if the demand does not adjust to changes in prices or adjusts very little (low price elasticity of demand) firms can increase prices and thus their profits indefinitely by reducing output. In a Cournot-Nash equilibrium with N identical firms output of an individual firm is given by $q = Q / N = D(P(a, N)) / N$ with profits given by:

$$\Pi = \Pi(a, N) = [P(a, N) - c(w, r, a)]D(P(a, N)) / N - F \tag{14}$$

This leads to a following proposition (Levine & Rickman 2007, p.10):

Proposition 1: *Assuming ϵ > 1 gives profits of Π(a, N), which decrease with respect to a and N (for proof see Appendix 3).*

The explanation for this is quite natural. As license prices increase so do the firm's total costs which shows as an increase in the retail price when at least part of the costs are transferred to prices. As a result demand decreases and assuming an elasticity of demand greater than unity implies that the overall revenue falls and decreases profit. As for the effect of N on profits, increase in the number of firms, i.e. stronger competition, lowers retail prices resulting in lower mark-ups, but also in higher demand. With constant elasticity of demand ϵ > 1 and sharing the revenue with more firms the overall effect on profits of individual firms is negative.

The number of firms competing in the market (N) has thus far been exogenous. We now move to consider the equilibrium with an endogenous number of firms. There are two alternative ways to endogenize this parameter (Levine & Rickman, 2007, p.10): by making N into a policy variable determined by the regulator while issuing the licenses, or assuming free entry of firms into the market with the participation constraint stating that profits cannot turn negative[20]. Levine and Rickman apply the latter principle. This can be seen as a clear distinction from reality, since there do exist significant barriers to entry in the markets utilizing frequencies as inputs. This is mainly due to the reason that in many such markets (e.g. mobile communications services) heavy

[20] Naturally this is also the result the market ends up with free entry, since the firms with negative profits eventually bail out as they go bankrupt. Profits are thus driven to zero.

investments in infrastructure (transmitter sites etc.) are required for setting up sufficiently large scale operations to be profitable. Thus the incumbents have strong positions against potential entrants, which would have to "start from scratch" or operate by leasing the incumbents readily available infrastructure. Actually the first approach of endogenizing N by making it a policy variable determined by the regulator makes much more sense in reality. As was discussed earlier in one of the main objectives of the regulator is to promote competition, which in practice is done e.g. though setting bidding caps in spectrum auctions to ensure that there will be a sufficient amount of license holders at the end of the assignment. On the other hand the same bidding requirements can be used to restrict the amount of players in the market. In effect, regulators can be seen as determining N through these types of measures. However, the choice of method in endogenizing the number of firms in the model by Levine and Rickman should have no effect in the upper-level conclusions or propositions provided by their analysis and thus we continue with the assumption that free entry to the market exists.

With free entry profits are driven to zero i.e. the number of firms in equilibrium N*, given license price a satisfies[21]

$$\Pi(a, N^*) = 0 \tag{15}$$

According to proposition 1 the profits are decreasing in N and they become negative for a sufficiently large N. If it is assumed that for a monopolist the profits are positive i.e. $\Pi(a,1) > 0$ then there exists a unique N* satisfying (15). In addition, differentiating (15) with respect to the license price a and remembering proposition 1 we get:

$$\frac{dN^*}{da} = -\frac{\frac{\partial \Pi}{\partial a}}{\frac{\partial \Pi}{\partial N}} < 0 \tag{16}$$

i.e. the number of firms in the market is negatively dependent on the size of license price a. From proposition 1 we can also conclude that $\Pi(0,1) > \Pi(a,1) > 0$. Also by remembering $\lim_{P\to\infty} PD(P) = 0$ we may deduce that $\Pi(a,1)$ becomes negative with a sufficiently large a. In other words, if the license price (also thought of as the access price to the market) is sufficiently

[21] Note that free entry will lead to suboptimal duplication of fixed costs F (Levine & Rickman, 2007, p.11). See also Perry (1984).

high even for a monopolist the profits may turn negative so that the monopolist eventually exits the market.

The total demand for channels (m), given the license price and the number of firms in the optimum N*(a), is determined by the product of the number of firms and their respective use of spectrum input Z:

$$m^*(a) = N^*(a)Z(a) \qquad (17)$$

Differentiating with respect to a using the Leibniz rule gives:

$$\frac{dm(a)}{da} = \frac{dN^*}{da} Z(a) + N^*(a)\frac{dZ}{da} \qquad (18)$$

and since both the number of firms and the amount of spectrum inputs used in production are negatively related to the spectrum (license) price it is obvious that d m(a)/d a < 0. The results given above are combined to give the following proposition (Levine & Rickman, 2007, p.11):

Proposition 2:

- *i)* *Given access price a and the demand curve D(P) there exists a unique number of firms N(a)* in the market.*
- *ii)* *N*(a) is decreasing in a .*
- *iii)* *A sufficiently high a exists to lead to all firms exiting the market.*
- *iv)* *Total demand for channels m*(a) is decreasing in a .*

Sub-proposition iii) demonstrates the power that the regulator has and on the other hand the cautiousness it has to show while determining the level of AIP. The aim is not only to ensure that frequencies are used, but also that the firms' operational prerequisites are not jeopardized by the additional costs incurred due to the AIP payment.

In addition, the final result iv) is crucial as will be shown in the next section. Levine and Rickman (2007, p.12) separate two effects leading to this result, which both have to do with demand of the end products produced using spectrum as input. Firstly, for a given number of firms the demand for end products decreases due to firms transference of increased costs (increased a) into their prices. Secondly, due to the increased costs some firms are forced to exit the market increasing the market power of the firms still left in the market. The increased market power is used to raise the mark-up on their end product pricing, which again decreases the demand for end products. Naturally the decreased demand of end products results in a decrease in the demand of inputs used to produce them i.e. spectrum (channels).

4.2.3 Optimal spectrum pricing

Next we extend the model to account for several sectors instead of just the one local market discussed in 4.2.2. The services and products offered within sectors are still assumed to be homogenous. In addition, services across sectors are assumed to be independent of each other i.e. neither substitutes nor complements. The correspondence between this assumption and the reality is two-fold. On one hand most of the services utilizing spectrum fulfill the assumption; it is difficult to find any substitutability or complementarity between commercial services such as TV broadcasting and governmental use of spectrum such as emergency frequencies or radar, or even between many of the commercial services (think of e.g. uses of unlicensed spectrum such as garage door openers and microwaves versus radio). On the other hand especially substitutability often occurs; a good example being TV versus radio broadcasts (unless one includes both of these under the wider term broadcasting, but it seems plausible to treat them as separate products).

The number of sectors in the economy equals p and in each sector there are l_i (i=1,2,…,p) local markets similar to the one discussed previously in the core economic model. The assignment of spectrum is still constrained by the interference constraints discussed previously in 4.2.1. Transmitters are required by each local market and the transmitters within the markets are too close to each other to allow for spectrum (channel) sharing. Thus, returning to our previous discussion, each node in the graph represents a group of these kinds of transmitters as well as a local market. To further simplify the notation each (local) group of transmitters can be considered as one transmitter on one site. The spectrum allocation problem can be embedded in the following wider economic allocation problem (Levine & Rickman, 2007, p. 12-14). Note that we now discuss allocation instead of assignment since we have multiple sectors providing different services and thus varying allocations of spectrum are possible. This is a significant improvement to the original Smith-NERA method, which presupposed allocation and only allowed optimization of assignment of spectrum.

Firstly, the total demand for channels in each local market $j=1,2,\ldots,l_i$ for each sector $i=1,2,\ldots p$ is calculated. Due to independence of services across sectors, and assuming the same across markets, demand in a local market j of sector i will depend only on the price P_{ij}, according to what was discussed previously in 4.2.2. Secondly, a social welfare function W is chosen to take the following form:

$$W = \sum_{i=1}^{p}\sum_{j=1}^{l_i} S_{ij} + (1+\Lambda)R + \sum_{i=1}^{p}\sum_{j=1}^{l_i} \Pi_{ij} \tag{19}$$

where S_{ij} is the consumer surplus in market j of sector i given by:

$$S_{ij} = \int_{P_{ij}}^{\infty} D_{ij}(p)dp \ . \tag{20}$$

R in turn is the revenue accrued to the regulator from access prices (license fees) and it naturally depends on the demand for channels as well as channel price, i.e.

$$R = \sum_{i=1}^{p}\sum_{j=1}^{q_i} m_{ij}(a_{ij})a_{ij} \ . \tag{21}$$

$1+\Lambda$ represents the cost of public funds where $\Lambda>0$ captures "the distortionary effects of taxes that would otherwise be required in the absence of this revenue (Levine & Rickman 2007, p.13)". In other words Levine and Rickman include both the AIP payment (the actual spectrum usage fee) and the cost recovery fee in the access price a and assume that without an access price the costs would be covered through taxation, which would have distortionary effects on social welfare. The right-most term in (19) is naturally the producer surplus which consists of the profits earned by all firms across the economy.

Alternatively (19) can be interpreted as the regulator's ("social planner") objective function incorporating any restrictions imposed by law. Since by UK and EU law spectrum should be priced according to spectrum management considerations (as discussed previously) instead of in order to raise taxes, it is implied that the revenue term should be ignored. Thus, Λ should equal -1. In the following analysis the revenue term is kept and $\Lambda>0$ or $\Lambda= -1$ substituted only after the optimization is completed.

The social planner maximizes (19) with respect to access prices a_{ij} for each of the sectors and the markets within them ($i=1,2,\ldots,p$ and $j=1,2,\ldots,l_i$), and the number of firms providing each service in each market N_j ($j=1,p$). The maximization is done subject to the engineering constraints discussed in 4.2.1 and the spectrum resource constraint:

$$\sum_{i=1}^{p}\sum_{j=1}^{l_i} m_{ij}(a_{ij}) \le Z^{\max} \qquad , \tag{22}$$

which simply states that the amount of spectrum allocated or assigned cannot exceed the total amount of spectrum available (implying scarcity of the resource). Since licenses are admitted to a limited number of firms contradicting the free entry condition discussed in 4.2.2 (while endogenizing N) the profits would not be driven to zero and they should be included in the welfare function as was done in (19).

If however the regulator is not responsible for market structure (i.e. N is exogenously determined), (19) would be maximized with respect to only the equilibrium prices subject to the free entry equilibrium condition

$$\Pi_{ij}(a_{ij}, N_{ij}) = 0; i = 1,2,...p; j = 1,2,...,l_i \quad , \tag{23}$$

the engineering constraints of 4.2.1 and the spectrum resource constraint.

In principle welfare maximization requires that in order to take into account the differing demands and cost conditions of different markets as well as the inter-market interference constraints, each market should have a unique spectrum price a. In practice this might be too strict a requirement, especially since the information needs (of the regulator setting the price) increase with increasingly individual prices for different markets. Thus, applying equal spectrum prices within a sector ($a_{ij} = a_i$) might be plausible – note that this was assumed previously as well while discussing the Smith-NERA method. This approach is further supported if there exists channel trading between markets, since restricting the possibility of arbitrage requires equal pricing.

The number of firms in this framework varies across markets in each sector (and thus across sectors). These firms can be thought of as local operators, with each firm providing a service in a single market. An alternative the approach, which Levine and Rickman (2007) adopt, is that each firm is a *network operator* providing its service across all markets within the sector. Thus, in sector i and market j $N_{ij}=N_i$ holds. This assumption is adopted in the example discussed in 4.2.4 and 4.2.5.

4.2.4 Optimal spectrum pricing with linear technology

In this section the optimal *network license price* across sectors consisting of N_i network operators providing services across l_i local markets is studied. Before constructing the general case the optimization task is discussed with the help of an example where $l_i=3$ i.e. there exists three markets within a sector i. With the help of this example the relationship between the interference constraints and firms' pricing and channel use decisions is described. This discussion is further developed leading to determination of the optimal AIP payments and their dependence on spectrum sharing.

Firms and markets are considered to be identical except for the interference constraints. The radio channels can be used in all l_i markets taking into account the interference constraints. Each channel holds a fixed license fee a_i payable to the regulator per unit of time (e.g. once a year). Due to the proximity of channels used within a local market, spectrum sharing is not possible within a local market, but may be possible between local markets (and thus sectors).

As stated above, Levine and Rickman start with an example with sector i, which consists of three local markets (l_i=3). Thus, there are four interference graphs to consider with respect to such an example: two homogenous graphs and two inhomogenous graphs. For clarity these special cases are depicted with letters A and B respectively in this subsection. The two homogenous graphs consist of a complete graph with edges between every pair of nodes and a graph with no edges. These two cases are depicted in appendix 1. The two inhomogenous cases are a graph with one edge and a graph with two edges, which are depicted in appendix 2.

The focus in this section is on the network operators' pricing and output decisions made with respect to the license price. First this is examined within a particular sector and thus we drop the sector subscript i from the equations (sector-specific prices are discussed later on in this chapter). For simplicity the production function is specialized to equal:

$$q = \left[\gamma_1 L^{\eta} + (1-\gamma_1)K^{\eta}\right]^{\frac{1}{\eta}}; z \geq q \qquad (24)$$

In other words, Levine and Rickman consider a linear technology for spectrum, for which spectrum is a pure complement, i.e. it cannot be substituted by capital or labor. Output q is thus produced using a Leontief technology with fixed proportions of inputs. For notational reasons units are chosen in a way that one radio channel is required for one unit of output and output capacity equals the number of channels available. This can be assumed without any loss of generality.

Thus, a firm k producing an output of q_{kj} in a particular local market j=1,2,3 requires a total of Z_k radio channel licenses with $Z_k \geq z_{kj} \geq q_{kj}$, where z_{kj} are the channels available for firm k in market j, which will naturally depend on the nature of the interference graph (the four different cases of spectrum sharing mentioned above). It is assumed that the license fee a is independent of the firm's location and the firm itself.

The total costs for firm k can be expressed as:

$$C_k(\{q_{kj}\}, Z_k, a) = F + aZ_k + c\sum_{j=1}^{3} q_{kj} \tag{25}$$

where F represents the set-up costs or fixed costs of operations, aZ_k the license fees paid and $c\sum_{j=1}^{3} q_{kj}$ the costs of producing the total output across the three markets, where c = (w,r) equals the cost function associated with the CES production function of labor and capital in (24).

Again the regulator first sets the fee for a particular sector i and firms then compete in the market. By backward induction firm k chooses labor and capital to minimize the costs $c(w,r)q_{kj}$ of producing q_{kj} in markets j=1,2,3 given the factor prices (w,r). Through this choice of units and the fixed proportions Leontief function the firm then knows its demand for channels; it requires q_{kj} channels in market j, since it was assumed that one unit of production requires one channel. Prices for the end products equal P_j in each market j=1,2,3. The firm acquires a license for Z_k channels at price a. According to the relevant interference graph the firm k is then able to utilize $z_{kj} \leq Z_k$ channels in market j with this license. In a Cournot-Nash equilibrium in each market this firm then maximizes profits by choosing outputs q_{kj} and channels Z_k:

$$\max \Pi_k = \sum_{j=1}^{3} (P_j - c)q_{kj} - aZ_k - F \tag{26}$$

given the spectrum (channel) constraint

$$q_{kj} \leq z_{kj} \leq Z_k \tag{27}$$

and the outputs $\tilde{q}_{kj}$ of the other firms in markets 1,2 and 3. The four cases with the four alternative interference graphs are considered next.

A) Homogenous interference graphs

Levine and Rickman start by introducing the cases with the two homogenous graphs i.e. the complete graph and the graph with no edges. An interference graph with no edges implies that all channels are available in each market and thus $z_{kj}=Z_k$. In other words it is assumed that all channels can be shared. Define $\lambda_j \geq 0$ as the shadow price for the spectrum constraint (27). Then the firm k maximizes the Lagrangian:

$$L_k = \Pi_k - \sum_{j=1}^{3} \lambda_j (q_{kj} - Z_k) \tag{28}$$

with respect to output q_{kj}, spectrum/channels Z_k and λ given the decisions of other firms. Remembering the definition of profits Π_k from (26) the first order conditions for j=1,2,3 are:

$$q_{kj} : (P_j - c) + q_j P'_j - \lambda_j = 0 \tag{29}$$

$$Z_k : -a + \sum_{j}^{3} \lambda_j = 0 \tag{30}$$

$$CS : \lambda_j (q_{kj} - Z_k) = 0 \tag{31}$$

The first FOC equates the marginal return of providing q in each market with the shadow price of spectrum in that particular market. The second FOC describes the license fee a as the shadow price of spectrum across the three markets i.e. correctly equating it with the network shadow price of the network operator. The third FOC simply states that if the spectrum constraint is not binding, i.e. there are no interference constraints and thus basically no scarcity in some market, the shadow value or price λ_j equals zero. This gives the Kuhn-Tucker complementary slackness conditions for each market.

The above calculation implies the following solution. By symmetry $\lambda_j = \lambda = a/3 > 0$ and thus the constraints are binding. In addition, in a symmetric Nash equilibrium the Lerner price is given by:

$$P_j = P = \frac{c + \frac{a}{3}}{1 - \frac{1}{\in N}} \tag{32}$$

Output is given by $Q = AP^{-\epsilon}$ (when $D(P) = AP^{-\varepsilon}$) in each of the markets and profits by $\Pi(a, N) = (P - c - a/3)\, Q/N - F$ per market as defined by (26). The number of firms N is, as was previously discussed, determined either by the free entry condition by which profits equal zero ($\Pi(a, N) = 0$) or through policies imposed by the regulator.

The second homogenous case where all nodes are joined, i.e. there is no possibility for channel sharing between markets, is quite similar to the no-edges case. Given Z_k there are now $z_{kj} = Z_k/3$

channels available per market. Through an analysis similar to that above the price in each market becomes:

$$P_j = P = \frac{c + a}{1 - \frac{1}{\in N}} \tag{33}$$

It is noteworthy that this equilibrium price is higher than in (32). Thus, since channel sharing does not exist across markets and firms cannot share the costs of spectrum, the corresponding equilibrium price for the end products and services increases. Proposition 1 stating that profits decrease with increases in a or N also implies that now profits are lower than in the previous case and less firms will enter the market in free entry equilibrium.

B) Inhomogenous interference graphs

Levine and Rickman then move on to discuss the two inhomogenous cases where spectrum sharing is partially restricted; the graph with two edges and the graph with one edge. These cases are of more interest since they imply differing prices between markets due to differing interference constraints. Thus, they resemble the actual market conditions for markets utilizing spectrum better than the homogenous cases; as we discussed in chapter 2 previously the demand and thus congestion between spectrum bands (i.e. here markets or sectors offering different services) differ implying different restrictions on usage and spectrum sharing.

Let us start with the graph with two edges as depicted in appendix 2 on the right. The markets j=1,2,3 are located in the three nodes marked with A,B and C. The form of the interference graph now indicates that markets B and C are able to share channels, but sharing is otherwise restricted. Thus, if we first assume that for a firm k, i.e. a network operator operating in all of the markets j=1,2,3, all Z_k channels are available in market 1 (=A) and q_{k1} of them are used, then $z_{k2} = z_{k3} = Z_{k1}$-q_{k1} are available in markets 2 (=B) and 3(=C). When making its production decisions as well decisions on how much to require spectrum resources firm k maximizes a Lagrangian:

$$L_k = \Pi_k - \lambda_1(q_{k1} - Z_k) - \lambda_2(q_{k2} - Z_k + q_{k1}) - \lambda_3(q_{k3} - Z_k + q_{k1}) \tag{34}$$

with respect to q_{kj}, Z_k and λ_k given the corresponding choices of other firms. Thus, the first order conditions are:

$$q_{k1}:(P_1-c)+q_1P'_1-\sum_j^3\lambda_j=0 \tag{35}$$

$$q_{k2}:(P_2-c)+q_2P'_2-\lambda_2=0 \tag{36}$$

$$q_{k3}:(P_3-c)+q_3P'_3-\lambda_3=0 \tag{37}$$

$$Z_k:-a+\sum_j^3\lambda_j=0 \tag{38}$$

$$CS:\lambda_{kj}(q_{kj}-z_{kj})=0 \tag{39}$$

Since market 1 releases spectrum resources to be used in the two other markets the spectrum constraint (CS) for market 1 does not bind and $\lambda_1=0$. Spectrum is fully utilized in markets 2 and 3 so that λ_2, $\lambda_3>0$. Following the same reasoning as before, in symmetric Cournot-Nash equilibria the equilibrium prices become:

$$P_1=\frac{c+a}{1-\frac{1}{\in N}} \tag{40}$$

$$P_2=P_3=\frac{c+\frac{a}{2}}{1-\frac{1}{\in N}} \tag{41}$$

It is noteworthy that the prices P_j are lower in markets where spectrum sharing is possible i.e. $P_2=P_3<P_1$. This is logical since the firms can share costs of spectrum use across these markets. Comparing these prices to the equilibrium prices attained in sectors with homogenous graphs reveals that the price is the lowest with full spectrum sharing (the first case discussed, an interference graph with no edges), mostly due to the same reason of cost sharing across markets. Compared to a sector with full-edge interference graphs with no spectrum sharing the prices are lower in the markets allowing spectrum sharing (2 and 3), but according to proposition 1 profits will initially be higher. This is due to the reason that spectrum sharing allows for lower license fee a, which in turn increases the profits earned by the spectrum users (firms). Thus, in a less congested sector (in the radio interference sense) more firms will enter the market attracted by the higher profits making the market more competitive (Levine & Rickman 2007, p.18).

We now move on to discussing a sector with an interference graph with one edge as depicted in appendix 2 on the left. The markets j=1,2,3 are again located in nodes A,B and C. The form of the interference graph now indicates that markets A and B as well as A and C are able to share channels, but sharing is restricted between markets B and C. For a firm k all Z_k channels are available in market 1 (=A) implying $Z_k = z_{k1}$. These available channels A can be shared in markets 2 and 3, but not between markets 2 and 3, and thus $z_{k2} = z_{k3} = Z/2$. The Lagrangian to be maximized with respect to q_{kj}, Z_k and λ_k, given other firms' actions, by firm k now takes the form:

$$L_k = \Pi_k - \lambda_1(q_{k1} - Z_k) - \lambda_k(q_{k2} - \frac{Z_k}{2}) - \lambda_3(q_{k3} - \frac{Z_k}{2}) \tag{42}$$

In the familiar way the FOC's are the following:

$$q_{kj} : (P_j - c) + q_j P'_j - \lambda_j = 0 \tag{43}$$

$$Z_k : -a + \lambda_1 + \frac{1}{2}(\lambda_2 + \lambda_3) = 0 \tag{44}$$

$$CS : \lambda_{kj}(q_{kj} - z_{kj}) = 0 \tag{45}$$

Levine and Rickman first solve for type I equilibrium where all the spectrum resource constraints bind ($\lambda_j>0$, j=1,2,3). Firstly, by symmetry of markets B and C it must hold that $\lambda_2 = \lambda_3$. Based on the FOC's the solution must then satisfy:

$$P_1 = \frac{c + \lambda_1}{1 - \frac{1}{\in N}} \tag{46}$$

$$P_2 = P_3 = \frac{c + a - \lambda_1}{1 - \frac{1}{\in N}} \tag{47}$$

$$D(P_1) = 2D(P_2)\text{, where } D(P) = AP^{-\in} \tag{48}$$

$$\lambda_2 = \lambda_3 = a - \lambda_1 \tag{49}$$

Solving for P_1, P_2, λ_1 and λ_2 we get:

$$P_1 = \frac{1}{(1 + 2^{\in})}\left[\frac{2c + a}{1 - \frac{1}{\in N}}\right] \tag{50}$$

$$P_2 = P_3 = \frac{2^{\epsilon}}{(1+2^{\epsilon})} \left[\frac{2c+a}{1-\frac{1}{\epsilon N}} \right] \tag{51}$$

$$\lambda_1 = \frac{c(1-2^{\epsilon})+a}{1+2^{\epsilon}} \tag{52}$$

$$\lambda_2 = \lambda_3 = \frac{(2^{\epsilon}-1)(a+c)}{(1+2^{\epsilon})} \tag{53}$$

From $\lambda_2 = \lambda_3 > 0$ we get $\epsilon > 1$, which was previously imposed. In addition, $\lambda_1 > 0$ requires that the license price a satisfies the following:

$$a > c(2^{\epsilon} - 1) \tag{54}$$

If the condition provided by (54) does not hold we have a type II equilibrium where the spectrum capacity constraint for market 1 does not bind i.e. $\lambda_1 = 0$ meaning that there exist spare spectrum channels. In that case the pricing decisions in the optimum satisfy:

$$P_1 = \frac{c}{1-\frac{1}{\epsilon N}} \tag{55}$$

$$P_2 = P_3 = \frac{c+a}{1-\frac{1}{\epsilon N}} \tag{56}$$

$$\lambda_2 = \lambda_3 = a \tag{57}$$

The results given above emphasize the relationship between interference constraints, license fees and the pricing and channel usage decisions of firms. If and only if the regulator imposes a sufficiently high license fee (AIP) so that (54) holds, all channels will be in full utilization in each market (Levine and Rickman 2007, p.20). Thus, Levine and Rickman's analysis clearly implies that in order to guarantee technical efficiency a sufficiently high AIP is needed; preventing any firm from acquiring a license and still leaving the spectrum un- or underutilized. On the other hand, prices are increased directly through the effect of the of the license fee on retail Lerner index (measuring market power) and indirectly through increased concentration of competition in a free-entry equilibrium. This in turn verifies that neither too high AIP payments impair overall welfare as well by decreasing consumer surplus.

The framework above can be extended by developing software capable of handling larger problems, but the small node number examples considered are sufficient in identifying the key issues behind AIP payments and their effects on market outcomes, which we will continue studying next. Levine and Rickman (2007, p.20) especially state that these examples demonstrate that the "spatially distributed aspects of channel assignment problems provide new challenges for analysis that go beyond standard economic treatments."

4.2.4.1 Optimal AIP

In the analysis above the sector subscript i was dropped since only a single sector with three local markets was considered. Next Levine and Rickman move on to discuss the regulator's choice of an optimal price for a particular sector i, i.e. sector-specific prices are introduced. The equilibrium concept considered is the free entry equilibrium and thus firms enter the market until profits are driven to zero. For analytical convenience only cases with homogenous graphs are considered, but instead of l_i=3 the case is generalized to account for any number of local markets l_i in sector i. As previously, the firms act as network operators providing a homogenous service across the local markets within the sector. It is however noteworthy that with homogenous graphs (no or full spectrum sharing) this is equivalent to assuming local operators: if we let a^L be the licence price for local operators and put a = la^L we arrive at an identical optimization problem described below (Levine & Rickman 2007, p.20). Within a sector i the N_i network operators are identical and they demand Z_i spectrum channels at a license price a_i to be determined here. The revenue for the regulator in sector i is thus $N_i Z_i a_i$ and in case of homogenous graphs, as was shown previously, the retail prices are identical across markets in a particular sector i, i.e. P_{ij}=P_i. Remembering the welfare function (19) previously and the definition for the consumer surplus S_{ij} in (20) as well as the fact that due to homogeneity of the interference constraints symmetry between markets requires S_{ij}=S_i, the regulator's problem set out in general form becomes:

$$\max W = W(\underline{a}) = \sum_{i=1}^{p} l_i \int_{P_i}^{\infty} D_i(P)dP + (1+\Lambda)\sum_{i=1}^{p} N_i Z_i a_i \qquad (58)$$

with respect to the license prices $\underline{a} = (a_1, a_2, ...a_p)$ in each sector from 1 to p and subject to the spectrum resource constraint

$$\sum_{i=1}^{p} N_i Z_i \le Z^{\max} \qquad (59)$$

and the interference constraints given by the graphs.

Since the analysis is restricted to homogenous graphs there are two types of relevant sectors: the ones with l_i nodes (local markets) all connected to each other, i.e. sectors without any spectrum sharing or re-use and those with full spectrum sharing. For the sectors without the re-use property the demand for spectrum is given by $N_i Z_i = l_i D_i(P_i)$ and based on the previous analysis the retail price is given by $P_i = \dfrac{c_i + a_i}{1 - \dfrac{1}{\in_i N_i}}$. For sectors with spectrum sharing the equivalent demand is determined by $N_i Z_i = D_i(P_i)$ and the retail price by $P_i = \dfrac{c_i + \dfrac{a_i}{l_i}}{1 - \dfrac{1}{\in_i N_i}}$. If we define k_i=1 for sectors without re-use and $k_i=1/l_i$ for sectors with re-use, the welfare maximization problem (58) can be written as:

$$W = W(\underline{a}) = \sum_{i=1}^{p} l_i \left[\int_{P_i}^{\infty} D_i(P)dP + (1+\Lambda) k_i D_i(P_i) a_i \right], \tag{60}$$

where

$$P_i = \frac{c_i + k_i a_i}{1 - \dfrac{1}{\in_i N_i}} \tag{61}$$

The regulator now maximizes (60) with respect to the spectrum prices (a vector) $\underline{a} = (a_1, a_2, ...a_p)$ subject to the interference constraints and the resource constraint

$$\sum_{i=1}^{p} k_i l_i D_i(P_i) \leq Z^{\max} \tag{62}$$

The interference constraints state that due to their (geographical) proximity channels cannot be shared within a local market but a firm, i.e. a network operator, can share channels between markets in a given sector provided that this sector supports channel sharing (i.e. $k_i=1/l_i$). Channel sharing between sectors is assumed to be prohibited by international harmonization agreements. In other words, certain channels are by agreement allocated to a certain use (i.e. sector, such as mobile communications), which cannot vary. This is often the case in reality as well; remember the discussion in chapter 3.1 regarding the often pre-determined or internationally coordinated nature of spectrum uses leaving only the assignment to be determined by the (national) regulators.

The regulator then assigns channels to each firm within different sectors. Taking the license price as given firms compete by making entry and exit decisions. This results in a retail price for each sector given by (61). The regulator's optimization problem is given by a Lagrangian:

$$L = W(\underline{a}) - \mu\left(\sum_{i=1}^{p} k_i l_i D_i(P_i) - Z^{\max}\right) \tag{63}$$

where μ as the resource constraint multiplier represents the shadow value of spectrum. If we write the retail price (61) as $P_i = P_i(a_i, N_i(a_i))$, the first order condition with respect to the spectrum price a_i optimized by the regulator equals:

$$-D_i(P_i)\frac{dP_i}{da_i} + (1+\Lambda)k_i\left(D_i(P_i) + a_i D'(P_i)\frac{dP_i}{da_i}\right) = \mu k_i D'(P_i)\frac{dP_i}{da_i} \tag{64}$$

On the left-hand-side of (64) the first term depicts the negative impact (marginal cost) of a (marginal) increase in spectrum license fee on consumer surplus through increased retail prices (given that $\frac{dP}{da} > 0$) and the second term captures the positive effect (marginal benefit) the change in a_i has on welfare through increased revenues. On the right-hand-side we have the marginal cost of spectrum evaluated at its shadow price μ. Levine and Rickman (2007, p.22) point out, that the expression for the optimal AIP in each sector does not depend on conditions of other sectors, but only on the sector-specific demand and supply conditions as well as the shadow price of spectrum. This is due to the assumed independence of sectors, i.e. services between sectors are neither substitutes nor complements as was previously imposed. Note that the value of the shadow price reflects spectrum scarcity: μ > 0 suggests scarcity exists whereas μ = 0 implies there is no shortage of spectrum.

Next Levine and Rickman proceed to evaluate dP_i/da_i using the free-entry condition

$$\Pi_i = \frac{l_i P_i D_i(P_i)}{\in_i N_i^2} - F_i = 0 \tag{65}$$

as calculated in more detail in appendix 3. Thus differentiating $P_i = P_i(a_i, N_i(a_i))$ with respect to a_i gives:

$$\frac{dP_i}{da_i} = \frac{\partial P_i}{\partial a_i} + \frac{\partial P_i}{\partial N_i}\frac{dN_i}{da_i} \tag{66}$$

On the right-hand-side we have the direct effect of license price changes on retail price accompanied with the indirect effect caused by the firms exiting the market as the cost of acquiring and holding spectrum increases. Solving the free-entry condition for the number of firms N_i and differentiating it with respect to the spectrum price in turn gives (for $\epsilon > 1$ assumed throughout the analysis):

$$\frac{dN_i}{da_i} = -\frac{(\epsilon_i - 1)(\epsilon_i N_i - 1)}{(2\epsilon_i N_i - 1 - \epsilon_i)} \frac{N_i \partial P_i}{P_i \partial a_i} < 0 \tag{67}$$

which is in accordance with what was discussed earlier; increasing operational costs (license fees) drive some firms out of business thus decreasing the number of firms in a given sector (N_i) and establishing a negative connection between license fees and the number of firms.

Substituting this into the previous equation (66) we arrive at:

$$\frac{dP_i}{da_i} = \rho_i \frac{\partial P_i}{\partial a_i} \tag{68}$$

where ρ_i is defined as

$$\rho_i = \rho_i(\epsilon_i, N_i) = \frac{2(\epsilon_i N_i - 1)}{(2\epsilon_i N_i - 1 - \epsilon_i)}. \tag{69}$$

In other words, the size of ρ_i is dependent on the number of firms and the price elasticity of demand (of the end products). From (69) it can be seen that as the number of firms tends towards infinity ($N_i \to \infty$), ρ_i approaches unity. Given N_i and letting ϵ_i vary between one and infinity, i.e. $\epsilon \in [1, \infty]$, then it must hold that $\rho_i \in \left[1, \frac{2N_i}{2N_i - 1}\right]$.

Dividing the FOC $dL / da_i = 0$ in (64) by the demand $D_i(P_i)$ and remembering the basic definition for the elasticity of demand $\epsilon_i = -\frac{P_i dD_i}{D_i dP_i}$, it is possible to write the FOC as:

$$a_i = \frac{\mu \epsilon_i \rho_i (1 - \frac{1}{\epsilon_i N_i}) + \frac{c_i}{k_i}((1+\Lambda)(1 - \frac{1}{\epsilon_i N_i}) - \rho_i)}{\rho_i + (1+\Lambda)(\epsilon_i \rho_i - 1)(1 - \frac{1}{\epsilon_i N_i})} \tag{70}$$

This equation along with the free-entry condition, where profits are driven to zero:

$$\Pi_i = \frac{l_i P_i D_i(P_i)}{\epsilon_i N_i^2} - F_i = 0 \tag{71}$$

The Kuhn-Tucker complementary slackness conditions:

$$\mu(\sum_{i=1}^{p} k_i D_i(P_i) - Z^{\max}) = 0 \tag{72}$$

and the retail price equation previously stated in (61):

$$P_i = \frac{c_i + k_i a_i}{1 - \frac{1}{\in_i N_i}}$$

determine the four equations for N_i, a_i, P_i and μ_i at the optimum, given the parameters $\in_i, c_i, k_i$ and A_i. This completes the solution for the optimal license price.

Although this method by Levine and Rickman gives a more sophisticated solution to determining AIP than any of the alternative methods suggested (see 4.1. and 4.3) it also imposes heavy information requirements for the regulator responsible for such a task. Namely, the price elasticity of demand, information on cost properties of firms within the market as well as accurate knowledge of the interference constraints must be possessed in order to solve such an optimization problem.

The first order condition (70) highlights an important result, which connects the optimal license price to the amount of congestion captured by k_i. It suggests that given the number of firms N_i, as congestion (measured by k_i) increases the optimal license fee (AIP) decreases and $a_i(N_i)$ shifts downwards. In addition, from the retail price equation (61) it is evident that congestion increases the retail price (or spectrum price per unit of output) so that the downward-sloping free-entry relationship $N_i(a)$ depicted in appendix 4 shifts to the left. This result is summarized by Levine and Rickman (2007, p.24) as

Proposition 3:

Assuming linear technology for spectrum and homogeneous interference graphs, ceteris paribus the optimal license fee (AIP) in sectors without spectrum sharing (i.e. re-use) is lower than that in sectors with spectrum sharing.

In other words, proposition three states that the constraint on spectrum sharing does not imply any "congestion tax" with the form of an increased license price. Levine and Rickman (2007, p.24) explain this perhaps counter-intuitive result in the following way. A higher license price is not necessary to reduce demand of spectrum in areas without spectrum re-use (or sharing) since firms do this themselves through increasing the retail prices. This is illustrated in the examples in appendices 5 and 6, of which 5 describes the situation where spectrum is scarce (i.e. the shadow value of spectrum $\mu > 0$) and 6 the situation without spectrum scarcity ($\mu=0$)[22]. With scarcity the retail price increases as k_i reflecting the congestion level rises from $k_i=1/l_i$ for an interference graph without edges (full spectrum sharing allowed) to $k_i=1$ for graphs with full edges (no channel sharing allowed).

The issue presented above can be further explained with a help of an example. Let the number of local markets in a given sector be five i.e. $l_i=5$.Then we have $k_i=1/5$ in the sector with spectrum sharing and $k_i=1$ for sectors without it. If spectrum is scarce ($\mu > 0$, the figure in appendix 5) comparing the first with the second shows that in the illustrative example the retail price more than doubles and the number of firms drops from approximately 5 to 3. Thus spectrum sharing allows for more firms to enter the market increasing competition. The regulator is naturally equally concerned with the welfare of both of these sectors and mitigates the effect of increasing market power (and thus increasing retail prices) by lowering the license price in the sector without spectrum sharing relative to the one with spectrum sharing. The previously discussed regulator's strive for equality (between consumers in each market) can be seen here; without the mitigation the spectrum license fee would further increase retail prices in the more concentrated market decreasing consumer surplus relative to the more competitive market.

If there is no scarcity of spectrum resources ($\mu=0$, the figure in appendix 6), equation (70) implies that $a_i k_i$ and thus the retail price in (61) is independent from k_i. The free-entry condition (71) further implies that this is also true for the number of firms N_i. This is depicted in the figure by the horizontal lines for the retail price P and the number of firms N. Thus, it must be that a_i is only proportional to $1/k_i$, which is also depicted in the figure in appendix 6.

Furthermore, if N_i increases substantially implying increased competition from (69) we get $\rho \to 1$ and (70) becomes

$$a_i \to \frac{\mu \in_i + \frac{c_i}{k_i}\Lambda}{(1+\Lambda)\in_i - \Lambda} \tag{73}$$

[22] These illustrative examples have been defined in a way that the baseline parameter values equal: c_i = Ai = 1, Λ= 0,3, ϵ_i= 2 and l_i = 3. Fixed entry costs F_i are chosen so that at baseline parameter values in sectors without spectrum re-use (k_i = 3), Ni = 4 (Levine & Rickman 2007, p.24).

and thus for Λ > 0, i.e. when taxes are regarded as distortionary, ${da_i}/{d \in_i} < 0$. This can also be seen in the figure in appendix 7, and in the numerical example this seems to hold also for smaller N_i. In other words, the inverse relationship between the license fee and elasticity of demand implies that the spectrum license fee acts as a Ramsey price. The result is summarized in the following proposition (Levine & Rickman 2007, p.25):

Proposition 4.

For large Ni the optimal spectrum fee acts as a Ramsey tax across the sectors. In other words, the optimal AIP in a sector is inversely related to the elasticity of demand. Based on a numerical analysis this may also hold for small N_i. This result presupposes Λ > 1, i.e. that taxes have distortionary effects.

If however Λ = 0 implying that there are no distortionary effects from taxation, the optimal license fee equals the shadow value of spectrum $a_i = \mu$, and is thus independent of any sector-specific features.

As was discussed previously, the (national) regulator is in many ways constrained by international harmonization and other regulations. So far we have assumed that license fees affect the total welfare positively by increasing the "revenue term" in (19). The situation may however be such that the regulator is by law constrained to ignore these benefits; for example laws forbidding any kind of revenue-raising objectives of AIP might constrain the regulator in this way. The general framework by Levin and Rickman handles this by assuming Λ = -1 in (19). When N_i is large (the competitive case) (73) becomes

$$a_i \to \mu \in_i - \frac{c_i}{k_i} \tag{74}$$

In this case the relationship between optimal AIP and the elasticity of demand is reversed. Also the relationship between the spectrum re-use parameter k_i and optimal AIP is the opposite to what was discussed in propositions 3 and 4. The results are thus heavily dependent on revenue generation considerations. This is summarized as (Levine & Rickman 2007, p.26):

Proposition 5

If welfare benefits of revenue collection through AIP payments are ignored due to legal or other reasons, the constrained optimal license fee in sectors without spectrum re-use is greater. For large

Ni, the optimal AIP in sectors allowing spectrum re-use is positively related to the elasticity of demand.

It is important to note that given proposition 5, the optimal AIP is interdependent on things usually beyond the control of any national regulator. Unless there is true international coordination between national and international telecommunication bodies, the level of welfare maximization that a national regulator can achieve while considering spectrum allocation, assignment and the possibility of AIP payments is restricted. The results regarding proposition 5 are graphically depicted in appendices 8 and 9.

4.2.4.2 Incorporation of Costs of Adjusting Licence Prices

As was discussed in subchapter 4.1, the Smith-NERA method of calculating AIP payments views the license fee formation as an incremental process. It suggests that the AIP payments need to be checked and altered in response to changes in opportunity costs caused by altered allocations of spectrum resources as well as due to technical development. Thus, only over time can the optimal solution (or a solution sufficiently near to the optimum) in allocation of spectrum be achieved. Levine and Rickman (2007) include this property in their model by introducing costs for adjusting the license prices. In this way the optimization problem becomes dynamic (instead of static).

The vector of optimal AIP payments solving the static problem set out previously is given by $\underline{a}*$. Now Levine and Rickman consider an intertemporal welfare loss function:

$$\Omega = \sum_{t=1}^{\infty} \beta^t \sum_{i=1}^{p} \left[(a_i * - a_i)^2 + \Phi (a_i(t) - a_i(t-1))^2 \right] \tag{75}$$

where β ϵ (0,1] is the discount factor and Φ a parameter representing the cost of adjustment towards the optimum (a_i* in a given sector or $\underline{a}*$ for the economy as a whole). The term $a_i * - a_i$ naturally expresses the distance from the optimum at any given time. Equation (75) can be interpreted in the following way (Levine and Rickman 2007, p. 26). The regulator would like to be at the static optimum with an optimal set of prices for all sectors. However, as the attainment of the optimum is an iterative process towards this optimum, the regulator faces costs of changing prices, which are proportional to $(\Delta a(t))^2$, where $\Delta a_i(t) = a_i(t) - a_i(t-1)$ expresses the change in the AIP payment over a time interval [t-1,t] in sector i. As $\Phi \to 0$, this problem approaches the static optimization problem discussed above, where the regulator jumps to the static optimum right away, i.e. $\underline{a} = \underline{a}*$.

At t=1 the regulator minimizes the welfare loss Ω with respect to the license prices $\underline{a}$ given historical prices $\underline{a}(0)$. The first order condition is expressed as:

$$-(a_i*-a_i(t))+\Phi(a_i(t)-a_i(t-1)-\Phi\beta(a_i(t+1)-a_i(t))=0 \qquad (76)$$

If the deviation of the current license $a_i(t)$ fee from the static optimum a_i* is denoted by $\hat{a}_i(t)$, i.e. $\hat{a}_i(t)=a_i(t)-a_i*$, equation (76) can be written as

$$\beta\Phi\hat{a}_i(t+1)-(1+\Phi(1+\beta))\hat{a}_i(t)+\Phi\hat{a}_i(t-1)=0 \qquad (77)$$

which is a second-order difference equation of $\hat{a}_i(t)$. To solve this z-transforms are taken to get the characteristic equation:

$$\beta\Phi z^2-(1+\Phi(1+\beta))z+\Phi=0 \qquad (78)$$

This second order equation has two positive roots, $z_1 < 1$ and $z_2 > 1$ and thus the system is saddle-path stable with a solution:

$$\hat{a}_i(t)=z_1^t\hat{a}_i(0) \qquad (79)$$

The adjustment process for a specific sector is illustrated with a numerical example in appendix 10. In this example the optimal AIP is assumed to equal unity ($a_i*=1$) and the starting point is at zero ($a_i(0)=0$). The example shows that the lower the adjustment costs the quicker the adjustment to the optimum. The optimization problem with adjustment costs illustrates the actions of a rational regulator, who adjusts the license fee towards the optimum with the speed of transition depending on the level of adjustment costs. This can be paralleled with the adjustment process described by the Smith-NERA method, where adjustments were to be made gradually.

4.2.5 Optimal pricing with general technology

In 4.2.4 the production technology was assumed to be linear restricting substitutability between spectrum resources and other factors of production. This was mainly done for reasons of tractability. As this may be too strict a restriction in reality, Levine and Rickman (2007) also set out the pricing regime with more general technology, which allows for substitution between spectrum and other inputs. In this way production function such as (3) or (4) depicted earlier can be considered.

The total costs of producing output q_i in sector i with a spectrum license fee a_i become:

$$C_i(q_i, a_i) = F_i + c_i(w_i, r_i, k_i a_i) q_i \tag{80}$$

F again represents the fixed (or the set-up) costs and c_i the costs per unit of output, where $k_i a_i$ is the effective cost of spectrum for firms. When these costs are minimized with respect to the license price, by Shephard's Lemma we end up with the demand for spectrum in sector i:

$$Z_i = \frac{\partial C_i}{\partial a_i} = l_i k_i c_{3i} \frac{D_i(P_i)}{N_i} \quad , \tag{81}$$

in which

$$c_{3i} = \frac{\partial c_i(w_i, r_i, x)}{\partial x} \tag{82}$$

The regulator's optimization problem for linear technology given by (60) and (62) can be generalized as:

$$\max W = W(\underline{a}) = \sum_{i=1}^{p} l_i \left[\int_{P_i}^{\infty} D_i(P) dP + (1+\Lambda) k_i c_{3i}(w_i, r_i, k_i a_i) D_i(P_i) a_i \right] \tag{83}$$

with respect to license prices $\underline{a}$ subject to the resource constraint

$$\sum_{i=1}^{p} k_i c_{3i}(w_i, r_i, k_i a_i) D_i(P_i) \le Z^{\max} \tag{84}$$

and the interference constraints given by the interference graphs. In (84) the retail price is naturally given by

$$P_i = \frac{c_i(w_i, r_i, k_i a_i)}{1 - \dfrac{1}{\in_i N_i}} \quad . \tag{85}$$

To be able to perform the procedure a production function for the service utilizing spectrum or the cost function (80) is needed. Even in that case the FOC's are not analytically tractable as was the case with a linear production function, and they would require a numerical solution. Thus, the solution to the problem with general technology is left here by Levine and Rickman (2007). It is however noteworthy that important information regarding the setup of optimal AIP is still achieved as was brought forward by the analysis; especially the propositions 1-5 and the group of equations (70-72 and 62 restated) providing the solution for optimal AIP in case of linear technology. In the next section we move on to discuss and compare the alternative approaches of determining AIP discussed above.

4.2.6 Critique and comparisons to the Smith-NERA method

This section draws comparisons between the different approaches for determining the optimal AIP payment discussed in the previous sections 4.1 and 4.2, and discusses the benefits and pitfalls of each of them. For notational purposes the approach developed by Smith and NERA (1996) and further enhanced by Indepen et al. (2007) is referred to as the Smith-NERA method and the approach proposed by Levine and Rickman (2007) the L-R method. As the Smith-NERA method's limitations were already discussed in detail in 4.1.3 this section concentrates on comparing the two methods. In addition, this section reflects on the key ways in which the L-R method accounts for some of the issues identified in Smith-NERA and what pitfalls may still remain.

First of all, the Smith-NERA approach can be made comparable to the L-R method by stating it as follows (Levine & Rickman 2007, p.28-31): Assume again as in Smith-NERA that there are two sectors in the economy, i.e. in the L-R context p=2. The cost function per unit of input q_i is given by

$$c_i(w,r,a) = wL_i(w,r,a) + rK_i(w,r,a) + aZ_i(w,r,a); i = 1,2 \tag{86}$$

If no constraints are imposed on firms regarding the inputs they can use then (86) represent the minimum costs chosen by firms given input prices (w,r,a). However, assume that the license fee is too low to clear the market in a way that that there exists at least one firm which lacks the spectrum to achieve the minimum costs. In other words, excessively low pricing of frequency resources leads to significant excess demand, which cannot all be satisfied by the current, limited amount of the resource.

Assume that firm 1 lacks spectrum whereas firm 2 is unconstrained. If we then consider a small incremental increase in spectrum for firm 1, ΔZ_1, which the firm substitutes for labor ΔL_1 and capital ΔK_1 the change in cost is given by

$$\Delta c_1 = -w\Delta L1 - r\Delta K_1 + a\Delta Z_1 < 0 \tag{87}$$

and for the unconstrained firm the change in cost equals

$$\Delta c_2 = w\Delta L_2 + r\Delta K_2 - a\Delta Z_2 > 0 \tag{88}$$

knowing that $\Delta Z_2 = - \Delta Z_1$ and firm 2 has minimized costs before the re-allocation. The absolute value $|\Delta c_1|$ represents the opportunity cost discussed in the Smith-NERA methodology. This value represents what the society forgoes by allocating spectrum to sector 2 rather than sector 1. As was discussed earlier, the opportunity cost estimates are based on costs of alternative uses for spectrum resources. In this way the Smith-NERA ideology can be stated in terms of the L-R methodology as well.

We can now move on to conclusively discuss the key strengths and weaknesses of the methods. The obvious benefit of the Smith-NERA method is its simplicity; the concept of opportunity cost is easy to understand and communicate to regulators as well as firms to which the spectrum license fee is applied. In addition, there already exist some estimates for these costs, as the methodology has been applied in some countries (namely the UK and New Zealand, see e.g. the reports by Ofcom). Naturally these costs are for a particular point in time with certain spectrum allocations, but they can be used as a benchmark when adjusting the AIP payment.

As already discussed in 4.1.3 and above, the Smith-NERA method however has several shortcomings. The key issues have to do with market structure (assumes perfect markets), the models accountancy for interference (does not explicitly take these constraints into account), restrictions on allocation vs. assignment of frequencies (presupposes allocation forgoing significant efficiency gains) and emphasizing private surplus (omits e.g. the effect of revenue generation for the government and on welfare).

In the L-R method the AIP construction is regarded as the regulator's optimization problem: the regulator maximizes overall welfare consisting of consumer surplus, producer surplus and the revenue streams for the government (i.e. society), with respect to the spectrum fee a given the interference and resource constraints. Thus, the model takes into account total welfare and the scarcity imposed congestion of spectrum resources.

In addition, the L-R model accounts for market structure by, rather than perfect markets, assuming several oligopolistic markets and sectors within a given economy. A crucial property in the L-R model is also that it allows for re-allocations of spectrum since, as was discussed earlier already in the introduction of the thesis, re-allocation to more optimal uses (in view of economic efficiency) may offer significant efficiency improvement compared to or in addition to re-assigning spectrum to users in a way that the user valuing the frequency most gets the license.

Both the Smith-NERA and L-R method include dynamic aspects agreeing that setting the AIP to its optimal level is dependent on the point of time and place as well as the surrounding circumstances at that time, such as the level of technical advancement. Changes in allocation of spectrum also change the underlying calculation assumptions of AIP and thus the optimal AIP varies in time and should be re-evaluated whenever the circumstances change. However, neither one of the approaches explicitly discusses how AIP may be used to ensure dynamic efficiency, i.e. encouraging innovation and especially investments in spectrum-efficient technologies. Both methods bring about improvements compared to the pure administrative methods, which do not impose price on frequencies implying significant economic rents for firms utilizing them and undermining the guiding effect the pricing may have on economic efficiency through allocating the resources to agents valuing them the most.

Although the method by Levine and Rickman gives a more sophisticated solution to determining AIP than any of the alternative methods suggested (see 4.1. and the upcoming chapter 4.3) it also imposes heavy information requirements for the regulator responsible for such a task. Namely, the price elasticity of demand, information on cost properties of firms within the market as well as accurate knowledge of the interference constraints must be possessed in order to solve such an optimization problem. In addition to this, the method proposed by Levine and Rickman is more complicated than Smith-NERA to apply in practice requiring significant computational power in order for the larger real life problems (with more than just three local markets to consider).

A further limitation of the model by Levine and Rickman can be stated to be omitting any considerations for strategic bidding among the spectrum users; a phenomenon often seen in real life and carefully accounted for e.g. in the auction literature. However, if we were to assume that policy measures, that are not explicitly captured by the model but which still exist in the market setting to which the model is embedded, take into account or restrict such strategic bidding the model remains evermore relevant.

4.3 AIP payments based on market transactions

AIP payments have also been suggested to be based on realized market transactions, i.e. auction prices or trading prices from the secondary markets when such are available. Applying market transaction prices is naturally straightforward and compared to the two alternatives for determining AIP presented in this thesis (subchapters 4.1 and 4.2) requires only moderate effort from the regulator's part in terms of price and relevance determination.

Using realized market transactions as a basis for AIP is however troublesome since there have been quite few market transactions and, as was pointed out earlier, no country has implemented a fully functioning free market for spectrum. Auction prices for same frequencies across different countries and markets vary significantly (nearly all national regulators collect and report the results of national auctions, but for a cross-country comparisons see e.g. Rayal[23]), which suggests that spectrum value is highly dependent on market structure as well as restrictions and conditions imposed in an auction. Thus, applicability across geographical borders is likely to be weak.

In addition, independent of market structure, the results of market transactions may be distorted for other reasons, such as the previously discussed concept of winner's curse, due to which the licensee may end up paying a significant premium on the actual value of the frequency. Imposing absurdly high AIP payments in turn may result in frequencies remaining completely unused. A fitting example of this is the 3G licenses granted by the national regulator Comisión del Mercado de las Telecomunicaciones (CMT) in Spain in March 2000, where the licenses were granted using beauty contest combined with AIP like payment that were imposed on the licenses. The regulator used the previous 3G auction results (pricing) from the United Kingdom Germany as benchmarks for setting the fees. As a result it was later on forced to lower the amount of the annual incentive payments as they no longer reflected the valuation of spectrum after the IT bubble had burst (ITU News Magazine, 2003).

Discrepancy between the private value of spectrum (according to which possible licensees bid in an auction) and social value for the entire society may further distort market transaction prices making them non-optimal as the basis for AIP. As there are also several constraints and pitfalls in determining AIP payments based on administrative judgment by the regulator, the realized

[23]Frank Rayal: http://frankrayal.com/2012/11/06/summary-of-select-spectrum-auction-results/

transaction prices can and should, however, be used as benchmarks for spectrum value, constructed using a more fine-tuned approach (such as the ones presented previously in this chapter)

Chapter 4 introduced the different methods that exist in determining AIP payments. These include the core method, called the Smith-NERA method, based on pricing at the level of opportunity cost, a method by Levine and Rickman (2007) which extends the Smith-NERA methodology to account for market structure and interference constraints, and a method which bases AIP payments on prices realized in market transactions.

The different methods were shown to impose very differing informational requirements for the regulator responsible for setting the AIP with the method by Levine and Rickman requiring most sophisticated knowledge whereas also giving most accurate estimates of welfare maximizing spectrum fees. The method by Levine and Rickman was shown to be a regulator's optimization problem where the regulator maximizes overall welfare with respect to the spectrum fee given the interference and resource constraints at the same time accounting for many of the limitations of the Smith-NERA model. Spectrum fees based on realized market prices were shown to be good only for benchmarking, not to be used for AIP determination as such due to their dependence on market characteristics and thus poor replicability across geographical or segment borders. The final chapter 5 summarizes the thesis and its main conclusions as well as makes suggestions for further studies.

5. Conclusions

The purpose of this thesis has been to offer an overview on spectrum management methods and to provide a detailed analysis of one of them, namely the Administrative Incentive Pricing (AIP). AIP is a pricing mechanism utilized in connection with administrative spectrum allocation and assignment methods to promote efficient use of spectrum. This section summarizes the key points in each previous chapter and draws conclusions on the applicability and formation of AIP as well as introduces possibilities for further studies.

Our society is becoming more and more wireless and in many other ways dependent on applications and services utilizing frequencies as inputs of production. Despite the growing demand and thus significance of spectrum resources in the society, as well as the vast economic benefits optimal utilization of such resource may bring, there have been very few new approaches to spectrum management during the last couple of decades. Traditional division and pricing methods such as beauty contests and auctions still dominate academic research. Given the scarcity of spectrum resources this should be an issue on every communication regulator's priority list.

Frequencies are managed and governed through a combination of international (e.g. ITU), multi-national or regional (e.g. the EU) and national authorities (national communications regulators such as MINTC in Finland). The need for spectrum management stems from the externality of interference, which imposes restrictions on spectrum usage. The main objective of spectrum management, alongside securing national safety and public interests, is efficient use of spectrum resources. Efficiency in the spectrum context is defined as economic efficiency, which can be shown to include the concepts of allocative, productive and dynamic efficiency.

Currently used spectrum management methods can be divided into three distinct categories, namely the administrative methods, i.e. lotteries, first-come-first-serve methods and beauty contests, market-based methods, i.e. auctions and secondary markets for spectrum, as well as the newer approaches of AIP, the viewpoint of frequencies as natural resources and the special case of unlicensed spectrum. Traditional assignment methods rely completely on the regulator to assign the frequency resources, without any market driven processes involved. As such they impose no price on frequencies (although it is common for the regulators to impose small cost recovery fees), but can be combined with pricing schemes such as the AIP. Market based methods in turn primarily rely on market processes to equalize supply and demand. The methods termed as new pursue to either find fundamental ways of treating and regulating spectrum differently than before (spectrum as a natural resource and unlicensed spectrum) or to complement more traditional methods in order to enhance efficiency of the resource use (AIP).

With respect to economic efficiency the market-based methods are preferred as they include bidding for the scarce resource and impose a price on spectrum. However, in some situations market based or even market mimicking methods (such as the AIP) cannot be utilized or other objectives than economic efficiency necessitate the use of other methods; the most important case being unlicensed spectrum, which existence is claimed to encourage innovation increasing dynamic efficiency. Auctioning can be shown to be preferred for high demand and value frequencies under competitive settings whereas AIP can also be utilized in connection with administrative assignment methods. It can also be used for public service frequencies, which traditionally have not been priced or subjected to any competitive bidding. As AIP is an annual (and changeable) fee imposed on spectrum users it effectively pushes spectrum users (firms) to regard spectrum as any other input of production when making production decisions. Auctioning fees in turn are imposed as a lump sum fee and thus may be regarded as sunk cost by the firms with weaker guiding effects from the viewpoint of the regulator.

Independent of the management methods used, the regulator is often bound to face a contradiction between private and social value of spectrum. Private spectrum value can be broken down into project and defensive values (forming the total returns from spectrum usage) and the option value. This private value in turn is likely to divert from the social value of spectrum whenever market distortions are present. Thus, the regulators' task often consists of counterbalancing somewhat contradictory objectives while also ensuring that efficiency is pursued.

When determining an optimal amount of AIP, the regulators currently have three methodologies at their disposal. The fee is usually based on the opportunity cost of spectrum use, since it is regarded to be a clear, relatively easily attainable measure in line with efficient outcomes, at least under perfect competition assumptions. The core method, called the Smith-NERA method is based on this principle. However, there are also two alternative ways proposed to calculate the optimal price; a method by Levine and Rickman (2007) which extends the Smith-NERA methodology to account for market structure and interference constraints, and a method which bases AIP payments on prices realized in market transactions.

First of all, the advantages and pitfalls of basing AIP payments on market transaction data are relatively straightforward. Applying this method is limited by the fact that there have been quite few market transactions and no country has implemented a fully functioning free market for spectrum. In other words, there is a lack of relevant pricing information to base the AIP estimates on. In addition, spectrum value is likely to be highly dependent on market structure as well as restrictions and conditions imposed in auctions implying weak applicability of results across geographies or sectors. All the downsides of auctioning are also at play hindering the use of realized transaction prices as anything more than benchmarks against AIP fees determined through more refined methods. Such downsides include overpaying for the auctioned resources (the winners curse), as well as the discrepancy between private value (according to which bids are submitted in auctions) and social value or the value maximizing overall welfare.

The two more refined models of determining AIP introduced in this thesis are of various degrees of complexity. The Smith-NERA model relies on opportunity costs and in its most simplistic form proposes license prices to be based on available estimates of the costs of alternative uses of the radio spectrum. Thus, in this relatively straightforward procedure the regulator first needs to identify the next best use/user for the spectrum resource and then determine the alternatives that the next best user has if denied the frequency input. Out of the possible alternatives the least expensive one is regarded as the alternative cost for spectrum. The model by Levine and Rickman in turn is a combination of economic modelling and information theory (in particular graph theory), which essentially treats the AIP formation as a welfare optimization problem. In such an optimization problem the regulator maximizes overall welfare consisting of consumer surplus, producer surplus and the revenue streams for the government with respect to the spectrum fee given the interference and resource constraints. When solved, the model gives a group of equations determining optimal AIP (see 4.2.4.1) assuming linear technology. The same can be repeated for general technology, but it requires computational power and is not analytically tractable (i.e. can only be shown with a numerical example). Key conclusions of the model by Levine and Rickman regarding optimal AIP can be summarized as follows: In a setting where interference, market structure and overall welfare including consumer and producer surplus as well as the revenue impact for the government from imposing AIP, the optimal AIP should be higher whenever spectrum sharing is possible and that it acts as Ramsey tax across sectors of the economy being inversely related to the elasticity of demand.

Based on the research and studies discussed in this thesis (especially Smith-NERA 1996; Indepen et al. 2004; Cave, Doyle and Webb 2007; Levine and Rickman 2007; Doyle 2007 and 2010; Aegis and Plum 2008) I have summarized below the key components that any model for determining optimal AIP payments should incorporate in order to promote efficiency. In connection with each component the feasibility of both the (more refined) models introduced in this thesis (Smith-NERA, Levine and Rickman) is assessed.

Firstly, to be of any real relevance the model needs to account for the actual market structure, which in many of the spectrum utilizing markets is far from perfect competition. The model by Levine and Rickman does this by assuming several oligopolistic markets whereas the Smith-NERA methodology relies on the unrealistic notion of perfect competition. However, neither of the models explicitly considers strategic bidding by the spectrum holders; a phenomenon often leading to market distortions. In order for the models to sufficiently resemble reality this kind of bidding is assumed to be taken into account (restricted) by policy measures, i.e. it is already accounted for by the market setting to which the model is embedded.

Secondly, the model needs to take into account interference, as interference is the main justification for spectrum management in general and the use of licensing as opposed to unlicensed spectrum. Levine and Rickman account for this by introducing interference constraints into the model utilizing graph theory while the Smith-NERA model in turn refrains from explicitly discussing interference. Thirdly, the model should allow for allocation as well as assignment of frequencies, as there are significant efficiency gains to be achieved in allocating spectrum to the welfare maximizing uses in addition to assigning them to users. Again, the model by Levine and Rickman satisfy this condition while the Smith-NERA method focuses on assignment of spectrum presupposing allocation (although the model's extension by Indepen et al. in 2004 in practice allows for considerations of re-allocation).

Fourthly, after it has been made sure that the model's assumptions are in line with reality (actual market structure and conditions) it should also account for all aspects of economic efficiency, namely allocative, productive and dynamic efficiency. This requirement can also be stated in terms of welfare maximization; the model must take into account the overall welfare, not for example only the private value (or producer surplus). As the objective of AIP is to impose a fee on spectrum users which encourages giving up un- or underutilized spectrum any AIP method inherently strives for technical efficiency (full utilization of spectrum) and thus productive efficiency. The Smith-

NERA method however focuses purely on productive efficiency, not acknowledging consumers' willingness to pay or the revenue impact that the collection of AIP has for the government. As was stated previously Levine and Rickman in turn consider optimizing overall welfare consisting of consumer and producer surplus and the revenue generation for the government.

As part of the efficiency determination, dynamic efficiency is again a concept inherently promoted by the idea of AIP, as it is an annually imposed and thus modifiable charge acting as a factor in the production decisions of the spectrum users. This differs from e.g. auction prices, which are regarded as sunk costs by the players when they make production decisions. Whenever AIP is optimally altered as time passes to reflect changes in the market circumstances or the development of technology (which obviously affect spectrum value) dynamic efficiency can be achieved. However, neither of the models presented takes a stand on dynamic efficiency and its implications on static efficiency leaving it a clear subject for future research regarding AIP.

In addition, this thesis has identified two main directions for further studies outside the application of AIP. On one hand, a study on the possible deregulation of spectrum as a result of the end of scarcity as technologies develop could be conducted to further challenge the main presumptions and conclusions of this thesis as well as spectrum management in general. On the other hand, the depicted viewpoint of spectrum as a natural resource and what this would imply for regulating and pricing spectrum resources is an interesting subject as well. Both of these studies would likely have profound effect on how we perceive, handle and manage the existing spectrum resources.

References

Literature – books and publications:

Aegis Systems Ltd, Plum Consulting Ltd (2008), "Administrative Incentive Pricing for Radiofrequency Spectrum", prepared for the Australian Communications and Media Authority (ACMA). Available at: http://www.acma.gov.au/webwr/_assets/main/lib310867/ifc12-09_app_a_1_plum_report_to_acma.pdf

Australian Communications and Media Authority, ACMA (2009), "Opportunity Cost pricing of Spectrum – Public consultation on administrative pricing for spectrum based on opportunity cost". Available at: http://www.acma.gov.au/webwr/_assets/main/lib310867/ifc12-09_final_opportunity_cost_pricing_of_spectrum.pdf

Bajari, Patrick; Hortacsu, Ali (2003), "The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay", *The RAND Journal of Economics*, Vol. 34, No. 2 (Summer, 2003), pp. 329-355

Buchanan, James (1969), "Cost and Choice: An Inquiry in Economic Theory", Markham Publishing Company, Chicago. Available at: http://www.econlib.org/library/Buchanan/buchCv6Cover.html

Calabrese, Michael (2009), "The End of Spectrum 'Scarcity': Building on the TV Bands Database to Access Unused Public Airwaves", New America Foundation, Wireless Future Program, working paper #25

Cave, Martin; Doyle, Chris; Webb, William (2007). "Essentials of Modern Spectrum Management". Cambridge University Press. Available at: http://www.knovel.com/web/portal/browse/display?_EXT_KNOVEL_DISPLAY_bookid=2334&VerticalID=0

Cramton, Peter (2002), "Spectrum Auctions", Handbook of Telecommunications Economics. Ed. Martin Cave, Sumit Majumdar, and Ingo Vogelsang. Amsterdam: Elsevier Science, p.605-639. Available at: http://works.bepress.com/cramton/50

Coase, R.H. (1959), "The Federal Communications Commission", *Journal of Law and Economics* vol. 2, p.1–40.

Doyle, Chris (2006), "The Economics of Pricing Radio Spectrum", Warwick Business School, prepared for the International Union of Radio Science (URSI). Available at: http://www.ursi.org/Proceedings/ProcGA05/pdf/C09.7%2802006%29.pdf

Doyle, Chris (2007), "The pricing of radio spectrum: using incentives mechanisms to achieve efficiency", Department of Economics, Warwick Business School, prepared for ITU workshop "Market mechanisms for spectrum management". Available at: http://www.itu.int/osg/spu/stn/spectrum/workshop_proceedings/Background_Papers_Final/Chris%20Doyle%20-%20Incentive%20based%20spectrum%20prices.pdf

Doyle, Chris (2010), "The need for a conservative approach to the pricing of radio spectrum and the renewal of radio spectrum licences", Apex Economics, prepared for the Australian Mobile Telecommunications Association. Available at http://www.amta.org.au/files/Doyle.Report.-.Spectrum.Licence.Renewal.pdf

Freyens, B. P., & Yerokhin, O. (2011), "Allocative vs. technical spectrum efficiency", *Telecommunications Policy*, vol. 35, issue 4, p.291-300

Indepen Consulting Ltd, Aegis Systems Ltd and Warwick Business School (2004), "An economic study to review spectrum pricing", prepared for the Radiocommunications Agency (RA). Available at: http://www.ictregulationtoolkit.org/en/Document.3044.pdf

ITU, International Telecommunication Union (2010), "Report ITU-R SM.2012-3 - Economic aspects of spectrum management". Available at: http://www.itu.int/pub/R-REP-SM.2012-3-2010

ITU, International Telecommunication Union (2012), “Exploring the Value and Economic Valuation of Spectrum”, Broadband Series Publications. Available at: http://www.itu.int/ITU-D/treg/broadband/ITU-BB-Reports_SpectrumValue.pdf

Levin, Harvey (1970), “Spectrum Allocation without Market”, *The American Economic Review*, Vol. 60, No. 2, p.209-218.

Levine, Paul; Rickman, Neil (2007), “Optimal Administered Incentive Pricing of Spectrum”, Discussion Papers in Economics, University of Surrey. Available at: http://www2.surrey.ac.uk/economics/files/apaperspdf/2007_DP10-07.pdf

MINTC, Ministry of Transport and Communications (2009), “Taajuusmaksukäytännöt eräissä EU-maissa”. Available at: http://www.MINTC.fi/c/document_library/get_file?folderId=339549&name=DLFE-8516.pdf&title=...

Marks, Phillippa; Williamson, Brian (2007), “Spectrum allocation, spectrum commons and public goods: The role of the market.”, *Communications and Strategies,* vol. 67, p.65-84.

McMillan, John (1994), “Why Auction the Spectrum”, *Telecommunications Policy*, Vol. 19, No. 3, p. 191-199.

McMillan, John (1995), “Selling Spectrum rights” *The Journal of Economic Perspectives*, Vol. 8, No. 3, p. 145-162.

Melody, William H. (1980), “Radio Spectrum Allocation: Role of the Market”, *The American Economic Review*, vol. 70, No. 2, Papers and Proceedings of the Ninety-Second Annual Meeting of the American Economic Association, pp. 393-397

Milgrom, Paul; Levin, Jonathan; Eilat, Assaf (2011), “The Case for Unlicensed Spectrum”, prepared for Google. Available at: http://ssrn.com/abstract=1948257

Ministry of Economic Development of New Zealand (2006), “Renewal of Management Rights for Cellular Services”. Available at: http://www.med.govt.nz

National Economic Research Associates Economic Consultants (NERA) and Smith System Engineering Ltd (1996), "Study into the use of Spectrum Pricing", prepared for the Radiocommunications Agency (RA). Available at: http://www.ofcom.org.uk/static/archive/ra/topics/spectrum-price/documents/smith/smith1.htm

Neto, Isabel; Wellenius, Bjorn. (2005), "The radio spectrum: Opportunities and challenges for the developing world", *The World Bank, Policy Research Working Paper Series:* 3742, Available at: http://search.proquest.com/docview/56604146?accountid=27468

Ofcom (2010), "SRSP: The revised Framework for Spectrum Pricing. Our policy and practice of setting AIP spectrum fees". Available at: http://stakeholders.ofcom.org.uk/binaries/consultations/srsp/statement/srsp-statement.pdf

Ryan, Patrick (2004), "Application of the Public-Trust Doctrine and Principles of Natural Resource Management to Electromagnetic Spectrum", *Michigan Telecommunications and Technology Law Review*, vol 10, pp. 285-372. Available at: http://www.mttlr.org/volten/Ryan.pdf

Ryan, Patrick (2005), "Treating the Wireless Spectrum as a Natural Resource", *Environmental Law Reporter*, Vol. 35, p. 10620. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=793526

Santivanez, C.; Ramanathan, R.; Partridge, C.; Krishnan, R.; Condell, M.; Polit, S. (2006), "Opportunistic Spectrum Access: Challenges, Architecture, Protocols", Proceedings of the 2nd annual international workshop on Wireless internet, Article No. 13

Staple, Gregory; Werbach, Kevin (2004), *IEEE Spectrum*. Vol. 41 Issue 3, pp.48-52

Telecom regulators' homepages:

ComReg (The Commission for Communications Regulation)
http://www.comreg.ie/ , last accessed 10.11.2014

CMT (Comisión del Mercado de las Telecomunicaciones)
http://www.cmt.es/, last accessed 23.12.2014

CRTC (Canadian Radio-television and Telecommunications Commission)
http://www.crtc.gc.ca/eng/home-accueil.htm, last accessed 10.11.2014

European Commission
http://ec.europa.eu/index_en.htm, last accessed 10.11.2014

FCC (The Federal Communications Commission)
http://www.fcc.gov/, last accessed 26.9.2014

ITU (The International Telecommunication Union)
http://www.itu.int/en/Pages/default.aspx, last accessed 25.10.2014

Ofcom (The Office of Communications)
http://www.ofcom.org.uk/, last accessed 26.9.2014

Other Internet sources:

Arstechnica (2012), "FCC to make spectrum sharing reality, whether carriers want it or not" article published 14.9.2012, last accessed 29.9.2014. Available at: http://arstechnica.com/business/2012/09/fcc-to-make-spectrum-sharing-reality-whether-carriers-want-it-or-not/

ITU, International Telecommunication Union News Magazine (2003)," Third-Generation (3G) Mobile. A status report: Evolution of wireless technologies", last accessed 23.12.2014. Available at: https://www.itu.int/itunews/issue/2003/06/thirdgeneration.html

Rayal, Frank, "Summary of Select Spectrum Auction Results", last accessed 23.12.2014. Available at: http://frankrayal.com/2012/11/06/summary-of-select-spectrum-auction-results/

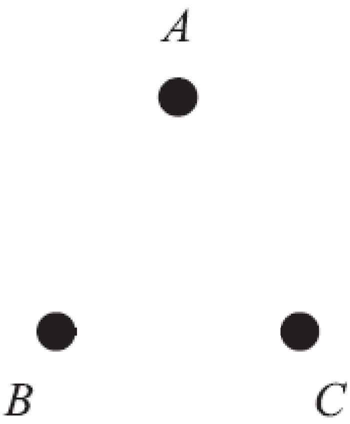
Taloussanomat (25.3.2014), " Datakatot kuristavat mobiilimarkkinoita, Suomessa hyvä tilanne". Last accessed 24.12.2014. Available at: http://www.digitoday.fi/mobiili/2014/03/25/datakatot-kuristavat-mobiilimarkkinoita-suomessa-hyva-tilanne/20144206/66

The Economist (2004), "Spectrum policy: On the same wavelength", article published 12.8.2004 (from the print edition), last accessed 29.9.2014. Available at: http://www.economist.com/node/3084475

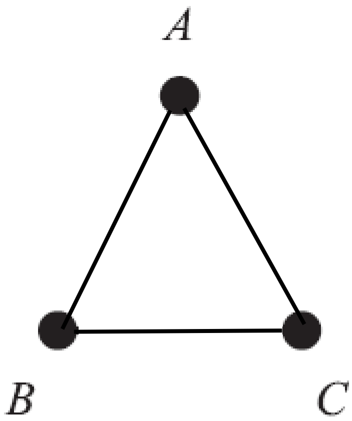
Appendices

Appendix 1. The two homogenous graphs in case l_i=3

Source: Levine and Rickman (2007)



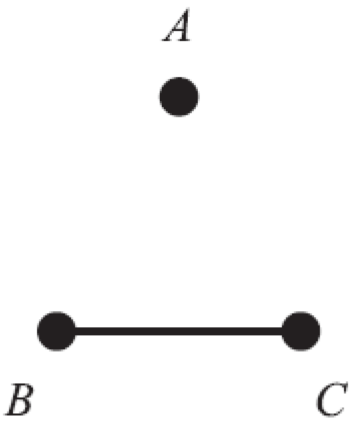
A graph with no edges



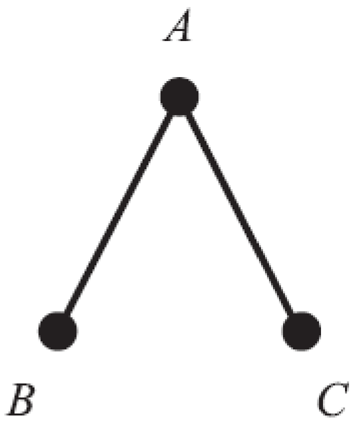
A complete graph with three edges

Appendix 2. The two inhomogenous graphs in case l_i=3

Source: Levine and Rickman (2007)



Single edge graph



Two edge graph

Appendix 3. Proof of proposition 1.

Source: Levine and Rickman (2007)

Using (7) we may write (11) as

$$\Pi = \Pi(a, N) = \frac{P(a, N)D(P(a, N)}{\epsilon N^2} - F \quad \text{(A.1)}$$

Then, differentiating (A.1) partially with respect to a, we arrive at

$$\frac{\partial \Pi}{\partial a} = \frac{[PD' + D]}{\epsilon N^2} \quad \text{(A.2)}$$

Hence $\frac{\partial \Pi}{\partial a} < 0$ if $PD' + D < 0$ ie $\epsilon > 1$.

Similarly differentiating (A.1) partially with respect to N:

$$\frac{\partial \Pi}{\partial N} = \frac{\partial P}{\partial N}\frac{1}{\epsilon N^2}[D + PD'] - \frac{2PD}{\epsilon N^3} \quad \text{(A.3)}$$

Again from (7) we have

$$\frac{\partial P}{\partial N} = -\frac{P}{N(\epsilon N - 1)} \quad \text{(A.4)}$$

Combining (A.3) and (A.4) we arrive at

$$\frac{\partial \Pi}{\partial N} = -\frac{P[(2N-1)\epsilon - 1]}{N^3(\epsilon N - 1)} \quad \text{(A.5)}$$

Since we assume that $\epsilon > 1$ we have that $(2N - 1)\epsilon > N(\epsilon - 1) > 0$ for $N \geq 1$. It follows that $\frac{\partial \Pi}{\partial N} < 0$. □

Appendix 4. Optimal AIP and the number of firms in the market

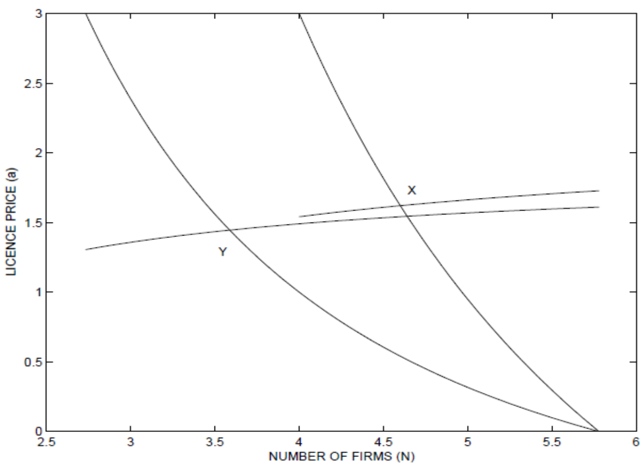


Figure 4: The Optimal Licence Price and Firm Numbers: $\Lambda = 0.3$, $\ell_i = 3$, $\mu = 3c$. X=sector with spectrum re-use. Y=sector without spectrum re-use.

Appendix 5. Optimal AIP and spectrum re-use with spectrum scarcity

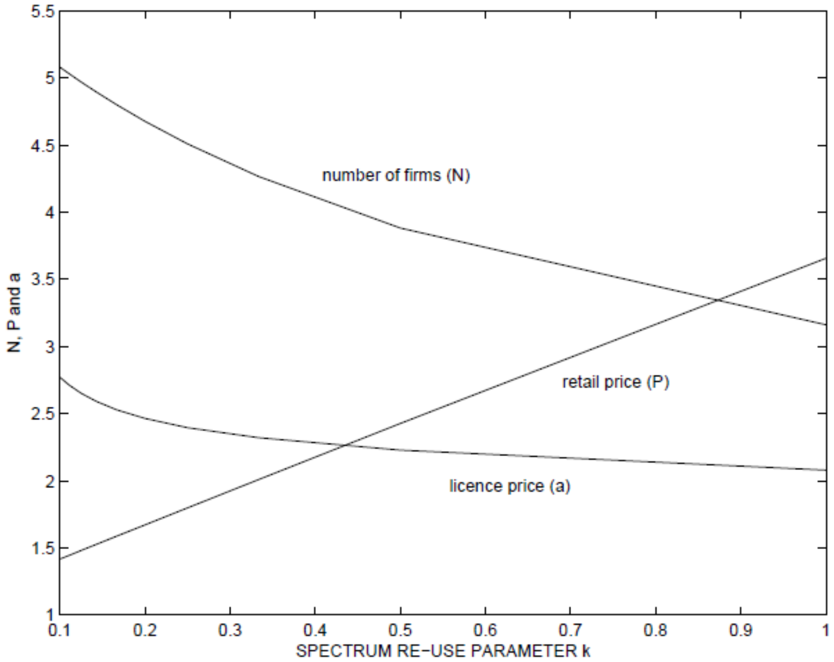


Figure 5: The Optimal Licence Price and Spectrum Re-Use: $\Lambda = 0.3, \mu = 3c$.

Appendix 6. Optimal AIP and spectrum re-use without spectrum scarcity

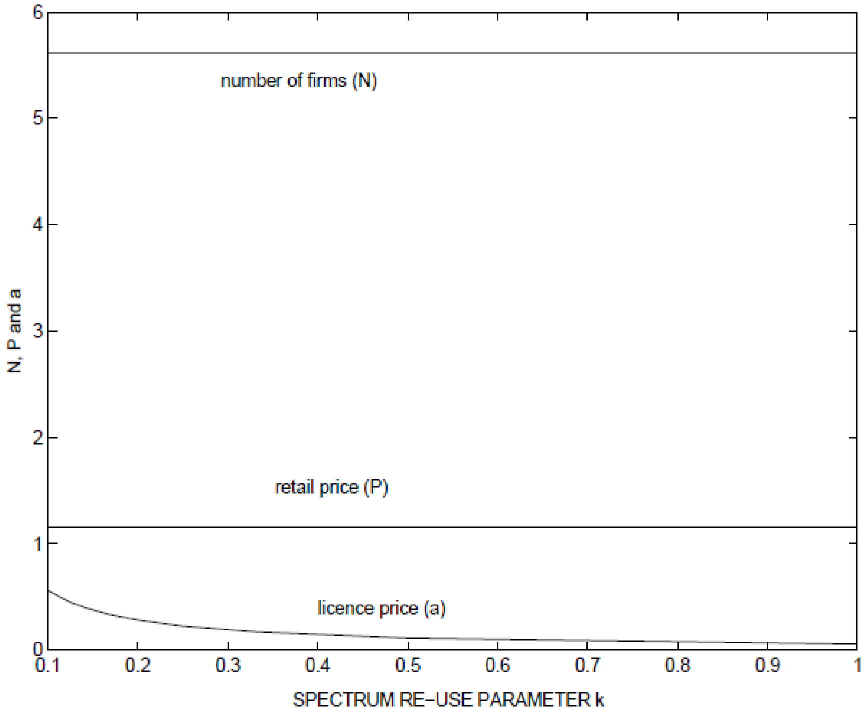


Figure 6: The Optimal Licence Price and Spectrum Re-Use: $\Lambda = 0.3$, $\mu = 0$.

Appendix 7. Optimal AIP and elasticity of demand for the end products

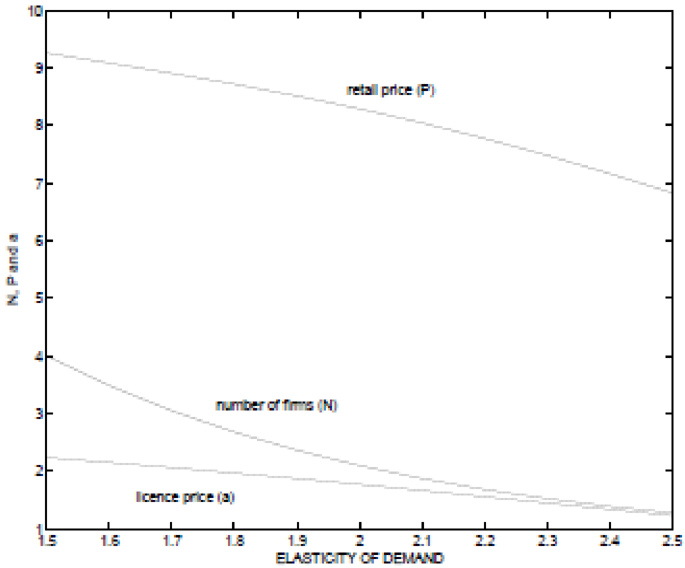


Figure 7: The Optimal Licence Price and Elasticity of Demand (ϵ): $\Lambda = 0.3$, $\mu = 3c$, $\ell_i = 3$.

Appendix 8. Constrained optimal AIP and the number of firms

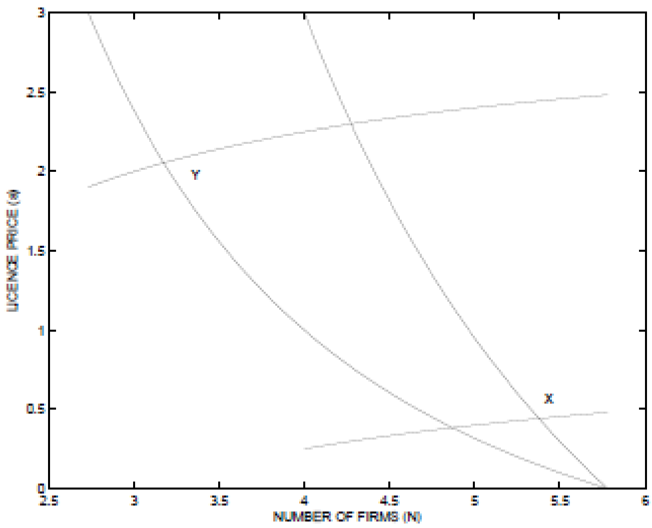


Figure 8: The Constrained Optimal Licence Price and Firm Numbers: $\Lambda = -1$, $\ell_i = 3$. X=sector with spectrum re-use. Y=sector without spectrum re-use.

Appendix 9. Constrained optimal AIP and the elasticity of demand

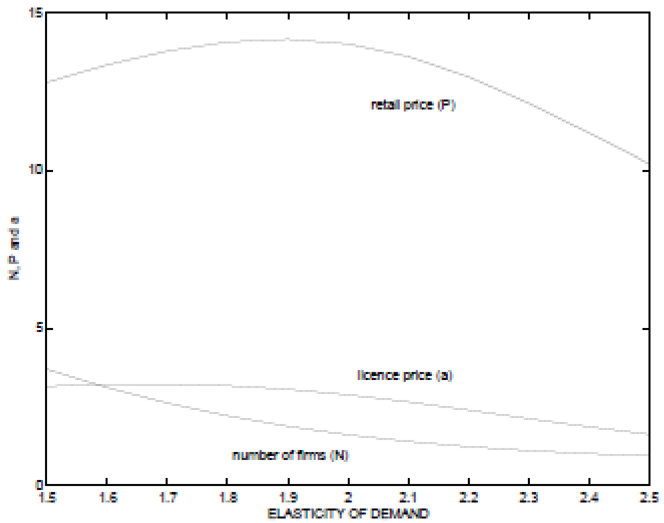


Figure 9: **The Constrained Optimal Licence Price Elasticity of Demand (ϵ):** $\Lambda = -1$, $\mu = 3c$, $\ell_i = 3$.

Appendix 10. Adjustment to optimal AIP

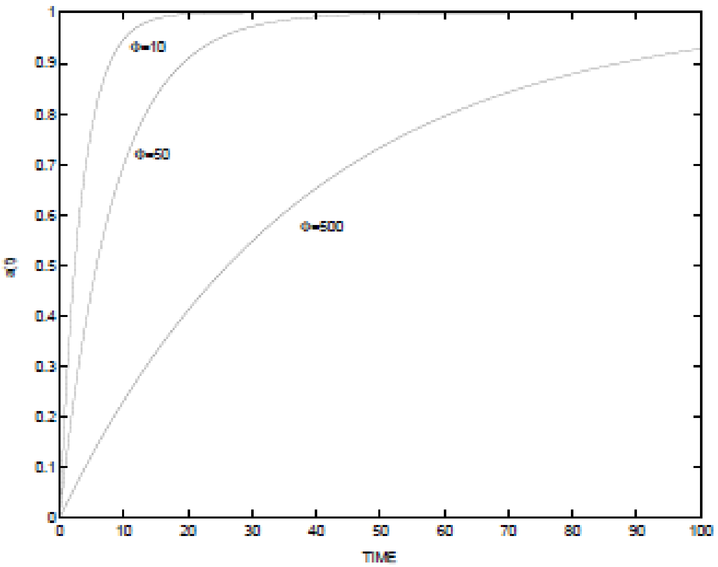


Figure 10: **Dynamic Adjustment.** $a_i^* = 1$; $a_i(0) = 0$.