Pricing Decisions by Australian General Practitioners

Elina Gilbourd

October, 2008

University of Sydney

Thesis submitted in partial fulfilment of the award course requirements of

the Bachelor of Commerce with Honours in Economics

Supervised by Dr. Donald J. Wright (Semester 1) and Dr. Nicholas De Roos (Semester 2)

Acknowledgements

I would like to thank my supervisors, Nick De Roos and Don Wright, for their time and invaluable guidance. I also thank Associate Professor Binh Tran-Nam and Dr. Michael Gilbourd for their helpful comments and critique. Thank you to my good friends in the honours class, for keeping me rosy cheeked and smiling throughout the year, and to Ashley Craig (the honorary honours student) for his suggestions. Also, a very special thank you to my boyfriend Asher Bender, for his patient help with Matlab and sub-editing, and for being a constant friend and kind supporter. Finally, my lovely family and wonderful friends in the real world, I thank them for their patience, love and support through-out the year.

Abstract

In the Australian market for primary healthcare, some General Practitioners (GPs) bulk bill patients while others charge them a direct fee. The prevalence of these two fee structures means that patients are paying different prices for the same service. From a policy perspective, universal bulk billing is preferred since this leads to more equitable access to free primary healthcare. The empirical observation that bulk billing and fee charging GPs can exist in a single location has not been explained in the literature. This thesis seeks to explain the observation by differentiating GPs into those who provide long consultations and those who provide short consultations. Given the nature of policy in the market, these two types of GPs have different incentives when deciding whether to bulk bill or charge a fee. The hypothesis put forward in this thesis is that GPs who have long consultations will prefer to charge a fee while GPs who have short consultations will prefer to bulk bill.

The model is characterised by a single-period interaction between the two representative GP types. These agents compete by choosing whether to bulk bill or charge a fee. The results show that if GP capacities are unconstrained, they will all choose to bulk bill in order to maximise the number of patients who come to see them. However, if the capacity of GPs is constrained, there is a range of model parameters for which the above hypothesis is confirmed. The introduction of capacity constraints is consistent with the general shortage of GPs in the Australian primary healthcare market.

Furthermore, the model suggests that universal bulk billing is only achievable when the available GP hours are significantly greater than the needs of the population. An implication of this outcome is that GPs will not be deterred from charging a fee unless there is a substantial over supply of GP hours. Increasing the number of GPs is becoming more difficult as medical graduates are becoming less likely to choose to train as GPs. This suggests that policy aimed at increasing the number of available GPs is likely to be inefficient. Other policies such as increasing the monetary incentive to bulk bill or promoting the role of practice nurses under Medicare may prove to be more successful at increasing the number of GPs who bulk bill.

CONTENTS

TABLES AND FIGURES

Figures

Tables

Chapter 1

Introduction

General Practitioners (hereafter GPs) in the Australian primary healthcare market are able to choose a fee structure; they have the option of bulk billing or charging a fee. This is essentially a pricing decision, where charging a fee results in a higher price being paid by patients. One would expect that the competition in this market would lead the prices of all GPs to converge to a single price. However, since the introduction of bulk billing in 1984, the participation rate of GPs in bulk billing has scarcely exceeded seventy-five per cent. This means that both bulk billing and fee charging are prevalent in the market.

Savage and Jones (2004) justified this by positing that in urban areas, where there are many GPs, competition forces them all to bulk bill. While in rural areas, where there are few of them, GPs are afforded market power and are thus able to charge a fee. This explanation does not account for the empirical evidence that GPs, in a given location, are observed to be choosing different fee structures. This thesis seeks to explain the dispersion of prices in the market without reference to the geographic density of GPs. In this way, the findings will extend the results of Savage and Jones (2004) by explaining situations where two GPs in the same location make different decisions about whether to bulk bill or charge a fee.

The explanation in this thesis will involve differentiating GPs into types that provide short or long consultations. This is based on the conjecture that, on average, GPs have a propensity towards providing a consultation of a certain length. Thus, GPs are not able to alter the

consultation length that they provide as readily as they can choose whether to bulk bill or charge a fee.

In this framework, the hypothesis is that GPs who provide longer consultations will be more likely to charge a fee while those that have shorter consultations will be more likely to bulk bill. This is due to three main features in this market. First, if a GP chooses to charge a fee rather than bulk bill, he bears a cost. This cost is the same regardless of the type of consultation that the GP provides. Hence, it will have a different impact on different types of GPs. Second, shorter and longer consultations are associated with different fee levels. And finally, GPs with long consultations are able to see fewer patients in a given length of time than those with short consultations.

There is currently no empirical study in Australia that tests the direct effect of average consultation length on GPs' decisions about fee structure. Although such an empirical study would be worthwhile, the aim of this thesis will be to use a theoretical framework to link a GP's average consultation length to his decision about fee structure. It will be shown that the choice of fee structure made by the two types of GPs will only differ once capacity constraints are imposed on them.

From a policy perspective, universal bulk billing is preferred since it leads to more equitable access to primary healthcare. It has been suggested that GPs have the market power to charge a fee because there is a general shortage of them in the market. For this reason, it is

argued that policy should be directed towards increasing the number of GPs in the market such that it remains in line with demand. However, the results of the model in this thesis show that, in order to induce all GPs to bulk bill, the supply of GPs needs to be well in excess of the needs of the population. This means that increasing the supply of GPs is not a sufficient measure to induce universal bulk billing, unless GPs are oversupplied. Hence, other policies may be more appropriate.

The remainder of this thesis will be structured in the following manner. In chapter 2, some literature will be discussed, this will broadly include papers about consultation lengths, bulk billing rates and existing economic models of physicians and the primary healthcare market. Chapter 3 will follow with an outline of the institutional setting of the primary healthcare market, namely the role of Medicare, the structure of GP fees and some contemporary policy issues.

In light of the literature and the institutional setting, the model will be constructed in Chapter 4. This will comprise of a simple model with two representative GPs competing based on a choice of fee structure. This will be followed by an extension that incorporates the fact that GPs have constrained capacities. In chapter 5, the resulting equilibria will be derived, some comparative statics will be undertaken and the implications of these findings will be discussed. Finally, in Chapter 6, the thesis will be concluded with some suggested extensions to the model that would serve as worthwhile future work in the area.

Chapter 2

Literature Review

The following is a critical overview of the literature associated with the aim of this thesis. Section 2.1 includes selected empirical studies which investigate consultation lengths of GPs. It establishes that GPs provide heterogeneous consultation lengths and that patients have preferences over consultation length. Section 2.2 explores the manner in which GPs and patients gain utility from their interaction (characterised as principal-agent interaction). In Section 2.3 the incentives of GPs associated with different payment schemes are described along with literature related to bulk billing rates in Australia. Section 2.4 deals with existing theoretical models of competition and pricing in the market for primary healthcare. Finally, in Section 2.5, the aim of this paper is described in light of the associated literature.

2.1 Consultation Length

In this section, empirical literature outlining the determinants and effects of GP consultation lengths will be discussed. Ogden et al (2004) compared patients' perceived length of consultation (relative to actual length) to their satisfaction with length and content across eight practices in the UK. They found that patients were split between preferring a longer or shorter consultation. Patients may prefer a shorter consultation because they are time constrained or have an uncomplicated ailment. Preference for a longer consultation was correlated with dissatisfaction about the emotional content of the consultation. Presumably a longer consultation would lead to these emotional needs being met. Deveugele (2003) conducted an observational study across six European countries. They investigated the

content of consultations and found that longer consultations were associated with more 'social talk'. During this time GPs could provide information and listen more extensively to patient concerns, this may have improved patients' emotional satisfaction. The above studies found that longer consultation lengths were correlated with patients that were female, older, had psychosocial problems to discuss or that had a new and complicated ailment. Conversely, Deveugele et al (2002) found that consultation length was not significantly correlated with the age or sex of the patient. It is important to note that the GPs who took part in this study had a lighter workload and were likely to be more interested in effective communication than a random sample of GPs. Interestingly, Deveugele et al (2002) found that these GPs reduced their consultation length as their workload increased, suggesting that GPs have control over consultation length.

Britt et al (2005) conducted a similar empirical study using data on Australian GPs; they found that GPs who were older or female tended to perform longer consultations. This offers an alternative result to Devegeule et al (2002), and suggests that consultation length varies with inherent characteristics of the GP. This result will be favoured in the construction of the model in this thesis for two main reasons. First, the findings represent Australian GPs whereas Devegeule et al (2002) described European GPs. Second, it allows for GP consultation lengths to be fixed, so that GPs are only able to influence their income through pricing decisions. By fixing consultation lengths, the incentives behind different pricing decisions can be isolated.

Britt et al (2005) also examined consultation length in a dynamic setting. They postulated that Australia's ageing population and the increasing proportion of female GPs is likely to lead to longer average consultations. This insight establishes the fact that the types of agents interacting in this market affect the market outcome. The literature discussed so far does not extend itself to a theoretical examination of GP propensities and patient preferences for particular consultation lengths. It is clear from the empirical literature that patients have varying preferences over consultation lengths which may reflect their underlying preferences for thorough treatment and emotional satisfaction as discussed in Deveugele (2003). It is also apparent that there is heterogeneity among GPs in terms of the consultation lengths that they provide. In this thesis, the consultation lengths of GPs and preferences of patients will be based on the above empirical results.

2.2 Doctor-Patient Relationship

The doctor-patient relationship has been described as one of agency by Scott and Vick (1999). The doctor acts as agent since he has more information about the link between health and healthcare than the patient. Ideally the doctor makes decisions that the patient would have made if the patient had the same information as the doctor. That is, the doctor needs to understand the patient's utility function and act to maximise it. The complicating factor is that the GP has his own utility function to maximise which may not coincide with the patient's best interests. Further complications include the degree to which the patient is able to communicate his health concerns and his utility function to the doctor and the ability of the doctor to communicate information to the patient. Rochaix (1989) created a physician-agent search model, concluding that the ability of informed patients to seek a second opinion led doctors to become better agents due to competition for patients (or

principals). Furthermore, Scott (2000) stated that a principal-agent relationship is made more successful through better communication which can be achieved through longer consultation. This involves a higher cost in terms of time spent but can improve the outcome of the relationship. This suggests that patients who have a strong preference for effective communication may be willing to bear some of the associated cost if this is outweighed by the benefit that they receive.

2.3 Physician Payment Schemes and the Rate of Bulk Billing

GPs can be paid by salary, capitation, fee-for-service (the scheme used in Australia) or a mix of these. Scott (2000) described the incentives that GPs experience under each of these payment schemes. Salary payment involves payment for work over a period of time regardless of the quantity of patients seen and services provided. This creates an incentive to minimise effort and increase referral and drug prescription. Capitation involves a payment per patient registered with a given GP for a set time period, regardless of the services provided. It creates an incentive to compete for patients by improving service. However, it has the potential to lead GPs to favour lower cost patients because this increases the gap between cost of care and the capitation payment.

Fee-for-service, used in Australia, involves a payment for each consultation or medical service provided. This creates an incentive to increase the volume of services. This increase is attributed to two effects as described in Scott (2000). First, GPs are able to induce demand for healthcare since they are able to encourage follow-up appointments and medical tests. This increase in care may exceed the needs of the patient and only be

motivated by an increase in the GP's revenue (physician-induced demand will be discussed in more detail in Section 2.4). Second, when the service fee is paid by a third party (such as Medicare– discussed below) there is a moral hazard problem. Since patients view each consultation as 'free' they might attend consultations in excess of their needs. They may overuse GP services for minor ailments which they would be deterred from doing if the consultation was associated with a fee.[1] These two effects make it difficult to restrain healthcare expenditure under a fee-for-service scheme.

In Australia, a percentage of the schedule consultation fee is paid by Medicare and a copayment is paid by the patient. If the GP chooses to bulk bill a patient, the copayment is zero. Since bulk billing was introduced, only a fraction of GPs have chosen to bulk bill all of their patients and the proportion of GPs that bulk bill has diminished (this is discussed more extensively in the Chapter 3). Hopkins and Speed (2005) described the decline in the proportion of bulk billing GPs in Australia from 78.6% in June 2000 to 65.7% in December 2003. They attributed this decline to two effects; the schedule fee not keeping pace with the cost of running a practice and the uneven geographical distribution of GPs. Hopkins and Speed (2005) also indicated that the annual percentage increase in the Medicare rebate has been outweighed by the percentage increase in the patient co-payment, resulting in the cost of primary healthcare being shifted from the government to patients.

[1] Note that, when a third party pays on behalf of the patient, this can also be a problem under the capitation and salary schemes.

The reason for this decreased rate of bulk billing amongst Australian GPs and the resulting effect on the overall primary care system have not been adequately assessed in a theoretical framework. Savage and Jones (2004) attributed the decreasing number of bulk billing GPs to the disproportionate distribution of GPs across urban and rural areas. They found that urban GPs were densely populated and price competition led them to bulk bill, whereas rural GPs were more dispersed which gave them the market power to set positive copayments. This conclusion is verified by an empirical study conducted by Khan et al (2004) on the relationship between the characteristics of GPs (and their practices) and bulk billing rates in New South Wales. Namely, they found that bulk billing rates were higher amongst practices in metropolitan areas than practices in rural areas. And that copayment levels were highest in locations where GP supply was lowest.

The results of Savage and Jones (2004) are a relatively straightforward description of the effects of geographical distribution on market power and fail to justify certain observable characteristics of the market. First, Hopkins and Speed (2005) stated that there was a decline in the proportion of GPs bulk-billing in urban (as well as rural) areas – this is not accounted for in the Savage and Jones model. Second, the Savage and Jones model describes GPs as agents who choose their fees based on geographical dispersion. Hence, it cannot be used to explain the incidence of GPs who are located near each other making different decisions about whether to bulk bill or charge a fee.

Khan et al (2004) made two additional findings that are of interest. First, they found that female GPs were less likely to bulk bill than male GPs. This is interesting to consider alongside the findings of Britt et al (2005) (discussed in Section 2.1) that female GPs tend to give longer consultations. Obviously one cannot infer, without further empirical study, that there is a direct link between longer consultations and choosing not to bulk bill. However, the existence of the indirect link points to a potential relationship between average consultation length and the decision to bulk bill. Second, they found that GPs who had a higher average patient load per week were more likely to bulk bill. Patient load can be affected by a GP's working hours and the number of patients that they see in a given space of time. Assuming that both these effects contribute to patient load, one may infer that GPs with a higher patient load are likely to have shorter consultation lengths on average. This is another finding that provides an indirect link between a GP's average consultation length and their decision about bulk billing.

Furler et al (2002) compared the rate of longer consultations in areas of different socioeconomic status. They found that shorter consultations dominated, representing approximately eighty-eight per cent of all consultations in their data set. They also showed that longer consultations were less common in areas of lower socioeconomic status. Their data suggested that in these areas patients were seeing GPs more frequently for shorter consultations. They postulated:

> "Consultation length may be determined to some extent by the fee structure that GPs work within. GPs in more disadvantaged areas are more likely to bulk bill...and it may be that those who bulk bill are best able to maximise their income through multiple shorter consultations rather than fewer longer ones" (p. 83)

This statement suggests a causal link where fee structure influences consultation length. The aim of this thesis is to show that the causal link operates in the opposite direction. This is based on the conclusion in Section 2.1 that, on average, GPs have a propensity towards consultations of a certain length. Furthermore, one would expect that it is easier for a GP to choose a fee structure than to control their consultation lengths.

2.4 Existing Models of Pricing in the Market for Primary Healthcare

The following section outlines the main focus of existing theoretical models of the primary care market. The broad categories that will be discussed are physician-induced demand, the level of competition, price discrimination and whether doctors should be modelled as profit maximisers. The purpose is to establish how the behaviour of agents in this market differs from other markets and to outline issues of contention in the literature.

There has been lengthy debate in the literature about whether the market for primary care is characterised by physician-induced demand (PID). This arose as an explanation for several empirical findings that showed quantity of services and prices increased when the GP to population ratio increased [for example, Evans (1974)]. The issue is important because, in the presence of PID, one cannot use traditional supply-demand analysis to determine the equilibrium prices and quantities in the market. Labelle et al (1994) summarised the existing definitions of PID as falling under two categories. First, PID has been defined in terms of physicians acting as imperfect agents. Due to self interest and asymmetry of information, a physician has an incentive to provide services in excess of a patient's needs. The GP has an incentive to do this so long as his marginal revenue for doing so exceeds his marginal cost.

Second, PID has been defined as physicians having the ability to shift consumers' demand curves outward by encouraging tests and procedures.

The concept of PID has been extensively discussed in the literature; however there is no consensus as to whether it is a plausible explanation for the empirical results (Labelle et al 1994). Hence, the issue of PID will not be incorporated as part the model in this thesis. The model will assume a uniformly distributed population of patients that attend one consultation each for one period; patients each receive a benefit for a consultation of a given length. The role of PID is not considered in this context. The issue may become more pertinent if the model were extended to incorporate multiple periods. One could investigate whether the GP is able to induce patients to return for subsequent consultations in future periods.

The ability of GPs to induce demand suggests that they have market power over patients. This raises the second point of contention; the literature is split in terms of defining the market as competitive or monopolistic. As discussed in Section 2.3, Savage and Jones (2004) modelled the market for primary care in terms of spatial competition amongst GPs. Newhouse (1970) suggested that patients' lack of information about price and quality led to low cross elasticities of demand. This forms a basis for the alternative view that the market is monopolistic. In this paper, it will be assumed that GPs compete for patients. This decision is based on the assumption that patients have not previously seen any of the GPs. This

means that a given GP cannot have market power over a portion of patients on the basis of an established relationship.

It has been suggested that GPs behave as price discriminating monopolists. Masson and Wu (1974) developed a model in which poor and rich patient types search for physician services. They concluded that physicians charged richer patients a higher price. That is, physicians were able to price discriminate based on income level. Hoerger (1990) described a model where physicians maximise their profits over the length of their relationship with a patient. He posited that, since medical care is an experience good, patients learn information about a doctor in their first visit. He concluded that physicians could engage in two-part pricing, charging higher prices to newer patients and lower prices to existing patients[2]. This discouraged patients from switching to a different physician since switching required paying the higher 'new patient' price.

These results serve to justify certain empirical observations. For example, Khan et al (2004) found that 34% of New South Wales GPs bulk bill selective patients. In light of the above discussion one can conclude that GPs do not choose which patients to bulk bill at random. Instead they price discriminate and only bulk bill lower income or long-established patients. This has an implication for the model in this paper. It will be assumed that patients are all of one type, meaning that GPs cannot price discriminate based on any observable

[2] To clarify, Hoerger actually found that the mark up over cost for new patients was actually lower than for established patients because the cost of service for subsequent consultations is lower. However, the price was higher for new patients than established patients.

characteristics of patients. Hence, when a GP chooses a fee structure, they will be charging all patients one price.

Finally, there is disagreement in the literature as to whether GPs are motivated by profit maximisation. Dionne et al (1985) explained that GPs should be defined as self-employed individuals that maximise net income and leisure time rather than as firms that maximise profit. The main purpose of this distinction is that models should have the capacity to factor in a GP's overall utility. That is, in addition to income one should be able to consider altruistic gain, prestige and the leisure time traded off when working (amongst other possible arguments in a GP's utility function). In this thesis, GPs will simply be described as profit maximisers. The reason for this is that the model aims to isolate the effect of average consultation length on fee structure. Thus the key elements are length of consultation and the revenue gained. A potential extension of the model could involve the two types of GPs having different utility functions. For example, a GP with longer consultations could provide longer consultations on the basis that he gains more utility from altruism or communication.

2.5 Revisiting the Aim of This Paper

To conclude this overview of the literature the aim of this paper needs to be revisited. Broadly the paper is about the different pricing decisions made by GPs in the Australian primary healthcare market. More specifically, the intention is to extend the findings of Savage and Jones (2004) and account for GPs in the same location making different decisions about whether to bulk bill or charge a fee. This will be achieved through a theoretical model that describes GPs as providing short or long consultation lengths and

having capacity constraints. This may cause different types of GPs to make different pricing decisions.

Chapter 3

Institutional Setting

This chapter deals with the institutional setting of the model. Section 3.1 describes the role of GPs in Australia's healthcare system and Section 3.2 outlines the role of Medicare in structuring GP fees. The purpose of these sections is to familiarise the reader with the market so that the construction of the model is contextualised. In Section 3.3 some associated policy issues will be discussed. These policies will later be addressed in light of the findings of the model.

3.1 The Role of GPs in the Australian Healthcare Market

GPs and hospital emergency rooms comprise primary healthcare services in Australia. This means that GPs are directly accessible to patients as a source of healthcare. GPs act as the gatekeepers to the rest of the healthcare system, choosing whether to treat a patient themselves or refer them to secondary or tertiary care (specialist physicians and specialist clinics). In Australia, patients are unable to access higher levels of healthcare without a referral from a GP. This is similar to countries such as the UK and Canada, but contrasts with the US where patients are able to refer themselves to specialists.

GPs influence the overall cost of the healthcare system, directly through their own activity, and indirectly through their decisions about referral. It is argued that GPs, in their role as gatekeepers, reduce overall healthcare expenditure. This is because they have better knowledge of the nature of intervention necessary and this leads to more efficient use of secondary care (Scott 2000). Also, as the primary providers of healthcare they impact on the

health level of the population through their role in both curative and preventative care. Therefore, GPs are central to the efficacy of Australia's healthcare system and it is pertinent to further the study of their economic incentives and behaviour so that policy may be appropriately constructed.

3.2 Medicare and GP Fees

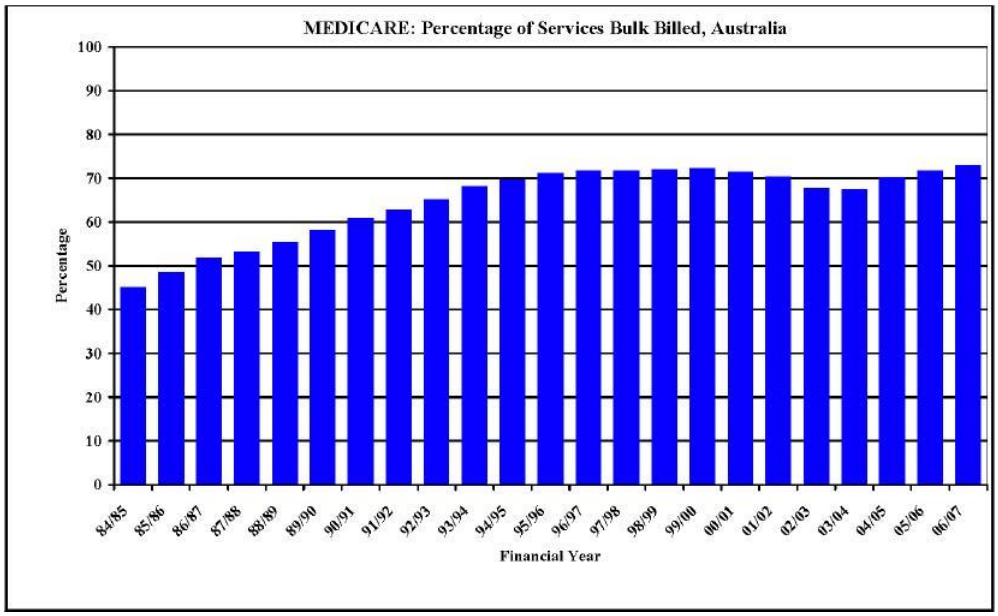
Medicare provides universal health insurance in Australia and is funded by a 1.5% tax levy. Medicare (in its current form) was introduced in 1984 in order to provide equitable access to quality healthcare (Biggs 2003). In Australia, GPs are paid on a fee-for-service basis – they receive a schedule fee based on the length of a consultation and the nature of services performed. A GP can choose to 'bulk bill', which means that the patient is not charged and the GP accepts a Medicare rebate as full payment of the schedule fee. Alternatively, a GP can choose to charge the patient a fee directly; this fee can exceed the schedule fee. The patient then claims the associated Medicare rebate from Medicare. The difference between the fee and the Medicare rebate is referred to as the patient's copayment. When a GP chooses to bulk bill, the copayment is zero.

It is pertinent to note that a GP's choice between bulk billing and fee charging does not directly influence government expenditure. This is because the Medicare rebate is the same for all services of a given type and length regardless of the GP's chosen fee structure. The decision will only affect the patient's copayment. If patients are deterred from attending consultations by higher copayments, there may be an indirect effect on government expenditure.

GPs choose to bulk bill for two reasons; the first is lower administration cost. Deciding to bulk bill means that the GP's surgery does not need to have facilities for accepting money from patients. On top of this, when a GP chooses to bulk bill, a larger portion of the administration of each consultation fee is done by Medicare. The second reason that GPs choose to bulk bill is that it attracts more patients. Lewis et al (2003) found that GPs felt pressured to bulk bill because if they introduced a fee, there would be plenty of other bulk billing practices that a patient could switch to.

Prior to 2005, the Medicare rebate was eighty-five percent of the schedule fee. It was hoped that GPs would universally choose to bulk bill all of their patients, however the proportion of GP services that have been bulk billed has not exceeded seventy-five per cent. Figure 3.1 shows the percentage of services which were bulk billed in each financial year from 1984/85 to 2006/07.

Figure 3.1: Bulk Billing Rate



Source: Department of Health and Ageing (2008)

It is clear from the figure that universal bulk billing has not been achieved. A proportion of GPs are bulk billing, while the remainder are charging a fee. The focus of this thesis is to understand why two different fee structures exist in this market.

Fee charging by GPs is not favoured by policymakers since it undermines equitable access to primary healthcare. For this reason, in 2004, the government introduced an incentive payment of around five to seven dollars on top of each bulk billed consultation for children under sixteen and concession card holders (ABS 2008). The purpose of this policy was to make bulk billed primary healthcare more accessible to those who have the lowest ability to pay.

Furthermore, due to declining participation in bulk billing, on January 1st 2005 the rebate was increased to one hundred percent of the schedule fee (ABS 2008). This means that bulk billing GPs now receive the full schedule fee from Medicare, while fee charging GPs are free to set a mark up over the schedule fee. The patient's copayment is equal to the mark up, since they receive the full schedule fee as a Medicare rebate. This policy will be emulated when the model is constructed in Chapter 4.

Due to the above policies, there has been a slight increase in the proportion of services which were bulk billed from 2004 onward; this can be seen in Figure 3.1. Such ad hoc increases will not have a long term effect. This is because GP remuneration levels need to

keep pace with inflation and the cost of running a practice; otherwise bulk billing rates will inevitably decline again in the future (Hingston 2006).

3.3 Policy issues

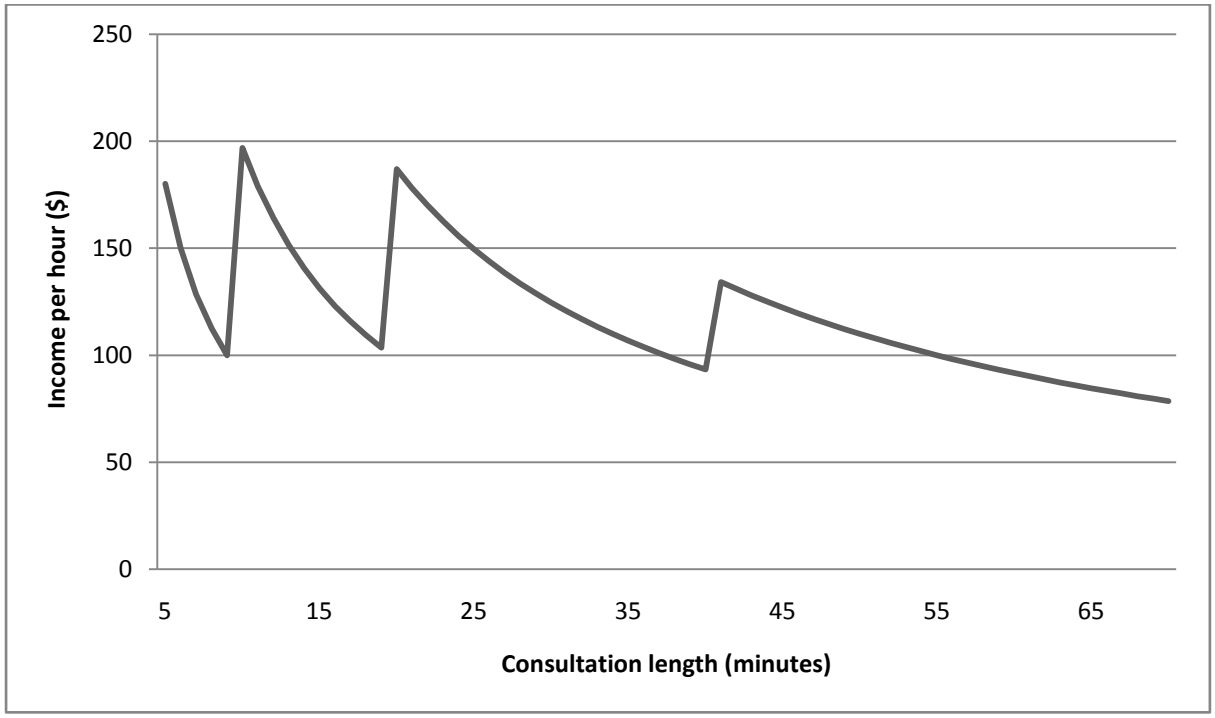
There are two contemporary policy issues that are relevant to the model presented in this thesis. First, the structure of the Medicare Benefit Schedule results in GPs who have shorter consultations being more profitable than GPs who have longer consultations. This is relevant because it gives GPs who have longer consultations an incentive to raise their fees. Second, there is a shortage of GPs relative to the healthcare needs of the population. This relates to the capacity constraints of GPs that will be introduced in Section 4.2.

The first issue is that the Medicare Benefit Schedule is structured such that it is more profitable for a GP to see many patients for short consultations than to see a few patients for longer consultations. Figure 3.2 shows the hourly income that a bulk billing GP can earn if they give consultations of a particular length. The incomes are calculated based on the assumption that GPs give consultations of identical length throughout the hour.

There are two features of note in the graph. The first notable feature is that the maximum attainable income diminishes as consultation length increases; this can be seen in the peaks of the graph. If a GP has 10 minute consultations they will earn $196.80 per hour; while if they have 20 minute consultations they will earn $186.80 per hour. And, if their consultations are over 40 minutes they can only earn a maximum of $134.20 per hour.

The second notable feature is that, within a particular rebate bracket, GPs are better off when consultation length is minimised. This is because the rebate level is constant for a particular bracket of time. For example, in the 10-20 minute bracket GPs earn $32.80 per consultation. This means that income peaks where consultations are 10 minutes in duration and then declines as the consultation length approaches 20 minutes; before the schedule fee increases to $62.30. The decline occurs because the schedule fee remains $32.80 and consultation length increases across the range. That is, the GP is able to see fewer patients in the hour but his income per patient does not increase. This pattern is repeated for each rebate bracket which causes the jagged appearance of the graph.

Figure 3.2: Hourly Income[3]



[3] Derived from Medicare Benefit Schedule Book (2007) surgery consultation items, these items are characterised by consultation length or the nature of services performed. Summary of relevant items:

Level 'A' (under 10 minutes)	$15.00
Level 'B' (under 20 minutes)	$32.80
Level 'C' (20-40 minutes)	$62.30
Level 'D' (over 40 minutes)	$91.70

The key point to be taken from Figure 3.2 is that, for a given fee structure, GPs are better off if they have shorter consultations. This means that a GP who has longer consultations is worse off bulk billing than a GP who has shorter consultations. This is one reason why GPs with longer consultations may prefer to charge a fee and may partially explain the fact that different GPs choose different fee structures. Hence, the construction of the model in this thesis will seek to explain the different fee structures by considering GPs with different average consultation lengths.

The second policy issue is the shortage of GPs relative to the primary healthcare needs of the Australian population. This shortage has come about due to two concurrent effects (Thistelthwaite et al 2008). The first is a demand effect; since chronic illness levels are increasing there is more demand for primary healthcare. Due to Australia's ageing population this effect is likely to become stronger over time. The second is a supply effect; GP remuneration and prestige levels are falling relative to more specialised positions in healthcare. This means that medical graduates are becoming less likely to pursue a career as a GP. Furthermore, GPs tend to prefer to practice in urban rather than rural areas. Rural areas are less appealing for various reasons such as cultural and social isolation, lack of career prospects for partners or a preference for raising children in the city. These effects result in an overall shortage of GPs which is more acute in rural areas. It is believed that this shortage is what gives GPs the market power to charge a fee.

Two broad policies have been suggested for dealing with the shortage of GPs. The first policy targets the incentives of medical graduates to become GPs (Thistelthwaite et al 2008). The most straightforward way to achieve this is by increasing GP remunerations to become in line with those of secondary care. This strategy requires a substantial increase in healthcare expenditure by the government. Another way to increase the likelihood of medical graduates becoming GPs is to promote entry into medical schools by students that are likely to choose to be GPs. Thistelthwaite et al (2008) found that females, those with a rural background or a history of community service were more likely to become GPs after graduating medical school. The second policy initiative is the promotion of practice nurses in order to supplement the shortage of GPs (Sahari 2007). The goal is to alleviate GPs from some of their simpler duties such as ongoing treatment of chronic problems, repeat prescriptions and preventative care. If this is achieved, GPs would have more time to deal with more complex patient concerns. Given the existing shortage of GPs, this would be a more efficient use of GPs' time.

The model in the next chapter follows from the institutional setting that has been outlined in this chapter. The hypothesis is that the GPs who have longer average consultations will choose to charge a fee, while those with shorter consultations will choose to bulk bill. If this hypothesis is true then it may explain the dispersion of prices in the market. Also, the factors that deter GPs from bulk billing will be identified and discussed in terms of policy.

Chapter 4

The Model

This chapter outlines the model which will be used in an attempt to explain the different fee structures observed in the primary healthcare market in Australia. In Section 4.1 the simple model is developed. Section 4.2 then extends the simple model to incorporate capacity constraints.

4.1 The Simple Model

This model deals with the supply and demand of primary healthcare. The supply side will be characterised in terms of two representative GPs. Then, the demand side will be defined as a population of patients. Finally, the interaction between supply and demand will be developed in a game theoretic framework.

4.1.1 General Practitioners (GPs)

There are two types of GPs competing in the market for primary healthcare, one supplies long consultations (type L) and the other supplies short consultations (type S). These types are assigned to GPs exogenously. In this model, each type is embodied by a representative agent, and these two agents compete by choosing a fee structure.

Assumption 1 A GP of a particular type provides identical length consultations to all patients that they see. The GP of type L provides longer consultations than the GP of type S.

This assumption is made for simplicity, in reality GPs provide varied consultation lengths to patients. The model can be considered as a study of average consultation lengths where the GP of type L gives longer consultations than the GP of type S, on average.

Assumption 2 GPs cannot change their type.

This assumption is based on the empirical findings of Britt et al (2005) discussed in Section 2.1. In essence, they found that average consultation lengths were correlated with certain inherent characteristics of GPs. This means that, in the model, consultation length can be treated as an exogenous parameter.

Consider a GP of type i ($i \in \{S, L\}$). This agent chooses a fee structure $P_i \in \{\text{BB, F}\}$ where BB is bulk billing and F is charging a fee. Choosing a fee structure is similar to choosing a price, but there are only two discrete options. Choosing BB means that a GP charges Medicare a schedule fee for a consultation and does not charge the patient. Choosing F means that the GP charges the patient the schedule fee with a mark up and the patient claims the schedule fee back from Medicare. When a GP chooses F they also incur some administrative costs.

It should be noted that in this model, the mark up over the schedule fee is assumed to be given exogenously and is constant for both GP types. In reality, GPs that choose F are also free to choose their prices. An endogenous treatment of the mark up is left to a future study.

Assumption 3 When a GP of type i chooses BB or F he commits to charging all patients that come to see him one price (that price is zero to the patient in the case of BB).

This assumption is based on the discussion in Section 2.4 where it was concluded that variation in a GP's prices occur due to price discrimination by the GP. In this model, GPs deal with one type of patient and hence cannot price discriminate based on observable characteristics of the patients.

For each choice of fee structure, the GPs have the following payoffs:

If they choose BB: $m_i F_i$

If they choose F: $m_i(1+x)F_i - m_i c - C$

Where:

m_i Proportion of patients that see the GP of type i; $(m_i \leq 1)$

F_i Schedule fee for an appointment of type $i \in \{S, L\}$. Where $F_L > F_S$;

F_L is the fee for a long consultation and F_S is the fee for a short consultation

x The mark up over the schedule fee charged by a GP that chooses F.

x is assumed to be exogenous.

C Fixed cost of charging a fee

The fixed cost is incurred when establishing the facilities to charge a fee (cash float, EFTPOS machines, staff training etc).

c Variable cost of charging a fee

The variable cost is the opportunity cost of the time spent processing transactions with patients by administrative staff. It reflects the fact that it takes administrative staff longer to charge patients a fee than to bulk bill them.

All costs that are incurred regardless of the decision about fee structure are normalised to zero. These costs include the cost of providing consultations and the cost of communicating with Medicare[4]. This is to simplify the model, since these costs will not bear on a GP's pricing decision which the model aims to investigate.

4.1.2 Patients

There is a population of N patients; each of them attends exactly one consultation. The use of the convention that all patients attend a consultation is plausible when speaking about the primary healthcare market, since healthcare is a necessity for a vast majority of people at some point throughout their life. The population is normalised such that $N = 1$.

Patients receive a payoff for attending a consultation. If they attend a bulk billing GP the price, for the patient, is zero. If they attend a fee charging GP the price, for the patient, is the copayment which is the GP's mark up over the schedule fee. Patient payoffs are:

[4] GPs need to convey information to Medicare about the consultations and services that they provide. This is so that the appropriate Medicare rebate can be paid to the GP (in the case of bulk billing) or the patient (in the case of fee charging).

If they attend a GP of type i that chose BB: β_i

If they attend a GP of type i that chose F: $\beta_i - xF_i$

where:

β_i The patient's benefit from a consultation of type $i \in \{S, L\}$

$\beta_S = 1$ and $\beta_L \sim U[0,2]$ (explained below)

Since all patients attend one consultation, they can only decide which type of consultation to attend. That is, they do not have the option of not attending any consultation. This means that only the relative benefit from the consultations is pertinent to patient decisions. Hence, the benefit of a short consultation is normalised to one.

By assumption, patients are distributed in such a way that half of the population prefers each type of consultation. This stems from the empirical findings of Ogden et al (2004) discussed in Section 2.1, they found that patients were split between preferring a shorter or longer consultation[5]. Furthermore, it is important that the benefit of a consultation is non-negative for all patients. In order to satisfy these conditions it must be the case that $\beta_L \geq 0$ and $E(\beta_L) = \beta_S$. Hence, the result is that patients are uniformly distributed between zero and two in terms of the benefit that they receive from a long consultation.

[5] To clarify, in Ogden et al (2004) patients had varied preferences for shorter or longer consultations, the statement that they are split *evenly* between preferring longer or shorter consultations is an assumption of this model.

Another interpretation of this distribution of patients is to consider patients as having a trade-off between their health level and time. If they have a complicated or serious ailment they will want to spend a long time with the GP, whereas, if they have a straightforward ailment they would prefer to minimise time spent with the GP. This interpretation means that a given patient, depending on their health level, could have different preferences for consultation length at different points in time.

The approach adopted in this model is that consultation length is horizontally differentiated. However, this is not the only way that the market may be described. Another approach could involve vertical differentiation where a longer consultation length provides a signal of higher quality; this will be discussed in more detail in Chapter 6.

4.1.3 GP and Patient Interaction

The interaction between GPs and patients is defined as a single period two-stage game that proceeds as follows,

Stage 1 The two GPs (type S and type L) simultaneously choose their fee structure, P_i.

Stage 2 Patients choose which of the two GPs to see for their consultation. This choice determines the payoffs to all players.

Assumption 4 GP of type i does not reveal his type to any patient prior to a consultation. GP of type i does reveal his choice of P_i to a patient prior to a consultation.

That is, patients cannot observe the GP types (S or L) ex ante, but they can observe the fee structures (BB or F). This assumption reflects the fact that medical practices typically advertise their fee structure but do not explicitly state the average consultation length provided by the GP.

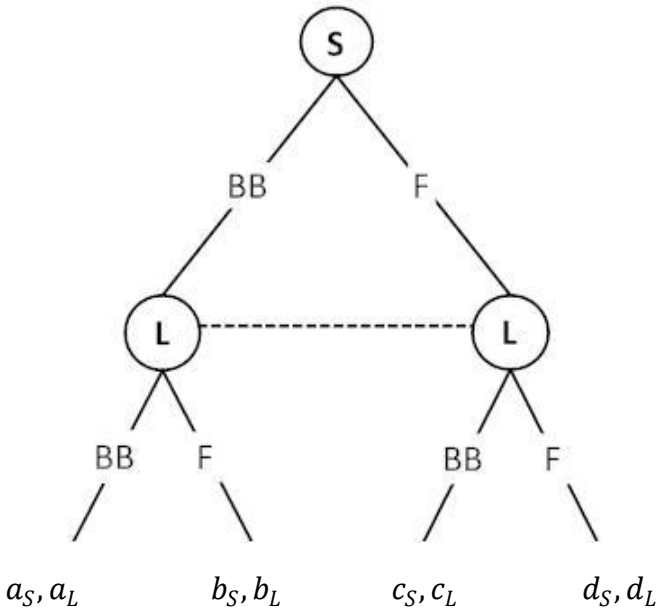
Assumption 5 GPs have unconstrained capacities; they are able to provide a consultation to all patients that decide to come to them.

This assumption means that the proportion of patients that see each GP is entirely based on the decisions made by patients. This assumption will be relaxed in Section 4.2.

Rationing Rule 1 If $P_S = P_L$ then patients split evenly between GPs (that is, for both GP types $m_i = 0.5$).

Since, if patients only observe one fee structure in the market, they have no information to help them find their desired GP. Hence, patients will go to a GP at random in the hope of seeing their preferred type. Since there are two GPs, by the law of large numbers, one would expect patients to split evenly between them. The game can be represented graphically; this is shown in Figure 4.1.

Figure 4.1: Game Tree



The nodes at the bottom of the tree represent the payoffs to GPs when each node is reached. That is, z_i is the payoff to GP of type $i \in \{S, L\}$ at node (z) where $z \in \{a, b, c, d\}$. As shown in the payoff equations above, the GP payoffs depend on the proportion of patients that see each GP at a given node. This means that GP payoffs need to be determined endogenously at each node based on how the population of patients splits between the two GPs.

4.2 Extending the Model to Incorporate Capacity Constraints

In this extended model Assumption 5 is relaxed. GPs are no longer able to treat any proportion of patients that choose to see them. This extension reflects the fact that GPs have limited working hours, which means that their capacity to see patients is constrained.

Rationing Rule 2 If a GP has reached his capacity, the proportion of patients that chose to see him in excess of his capacity are crowded out to the other GP (assuming that the other GP has not reached capacity).

This rule is a product of the construction of the model, namely the fact that all patients attend one consultation. If patients are not seen by the GP that they chose initially, they will attempt to see the other GP. The process stops once both GPs have reached their capacity, and this can occur before the entire population of patients recieves a consultation.

Assumption 5A $0 < l < s < N = 1$

Where s and l are the maximum proportions of the population that the GP of type S and the GP of type L can see in the available time (respectively). GPs in this model have only one decision variable (fee structure), this implies that they cannot change their working hours. It is assumed that both GP types have the same working hours. This means that both have the same amount of time to see the proportion of patients that come to them. Remembering that type L spends longer with each patient, it is clear that he will take longer than type S to see a given proportion of the population. From this it can be concluded that type L has a lower capacity to see patients than type S. Furthermore, it is assumed that type S is not able to see the entire population of patients in the given amount of time. This assumption is justified because there are two GPs in the market. If type S were able to serve the entire population, type L may not have an incentive to enter the market in the first place.

Chapter 5

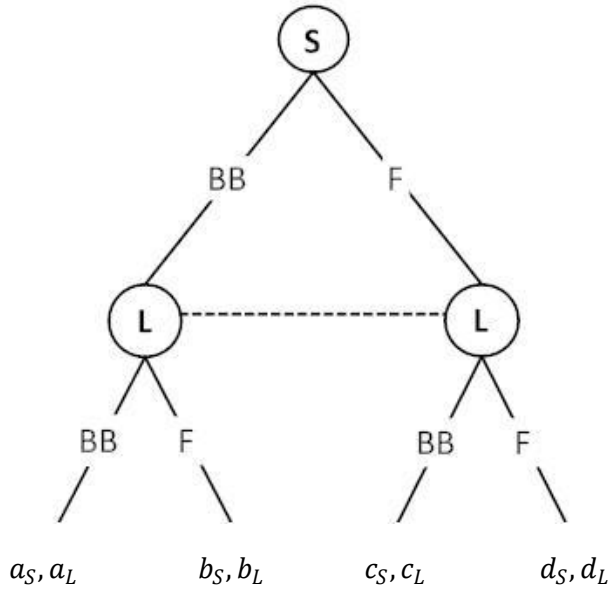
Results

In this chapter, the results of the model which was constructed in the previous chapter are derived and their implications are discussed. Section 5.1 describes the results of the simple model outlined in Section 4.1 and Section 5.2 contains a short discussion of these results. Section 5.3 outlines the results of the extended model described in Section 4.2, and finally, Section 5.4 deals with the implications of the extended model results.

5.1 Results of the Simple Model

In order to find the equilibrium of this model, it is necessary to derive the payoffs received by GPs at each of the nodes $(a), (b), (c)$ and (d). The game tree is reproduced in Figure 5.1 for convenience. These payoffs are determined endogenously based on how patients respond to the fee structures that they observe. The payoffs are derived via backward induction, at each node $(a) - (d)$ the choice of P_i (fee structure) is fixed for both GP types, patients observe the fee structures and make their decision about which GP to attend.

Figure 5.1: Game Tree (reproduced from Figure 4.1)



At nodes (a) and (d) patients observe that $P_S = P_L$. Hence, by Rationing Rule 1, the patients split equally between type S and type L. Thus the proportions that see each GP type are $m_S = m_L = 0.5$, the payoffs to GPs at each node are:

GP payoffs at node (a): $a_S = 0.5F_S$

$$a_L = 0.5F_L$$

GP payoffs at node (d): $d_S = 0.5(1 + x)F_S - 0.5c - C$

$$d_L = 0.5(1 + x)F_L — 0.5c - C$$

At nodes (b) and (c) patients cannot distinguish which of these nodes has been reached, they can only observe the fact that $P_S \neq P_L$. Their decision about which GP to attend must be based on their beliefs about the likelihood that their preferred type has chosen a particular fee structure. Let λ denote a patient's belief that he has reached node (b) when he observes that $P_S \neq P_L$. The value of λ can be defined by any of the following conditional probabilities:

$$\lambda \equiv P(b|b \cup c)$$

$$= P(S|BB)$$

$$= P(L|F)$$

It is assumed that beliefs are uniform for all patients. A patient's expected utility for going to a GP when they observe $P_S \neq P_L$ is:

If they go to the GP that chose BB: $E(U|BB) = \lambda\beta_S + (1 - \lambda)\beta_L$

If they go to the GP that chose F: $E(U|F) = \lambda(\beta_L - xF_L) + (1-\lambda)(\beta_S - xF_S)$

To find the proportion of patients that attend each GP type one needs to identify β_L^*, which represents the patient who is indifferent between seeing the GP that chose BB and the GP that chose F. For this indifferent patient:

$$\lambda\beta_S + (1-\lambda)\beta_L = \lambda(\beta_L - xF_L) + (1-\lambda)(\beta_S - xF_S)$$

$$\beta_L^* = 1 - \frac{x}{1-2\lambda}[(1-\lambda)F_S + \lambda F_L] \qquad (\beta_S = 1) \qquad (1)$$

Note that if $\lambda = 0.5$ then β_L^* is undefined.

This β_L^* represents the cut off point where patients split between the two GP types. Patients who are positioned at $\beta_L > \beta_L^*$ prefer to attend the GP of type L, while patients who are positioned at $\beta_L < \beta_L^*$ will prefer to attend the GP of type S. Although patients do not observe the type of their preferred GP, they are able to infer which fee structure their preferred GP type is more likely to choose based on their belief, λ.

The equilibrium needs to be consistent with patient beliefs and since patient beliefs are unknown, all values of λ need to be considered, Figure 5.2 shows the relationship between λ and β_L^* that is represented by equation (1).

Figure 5.2: Indifferent Patient

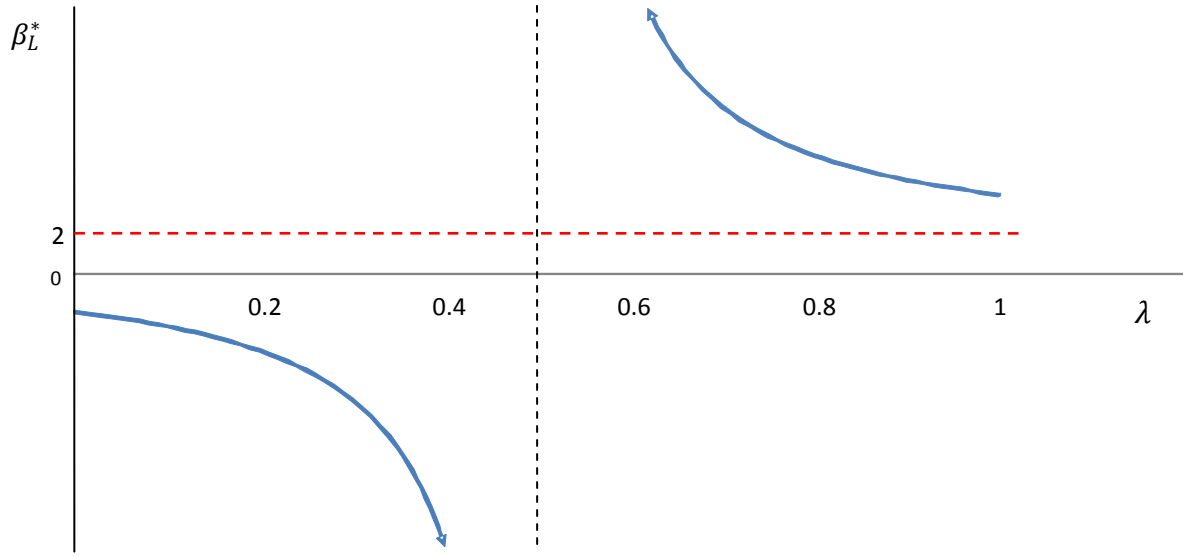


Figure 5.2 holds for any values of x, F_S and F_L that satisfy the following two conditions:

$$xF_S > 1$$

$$xF_L > 1$$

These conditions are quite reasonable if one assumes that the fee and mark up recieved by GPs for each consultation are sizeable[6].

In Figure 5.2, the dashed red line represents $\beta_L^* = 2$, recall that the preferences of the patient population have been defined by $\beta_L \sim U[0,2]$. A point to note in the graph is that $\beta_L^* \notin [0,2]$ for any $\lambda \in [0,1]$. This means that patients are not divided between GPs for any value of the belief λ, instead, the entire patient population will always choose to attend the same GP. There are three cases for patient beliefs:

[6] Later in the results, reasonable values for these parameters will be discussed. The parameters will be set such that $F_S = 32$, $F_L = 62$ and $x \in (0.4, 1.1)$. The above conditions are always satisfied for these values.

I. $0 \leq \lambda < 0.5$

For these values of λ, one can see from Figure 5.2 that $\beta_L^* < 0$. This means that all patients prefer to see the GP of type L, since $\beta_L > \beta_L^*$ for the entire patient population. They also believe that the GP of type L is more likely to choose BB since

$$\lambda = P(L|F) < 0.5$$

Hence, for this set of beliefs, all patients will choose to attend the GP that chose BB.

II. $\lambda = 0.5$

From Figure 5.2, or by considering equation (1), one can see that for $\lambda = 0.5$, β_L^* is undefined. This means that no patient is indifferent between seeing the GP of type L and the GP of type S. When this is the case, price is the only factor that influences patients. All patients will attend the bulk billing GP because this increases their utility, since the price paid by the patient is zero. Hence, for the belief $\lambda = 0.5$, all patients will choose to attend the GP that chose BB.

III. $0.5 < \lambda \leq 1$

For these values of λ, one can see from Figure 5.2 that $\beta_L^* > 2$. This means that all patients prefer to go to the GP of type S since $\beta_L \leq \beta_L^*$ for the entire patient population. Patients also believe that type S is more likely to choose BB since

$$\lambda = P(S|BB) > 0.5$$

Hence, for this set of beliefs, all patients will choose to attend the GP that chose BB.

Although the underlying motivation is different for each set of beliefs, in all three of the above cases, the entire population of patients attends the bulk billing GP. Hence, the payoffs to GPs at node (b) and (c) are the same for all values of λ, with the proportion of

patients that see each GP being $m_i = 1$ for the GP that chose BB and $m_i = 0$ for the GP that chose F. The payoffs to GPs are:

GP payoffs at node (b): $b_S = F_S$

$b_L = -C$

GP Payoffs at node (c): $c_S = -C$

$c_L = F_L$

GP payoffs at each of the nodes can be summarised by Table 5.1:

Table 5.1: Payoff Matrix 1

		L	
		BB	F
S	BB	$0.5F_S, 0.5F_L$	$F_S, -C$
	F	$-C, F_L$	$0.5(1+x)F_S - 0.5c - C, 0.5(1+x)F_L — 0.5c - C$

Proposition 1 The simple model has a unique Nash equilibrium in pure strategies; this equilibrium is {BB, BB}.

Proof: Suppose both GPs choose BB, if a GP deviates and chooses F the other GP can attract the entire patient population by retaining BB. This will result in the GP that chose F having no patients and therefore no revenue. Thus, no GP has an incentive to unilaterally deviate from BB. Furthermore, since BB is a dominant strategy for both type S and type L, this equilibrium is unique. □

Proposition 2 There is no equilibrium in mixed strategies which satisfies Bayes' consistency in the simple model.

Proof: Denote μ_S as the probability that S chooses F and μ_L as the probability that L chooses F. For GP of type L the expected payoffs for choosing each fee structure are:

If he chooses BB: $\mu_S c_L + (1-\mu_S)a_L$

If he chooses F: $\mu_S d_L + (1-\mu_S)b_L$

The expected payoffs of each fee structure for GP of type S are:

If he chooses BB: $(1-\mu_L)a_S + \mu_L b_S$

If he chooses F: $(1-\mu_L)c_S + \mu_L d_S$

GPs are indifferent between their strategies when the following expressions hold:

$$\mu_S c_L + (1-\mu_S)a_L = \mu_S d_L + (1-\mu_S)b_L \quad (1)$$

$$(1-\mu_L)a_S + \mu_L b_S = (1-\mu_L)c_S + \mu_L d_S \quad (2)$$

In order to satisfy Bayes' consistency, patient beliefs need to be consistent with true proportions, using Bayes' rule:

$$\lambda = \frac{(1-\mu_S)\mu_L}{(1-\mu_S)\mu_L + \mu_S(1-\mu_L)} \quad (3)$$

Also, μ_S, μ_L and λ need to be restricted such that:

$$0 \leq \mu_i \leq 1 \quad \text{with } 0 < \mu_i < 1 \text{ for some } i \in \{S, L\} \quad (4)$$

$$0 \leq \lambda \leq 1 \quad (5)$$

Subbing the payoffs from Table 5.1 into equations (1) and (2) and rearranging, the result is:

$$(1): \quad \mu_S = \frac{F_L + 2C}{xF_L - c}$$

$$(2): \quad \mu_L = \frac{F_S + 2C}{xF_S - c}$$

Since, $F_i > 0$, $C \geq 0$, $c \geq 0$ and $0 \leq x \leq 1$ it is always true that:

$$F_L + 2C \geq xF_L - c \text{ and,}$$

$$F_S + 2C \geq xF_S - c$$

Thus, $\mu_S \geq 1$ and $\mu_L \geq 1$ which violates condition (4). Therefore the system of equations $(1) - (5)$ cannot be satisfied and there is no equilibrium in mixed strategies. □

Proposition 2 is consistent with the proof of uniqueness in Proposition 1, however it is important to establish how one would seek to prove a mixed strategy equilibrium if it did exist.

5.2 The Implications of the Simple Model Results

Under the assumptions of the simple model, it has been found that all GPs will choose to bulk bill. Both GPs choose to bulk bill due to the information asymmetry in the market. Since patients are unaware of the type of consultation provided by each GP, the price has a strong effect on patient decisions. If patients observe different fee structures they will prefer to go to the cheaper GP. This gives GPs an incentive to bulk bill in order to attract more patients. Since no GP has an incentive to choose to charge a fee, the only equilibrium occurs where

both GPs bulk bill. This outcome is analogous to the Bertrand Paradox, in which two competitors are sufficient to lower the market price to marginal cost. This is because the assumptions of this model are consistent with those of the Bertrand game, GPs compete only once by choosing a price, there is no entry by other GPs and patients cannot distinguish between the GP types.

This result is what one expects to find in a competitive market. However, it does not explain the fact that, in reality, a proportion of Australian GPs choose to charge a fee. This means that the simple model does not adequately reflect the Australian primary healthcare market.

Inspecting the result of the simple model more closely, it becomes clear that it is not realistic. Considering the equilibrium {BB, BB}, one should note that each GP sees half of the population. This does not capture the fact that it would take the GP of type L far longer than the GP of type S to see the same proportion of the patient population. This means that it is possible that the GP of type L will not be able to see all of the patients that he is assigned in equilibrium. Also, consider the payoffs for {BB,F} and {F,BB}, in each case one GP type sees the entire population of patients. It seems dubious to assume that one GP type would be able to handle the entire patient population alone. In the following section, the impact of capacity constraints is considered as a possible resolution to these inconsistencies.

5.3 Results of the Extended Model

In this section, Assumption 5A and Rationing Rule 2 are applied in order to incorporate the fact that, due to their limited working hours, GPs have constrained capacities. This will alter the payoffs to GPs from those in the simple model. Recall that s and l denote the capacity constraints of the GP of type S and the GP of type L (respectively). The changes to GP payoffs are summarised in Table 5.2, the explanations for each change appear below.

Table 5.2: Summary of Changes

Payoff	The Simple Model	The Extended Model
a_S	$0.5F_S$	$(1-l)F_S$
a_L	$0.5F_L$	lF_L
b_S	F_S	sF_S
b_L	$-C$	$(1-s)(1+x)F_L-(1-s)c-C$
c_S	$-C$	$(1-l)(1+x)F_S-(1-l)c-C$
c_L	F_L	lF_L
d_S	$0.5(1+x)F_S-0.5c-C$	$(1-l)(1+x)F_S-(1-l)c-C$
d_L	$0.5(1+x)F_L—0.5c-C$	$l(1+x)F_L—lc-C$

In the simple model, at nodes (a) and (d) patients were split evenly between the two GP types. In the extended model, Assumption 5A states that $l < s$. This means the GP of type L may not be able to see the same number of patients as the GP of type S. It will also be assumed that $l < 0.5$, which means that by Rationing Rule 2, $0.5 - l$ patients must switch from the GP of type L to the GP of type S.

In the simple model, at node (b) all patients attended the GP of type S. In the extended model, Assumption 5A states that $s < 1$. This means that by Rationing Rule 2, $1 - s$ patients must switch from the GP of type S to the GP of type L. It is pertinent to note the fact that if

$1 - s > l$, the type L cannot see all of the patients that are transferred to him from type S. If this is the case, then there will be a portion of the patient population that are not seen by any GP.

Finally, in the simple model, at node (c) all patients attended the GP of type L. In the extended model, it has been assumed that $l < 0.5$. This means that by Rationing Rule 2, $1 - l$ patients must switch from the GP of type L to the GP of type S. As above, if $1 - l > s$ there will be a portion of the patient population that will not be seen by any GP.

The new payoffs can be summarised by Table 5.3:

Table 5.3: Payoff Matrix 2

		L	
		BB	F
S	BB	$(1-l)F_S, lF_L$	$s(F_S), (1-s)F_L(1+x) - (1-s)c - C$
	F	$(1-l)F_S(1+x) - (1-l)c - C, l(F_L)$	$(1-l)F_S(1+x) - (1-l)c - C, lF_L(1+x)$— $lc - C$

From Table 5.3, one can define two inequalities that need to be satisfied so that each combination of the GPs' strategies represents an equilibrium in pure strategies. Table 5.4 shows these inequality conditions, for each equilibrium there is a condition based on the payoffs of each GP type. Essentially, each of these inequalities shows the relation that needs to be satisfied so that a GP prefers a particular strategy, when he treats the other GP's strategy as given.

Table 5.4: Equilibria

Equilibrium	Type S	Type L
{BB, BB}	$(1-l)F_S \geq (1-l)F_S(1+x) - (1-l)c - C$	$lF_L \geq (1-s)F_L(1+x) - (1-s)c - C$
{BB, F}	$sF_S \geq (1-l)F_S(1+x) - (1-l)c - C$	$(1-s)F_L(1+x) - (1-s)c - C \geq lF_L$
{F, BB}	$(1-l)F_S(1+x) - (1-l)c - C \geq (1-l)F_S$	$lF_L \geq lF_L(1+x) - lc - C$
{F, F}	$(1-l)F_S(1+x) - (1-l)c - C \geq sF_S$	$lF_L(1+x) - lc - C \geq lF_L$

The most meaningful way to assess the impact of capacity constraints on the equilibria is to consider how the capacity constraints of each of the GP types interact. For this reason, it is pertinent to rearrange the above conditions so that they show the relationships between s and l that are required for each equilibrium to be satisfied. The rearranged inequalities are shown in Table 5.5.

Table 5.5 Capacity Constraints and Equilibria

Equilibrium	Type S	Type L
{BB, BB}	$l \geq \frac{xF_S - c - C}{xF_S - c}$	$s \geq \frac{(1+x-l)F_L - c - C}{(1+x)F_L - c}$
{BB, F}	$l \geq \frac{(1+x-s)F_S - c - C}{(1+x)F_S - c}$	$s \leq \frac{(1+x-l)F_L - c - C}{(1+x)F_L - c}$
{F, BB}	$l \leq \frac{xF_S - c - C}{xF_S - c}$	$l \leq \frac{C}{xF_L - c}$
{F, F}	$l \leq \frac{(1+x-s)F_S - c - C}{(1+x)F_S - c}$	$l \geq \frac{C}{xF_L - c}$

The equations in Table 5.5 contain several parameters in addition to the capacity constraints of both GP types, namely F_S, F_L, c, C and x. These parameters need to be fixed in order to isolate and evaluate the relationship between s, l and each of the equilibria. The simplest parameters to fix are the schedule fees. For the remainder of the results they will be fixed such that:

$$F_S = 32$$

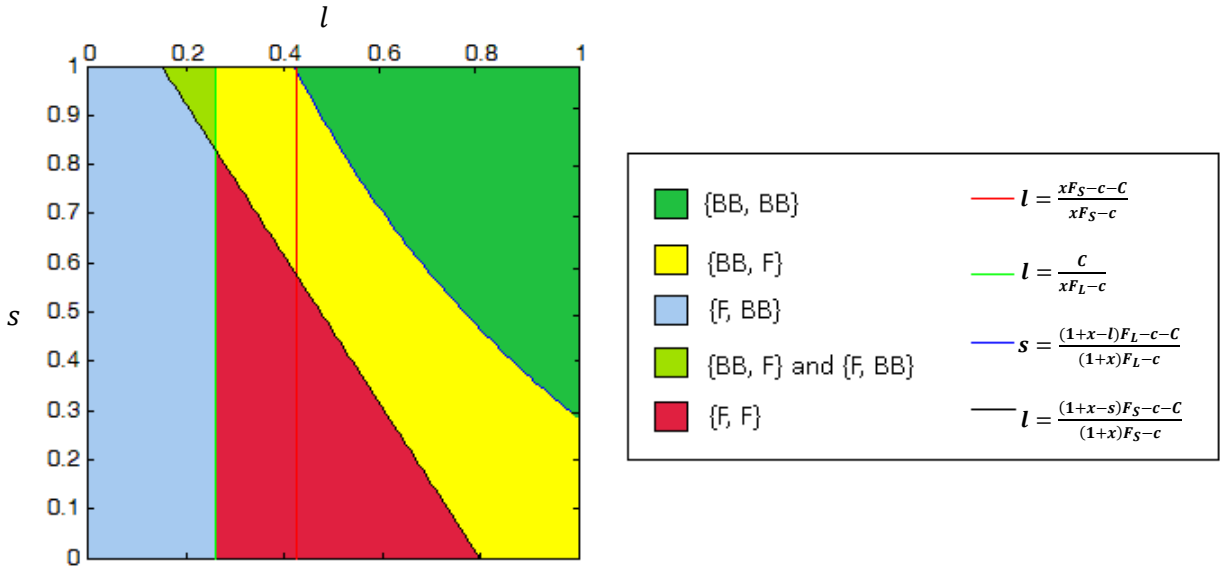
$$F_L = 62$$

This is done for empirical reasons, the values reflect the level 'B' and level 'C' surgery consultation fees in the Medicare Benefit Schedule Book (MBS 2007). Level 'B' is a consultation between 10 and 20 minutes in duration and the associated fee is \$32.80. Level 'C' is a consultation of up to 40 minutes in duration and the associated fee is \$62.30. The fees are rounded down to the nearest dollar for simplicity. Results were derived for a variety of values of F_S and F_L. It was found that their exact value was not pertinent, and only their value relative to other parameters had an impact on the result. Hence, for the remainder of the results, F_S and F_L will remain fixed, while comparative statics will be employed in order to test the effect of changes in the other parameters on the equilibria.

Prior to examining the results, it is pertinent to describe how to interpret them. Figure 5.3 represents a sample of the results that were found. The axes are the capacity constraints of each GP type. Each shaded area shows the range of the GPs' capacities for which a given equilibrium holds. This is based on the inequalities in Table 5.5. The colour which identifies each equilibrium is shown beside the graph in Figure 5.3. Note that the lime green area

represents the overlap of two equilibria {BB, F} and {F, BB} while the rest of the colours represent a unique equilibrium in pure strategies. The final point of note is that the boundaries each have a unique colour, this is so that one can see how the boundaries shift as the parameters are changed. Each coloured line is defined by an equality in the table beside the graph.

Figure 5.3: Results [7]



In Figure 5.3, the result is shown for all combinations of s and l. The graph is not consistent with Assumption 5A which states that $l < s$. This assumption is a product of the definitions of each type, since the GP of type L spends longer with each patient, he can see less patients in a given period of time. Hence, for the remainder of the results, the capacities of each GP type will be limited to the range where $l < s$. In so doing, the more realistic area where the GP of type L has a lower capacity constraint than the GP of type S can be focused on. It has

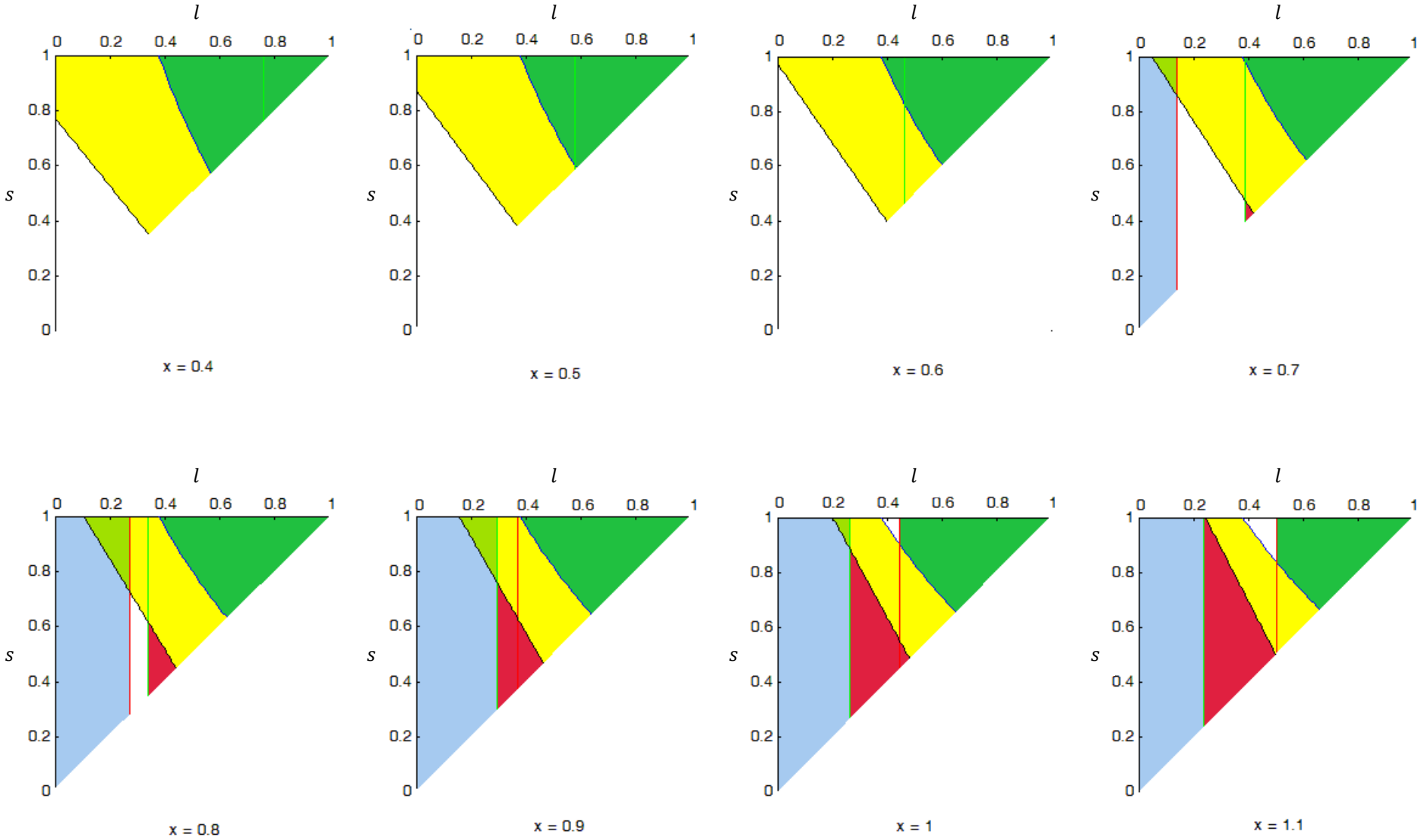
[7] Figure 5.3 is shown for explanatory purposes, it represents the result when $x = 0.7$, $C = 10$ and $c = 5$.

also been assumed that $l < 0.5$, however, for the purpose of the comparative statics it is worthwhile to consider the entire range of l.

The remainder of this section comprises of two parts. First, the effect of changes in each of x, c and C on each equilibrium are described. Following this, the observed relationship between GP capacities and each equilibrium will be outlined.

To begin with, comparative statics are considered in order to observe the effect of changes in each of the parameters on the conditions under which each equilibrium holds. This analysis is undertaken by simulation because there are many parameters that are easier to assimilate in a visual format. Figure 5.4 shows the effect of variation in the mark up x on the equilibria. The parameters for cost are fixed at $C = 15$ and $c = 5$, these values represent relatively low costs compared to the potential revenue of each GP. In general, the patterns that are discussed hold for the entire ranges of c and C that were tested.

Figure 5.4: Varying x



The overall pattern is that, as x becomes larger, both GP types are more likely to charge a fee. This effects the GP of type L for lower values of x than the GP of type S. As x increases, {BB, BB} only holds as an equilibrium for higher levels of both s and l. This is because an increase in x makes charging a fee more attractive relative to bulk billing. Thus, GPs will avoid bulk billing unless both of their capacities are sufficiently high to induce price competition as they seek to attract patients.

The GP of type L has an incentive to charge a fee for the entire range of x. That is, {BB, F} occurs at low levels of x, including the lowest value $x = 0.4$. It is likely that this is due to the low capacity constraint, l, since the GP of type L cannot be profitable from bulk billing when he is not able to see many patients. As x increases, {BB,F} stops occuring at lower values of l, since it becomes more appealing for the GP of type S to charge a fee as well. This means that the equilibrium {F, F} expands across higher values of s as x increases, although it only begins to appear where $x = 0.8$ which is a relatively high mark up.

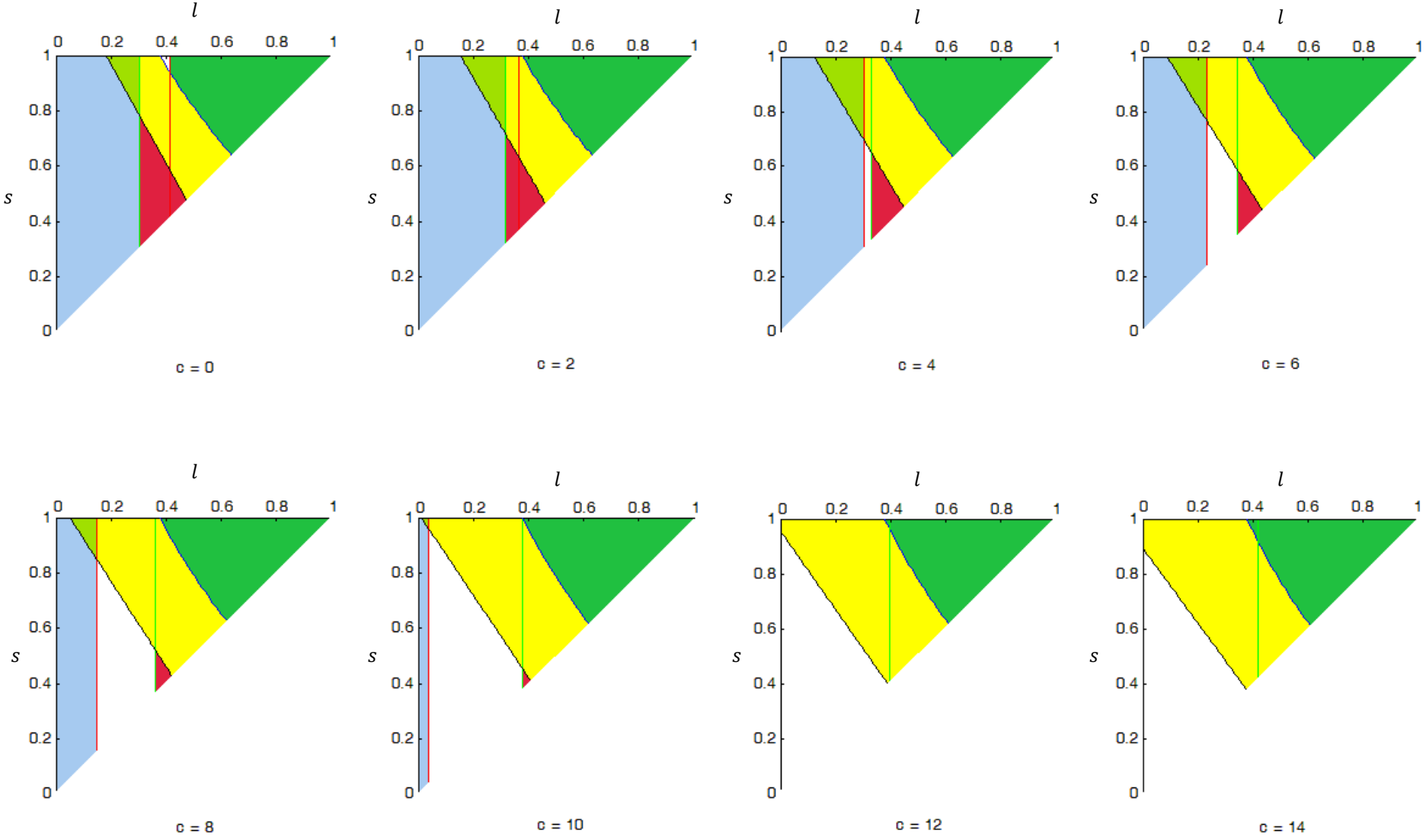
Intuitively, this happens because the mark up has a lower marginal benefit for the GP of type S than for the GP of type L. This means that the mark up that is sufficiently high to induce the GP of type S to charge a fee must be greater than the equivelant mark up for the GP of type L.

Finally note that, unlike the other equilibria which all increase or decrease monotonically with x, there is a non-monotonic relationship between x and the occurence of the

equilibrium {F, BB}. Initially, as x increases, {F, BB} occurs for higher levels of l, eventually after $x = 0.9$ the values of l for which {F, BB} holds begin to decline. This occurs because there are two concurrent effects, one effect is that l is not sufficiently high for the GP of type L to cover the administrative cost of charging a fee with his revenue from patients. The other effect is that, for higher levels of x, type L recieves more revenue from each patient that he sees. Initially the first effect dominates which is seen in the expansion of the {F, BB} region, after $x = 0.9$ the second effect dominates and the region contracts.

The next step is to consider the effect of variation in the variable cost, c. The variable cost is assumed to be small relative to the schedule fee, since it only reflects the extra administrative cost associated with charging a patient. For this reason, in Figure 5.5, c is considered over the range zero to fourteen. The range includes $c = 0$, in order to account for the possibility that there is no variable cost associated with charging a fee. Also, note that changes in both of the cost paramaters are considered relative to the fixed values of F_S and F_L. For this reason, a decline in c or C is equivalent to an increase in the schedule fees.

Figure 5.5: Varying c



In Figure 5.5, $x = 0.8$ and $C = 15$ which represents a sizeable mark up and a relatively low fixed cost for both GP types. The patterns that are discussed hold for the entire ranges of x and C that were tested. As c increases the GP of type S becomes less likely to charge a fee. This is shown by the fact that the {F, F}, and {F, BB} regions become less prominent. Note that, as c increases, {BB, F} becomes larger. This is because this range of c is not sufficient to discourage the type L from charging a fee, because his revenue is higher relative to type S and hence he can bear a larger cost. As c increases, the {BB, BB} region remains largely unchanged. This is because the type S is becoming more likely to bulk bill, while the type L is still prefers to charge a fee and thus it is the {F, BB} region that expands. By the above reasoning, if the test were extended to higher values of c, the type L would eventually also become more likely to bulk bill and the {BB, BB} region would expand.

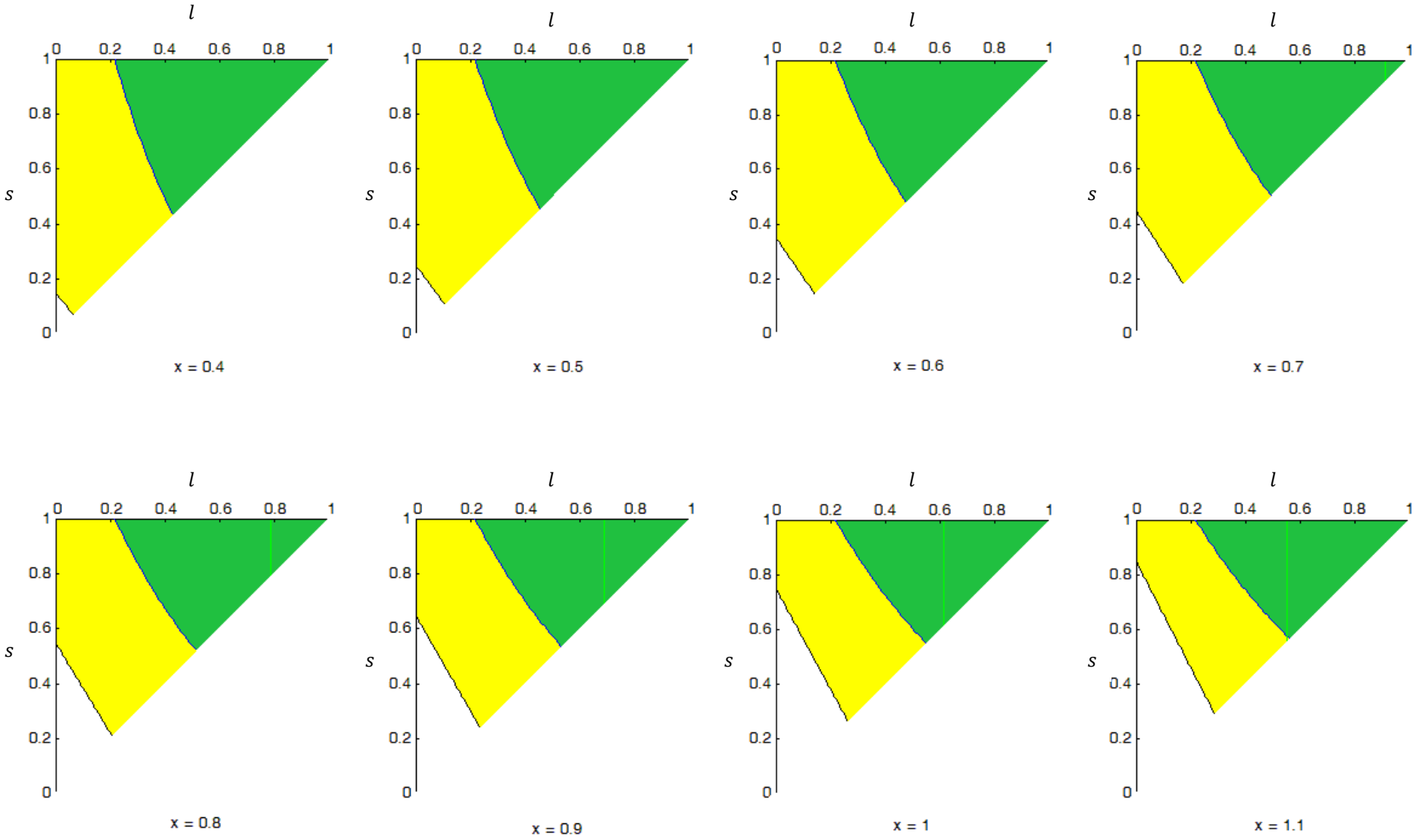
Having looked at what happens as c and x vary, the final step is to consider what happens when C changes. The fixed cost, C, was tested over the range five to forty. The fixed cost is the most difficult parameter to justify based on empirical grounds as the expense incurred by GPs when they choose to charge a fee is unknown. Hence, it was tested over a relatively large range. However, it was not tested for $C = 0$, since it is important to emphasise that the construction of the model states that there is definitely some cost imposed on GPs for charging a fee.

The observed trend is that, as C increases, GPs became more likely to bulk bill, the effect on the equilibria is the same as for increases in c. This is because, as the cost of charging a fee

increases, choosing to do so becomes less appealing. Since Figure 5.5 already shows the pattern that occurs as the cost of charging a fee increases, the approach taken here is to show a level of C that is sufficient to completely discourage the type S from charging a fee. Figure 5.6 is an example of the results when $C = 35$ which is a relatively large fixed cost, also $c = 5$. The reason that the GP of type S will not choose F is self evident, since the cost of charging a fee is higher than any revenue he could gain from it. The GP of type L, on the other hand, will still charge a fee when l is relatively low. This is because it is worthwhile for him to charge patients a mark up when he is only able to see a small number of them.

In summary, the above comparative statics show that, firstly, as the cost of charging a fee increases both GPs become more likely to bulk bill. The GP of type S becomes discouraged from charging a fee for lower values of c and C than the GP of type L. This is because type S has relatively lower revenue per consultation than type L and is therefore more sensitive to changes in costs. The above results also show that as x increases both GPs become more likely to charge a fee. This effects the GP of type L for lower values of x relative to the GP of type S. The reason is that the marginal benefit of a given mark up is greater for type L than type S, since the schedule fee for a long consultation is higher. This higher marginal benefit serves as a stronger inducement to charge a fee for the GP of type L. These results provide some justification for different GP types, in the same location, making different decisions about whether to bulk bill or charge a fee.

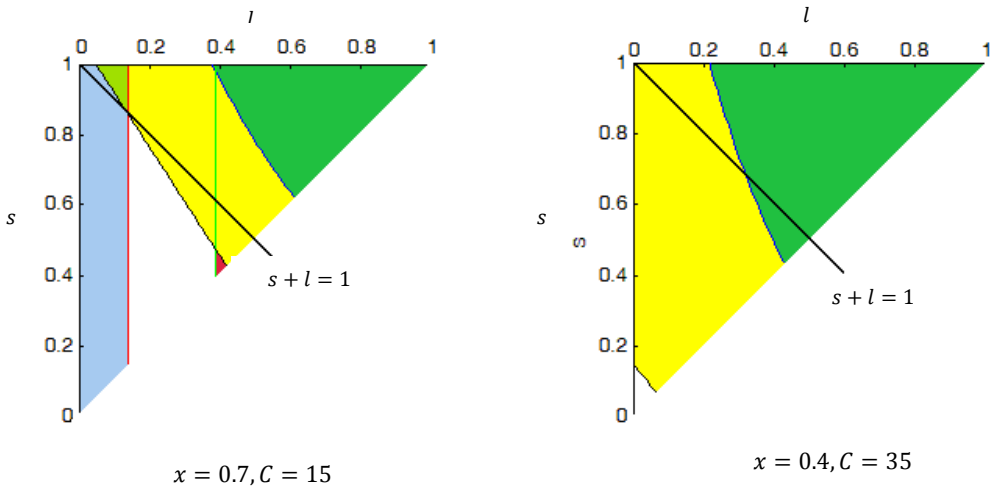
Figure 5.6: High Fixed Cost



Having considered the comparative statics of the other parameters, the relationship between s, l and each equilibrium can now be considered.

As a general rule, {BB, BB} only occurs where s and l are both very high, such that they satisfy the relation $s + l > 1$. The exception to this rule occurs where C is very high, the mark up x is low and both s and l are low simultaneously. Consider Figure 5.7, the relation $s + l = 1$ is shown on each graph. Figure 5.7(A) shows the case that holds for the majority of the results, the {BB, BB} region occurs entirely above the line $s + l = 1$. Figure 5.7(B) shows the exception to the rule, C is higher and x is lower than in Figure 5.7(A) (in both graphs $c = 5$). The result, in Figure 5.7(B), is that when both GPs have very low capacities, {BB, BB} occurs where $s + l < 1$. This is due to the fact that, the incentive to charge a fee is diminished by the low value of x and the high value of C. This is an extreme case, in general it is more likely that both GPs will need to have excess capacities in order for competition between them to be sufficiently strong to induce them both to choose to bulk bill. This has implications for policy that will be discussed in Section 5.4.

Figure 5.7: Conditions for {BB, BB}



$x = 0.7, C = 15$

(A)

$x = 0.4, C = 35$

(B)

The remaining equilibria can be analysed by observing Figures 5.4 and 5.5 above. Firstly, it can be seen that {BB, F} is characterised by a negative relationship between s and l. This means that for the GP of type L to be able to charge a fee, as his capacity increases, the capacity of the GP of type S needs to decrease. The reason for this is that total GP capacity must not be too high for the type L to be afforded the market power to charge a fee while the type S bulk bills.

Secondly, in the figures above, {F, BB} occurs where l is very low. The underlying reason for this is that the GP of type L cannot cover the cost of charging a fee with the revenue he recieves from seeing only a few patients, for this reason he chooses to bulk bill. Furthermore, a large proportion of the population will be crowded out to the GP of type S which gives him the market power to charge a fee. Since these patients have no choice other than to see the GP of type S.

Finally, in the figures above, {F, F} occurs in the region where l is sufficiently high for the GP of type L to cover the cost of charging a fee with the revenue from seeing patients. However, l must not be high enough to stop patients from being crowded out from the GP of type L to the GP of type S, since this affords the GP of type S the market power to charge a fee as well.

5.4 Implications of the Results of The Extended Model

In the Australian primary healthcare market, GPs are split between bulk billing and charging a fee. In terms of the model, it is difficult to conclude whether the outcome in the market is reflected by {F, BB} or {BB, F} . Intuitively {BB, F} seems more likely. This is because, on the one hand, the marginal benefit of charging a fee is higher for the GP of type L which encourages him to charge a fee. On the other hand, the cost of charging a fee acts as a stronger deterent for the GP of type S; this encourages him to choose to bulk bill. Also, {F, BB} occurs in regions where the capacity of the GP of type L is extremely low, which makes it a less likely outcome. The result shows that both outcomes {BB, F} and {F, BB} are possible under certain conditions in the context of the model. Both occur when there is a general undersupply of GP hours. This means that the model supports the conjecture that the different choices of fee structure by different GPs in the Australian primary healthcare market reflects a general shortage of GPs.

As discussed in Section 3.3 it has been suggested that, in order to entice all GPs to bulk bill, policy should be directed towards eliminating the shortage of GPs. This means increasing the supply of GPs, which translates to increasing the capacities of the GPs in the model. However, the results of the extended model have shown that, in general, {BB, BB} only occurs when the total capacity of GPs is significantly greater than the size of the population. This means that increasing the supply of GPs to encourage universal bulk billing can only be achieved if GPs are oversupplied. This is not an efficient way to achieve lower prices, since by the construction of the model, it requires practicing GPs to be underutilising their capacities. Furthermore, policy has thus far failed to attract a sufficient number of GPs to

meet the demand for primary healthcare, it is difficult to concieve that there will be any more success in creating an oversupply of GPs. The model has shown that for {BB, BB} to occur without necessitating an oversupply of GPs, the cost of charging a fee needs to be very high and the mark up needs to be low. The governement could achieve this through policy such as capping the mark up to a low level. However, such a policy would be detrimental to the incomes of GPs. Over time this would discourage medical graduates from entering the market, and hence exacerbate the issue of the overall shortage of GPs.

In Section 3.3, the possibility of expanding the role of practice nurses under Medicare in order to alleviate GPs of some of their more simple duties was discussed. Although nurses were not explicitly incorporated in the model in this thesis, there is some scope for such a policy to be succesful. This is because a practice nurse reduces the workload of the GP for a given patient, this means that the GP would have the ability to see more patients. In terms of the model, their capacity will be higher. The difference between adding nurses and GPs is that nurses are not direct competitors to GPs due to their relatively limited set of skills.

Overall, there is a trade-off between the provision of universal bulk billing and maintaining the incomes of GPs. This trade-off occurs due to the limited size of the Medicare budget. Under the current system, it seems that inducing universal bulk billing will benefit patients in the short run. However it will detriment them in the long run when the supply of GP services falls even further, in response to the reduced reimbursement. The option of increasing the reimbursement that GPs recieve for bulk billing is constrained by the

Medicare budget. For this reason, it may be more practical to accept the fact that some GPs will charge fees and focus the limited funding on reimbursing those patients with the least ability to pay.

Chapter 6

Concluding Remarks

The conclusion will begin with a short synthesis of the findings of this thesis. Following this, some extensions to the model will be suggested. These would be worthwhile avenues for future research into the economics of the Australian market for primary healthcare.

In summary, this thesis has achieved the aims of the hypothesis outlined in Chapter 1. The results demonstrate that, when GP capacities are constrained, the conjecture that a GP with long consultations will choose to charge a fee while a GP with short consultations will choose to bulk bill can be supported for certain ranges of the model parameters. This is because GPs with long consultations receive a higher marginal benefit from charging a fee and are less sensitive to the associated cost than GPs with short consultations.

The results have shown that if GPs have unconstrained capacities they will all choose to bulk bill. Since GPs do not all bulk bill in reality, this raises the issue of the general shortage of GPs in the Australian primary healthcare market. The result confirms that a shortage of GPs plays a role in allowing a proportion of GPs the market power to charge a fee. The more interesting result is that, in order to induce all GPs to bulk bill, their total capacity must significantly exceed the needs of the population. For this reason, the policy of encouraging an increase in the supply of GPs will not, on its own, lead all GPs to bulk bill unless GPs are oversupplied. It is the opinion of the author, that this is not an efficient method of achieving universal bulk billing since the vast set of skills embodied in practicing GPs will be underemployed. The policies of expanding the role of practice nurses under Medicare, or

improving the monetary incentives of bulk billing may be more effective in achieving the goal of universal bulk billing.

This thesis is entirely theoretical, for this reason it would be worthwhile to undertake an empirical study to find out whether average consultation lengths of GPs are linked to their pricing decisions. The results have shown that any combination of fee structures chosen by the two types of GPs can represent an equilibrium for certain values of the model parameters. For this reason, an empirical study would assist in confirming that GPs with long consultations tend to charge fees while GPs with short consultations tend to bulk bill.

The remainder of this conclusion will evaluate the theoretical model constructed in this thesis and consider several extensions and suggestions for future work that may make it a better reflection of the Australian market for primary healthcare.

First, the model in this thesis assumes that when GPs choose to charge a fee, the mark up over the schedule fee is exogenous. This does not reflect the fact that, in reality, when GPs choose to charge a fee they have the freedom to set their prices. It would be worthwhile to explore the mark up optimisation problem of a GP that charges a fee. It may be the case that one type of GP has more market power and can set a higher mark up, which would serve as further justification for different types choosing different fee structures.

Second, it is assumed in the model that there are equal proportions of each GP in the market. It would be valuable to explore whether changing the proportions of each type will

affect the outcome. Recalling the empirical findings of Britt et al (2005), discussed in Section 2.1, they concluded that female GPs were more likely to have longer consultations. If this is true, then increasing participation of females in the GP work force is likely to increase the proportion of GPs that provide long consultations. The question of interest is whether this will have an impact on GP decisions about fee structure.

Third, an assumption of the model is that patients have varied preferences for consultation length. That is, in the model GP services were horizontally differentiated. One could instead argue that all patients prefer a longer consultation because this represents a signal of quality. It would be worthwhile to test whether altering the model to incorporate vertical differentiation of GP services would affect the equilibrium outcome.

Finally, the model in this thesis comprises a single period interaction between GPs and patients. A significant extension would be to explore the scenario in a multiple period setting. Over time patients would be able to learn information about the GPs that they had seen in previous periods. However, it is still pertinent to consider a single period since there is information asymmetry in the primary healthcare market and patients do not typically see many GPs or switch between GPs often. This means that, the first period payoff is relevant because the period that it takes for patients to assess their GP and choose to switch may represent a sizeable length of time. Another potential extension, related to the multiple period setting would be to consider entry into the market by new GPs. It would be pertinent to explore how the fee structures of the established GPs would affect the choice of fee

structure made by an entering GP. Furthermore, whether the choice of an entering GP would differ depending on the length of consultations they provided.

This discussion of potential extensions to the model demonstrates that this thesis lays the ground work for a variety of areas for future research. However, the model has achieved the initial aim of this thesis which was to justify the different pricing decisions made by Australian general practitioners.

Reference list

ABS, (2008). “Year Book of Australia- 2008”. [Accessed 09/09/08] http://www.abs.gov.au/ausstats/abs@.nsf/7d12b0f6763c78caca257061001cc588/A50BD9743BF2733ACA2573D2001078D8?opendocument

Biggs, A. (2003) Medicare Background Brief [Accessed 08/09/08] http://www.aph.gov.au/library/intguide/SP/medicare.htm

Britt, H. Valenti, L. Miller, G. (2005). “Determinants of consultation length in Australian general practice.” Medical Journal of Australia **183**(2): 6.

Department of Health and Ageing (2007) Medicare Benefit Schedule Book [Accessed 08/09/08] http://www.health.gov.au/internet/main/publishing.nsf/Content/health-publicat.htm

Department of Health and Ageing (2008) Medicare Statistics - March Quarter 2008 [Accessed 08/09/08] http://www.health.gov.au/internet/main/publishing.nsf/Content/health-publicat.htm

Deveugele, M. (2003). “Doctor-patient communication in general practice: An observational study in six European countries.” Belgium, Universiteit Gent (Belgium).

Deveugele, M. Derese, A. Brink-Muinen, A. Bensing, J.Maeseneer, J. (2002). “Consultation length in general practice: Cross sectional study.” British Medical Journal **325**(7362): 472.

Dionne, G. Contandriopoulos, A. (1985). “Doctors and their workshops: A review article.” Journal of Health Economics **4**(1): 21-33

Elliot, A. (2003). "The Decline in Bulk-Billing: Explanations and Implications." [Accessed 03/05/08] http://www.aph.gov.au/library/pubs/CIB/2002-03/03cib03.htm

Evans, Robert G. “Supplier-Induced Demand: Some Empirical Evidence and Implications.” In The Economics of Health and Medical Care, Mark Perlman, ed. New York: Halsted Press: 1974.

Furler, J. S. Harris, E. Chondros, P. Davies, P.G.P. Harris, M.F. Young, D.Y.L. (2002). “The inverse care law revisited: Impact of disadvantaged location on accessing longer GP consultation times.” Medical Journal of Australia **177**(2): 80-83

Hingston, C. (2006). “Corporates behind bulk-bill rise: AMA” Australian Doctor (September 2006): 6

Hoerger, T.J. (1990). “Two-part pricing and the mark-ups charged by primary care physicians for new and established patient visits”. Journal of Health Economics **8**(4): 399-417

Hopkins, S. and Speed, N. (2005). “The decline in `free' general practitioner care in Australia: reasons and repercussions.” Health Policy **73**(3): 316-329

Khan, A. Hussain, R. Plummer, D. Minichiello, V. (2004). “Factors associated with bulk billing: experience from a general practitioners’ survey in New South Wales.” Australian and New Zealand Journal of Public Health **28**(2): 135 -139

Labelle, R. Stoddart, G. Rice, T. (1994). “A re-examination of the meaning and importance of supplier-induced demand.” Journal of Health Economics 13(3): 347-368

Lewis, J.M. Marjoribanks, T. Pirotta, M. (2003). “Changing professions: General Practitioners’ perceptions of autonomy on the front line”. Journal of Sociology **39**(1): 44-61

Masson, R.T. Wu, S. (1974). “Price discrimination for physician services.” Journal of Human Resources **9**(1): 63-79

Newhouse, J.P. (1970). “A model of physician pricing.” Southern Economic Journal 37(2): 174-183

Ogden, J. Bavalia, K. Bull, M. Frankum, S. Goldie, C. Gosslau, M. Jones, A. Kumar, S. Vasant, K. (2004). “I want more time with my doctor”: a quantitative study of time and the consultation.” Family Practice **21**(5): 479.

Rochaix, L. (1989). “Information asymmetry and search in the market for physicians' services.” Journal of Health Economics **8**(1): 53-84.

Sahari, S. (2007). “Nurses may save the day”. Australian Doctor (October 2007): 32

Savage, E. and Jones G. (2004). “An Analysis of the General Practice Access Scheme on GP Incomes, Bulk Billing and Consumer Copayments.” Australian Economic Review **37**(1): 31-40.

Scott, A., Culyer, A. J. Newhouse, J. P. (2000). Chapter 22 Economics of general practice. Handbook of Health Economics, Elsevier. **Volume 1, Part 2:** 1175-1200.

Scott, A. and Vick, S. (1999). “Patients, doctors & contracts: An application of principal-agent theory to the doctor-patient.” Scottish Journal of Political Economy **46**(2): 111.

Thistlethwaite, J. Leeder, S. Kidd, M. Shaw, T. (2008) “Addressing general practice workforce shortages: policy options”. Medical Journal of Australia 189(2): 118-121