Essays on the Regulation of Health Care Provision and the Economics of Chronic Diseases

Anderson Eduardo Stanciole

Ph.D. Thesis

The University of York

Department of Economics and Related Studies

2007

Abstract

This thesis comprises two main topics. The first part of the thesis focuses on the regulation of health care provision, while the second part provides empirical evidence of the economic consequences of chronic diseases in developing countries.

The first part examines the incentives that condition the relationship between hospitals and health care purchasers (Chapters 1 and 2) and the provision of preventive care in a competitive health insurance market (Chapter 3). There are three main areas of consideration. Chapter 1 presents an axiomatic bargaining model of prices and activity, examining the negotiation between hospitals and purchasers in situations where, as usually occurs, none of the parties hold all of the bargaining power. Chapter 2 estimates the effect of waiting times on hospital costs using a sample of 283 hospitals over the period 1995-2002 in the NHS. This analysis seeks to clarify the efficiency effects of imposing waiting times for elective surgeries. Chapter 3 extends Rothschild and Stiglitz's (1976) model of adverse selection in a competitive health insurance market by considering the incentives for prevention.

The second part of the thesis consists of two empirical investigations regarding the economic effects of chronic diseases in Brazil, India and Russia. Chapter 4 provides an analysis of the relationship between socio-economic inequality and chronic diseases, whilst Chapter 5 estimates the effect of chronic diseases on household economic performance, as measured by health expenditures, non-health expenditures and labour productivity. The results emphasise the relevance of

chronic diseases for developing countries, challenging the view that this problem is restricted to more developed societies.

Contents

List of Tables

List of Figures

Acknowledgments

I wish to extend my appreciation to all the people that contributed to make this thesis possible. First, I would like to thank my supervisor Luigi Siciliani for his guidance and patience in helping me to focus on the work and to overcome many difficulties. I would also want to thank Michael Kuhn, who supervised me during the first year of the PhD.

I am very grateful for the input from Andrew Jones and Karen Mumford, who were members of my Thesis Advisory Group. In particular, it was an invaluable privilege to benefit from the contact with Andrew Jones during the HEDG seminars. I also thank my two co-authors, Rowena Jacobs and Dele Abegunde. A special thanks for Dele Abegunde, not only for the intellectual assistance, but also for the constant encouragement and example of faith. Many individuals among the staff and students in the Department of Economics and Related Studies also deserve my appreciation. In particular, Sue Bowden and the colleagues from Genesis 6 and HEDG.

I thank my family for the emotional support they have offered over all these years. I thank my wife, which has always been at my side providing enormous back-up and helping me to make many difficult decisions. My father and my brothers and sisters, though they were not here, were always in my mind driving me to my objective. Finally, I dedicate this thesis to my mother. Unfortunately, she could not wait until it was finished, but she was probably the most important person that allowed me to be here today.

Financial support of the Ministry of Education of Brazil through the grant CAPES BEX 147402-2 is gratefully acknowledged.

Declaration

Chapter 1 is a joint paper with Luigi Siciliani. I contributed to the development of the original idea, prepared the first draft and contributed to subsequent revisions. I ran the numerical simulations using the software Maple 9.5. Earlier versions of the paper have been presented at the Micro Reading Group at the Department of Economics in York (York, 2006) and at the 69^{th} Health Economists Study Group (York, 2006). The paper has been submitted to *Health Economics*.

Chapter 2 results from a collaboration with Luigi Siciliani and Rowena Jacobs. I contributed to the original idea, analysed the data, wrote the first draft and contributed to latter revisions.

Chapter 3 is chronologically the first chapter to be written. It was started during the first year of my PhD, while I was under the supervision of Michael Kuhn.

Chapters 4 and 5 have been co-authored by Dele Abegunde from the World Health Organization. In both cases I was responsible for conceptualising the original idea, developing the methodological approach, preparing and analysing the data, writing the first draft and contributing for further revision in the text. Chapter 4 has been presented at the Health Econometrics and Data Group at the University of York (York, 2006) and at the 6^{th} World Congress of the International Health Economics Association (Copenhagen, 2007). The paper has been submitted to *Social Science and Medicine*.

Previous versions of Chapter 5 have been presented at the 67^{th} Health Economists Study Group (Newcastle-upon-Tyne, 2005), the 2^{nd} International Conference of Health Fi-

nancing in Developing Countries (Université d'Auvergne, 2005), the Health Econometrics and Data Group (York, 2005) and the 6^{th} European Conference of Health Economics (Budapest, 2006). It has been accepted for publication by *Social Science and Medicine*.

Introduction

This thesis comprises two main topics. The first part of the thesis focuses on the regulation of health care provision, while the second part provides empirical evidence of the economic consequences of chronic diseases in developing countries.

Chapters 1 to 3 investigate both theoretical and empirical issues associated with the provision of health care, considering two situations. Chapters 1 and 2 will firstly examine the incentives that shape the economic behaviour of hospitals in an environment similar to the UK National Health Service (NHS). In this environment, a single purchaser establishes contracts with individual hospitals for the provision of health care services for the population at large. The system is financed by general taxation and patients receive free care at the point-of-use.

Chapter 1 explicitly models the negotiation process between hospitals and purchasers using an axiomatic bargaining framework (see Nash 1950, Nash 1953, Osborne and Rubinstein 1990, Barros and Martinez-Giralt 2006). Bargaining occurs in terms of both prices and quantities. This approach allows the investigation of recent developments in health care financing, such as the policy of payment by results, currently under consideration in the UK.

The discussion about hospital incentives is extended in Chapter 2, which estimates the effect of waiting times on hospital costs in the UK NHS. In recent years, there has been substantial interest in waiting times for elective surgeries. Given the limitations in the amount of resources allocated to the NHS, waiting times have become an inescapable real-

ity. Moreover, there is an established view that waiting times increase economic efficiency by prioritising the cases that are treated, and decreasing the need to hold installed excess capacity to attend emergency cases (see Hughes and McGuire 2003). This chapter provides an empirical test for this view by directly estimating the effect of waiting times on hospital costs.

Chapter 3 concludes the first part on the regulation of the provision of health care. However, instead of considering the interplay between purchasers and hospitals, this chapter focuses on the incentives for the provision of health care by health insurance plans in a competitive market. More specifically, this chapter extends the study of adverse selection: explorations in this area normally focus only on curative care, rather than considering the effect on the provision of preventive services.

The second part of the thesis is essentially empirical, and considers the economic aspects associated with chronic diseases in developing countries. In recent years we have witnessed an increased awareness of the epidemiological burden of chronic diseases in the developing world (see Epping-Jordan, Galea, Tukuitonga and Beaglehole 2005, WHO 2005, World Bank 2005). These two chapters examine the channels through which chronic diseases have an impact on economic well-being in Brazil, India and Russia. These countries, particularly Brazil and Russia, currently display an intermediate level of economic development and possess large populations. Moreover, the prevalence of chronic diseases is projected to increase in the future, due to lifestyle changes, increases in life expectancy and associated population ageing.

Chapter 4 opens with a study of the association between socio-economic inequality and the prevalence of chronic diseases. The results suggest that the burden of chronic diseases disproportionately affects the poor in Brazil and Russia. Chapter 5 extends this analysis and evaluates how chronic diseases affect economic welfare at the household level in Brazil, India and Russia. The analysis shows that chronic diseases have significant economic consequences, increasing the level of health care expenditure and decreasing the ability to work and earn income. Ultimately, household non-health related consumption is negatively affected. The results from Chapters 4 and 5 suggest an important role for chronic diseases in determining the economic possibilities for households in developing countries.

Below, we provide a more detailed overview of the main topics covered in each chapter. Chapter 1 "Bargaining and the Provision of Health Services" is a theoretical investigation where we model and compare the bargaining process between a purchaser of a health service, such as a health authority, and a provider (the hospital) in three plausible scenarios: a) The purchaser sets the price, and activity is negotiated between the purchaser and the provider: *activity* bargaining; b) The price is negotiated between the purchaser and the provider, but the activity is chosen unilaterally by the provider: *price* bargaining; and c) Price and activity are simultaneously negotiated between the purchaser and the provider: *efficient* bargaining. We show that: 1) If the bargaining power of the purchaser is high (conversely, low), *efficient* bargaining leads to higher (lower) activity and purchaser's utility, and lower (higher) prices and provider's utility compared to *price* bargaining. 2) In *activity* bargaining, prices are lowest, the purchaser's utility is highest and the provider's utility is lowest; activity is generally lowest, but higher than in *price* bargaining for the high

bargaining power of the purchaser. 3) If the purchaser has higher bargaining power, this reduces prices and activity in *price* bargaining; it reduces prices but increases activity in *activity* bargaining; and it reduces prices but has no effect on activity in *efficient* bargaining.

Using a sample of 283 hospitals over the period 1995-2002 in the NHS, Chapter 2 empirically estimates the elasticity of hospital costs with respect to waiting times. Total hospital costs are estimated using pooled OLS, fixed effects and random effects model. In each case waiting times are entered in the regression in both linear and quadratic terms, assisting in the identification of the curvature of the cost function. The signs of the estimated coefficients are consistent with Iversen's (1993) model: the coefficient is negative for the linear effect and positive for the quadratic effect, suggesting a U-shaped relationship between hospital costs and waiting times. However, the coefficients are generally not statistically significant, suggesting that waiting times have no impact on hospital costs. This casts some doubt on the effectiveness of waiting times as a tool to control hospital costs.

Chapter 3 develops a theoretical model, focusing on the incentives for the provision of preventive care in a competitive health insurance market with adverse selection. This is an extension of the model proposed by Rothschild and Stiglitz (1976) with the introduction of prevention. We assume that patients differ with respect to the probability of illness and with the efficiency of prevention in reducing this probability.

Comparing the first best with the unregulated market equilibrium, we show that asymmetric information distorts the equilibrium levels of both preventive and curative care. Low-risk patients are separated by receiving less than optimal curative care. Moreover, the level of preventive care is also distorted, with the direction of distortion depending on the

relative efficiency of prevention for each risk type. Compared to high-risk patients, low-risk patients receive lower (higher) marginal benefit from preventive care if prevention is relatively more (less) efficient.

The second part of the thesis then extends the discussion about prevention by focusing on the economics of chronic diseases, more specifically in the context of low- and middle-income countries. There are two empirical chapters in this part. Chapter 4 provides evidence of the determinants of inequalities for chronic diseases in Brazil and Russia, focusing on three specific diseases: heart disease, hypertension and diabetes.

We show that in both countries, poorer households face a considerably higher probability of being affected by chronic diseases. The concentration index in Russia starts at -0.021 in 2000, peaks at -0.046 in 2003, and is equal to -0.035 in 2004. In Brazil the concentration changes from -0.028 in 1998 to 0.003 in 2003. The health inequity index is around -0.01 in Russia (with a rising trend) and -0.06 in Brazil. Standardising variables have opposite effects in each country, reducing the level of observed health inequality in Brazil, but increasing the level in Russia. This suggests a differential impact of recent economic changes on the socio-economic inequality among demographic groups in each country. Socio-economic status, co-morbidities and education are the most important factors explaining worsening health inequalities over the time, with the bulk of this effect due to changes in elasticities with respect to the determinants of chronic diseases, rather than changes in the concentration of such determinants. The results suggest that efforts to relieve the burden of chronic diseases from poorer households should aim at maximising the impact of externalities from other policy areas.

The study of the economic effect of chronic diseases is continued in Chapter 5. This paper presents some evidence of the effect of chronic diseases on household health expenditure, non-health expenditure, labour productivity (earned income and work days lost) and remittances from other households, using data from the Living Standard Measurement Surveys (LSMS) from Brazil, India and Russia.

The results indicate that each additional case of chronic disease in the household generates a conditional increase in health expenditure of 21%, 120% and 14% in Brazil, India and Russia respectively. Since the potential to work is affected by chronic diseases, labour income reduces by 7.9% in Brazil and 4.8% in Russia. This effect is partially offset by a conditional increase in remittance of 3.1% in Brazil and 8.7% in Russia. In Brazil, the net effect on non-health consumption is an increase of 3.9%, suggesting that households are able to insure non-health consumption against chronic diseases, possibly from remittances. In Russia, however, where unobserved heterogeneity is accounted for, there is a net reduction of non-health consumption of 3.2%, suggesting that chronic diseases actually reduce overall household welfare.

Chapter 6 concludes, discussing the policy implications, and presenting the limitations and possible extensions of the study.

Chapter 1
Bargaining and the provision of health services

1.1 Introduction

Prospective payment systems are used widely to remunerate health care providers. They usually take the form of Diagnosis Related Groups (DRGs) pricing or similar methods, such as Healthcare Resource Groups (HRGs) in the United Kingdom or Group Homogenes de Maladie (GMC) in France. Depending on the institutional context, purchasers and providers bargain on price, activity, or both. For example, in the US, Health care Maintenance Organisations (HMOs) or private health insurers bargain on price, and seldom activity, with the hospitals (see Barros and Martinez-Giralt 2006, Brooks, Dor and Wong 1997). In the UK, Health Authorities and Primary Care Trusts negotiate price and activity with NHS Trusts under "cost and volume" or "sophisticated" contracts. The government is discussing the implementation of the policy known as "Payment by Results", where prices are regulated, but activity is negotiated between the Primary Care Trust and the NHS Trust. Within the Medicare Programme in the US, prices are chosen by the purchaser (Medicare), while activity is either chosen or bargained with the provider. Similar arrangements exist throughout Europe (see Figueras, McKee, Mossialos and Saltman 2005, Le Grand, Mossialos and Saltman 1999).

Although we observe a substantial amount of bargaining between purchasers and providers, the theoretical literature on the relative merits of prospective payment systems normally assumes that payers are able to set the prices, and often activity, unilaterally, while providers choose the amount of quality and cost-containment effort (see, for example, Ma 1994, Chalkley and Malcomson 1998a, Chalkley and Malcomson 1998b, Mougeot and Naegelen 2005, de Fraja 2000). This implies that purchasers have all the bargaining power, which is a simplifying assumption, as the empirical evidence suggests that providers may hold at least some of it. Propper (1996) shows that in England purchasers with higher bargaining power could secure lower prices. Brooks et al. (1997) estimate that US hospitals hold on average 65% of the bargaining power when negotiating with private insurers. Melnick, Zwanziger, Bamezai and Pattison (1992) find a negative association between purchasers with greater market shares and prices charged by the providers.

This study models the bargaining process between a purchaser of health services (a health authority) and a provider (a hospital) in three plausible scenarios: a) the purchaser sets the price, and the activity is bargained between the purchaser and the provider: *activity* bargaining; b) the price is bargained between the purchaser and the provider, and the activity is chosen unilaterally by the provider: *price* bargaining; c) price and activity are bargained simultaneously between the purchaser and the provider: *efficient* bargaining.

The background context of the interaction between providers and purchasers differs across countries. It is important to consider the implications of the bargaining framework for the health care sector. Price bargaining is relatively more common in health systems based on private provision of healthcare and insurance. In this case, insurers and providers

bargain on the price of procedures, whilst providers decide on the amount of activity. On the other hand, under public health care programmes one usually finds the other two types of bargaining. In some cases, purchasers have the ability to set prices unilaterally, whilst having to bargain on the amount of activity supplied by the provider. In others, such as in the current situation in the UK, efficient bargaining occurs, meaning that purchaser and provider bargain simultaneously on both activity and price. The results of this analysis suggest some specific implications for the health care sector of the different bargaining scenarios. This points some directions that might help policy makers in determining the most appropriate setting according to the specific conditions of the market.

Our main objective is to compare prices, activity and the utility of provider and purchaser in each of the three different scenarios. The results are: 1) if the bargaining power of the purchaser is higher than a certain threshold, *efficient* bargaining leads to higher activity and purchaser's utility, and lower prices and provider's utility, compared to *price* bargaining. The results are reversed if the bargaining power is below the threshold. The threshold is higher when the marginal benefit function is steeper (ie the benefit function is more concave) and when the marginal cost function is flatter (ie the cost function is more convex). This result is surprising, as one would expect the purchaser to be better off when she can bargain with both instruments, price and activity. This intuition proves correct only when the bargaining power of the purchaser is high. When it is low, the purchaser would be better off contracting on prices only. 2) In *activity* bargaining, price and the provider's utility are lowest and the purchaser's utility is highest. The level of activity in *activity* bargaining is always lower than in *efficient* bargaining. It is also lower than in *price* bargaining, but

only if the bargaining power of the purchaser is below a certain value (which, according to numerical simulations is at least 0.59). This also implies that moving from *price* to *activity* bargaining might reduce activity. It will certainly be reduced if we move from *efficient* to *activity* bargaining. 3) In *price* bargaining, higher bargaining power of the purchaser reduces prices and activity; in *activity* bargaining it reduces prices, but increases activity; and in *efficient* bargaining it reduces prices but has no effect on activity.

This study contributes to the literature on purchaser-provider bargaining in healthcare (for a recent survey see Barros and Martinez-Giralt (2006)). The model focuses specifically on the interaction between purchaser and provider, and does not consider an active role for patients. Ellis and McGuire (1990) develop a model in which patients and doctors bargain about the intensity of treatment, and derive the optimal combination of patient's insurance and reimbursement for the provider which maximises consumer welfare.[1] Barros and Martinez-Giralt (2005a) show that, when bargaining with providers, purchasers may prefer to bargain with a professional association rather than a subset of more efficient providers. Barros and Martinez-Giralt (2000) analyse the bargaining process, in which the purchaser can choose whether to negotiate with each provider separately or jointly, or announce a contract that any provider is free to sign (the "any willing provider" clause). They show that if the total surplus is high, the purchaser prefers the system of "any willing provider", but if it is low she prefers either joint or separate negotiations. Gal-Or (1997) shows that purchasers (private insurers) might be willing to sign exclusive contracts with a subset of providers in order to secure more favourable terms during bargaining. Gal-Or (1999a) stud-

[1] Dor and Watson (1995) evaluate how different payment mechanisms affect the incentives in the relationship between hospitals and physicians.

ies whether vertical mergers between hospitals and physician practices might enhance their bargaining power with the insurers (see also Gal-Or (1999b)). Barros and Martinez-Giralt (2005b) explore the implications of the coexistence of a public and a private sector in the provision of health services. They argue that the public sector might choose to hold idle capacity in order to extract more beneficial conditions when bargaining with the private sector for the provision of services. There are other applications of bargaining in the health economics literature. Clark (1995) examines how to divide a budget between two patients with different health conditions and capacity to benefit. Pecorino (2002) models the effects of drug reimports from Canada on the profitability of US domestic pharmaceutical companies.[2]

The study is organised as follows. Section 1.2 presents the model. Section 1.3 provides a comparison of the different scenarios. Section 1.4 extends the model by adding quality and cost-containment effort. Section 1.5 offers concluding remarks and policy implications.

1.2 The model

We model the bargaining process between a purchaser of health services, such as a health authority, and a provider (a hospital). Define y as the number of patients treated and p as the price the provider receives for each patient treated. The provider's utility U is given by

[2] See also Wright (2004) for a model of price regulation in the pharmaceutical sector where the regulator and the pharmaceutical company bargain over a subsidy.

its surplus $U(p, y) = py - C(y)$, where $C(y)$ is the cost function of the provider, which satisfies $C_y > 0$, $C_{yy} > 0$ (increasing marginal cost).

The purchaser's utility (or health authority utility) is given by the difference between the benefit for the patients $B(y)$ and the transfer to the provider: $V(p, y) = B(y) - py$. The benefit function satisfies $B_y > 0$ and $B_{yy} \leq 0$.[3]

We analyse three plausible scenarios. 1) *Activity* bargaining: the purchaser sets the price, and activity is bargained between the purchaser and the provider. 2) *Price* bargaining: the price is bargained between the purchaser and the provider, but activity is chosen by the provider. 3) *Efficient* bargaining: price and activity are bargained simultaneously between the purchaser and the provider.

Define γ, with $0 \leq \gamma \leq 1$, as the bargaining power of the purchaser, $(1 - \gamma)$ as the bargaining power of the provider, $\overline{V}$ and $\overline{U}$ as the outside options for the purchaser and the provider respectively, and $\widetilde{V} = V - \overline{V}$ and $\widetilde{U} = U - \overline{U}$. The outside options represent the payoff that each part would attain if they failed to reach an agreement (see Muthoo 1999). Consider for example the case of negotiations on the price of services. If the negotiations fail, it might be the case that the purchaser would need to contract at current market prices. In this case, this is the purchaser outside option. For notational simplicity let $V^i = V(p^i, y^i), U^i = U(p^i, y^i)$, where $i = a, p, e$ denotes respectively activity, price and

---

[3] A more general objective function for the purchaser is $B(y) - (1+\lambda)py + \delta U$, where λ is the opportunity cost of public funds and δ is the weight attached to the utility of the provider. The main results of the analysis with this more general specification would be qualitatively similar as long as either $\lambda > 0$ or $\delta < 1$. We therefore focus on the special case where $\lambda = \delta = 0$.

efficient bargaining. In all the sections below we use Nash bargaining to solve for optimal conditions (see Nash 1950, Nash 1953, Kalai 1977, Osborne and Rubinstein 1990).[4]

1.2.1 Activity bargaining

In the first scenario, we assume that first the purchaser chooses the price, then the purchaser and the provider bargain on activity.[5] For a given price p, the bargained activity can be determined by solving:

$$\max_{y} \left[B(y) - py - \overline{V}\right]^{\gamma} \left[py - C(y) - \overline{U}\right]^{1-\gamma} \tag{1.1}$$

The First Order Condition (FOC) is:

$$y^a: \quad \frac{\gamma}{\widetilde{V}}\left(B_y - p\right) = \frac{1-\gamma}{\widetilde{U}}\left(C_y - p\right) \tag{1.2}$$

Proof. The result is obtained by differentiating $\gamma \log\left[B(y) - py - \overline{V}\right] +$

$(1-\gamma)\log\left[py - C(y) - \overline{U}\right]$ with respect to y. The Second Order Condition is

$\Gamma = \gamma \frac{B_{yy}V - (B_y - p)^2}{\widetilde{V}^2} - (1-\gamma)\frac{C_{yy}\widetilde{U} + (p - C_y)^2}{\widetilde{U}^2} < 0$, which is always satisfied. ■

To interpret the optimal condition on the bargained activity it is useful to distinguish two cases, low price and high price (see Figure 1.1). 1) If the exogenous price p is low ($B_y(y^a) > p$ and $C_y(y^a) > p$), the desired activity for the purchaser is *higher* than the desired activity for the provider. The bargained activity lies somewhere between the desired activity of the two parties. The LHS of Eq.(1.2) is the net marginal benefit of activity

---

[4] The Nash bargaining solution has been used extensively in labour economics to examine negotiations between trade unions and firms with respect to wages and employment. See, for example, Oswald (1985) for a survey of the literature, and (and Manning 1987, McDonald and Solow 1981, Sampson 1993, Bulkley and Myles 1997).

[5] A different interpretation is that the Department of Health fixes the price, then the Health Authority and provider bargain on activity. The implicit assumption is that the Department of Health and the Health Authority share the same objective function.

for the purchaser, weighted by her utility and her bargaining power. The RHS is the net marginal cost for the provider, also weighted by his utility and his bargaining power. 2) If the exogenous price p is high ($p > B_y(y^a)$ and $p > C_y(y^a)$), the desired activity for the purchaser is *lower* than the desired activity for the provider. The FOC can be rewritten as $\frac{\gamma}{\widetilde{V}}(p - B_y) = \frac{1-\gamma}{\widetilde{U}}(p - C_y)$. Again, the bargained activity lies between the desired activity of the two parties.

Figure 1.1 illustrates different bargained activity levels ($y^a(p)$) for three different values of the bargaining power of the purchaser, equal to 0.3, 0.5 and 0.7 respectively. In equilibrium it is always the case that $\widetilde{U} \geq 0$ and $\widetilde{V} \geq 0$, so that the equilibrium lies in the area between the average and marginal benefit, and the area between the average and marginal cost.

Finally if $p = B_y(y^a) = C_y(y^a)$ (i.e. where the marginal benefit curve crosses the marginal cost curve), there is no disagreement between purchaser and provider, so that y^a is such that $B_y = C_y$.

By differentiating Eq.(1.2) with respect to γ we obtain $\frac{\partial y^a}{\partial \gamma} = \frac{(B_y - p)\widetilde{U} - (p - C_y)\widetilde{V}}{\widetilde{V}\widetilde{U}(-\Gamma)}$. If the price is low, a higher bargaining power of the purchaser increases activity ($\frac{\partial y^a}{\partial \gamma} > 0$). If the price is high it reduces activity ($\frac{\partial y^a}{\partial \gamma} < 0$).

The effect of a change of price on activity is:

$$\frac{\partial y^a}{\partial p} = \frac{1}{-\Gamma}\left((1-\gamma)\frac{C_y - \frac{C}{y}}{\widetilde{U}^2} - \gamma\frac{\frac{B}{y} - B_y}{\widetilde{V}^2}\right) \tag{1.3}$$

which in general is indeterminate. According to our assumptions, it is always the case that $C_y > \frac{C}{y}$ and $\frac{B}{y} > B_y$, since the marginal cost is higher than the average cost, and the

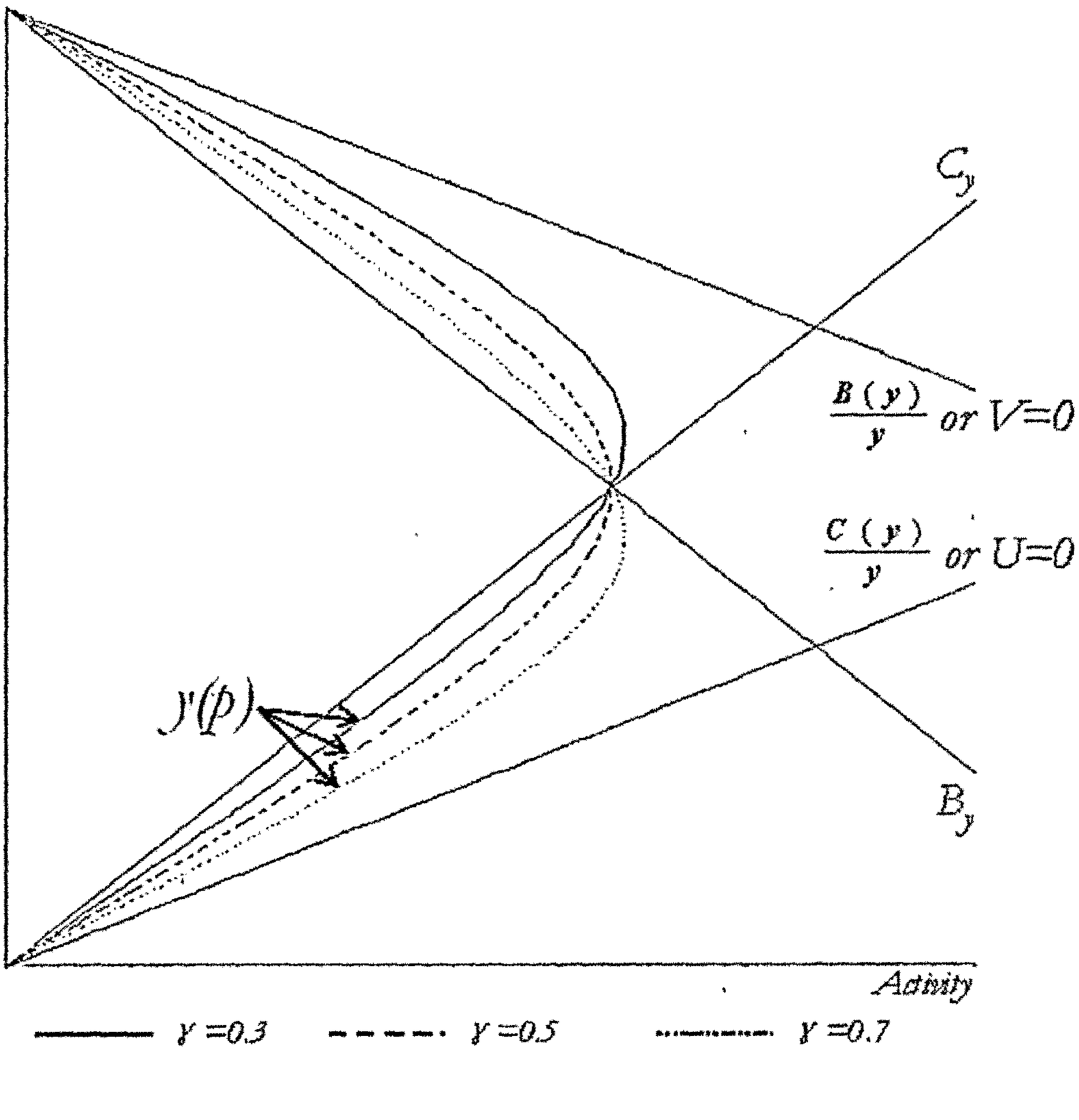


Fig. 1.1. Activity bargaining

average benefit is higher than the marginal benefit. For low levels of p the provider utility $\widetilde{U}$ is low (and the purchaser utility $\widetilde{V}$ is high) so that $\frac{\partial y^a}{\partial p} > 0$. In this case, the provider is interested in increasing activity because the equilibrium is below the marginal cost curve. The purchaser will also be interested in increasing activity since this brings the equilibrium closer to the marginal benefit curve. On the other hand, for high levels of p the purchaser utility $\widetilde{V}$ is low (and the provider utility $\widetilde{U}$ is high) so that $\frac{\partial y^a}{\partial p} < 0$ for low p. This result is consistent with the example shown in Figure 1.1.

The above analysis holds for a given price. The purchaser chooses the price to maximize:

$$\max_p B(y^a(p)) - py^a(p) \tag{1.4}$$

The FOC is:

$$p^a: \quad B_y y_p = y + py_p \tag{1.5}$$

The optimal price is determined such that the marginal benefit of higher activity equals the marginal cost. The SOC is: $B_{yy}y_p^2 + B_y y_{pp} - 2y_p - py_{pp}$.

1.2.2 Price bargaining

In the second scenario, we assume that first the purchaser and the provider bargain on price, then the activity is chosen unilaterally by the provider.[6] By backward induction, for a given price the hospital chooses the level of activity, which maximises $U = py - C(y)$, leading

[6] This setup is analogous to the model of bargaining between a firm and a union over wage and employment (McDonald and Solow, 1981; Manning, 1987), where the firm sets the employment, but the wage is bargained with the union.

to the FOC:

$$y^p: \quad p = C_y \tag{1.6}$$

with $\frac{\partial y^p}{\partial p} = \frac{1}{C_{yy}} > 0$ and $\frac{\partial^2 y^p}{\partial p^2} = 0$ (the SOC is $-C_{yy} < 0$). The bargained price can be determined by solving:

$$\max_p \left[B(y^p(p)) - py^p(p) - \overline{V}\right]^{\gamma} \left[py^p(p) - C(y^p(p)) - \overline{U}\right]^{1-\gamma} \tag{1.7}$$

Thanks to the envelope theorem, $U_p = y^p(p)$. The FOC for the bargained price is:

$$p^p: \quad \frac{\gamma}{\widetilde{V}} B_y y_p + \frac{(1-\gamma)}{\widetilde{U}} y = \gamma \frac{y + py_p}{\widetilde{V}} \tag{1.8}$$

Proof. By taking the log and differentiating with respect to p we obtain $\gamma \frac{B_y y_p - y(p) - py_p}{\widetilde{V}} + (1-\gamma) \frac{y(p) + py_p - C_y y_p}{\widetilde{U}} = 0$. From the FOC of the provider we know that $p = C_y$. By simplifying, we obtain: $\gamma \frac{B_y y_p - y(p) - py_p}{\widetilde{V}} + (1-\gamma) \frac{y(p)}{\widetilde{U}} = 0$. The SOC is $-\frac{\gamma \widetilde{U}^2 ((B_y - p) y_p - y)^2 + (1-\gamma) \widetilde{V}^2 y^2}{\widetilde{V}^2 \widetilde{U}^2} - \frac{\gamma \left(2 - \frac{B_{yy}}{C_{yy}}\right) \widetilde{U} - (1-\gamma) \widetilde{V}}{\widetilde{V} \widetilde{U}} y_p$. ■

The LHS of Eq.(1.8) is the marginal benefit of a higher price, and includes the marginal benefit for the purchaser of higher activity (weighted by her bargaining power, her utility and the responsiveness of supply) and the marginal benefit for the provider of a higher surplus (also weighted by his bargaining power and utility). The RHS is the marginal cost for the purchaser of a higher price and overall transfer (also weighted).

If the purchaser holds all the bargaining power ($\gamma = 1$), the optimal price is such that: $B_y y_p = y + py_p$. If the provider holds all the bargaining power, the optimal price is the highest possible compatible with the purchaser having a non-negative utility. The bargained price is an intermediate level between these two extremes.

1.2.3 Efficient bargaining

In the third scenario, purchaser and provider bargain simultaneously on activity *and* price.

This setting is called *efficient* bargaining, because it reduces the potential for unexplored opportunities from mutual gain.[7] The bargaining problem is:

$$\max_{p,y} \left[B(y) - py - \overline{V}\right]^{\gamma} \left[py - C(y) - \overline{U}\right]^{1-\gamma} \tag{1.9}$$

After obtaining the FOCs and rearranging, we obtain:

$$y^e : \; B_y = C_y \tag{1.10}$$

$$p^e = (1-\gamma)\frac{B(y^e) - \overline{V}}{y^e} + \gamma\frac{C(y^e) + \overline{U}}{y^e} \tag{1.11}$$

Proof. Define $\Omega = \left[B(y) - py - \overline{V}\right]^{\gamma} \left[py - C(y) - \overline{U}\right]^{1-\gamma}$. Then: $\frac{\partial \log \Omega}{\partial p} = -\frac{\gamma y}{B(y)-py-\overline{V}} + \frac{(1-\gamma)y}{py-C(y)-\overline{U}} = 0$ and $\frac{\partial \log \Omega}{\partial y} = \frac{\gamma(B_y-p)}{B(y)-py-\overline{V}} + \frac{(1-\gamma)(p-C_y)}{py-C(y)-\overline{U}} = 0$. From the first equation we obtain $p = \frac{\gamma\left[C(y)+\overline{U}\right]+(1-\gamma)\left[B(y)-\overline{V}\right]}{y}$, which, substituted into the second one, yields: $B_y = C_y$. The SOCs are: $\frac{\partial^2 \log \Omega}{\partial p^2} = -y^2\left(\frac{\gamma}{\bar{V}^2} + \frac{1-\gamma}{\bar{U}^2}\right) < 0$, $\frac{\partial^2 \log \Omega}{\partial y^2} = \gamma\frac{B_{yy}\bar{V} - (B_y-p)^2}{\bar{V}^2} - (1-\gamma)\frac{C_{yy}\bar{U} + (p-C_y)^2}{\bar{U}^2} < 0$, and $\frac{\partial^2 \log \Omega}{\partial p^2}\frac{\partial^2 \log \Omega}{\partial y^2} > \left(\frac{\partial^2 \log \Omega}{\partial p \partial y}\right)^2$. $\frac{\partial^2 \log \Omega}{\partial p \partial y} = -\frac{\gamma}{\bar{V}} + \frac{1-\gamma}{\bar{U}} + y(B_y-p)\left(\frac{\gamma}{\bar{V}^2} + \frac{1-\gamma}{\bar{U}^2}\right) = \frac{(1-\gamma)B+\gamma C-py}{\bar{V}\bar{U}} + y(B_y-p)\left(\frac{\gamma}{\bar{V}^2} + \frac{1-\gamma}{\bar{U}^2}\right) = y(B_y-p)\left(\frac{\gamma}{\bar{V}^2} + \frac{1-\gamma}{\bar{U}^2}\right)$, where the last simplification follows from the FOC for price. $\frac{\partial^2 \log \Omega}{\partial p^2}\frac{\partial^2 \log \Omega}{\partial y^2} > \left(\frac{\partial^2 \log \Omega}{\partial p \partial y}\right)^2 = -y^2\left(\frac{\gamma}{\bar{V}^2} + \frac{1-\gamma}{\bar{U}^2}\right)\left[\gamma\frac{B_{yy}\bar{V}}{\bar{V}^2} - (1-\gamma)\frac{C_{yy}\bar{U}}{\bar{U}^2} - (B_y-p)^2\left(\frac{\gamma}{\bar{V}^2} + \frac{1-\gamma}{\bar{U}^2}\right)\right]$

$-y^2(B_y-p)^2\left(\frac{\gamma}{\bar{V}^2} + \frac{1-\gamma}{\bar{U}^2}\right)^2 = -\gamma\frac{B_{yy}\tilde{V}}{\bar{V}^2} + (1-\gamma)\frac{C_{yy}\tilde{U}}{\bar{U}^2} > 0$. All three SOCs are always satisfied, since $B_{yy} \leq 0$. ■

---

[7] The outcome achieved in price bargaining is not efficient. As remarked by Aronsson, Lofgren and Wikstrom (1993), "there are unexplored profits and/or utility gains from bargaining".

The negotiated level of activity maximises the sum of the surplus for the purchaser and for the provider $U + V = B(y) - C(y)$. In this respect the level of activity is *efficient*. The optimal price is a weighted average of the average cost of the provider and the average welfare of the purchaser.[8]

1.3 Regime comparison

1.3.1 Constant marginal benefit

To gain some insights into how the different scenarios relate to each other, we consider the following functional forms: a) the benefit function is linear in activity: $B(y) = ay$; b) the cost function is quadratic: $C(y) = \frac{c}{2}y^2$ with $C_y = cy$; c) the outside options are normalised to zero ($\overline{V} = \overline{U} = 0$).

The equilibrium for the three scenarios is reported in Table 1.1. Proofs are in the appendix.

Table 1.1. Equilibrium with constant marginal benefit

Activity bargaining	Price bargaining	Efficient bargaining
$p^a = \frac{a}{2}$	$p^p = \frac{a(2-\gamma)}{2}$	$p^e = \frac{a(2-\gamma)}{2}$
$y^a = \frac{a}{c(2-\gamma)}$	$y^p = \frac{a(2-\gamma)}{2c}$	$y^e = \frac{a}{c}$
$V^a = \frac{a^2}{2c(2-\gamma)}$	$V^p = \gamma\frac{a^2(2-\gamma)}{4c}$	$V^e = \frac{\gamma a^2}{2c}$
$U^a = \frac{a^2(1-\gamma)}{2c(2-\gamma)^2}$	$U^p = \frac{a^2(2-\gamma)^2}{8c}$	$U^e = \frac{a^2(1-\gamma)}{2c}$

[8] This result is in line with the model of employment-wage bargaining analysed by Manning (1987) in the context of firm-union negotiations. The level of employment does not depend on the payoffs of firm and union. Consequently, they "can agree on this level and then bargain about the distribution of the rents" (Manning, 1987, p.131).

The following proposition compares prices, activity and utility under different regimes.

Proposition 1 *(a)* $p^e = p^p \geq p^a$*; (b)* $y^e \geq \{y^p; y^a\}, y^p \geq y^a$ *if* $\gamma \leq 0.59$*; (c)* $V^a \geq V^e \geq V^p$*; (d)* $U^p \geq U^e \geq U^a$.

Proof. (a) $p^p = \frac{a(2-\gamma)}{2} \geq \frac{a}{2} = p^a$ if $\gamma \leq 1$. (b) $y^a = \frac{a}{c(2-\gamma)} \leq y^e = \frac{a}{c}$ if $\frac{a}{c(2-\gamma)} \leq \frac{a}{c}$ or $\gamma \leq 1$; $y^p = \frac{a(2-\gamma)}{2c} \leq y^e = \frac{a}{c}$ if $\gamma \geq 0$; $y^p = \frac{a(2-\gamma)}{2c} \geq y^a = \frac{a}{c(2-\gamma)}$ if $(2-\gamma)^2 \geq 2$ or $\gamma \leq 0.59$. (c) $V^a = \frac{a^2}{2c(2-\gamma)} \geq V^e = \frac{\gamma a^2}{2c}$ if $2\gamma - \gamma^2 - 1 \leq 0$ or $-(\gamma-1)^2 \leq 0$; $V^e = \frac{\gamma a^2}{2c} \geq V^p = \frac{\gamma a^2(2-\gamma)}{4c}$ if $\gamma \geq 0$. (d) $U^p = \frac{a^2(2-\gamma)^2}{8c} \geq U^e = \frac{a^2}{2c}(1-\gamma)$ if $\frac{(2-\gamma)^2}{4} \geq (1-\gamma)$ or $4 + \gamma^2 - 4\gamma \geq 4 - 4\gamma$, or if $\gamma^2 > 0$; $U^e = \frac{a^2}{2c}(1-\gamma) \geq U^a = \frac{a^2}{2c}\frac{1-\gamma}{(2-\gamma)^2}$ if $(2-\gamma)^2 \geq 1$, which is always the case, since $0 \leq \gamma \leq 1$. ■

The price in *efficient* bargaining is equal to the price in *price* bargaining, which is higher than or equal to the price in *activity* bargaining. The activity in *efficient* bargaining is the highest. The activity in *price* bargaining is higher than in *activity* bargaining when the bargaining power of the purchaser is below 0.59.

The purchaser weakly prefers *activity* bargaining to *efficient* bargaining, and *efficient* bargaining to *price* bargaining. The provider weakly prefers *price* bargaining to *efficient* bargaining, and prefers *efficient* bargaining to *activity* bargaining.

In summary, the purchaser is better off in activity bargaining and the provider is better off in price bargaining. Activity is highest in efficient bargaining and prices are highest in efficient or price bargaining.

Figure 1.2 below displays the solution under different regimes. An arrow indicates increasing bargaining power of the purchaser. In *efficient* bargaining, a higher bargaining power of the purchaser reduces prices but has no effect on the level of activity. In *activity*

bargaining, higher bargaining power of the purchaser induces higher activity, but has no effect on prices. In *price* bargaining, higher bargaining power of the purchaser reduces both prices and activity.

The solution in price bargaining, where the purchaser holds all the bargaining power, coincides with the solution in activity bargaining, where the provider has all the bargaining power (point A). The solutions in price and efficient bargaining coincide when the provider holds all the bargaining power (point B). The solutions in activity and efficient bargaining coincide when the purchaser holds all the bargaining power (point C). Finally, the activity in *price* bargaining is higher than in *activity* bargaining only for low bargaining power of the purchaser.

Figure 1.2 also compares the solution when both parties have the same bargaining power ($\gamma = 0.5$). Prices are higher in *efficient* and *price* bargaining (points $E^{\gamma=0.5}$ and $P^{\gamma=0.5}$ respectively). Activity is highest in efficient bargaining and lowest in activity bargaining (point $A^{\gamma=0.5}$).

1.3.2 Decreasing marginal benefit

We extend the previous analysis, and assume a more general specification of the benefit function: $B(y) = ay - \frac{b}{2}y^2$, with decreasing marginal benefit, while we maintain the other assumptions: $C(y) = \frac{c}{2}y^2$, $\overline{V} = \overline{U} = 0$. Table 1.2 reports the solution in *price* and *efficient* bargaining. Proofs are in the appendix. The solution for *activity* bargaining is more involved, and is derived separately in section 3.2.1.

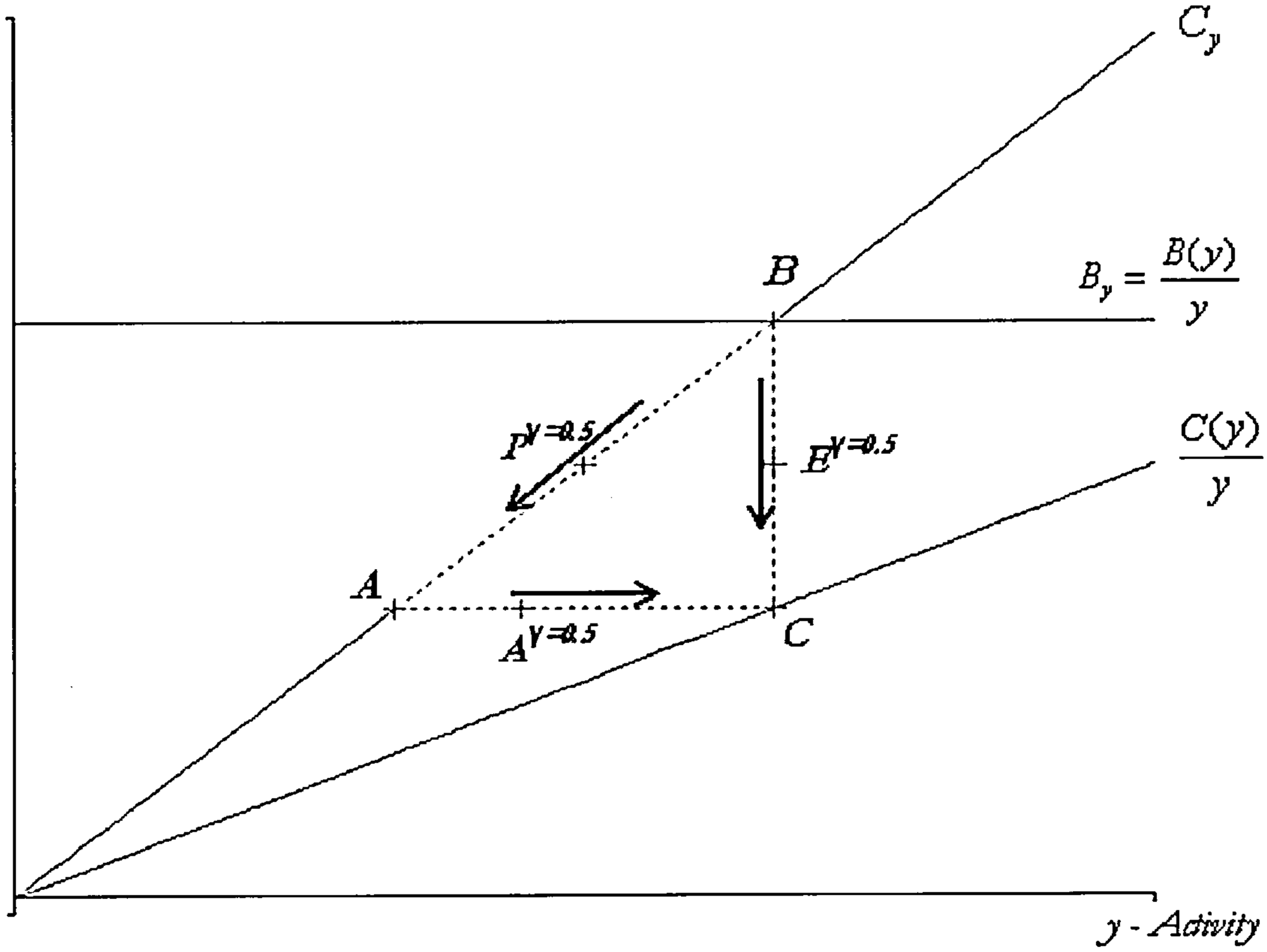


Fig. 1.2. Comparison of scenarios with constant marginal benefit

Table 1.2. Equilibrium with decreasing marginal benefit

Price bargaining	Efficient bargaining
$p^p = \frac{ac(2-\gamma)}{b+2c}$	$p^e = \frac{a((1-\gamma)b+(2-\gamma)c)}{2(b+c)}$
$y^p = \frac{a(2-\gamma)}{b+2c}$	$y^e = \frac{a}{b+c}$
$V^p = \frac{\gamma a^2(2-\gamma)}{2(b+2c)}$	$V^e = \frac{\gamma a^2}{2(b+c)}$
$U^p = \frac{a^2c(2-\gamma)^2}{2(b+2c)^2}$	$U^e = \frac{(1-\gamma)a^2}{2(b+c)}$

The following proposition compares the two regimes.

Proposition 2 *If* $\gamma > \frac{b}{b+c}$, *then (a)* $p^p > p^e$, *(b)* $y^e > y^p$, *(c)* $U^p \gtrless U^e$, *(d)* $V^e > V^p$.

Proof. (a) $p^p > p^e$ if $\frac{ac(2-\gamma)}{b+2c} > \frac{a((1-\gamma)b+(2-\gamma)c)}{2(b+c)}$ or $b(c\gamma + b\gamma - b) > 0$ or $\gamma > \frac{b}{b+c}$. (b) $y^e > y^p$ if $\frac{a}{b+c} > \frac{a(2-\gamma)}{b+2c}$ or $b + 2c - (2-\gamma)(b+c) > 0$ or $\gamma > \frac{b}{b+c}$. (c) $U^p > U^e$ if $\frac{a^2c(2-\gamma)^2}{2(b+2c)^2} > \frac{(1-\gamma)a^2}{2(b+c)}$ or $b^2\gamma + bc\gamma^2 + c^2\gamma^2 - b^2 > 0$ or $\gamma = \left\{-\frac{b}{c}, \frac{b}{b+c}\right\}$. (d) $V^e > V^p$ if $\frac{\gamma a^2}{2(b+c)} > \frac{\gamma a^2(2-\gamma)}{2(b+2c)}$ or $(b + 2c) - (b + c)(2 - \gamma) > 0$ or $\gamma > \frac{b}{b+c}$. ■

If the bargaining power of the purchaser is sufficiently high ($\gamma > \frac{b}{b+c}$) prices are higher in *price* bargaining, activity is lower, the provider is better off and the purchaser is worse off. If the bargaining power of the purchaser is sufficiently low ($\gamma < \frac{b}{b+c}$) all the results are reversed. The threshold $\frac{b}{b+c}$ increases with b and decreases with c. Note that if $b = 0$ we are back to the results of proposition 1. Therefore, if the purchaser has low bargaining power, *efficient* bargaining yields a lower utility for the purchaser than in *price* bargaining. This is a surprising result: we would expect the purchaser to be better off when she can bargain with more instruments, ie both prices and activity. But this holds true only if her bargaining power is high. If her bargaining power is low, having more instruments is counterproductive. The purchaser is better off when she cannot bargain on activity.

Figure 1.3 below displays the solution under the two regimes. The solutions in efficient and price bargaining are depicted by line BC and AD respectively. An arrow indicates increasing bargaining power of the purchaser. As before, in *efficient* bargaining activity is constant, irrespective of the distribution of bargaining power, and the price decreases as the

bargaining power of the purchaser increases. In *price* bargaining, both prices and activity decrease as the bargaining power of the purchaser increases.

It is useful to compare these results with those obtained in the previous section by assuming constant marginal benefit. When the bargaining power of the purchaser is low, the activity in *efficient* bargaining is lower than in *price* bargaining but with constant marginal benefit it is always higher. Furthermore, the prices under the two regimes differ, but are identical with constant marginal benefit.

Decreasing marginal benefit and activity bargaining

In this section we derive the solution in activity bargaining. For a given price, the optimal bargained activity is:

$$y^a(p) = \frac{\left(\frac{2-\gamma}{2}c(a-p) + bp\frac{1+\gamma}{2}\right) - \sqrt{\left(\frac{2-\gamma}{2}c(a-p) + bp\frac{1+\gamma}{2}\right)^2 - 2bcp(a-p)}}{bc} \quad (1.12)$$

See the appendix for the proof. The optimal price is given by the price which maximises $V = ay^a(p) - \frac{b}{2}y^a(p)^2 - py^a(p)$. Given the complexity of the solution, it is not possible to derive manageable expressions for price and activity. To compare the solutions for the three scenarios we resort to numerical simulations. Our strategy is to specify a grid of values for all the parameters of the model (a, b, c and γ), and compute the solution numerically. We fix $a = 1$, and specify a grid for $b \in \{0, 0.5, 1, 1.5, ..., 30\}$, $c = \{0, 0.5, 1, 1.5, ..., 30\}$ and $\gamma = \{0, 0.1, ..., 0.9, 1\}$.

For example, supposing that $a = b = c = 1$ and $\gamma = 0.5$, then $y^a(p) = \frac{3}{4} - \sqrt{(p-1)2p + \frac{9}{16}}$ and $V = (1-p)\left(\frac{3}{4} - \sqrt{(p-1)2p + \frac{9}{16}}\right) - \frac{1}{2}\left(\frac{3}{4} - \sqrt{(p-1)2p + \frac{9}{16}}\right)^2$,

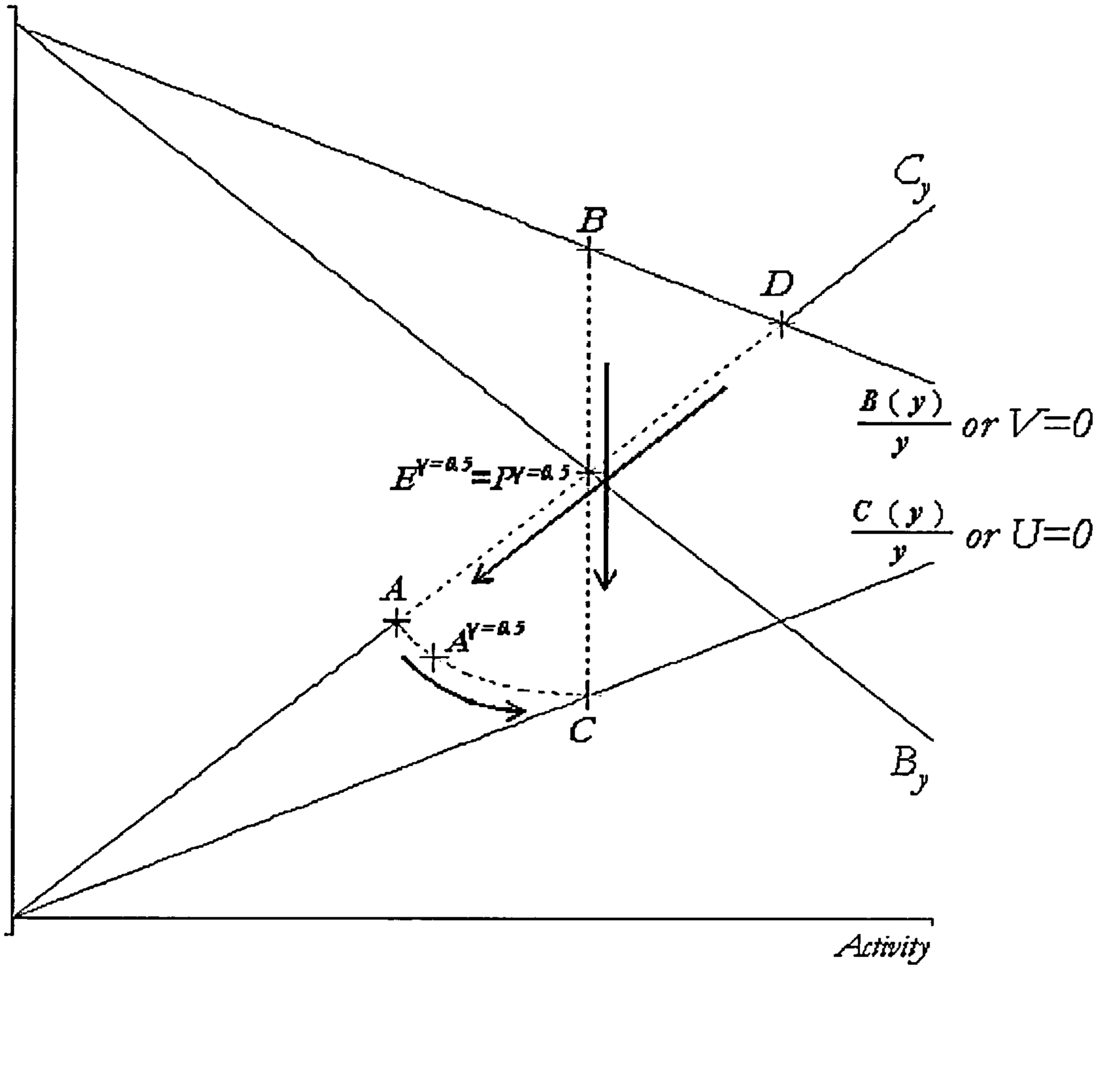


Fig. 1.3. Comparison of scenarios with decreasing marginal benefit

the solution of which is $p^a = 0.29$ and $y^a = 0.36$. Table 1.3 reports the solution for $a = b = c = 1$ and $\gamma = \{0, 0.1, 0.25, 0.5, 0.75, 0.9, 1\}$. The tables for the other values of b and c are reported in the appendix.

Overall, the numerical simulations suggest that in *activity* bargaining prices are lowest, the purchaser's utility is highest and the provider's utility is lowest (note the similarity with proposition 1). Activity is lower than in *efficient* bargaining. It is lower than in *price* bargaining when the bargaining power of the purchaser is below a certain threshold, which is between 0.7 and 0.95 in our simulations.

The solution in *activity* bargaining is displayed in Figure 1.3, on the line AC which was derived by plotting the numerical solution a thousand times. In contrast to the solution with constant marginal benefit, in *activity* bargaining the price is not fixed any longer. As the bargaining power of the purchaser increases, the price decreases and activity increases.

As in the previous section, the solution in *price* bargaining with $\gamma = 1$ coincides with *activity* bargaining when $\gamma = 0$ (point A), and the solution in *activity* and *efficient* bargaining coincide when $\gamma = 1$ (point C). However, when $\gamma = 0$ (points B and D) *efficient* and *price* bargaining yield different solutions. Finally, when both parties have the same bargaining power, the solutions in *efficient* bargaining and *price* bargaining coincide at the point where marginal cost equals marginal benefit.

Finally, in *price* bargaining an increase in the bargaining power of the purchaser reduces prices and activity, but in *activity* bargaining it reduces prices but *increases* activity.

Table 1.3. Numerical simulation of equilibrium with decreasing marginal benefit

Simulation based on the parameters a=1,b=1,c=1

	$\gamma = 0$	$\gamma = 0.1$	$\gamma = 0.25$	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 0.9$	$\gamma = 1$
y^a	0.33	0.33	0.34	0.36	0.40	0.44	0.50
y^e	0.50	0.50	0.50	0.50	0.50	0.50	0.50
y^p	0.67	0.63	0.58	0.50	0.42	0.37	0.33
p^a	0.33	0.32	0.31	0.29	0.27	0.25	0.25
p^e	0.75	0.70	0.63	0.50	0.38	0.30	0.25
p^p	0.67	0.63	0.58	0.50	0.42	0.37	0.33
V^a	0.17	0.17	0.18	0.19	0.21	0.23	0.25
V^e	0	0.03	0.06	0.13	0.19	0.23	0.25
V^p	0	0.03	0.07	0.13	0.16	0.17	0.17
U^a	0.06	0.05	0.05	0.04	0.03	0.02	0
U^e	0.25	0.23	0.19	0.13	0.06	0.03	0
U^p	0.22	0.20	0.17	0.13	0.09	0.07	0.06

1.4 Adding quality and effort

In this section we extend the model by introducing quality and cost containment effort, and we show that the results using this more general specification are qualitatively similar to the ones obtained above. We follow the approach suggested by Ma (1994) and Chalkley and Malcomson (1998b). Define q as the quality generated by the provider and e as the cost-containment effort. The cost function of the provider is $C(y,q,e) + \varphi(y,q,e)$. C includes the monetary cost, which increases with quality and activity but decreases with effort: $C(y,q,e)$, with $C_y > 0, C_q > 0$ and $C_e < 0$. φ is the non-monetary cost, or disutility, which increases with activity, quality and effort: $\varphi(y,q,e)$, with $\varphi_y > 0, \varphi_q > 0$ and $\varphi_e > 0$.

We also assume that the demand for treatment depends positively on quality so that $y = y(q)$ with $y_q > 0$. Of course, this assumption is only valid in case quality is an observable variable. This assumption implies $y = y(q) \Leftrightarrow q = q(y)$, $q_y > 0$. Therefore by contracting activity the purchaser can implicitly contract the level of quality. The benefit function of the patients is $B = B(y,q)$ with $B_y > 0$ and $B_q > 0$. Since quality is a positive

function of activity, we can also write $B = B(y, q(y))$ with $\frac{\partial B}{\partial y} = \frac{\partial B}{\partial y} + \frac{\partial B}{\partial q}\frac{\partial q}{\partial y} > 0$. The provider's utility is given by the surplus: $U = py - C(y, q(y), e) - \varphi(y, q(y), e)$. The purchaser's utility is $V = B(y, q(y)) - py$.

1.4.1 Activity bargaining

We assume that first, the purchaser sets the price; second, the purchaser and provider bargain on activity; third, the provider chooses effort. We solve by backward induction. For a given price and activity (stage three), the provider maximises the surplus U with respect to effort so that:

$$U_e(e^*) = 0 : -C_e(y, q(y), e^*) = \varphi_e(y, q(y), e^*) \tag{1.13}$$

The optimal effort for the provider $e^*(y)$ is such that the marginal benefit of lower cost is equal to the marginal disutility of effort. The indirect utility function of the provider is $U(p, y, q(y), e^*(y)) = py - C(y, q(y), e^*(y)) - \varphi(y, q(y), e^*(y))$.

For a given price (stage two), the activity bargaining problem between purchaser and provider is:

$$\max_y \left[V(p, y, q(y)) - \overline{V}\right]^{\gamma} \left[U(p, y, q(y), e^*(y)) - \overline{U}\right]^{1-\gamma} \tag{1.14}$$

whose FOC is:

$$y^a : \gamma \frac{B_y + B_q q_y - p}{\widetilde{V}} = (1-\gamma)\frac{C_y + \varphi_y + \left(C_q + \varphi_q\right) q_y - p}{\widetilde{U}} \tag{1.15}$$

The volume of activity is such that the difference between the marginal benefit and the price (weighted by the relevant factors) equals the difference between the marginal cost and the price (also weighted by the relevant factors). The condition is analogous to Eq.(2).

However, the marginal benefit and marginal cost also include the additional benefit and cost from higher quality. The marginal cost includes both the monetary and non-monetary cost.

In stage one the purchaser sets the price to maximise:

$$\max_p B(y^a(p), q(y^a(p))) - py^a(p) \tag{1.16}$$

The FOC is:

$$p^a: \quad B_y y_p + B_q q_y y_p = y + p y_p \tag{1.17}$$

The optimal price is such that the marginal benefit of higher activity and quality induced by a higher price is equal to the marginal cost.

1.4.2 Price bargaining

First the purchaser and the provider bargain on price, and then the provider chooses the level of activity and cost-containment effort. For a given price the provider maximises the surplus U with respect to activity and effort, so that:

$$U_y(y^*, e^*) = 0: \quad p = C_y + \varphi_y + \left(C_q + \varphi_q\right) q_y \tag{18}$$

$$U_e(y^*, e^*) = 0: \quad -C_e = \varphi_e \tag{19}$$

The provider chooses the level of activity which equates the price to the marginal monetary and non-monetary cost. The marginal cost also takes into account the indirect effect of activity caused by increased quality, which is captured by the last term on the RHS. The optimal effort is such that the marginal benefit of lower cost is equal to the marginal disutility of effort. The indirect utility function of the provider is $U(p, y^*(p), q(y^*(p)), e^*(p))$. Note that $\frac{\partial y^*}{\partial p} = \frac{-U_{ee}}{U_{yy}U_{ee}-U_{ye}^2} > 0$, $\frac{\partial e^*}{\partial p} = \frac{U_{ye}}{U_{yy}U_{ee}-U_{ye}^2} \gtrless 0$ and $\frac{\partial U}{\partial p} = y^*$ (by the envelope the-

orem). The price bargaining problem is given by:

$$\max_{p} \left[\begin{array}{c} B(y^*(p), q(y^*(p))) \\ -py^*(p) - \overline{V} \end{array} \right]^{\gamma} \left[\begin{array}{c} py^*(p) - C(y^*(p), q(y^*(p)), e^*(p)) \\ -\varphi(y^*(p), q(y^*(p)), e^*(p)) - \overline{U} \end{array} \right]^{1-\gamma} \tag{1.20}$$

The FOC is:

$$p^p : \frac{\gamma}{\widetilde{V}} \left(B_y + B_q q_y\right) y_p + \frac{(1-\gamma)}{\widetilde{U}} y = \frac{\gamma}{\widetilde{V}} \left(y + p y_p\right) \tag{1.21}$$

The optimal price is such that the weighted marginal benefit for the purchaser of higher activity and quality, plus the weighted marginal benefit for the provider in terms of higher surplus, is equal to the weighted marginal cost for the purchaser.

1.4.3 Semi-efficient bargaining

We now consider the scenario where the parties bargain on price and activity simultaneously. This is called semi-efficient bargaining because the effort variable is not bargained, but chosen by the provider. First the purchaser and the provider bargain on price and activity, then the provider chooses the cost-containment effort. For a given activity and price the supplier maximises the surplus U with respect to effort,

$$U_e\left(e^*\right) = 0 : \quad -C_e = \varphi_e \tag{1.22}$$

which provides $e^*(y)$. The bargaining problem is:

$$\max_{p,y} \left[B(y, q(y)) - py - \bar{V}\right]^{\gamma} \left[\begin{array}{c} py - C(y, q(y), e^*(y)) \\ -\varphi(y, q(y), e^*(y)) - \overline{U} \end{array} \right]^{1-\gamma} \tag{1.23}$$

whose FOCs are:

$$y^e : B_y + B_q q_y = C_y + \varphi_y + q_y \left(C_q + \varphi_q\right) \tag{1.24}$$

$$p^e = (1-\gamma) \frac{B - \overline{V}}{y} + \gamma \frac{C + \varphi + \overline{U}}{y} \tag{1.25}$$

The price equals the weighted sum of the average benefit to the purchaser and the average cost to the provider, which includes the non-monetary cost. The optimal activity balances the purchaser's marginal benefit with the provider's marginal cost.

1.5 Conclusions

Different countries have different institutional and bargaining settings for purchasers and providers. They usually follow one of three scenarios: the purchaser sets the price, but activity is bargained between purchaser and provider: *activity* bargaining; the price is bargained between purchaser and provider, but activity is chosen unilaterally by the provider: *price* bargaining; and price and activity are bargained simultaneously between purchaser and provider: *efficient* bargaining. We find that if the bargaining power of the purchaser is low, *efficient* bargaining leads to higher prices and provider's utility, and lower activity and purchaser's utility, compared to *price* bargaining. This result seems surprising, as one would expect the purchaser to be better off when she can bargain with more instruments, ie both price and activity. However, this intuition holds true only if the bargaining power of the purchaser is high. If her bargaining power is low, having more instruments is counterproductive. One policy implication is that purchasers with low bargaining power may be better off if restricted to bargaining on prices only, and not on price and activity. Future empirical work might quantify the bargaining power of the purchaser and the provider in health care markets. This might help governments to decide whether to encourage purchasers to bargain on prices only, or on price and activity simultaneously.

The analysis also confirms the intuition that if purchasers can set prices (*activity* bargaining), net consumer welfare (patient benefit, net of transfer to the provider) is highest. This result holds for any level of bargaining power of the purchaser. The analysis therefore supports policies such as "payment by results" in the UK, where prices are fixed by the purchaser or the regulator.

One less intuitive result is that by shifting from *efficient* and *price* bargaining (as in "cost and volume" or "sophisticated" contracts) to *activity* bargaining (as in "payment by results"), the level of activity is likely to decrease. More precisely, this study predicts that moving from efficient to activity bargaining will certainly reduce activity. This is in contrast to what is normally thought, i.e. that "payment by results" will encourage activity. When moving from *price* to *activity* bargaining, activity will decrease (increase) if the bargaining power of the purchaser is low (high). Further empirical work might test whether policies such as "payment by results" in the UK are likely to increase or decrease activity compared to previous policies.

Finally, most of the empirical work focuses on the effect of bargaining power on prices (Barros and Martinez-Giralt 2006). This study provides clear predictions of the effect of the bargaining power on activity as well as price. More precisely, under *price* bargaining a higher bargaining power of the purchaser reduces activity; under *activity* bargaining it increases activity; and under *efficient* bargaining it has no effect on activity. Further empirical work might test such predictions.

1.A Appendix

1.A.1 Constant marginal benefit

Activity bargaining. $p^a = \frac{a}{2}$, $y^a = \frac{a}{c(2-\gamma)}$, $V^a = \frac{a^2}{2c(2-\gamma)}$, $U^a = \frac{a^2(1-\gamma)}{2c(2-\gamma)^2}$.

Proof. The rule determining activity is, for a given price: $\gamma\frac{a-p}{(a-p)y} + (1-\gamma)\frac{p-cy}{\left(p-\frac{c}{2}y\right)y} = 0$, from which $y = \frac{2p}{c(2-\gamma)}$. The FOC for price is: $\frac{2a}{c(2-\gamma)} - \frac{4p}{c(2-\gamma)} = 0$, from which: $p^a = \frac{a}{2}$ (the SOC is $-\frac{4p}{c(2-\gamma)} < 0$). The bargained activity is therefore: $y^a = \frac{a}{c(2-\gamma)}$. The utility of the purchaser and the provider are: $V^a = (a-p)\,y = \frac{a^2}{2c(2-\gamma)}$ and $U^a = \left(p - \frac{c}{2}y\right)y = \frac{a^2(1-\gamma)}{2c(2-\gamma)^2}$.

■

Price bargaining. $p^p = \frac{a(2-\gamma)}{2}$, $y^p = \frac{a(2-\gamma)}{2c}$, $V^p = \frac{\gamma a^2(2-\gamma)}{4c}$, $U^p = \frac{a^2(2-\gamma)^2}{8c}$.

Proof. Since $y = \frac{p}{c}$ with $y_p = \frac{1}{c}$, the FOC for the bargained price is: $\gamma\frac{(a-p)\frac{1}{c} - \frac{p}{c}}{\frac{ap}{c} - \frac{p^2}{c}} + (1-\gamma)\frac{\frac{p}{c}}{\frac{p^2}{c} - \frac{p^2}{2c}} = 0$, which gives: $p^p = \frac{a(2-\gamma)}{2}$ (the SOC is $-\frac{1}{(a-p)^2p^2}\left((a-p)^2 + p^2\right) - \frac{2}{p^2}(1-\gamma) < 0$). Hence $y^p = \frac{a(2-\gamma)}{2c}$, $V^p = (a-p)\,y = \gamma\frac{a^2(2-\gamma)}{4c}$ and $U^p = \left(p - \frac{c}{2}y\right)y = \frac{a^2(2-\gamma)^2}{8c}$. ■

Efficient bargaining. $p^e = \frac{a(2-\gamma)}{2}$, $y^e = \frac{a}{c}$, $V^e = \frac{\gamma a^2}{2c}$, $U^e = \frac{a^2(1-\gamma)}{2c}$.

Proof. The FOC w.r.t. price implies: $p = (1-\gamma)\,a + \gamma\frac{c}{2}y$. The FOC w.r.t. activity implies: $y^e = \frac{a}{c}$. Therefore $p^e = \frac{a(2-\gamma)}{2}$ and $V^e = (a-p)\,y = \gamma\frac{a^2}{2c}$ and $U^e = \left(p - \frac{c}{2}y\right)y = (1-\gamma)\frac{a^2}{2c}$. ■

1.A.2 Decreasing marginal benefit

Price bargaining. $p^p = \frac{ac(2-\gamma)}{b+2c}$, $y^p = \frac{a(2-\gamma)}{b+2c}$, $V^p = \frac{\gamma a^2(2-\gamma)}{2(b+2c)}$, $U^p = \frac{a^2c(2-\gamma)^2}{2(b+2c)^2}$.

Proof. Since $y = \frac{p}{c}$ with $y_p = \frac{1}{c}$, the FOC for the bargained price is: $\gamma \frac{\left(a - b\frac{p}{c} - p\right)\frac{1}{c} - \frac{p}{c}}{\left(a\frac{p}{c} - \frac{b}{2}\frac{p^2}{c^2} - \frac{p^2}{c}\right)} + (1-\gamma)\frac{\frac{p}{c}}{\frac{p^2}{c} - \frac{p^2}{2c}} = 0$, which simplifies to $\gamma \frac{\left(a - b\frac{p}{c} - p\right) - p}{\left(a - \frac{b}{2}\frac{p}{c} - p\right)} + 2(1-\gamma) = 0$ or $\gamma\left(a - b\frac{p}{c} - p\right) - \gamma p + 2(1-\gamma)\left(a - \frac{b}{2c}p - p\right) = 0$, giving: $p^p = \frac{ac(2-\gamma)}{b+2c}$. Hence $y^p = \frac{a(2-\gamma)}{b+2c}$, $V^p = \left(a - \frac{b}{2}y - p\right)y = \frac{\gamma a^2(2-\gamma)}{2(b+2c)}$ and $U^p = \left(p - \frac{c}{2}y\right)y = \frac{a^2c(2-\gamma)^2}{2(b+2c)^2}$. ■

Efficient bargaining. $p^e = \frac{a((1-\gamma)b + (2-\gamma)c)}{2(b+c)}$, $y^e = \frac{a}{b+c}$, $V^e = \frac{\gamma a^2}{2(b+c)}$, $U^e = \frac{a^2(1-\gamma)}{2(b+c)}$.

Proof. The FOC w.r.t. price implies: $p^e = (1-\gamma)\left(a - \frac{b}{2}y\right) + \gamma\frac{cy}{2}$. The FOC w.r.t. activity implies: $y^e = \frac{a}{b+c}$. Therefore $p^e = \frac{a((1-\gamma)b + (2-\gamma)c)}{2(b+c)}$ and $V^e = \left(a - \frac{b}{2}y - p\right)y = \frac{\gamma a^2}{2(b+c)}$ and $U^e = \left(p - \frac{c}{2}y\right)y = \frac{a^2(1-\gamma)}{2(b+c)}$. ■

Activity bargaining

Proof. From FOC w.r.t. y we have $\gamma\frac{(a-by)-p}{ay - \frac{b}{2}y^2 - py} = -(1-\gamma)\frac{p-cy}{py - \frac{c}{2}y^2}$ or $\gamma(a - by - p)\left(p - \frac{c}{2}y\right) + (1-\gamma)(p-cy)\left(a - \frac{b}{2}y - p\right) = 0$. Upon expanding, we obtain $\frac{bc}{2}y^2 - y\left(c\frac{2-\gamma}{2}(a-p) + bp\frac{1+\gamma}{2}\right) + p(a-p) = 0$, with solution $y = \frac{\left(c\frac{2-\gamma}{2}(a-p) + bp\frac{1+\gamma}{2}\right) - \sqrt{\left(c\frac{2-\gamma}{2}(a-p) + bp\frac{1+\gamma}{2}\right)^2 - 4\frac{bc}{2}p(a-p)}}{bc}$. ■

1.A.3 Tables of simulations with decreasing marginal benefit

1. $a = c = 1$, different b

Simulation with $a = c = 1, b = 0.5$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.40	0.41	0.41	0.45	0.46	0.48	0.55	0.59	0.67
y^e	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67
y^p	0.80	0.76	0.72	0.64	0.60	0.56	0.48	0.44	0.40
p^a	0.40	0.39	0.38	0.37	0.36	0.35	0.34	0.33	0.33
p^e	0.83	0.78	0.73	0.63	0.58	0.53	0.43	0.38	0.33
p^p	0.80	0.76	0.72	0.64	0.60	0.56	0.48	0.44	0.40
V^a	0.20	0.21	0.21	0.23	0.24	0.25	0.28	0.31	0.33
V^e	0.00	0.03	0.07	0.13	0.17	0.20	0.27	0.30	0.33
V^p	0.00	0.04	0.07	0.13	0.15	0.17	0.19	0.20	0.20
U^a	0.08	0.08	0.07	0.07	0.06	0.05	0.04	0.02	0.00
U^e	0.33	0.30	0.27	0.20	0.17	0.13	0.07	0.03	0.00
U^p	0.32	0.29	0.26	0.20	0.18	0.16	0.12	0.10	0.08

Simulation with $a = c = 1, b = 1$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.33	0.33	0.34	0.36	0.36	0.37	0.41	0.44	0.50
y^e	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
y^p	0.67	0.63	0.60	0.53	0.50	0.47	0.40	0.37	0.33
p^a	0.33	0.32	0.31	0.30	0.29	0.28	0.26	0.25	0.25
p^e	0.75	0.70	0.65	0.55	0.50	0.45	0.35	0.30	0.25
p^p	0.67	0.63	0.60	0.53	0.50	0.47	0.40	0.37	0.33
V^a	0.17	0.17	0.17	0.19	0.19	0.20	0.22	0.23	0.25
V^e	0.00	0.03	0.05	0.10	0.13	0.15	0.20	0.23	0.25
V^p	0.00	0.03	0.06	0.11	0.13	0.14	0.16	0.17	0.17
U^a	0.06	0.05	0.05	0.04	0.04	0.04	0.02	0.02	0.00
U^e	0.25	0.23	0.20	0.15	0.13	0.10	0.05	0.03	0.00
U^p	0.22	0.20	0.18	0.14	0.13	0.11	0.08	0.07	0.06

Simulation with $a = c = 1, b = 1.5$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.28	0.28	0.29	0.29	0.30	0.31	0.33	0.35	0.40
y^e	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
y^p	0.57	0.54	0.51	0.46	0.43	0.40	0.34	0.31	0.29
p^a	0.28	0.28	0.27	0.25	0.24	0.24	0.22	0.21	0.20
p^e	0.7	0.65	0.60	0.50	0.45	0.40	0.30	0.25	0.20
p^p	0.57	0.54	0.51	0.46	0.43	0.40	0.34	0.31	0.29
V^a	0.14	0.15	0.15	0.15	0.16	0.16	0.18	0.19	0.20
V^e	0.00	0.02	0.04	0.08	0.10	0.12	0.16	0.18	0.2
V^p	0.00	0.03	0.05	0.09	0.11	0.12	0.14	0.14	0.14
U^a	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.01	0.00
U^e	0.20	0.18	0.16	0.12	0.10	0.08	0.04	0.02	0.00
U^p	0.16	0.15	0.13	0.10	0.09	0.08	0.06	0.05	0.04

Simulation with $a = c = 1, b = 2$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.25	0.25	0.25	0.25	0.26	0.26	0.28	0.29	0.33
y^e	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
y^p	'0.50	0.48	0.45	0.40	0.38	0.35	0.30	0.28	0.25
p^a	0.25	0.24	0.23	0.22	0.21	0.20	0.19	0.18	0.17
p^e	0.67	0.62	0.57	0.47	0.42	0.37	0.27	0.22	0.17
p^p	0.50	0.48	0.45	0.40	0.38	0.35	0.30	0.28	0.25
V^a	0.12	0.13	0.13	0.13	0.14	0.14	0.15	0.16	0.17
V^e	0.00	0.02	0.03	0.07	0.08	0.10	0.13	0.15	0.17
V^p	0.00	0.02	0.05	0.08	0.09	0.11	0.12	0.12	0.13
U^a	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.00
U^e	0.17	0.15	0.13	0.10	0.08	0.07	0.03	0.02	0.00
U^p	0.13	0.11	0.10	0.08	0.07	0.06	0.05	0.04	0.03

Simulation with $a = c = 1, b = 2.5$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.22	0.22	0.22	0.22	0.22	0.23	0.24	0.25	0.29
y^e	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29
y^p	0.44	0.42	0.40	0.36	0.33	0.31	0.27	0.24	0.22
p^a	0.22	0.21	0.21	0.20	0.19	0.18	0.16	0.15	0.14
p^e	0.64	0.59	0.54	0.44	0.39	0.34	0.24	0.19	0.14
p^p	0.44	0.42	0.40	0.36	0.33	0.31	0.27	0.24	0.22
V^a	0.11	0.11	0.11	0.12	0.12	0.12	0.13	0.13	0.14
V^e	0.00	0.01	0.03	0.06	0.07	0.09	0.11	0.13	0.14
V^p	0.00	0.02	0.04	0.07	0.08	0.09	0.11	0.11	0.11
U^a	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.00
U^e	0.14	0.13	0.11	0.09	0.07	0.06	0.03	0.01	0.00
U^p	0.10	0.09	0.08	0.06	0.06	0.05	0.04	0.03	0.02

Simulation with $a = c = 1, b = 3$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.20	0.20	0.20	0.20	0.20	0.20	0.21	0.22	0.25
y^e	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
y^p	0.40	0.38	0.36	0.32	0.30	0.28	0.24	0.22	0.20
p^a	0.20	0.19	0.19	0.18	0.17	0.16	0.15	0.13	0.12
p^e	0.63	0.58	0.53	0.43	0.38	0.33	0.23	0.18	0.13
p^p	0.40	0.38	0.36	0.32	0.30	0.28	0.24	0.22	0.20
V^a	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.12	0.12
V^e	0.00	0.01	0.03	0.05	0.06	0.08	0.10	0.11	0.13
V^p	0.00	0.02	0.04	0.06	0.08	0.08	0.10	0.10	0.10
U^a	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.00
U^e	0.13	0.11	0.10	0.08	0.06	0.05	0.03	0.01	0.00
U^p	0.08	0.07	0.06	0.05	0.05	0.04	0.03	0.02	0.02

Simulation with $a = c = 1, b = 5$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.17
y^e	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
y^p	0.29	0.27	0.26	0.23	0.21	0.20	0.17	0.16	0.14
p^a	0.14	0.14	0.13	0.13	0.12	0.12	0.10	0.09	0.08
p^e	0.58	0.53	0.48	0.38	0.33	0.28	0.18	0.13	0.08
p^p	0.29	0.27	0.26	0.23	0.21	0.20	0.17	0.16	0.14
V^a	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.08	0.08
V^e	0.00	0.01	0.02	0.03	0.04	0.05	0.07	0.08	0.08
V^p	0.00	0.01	0.03	0.05	0.05	0.06	0.07	0.07	0.07
U^a	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00
U^e	0.08	0.08	0.07	0.05	0.04	0.03	0.02	0.01	0.00
U^p	0.04	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01

Simulation with $a = c = 1, b = 30$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
y^e	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
y^p	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03
p^a	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02
p^e	0.52	0.47	0.42	0.32	0.27	0.22	0.12	0.07	0.02
p^p	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03
V^a	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
V^e	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.02
V^p	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.02
U^a	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U^e	0.02	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00
U^p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

2. $a = b = 1$, different c

Simulation with $a = b = 1, c = 0.25$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.67	0.66	0.66	0.66	0.66	0.66	0.68	0.71	0.80
y^e	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
y^p	1.33	1.27	1.20	1.07	1.00	0.93	0.80	0.73	0.67
p^a	0.17	0.16	0.16	0.15	0.14	0.13	0.12	0.11	0.10
p^e	0.60	0.55	0.50	0.40	0.35	0.30	0.20	0.15	0.10
p^p	0.33	0.32	0.30	0.27	0.25	0.23	0.20	0.18	0.17
V^a	0.33	0.34	0.34	0.34	0.35	0.35	0.37	0.38	0.40
V^e	0.00	0.04	0.08	0.16	0.20	0.24	0.32	0.36	0.40
V^p	0.00	0.06	0.12	0.21	0.25	0.28	0.32	0.33	0.33
U^a	0.06	0.05	0.05	0.04	0.04	0.03	0.02	0.02	0.00
U^e	0.40	0.36	0.32	0.24	0.20	0.16	0.08	0.04	0.00
U^p	0.22	0.20	0.18	0.14	0.13	0.11	0.08	0.07	0.06

Simulation with $a = b = 1, c = 0.5$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.50	0.50	0.50	0.51	0.52	0.52	0.56	0.59	0.67
y^e	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67
y^p	1.00	0.95	0.90	0.80	0.75	0.70	0.60	0.55	0.50
p^a	0.25	0.24	0.23	0.22	0.21	0.20	0.19	0.18	0.17
p^e	0.67	0.62	0.57	0.47	0.42	0.37	0.27	0.22	0.17
p^p	0.50	0.48	0.45	0.40	0.38	0.35	0.30	0.28	0.25
V^a	0.25	0.25	0.26	0.27	0.27	0.28	0.30	0.31	0.33
V^e	0.00	0.03	0.07	0.13	0.17	0.20	0.27	0.30	0.33
V^p	0.00	0.05	0.09	0.16	0.19	0.21	0.24	0.25	0.25
U^a	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.02	0.00
U^e	0.33	0.30	0.27	0.20	0.17	0.13	0.07	0.03	0.00
U^p	0.25	0.23	0.20	0.16	0.14	0.12	0.09	0.08	0.06

Simulation with $a = b = 1, c = 0.75$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.40	0.40	0.40	0.42	0.43	0.44	0.47	0.50	0.57
y^e	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57
y^p	0.80	0.76	0.72	0.64	0.60	0.56	0.48	0.44	0.40
p^a	0.30	0.29	0.28	0.27	0.26	0.25	0.23	0.22	0.21
p^e	0.71	0.66	0.61	0.51	0.46	0.41	0.31	0.26	0.21
p^p	0.60	0.57	0.54	0.48	0.45	0.42	0.36	0.33	0.30
V^a	0.20	0.20	0.21	0.22	0.22	0.23	0.25	0.26	0.29
V^e	0.00	0.03	0.06	0.11	0.14	0.17	0.23	0.26	0.29
V^p	0.00	0.04	0.07	0.13	0.15	0.17	0.19	0.20	0.20
U^a	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.02	0.00
U^e	0.29	0.26	0.23	0.17	0.14	0.11	0.06	0.03	0.00
U^p	0.24	0.22	0.19	0.15	0.14	0.12	0.09	0.07	0.06

Simulation with $a = b = 1, c = 1.25$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.28	0.29	0.29	0.31	0.32	0.33	0.37	0.39	0.44
y^e	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
y^p	0.57	0.54	0.51	0.46	0.43	0.40	0.34	0.31	0.29
p^a	0.35	0.35	0.34	0.32	0.31	0.31	0.29	0.28	0.27
p^e	0.78	0.73	0.68	0.58	0.53	0.48	0.38	0.33	0.28
p^p	0.71	0.68	0.64	0.57	0.54	0.50	0.43	0.39	0.36
V^a	0.14	0.15	0.15	0.16	0.17	0.17	0.19	0.20	0.22
V^e	0.00	0.02	0.04	0.09	0.11	0.13	0.18	0.20	0.22
V^p	0.00	0.03	0.05	0.09	0.11	0.12	0.14	0.14	0.14
U^a	0.05	0.05	0.05	0.04	0.04	0.03	0.02	0.01	0.00
U^e	0.22	0.20	0.18	0.13	0.11	0.09	0.04	0.02	0.00
U^p	0.20	0.18	0.17	0.13	0.11	0.10	0.07	0.06	0.05

Simulation with $a = b = 1, c = 1.5$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.25	0.25	0.26	0.27	0.28	0.29	0.33	0.35	0.40
y^e	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
y^p	0.50	0.48	0.45	0.40	0.38	0.35	0.30	0.28	0.25
p^a	0.37	0.36	0.35	0.34	0.33	0.33	0.31	0.30	0.30
p^e	0.80	0.75	0.70	0.60	0.55	0.50	0.40	0.35	0.30
p^p	0.75	0.71	0.68	0.60	0.56	0.53	0.45	0.41	0.38
V^a	0.12	0.13	0.13	0.14	0.15	0.15	0.17	0.18	0.20
V^e	0.00	0.02	0.04	0.08	0.10	0.12	0.16	0.18	0.20
V^p	0.00	0.02	0.05	0.08	0.09	0.11	0.12	0.12	0.13
U^a	0.05	0.04	0.04	0.04	0.03	0.03	0.02	0.01	0.00
U^e	0.20	0.18	0.16	0.12	0.10	0.08	0.04	0.02	0.00
U^p	0.19	0.17	0.15	0.12	0.11	0.09	0.07	0.06	0.05

Simulation with $a = b = 1, c = 1.75$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.22	0.22	0.23	0.24	0.26	0.27	0.30	0.32	0.36
y^e	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36
y^p	0.44	0.42	0.40	0.36	0.33	0.31	0.27	0.24	0.22
p^a	0.38	0.38	0.37	0.35	0.35	0.34	0.33	0.32	0.31
p^e	0.82	0.77	0.72	0.62	0.57	0.52	0.42	0.37	0.32
p^p	0.78	0.74	0.70	0.62	0.58	0.54	0.47	0.43	0.39
V^a	0.11	0.11	0.12	0.13	0.13	0.14	0.16	0.17	0.18
V^e	0.00	0.02	0.04	0.07	0.09	0.11	0.15	0.16	0.18
V^p	0.00	0.02	0.04	0.07	0.08	0.09	0.11	0.11	0.11
U^a	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.01	0.00
U^e	0.18	0.16	0.15	0.11	0.09	0.07	0.04	0.02	0.00
U^p	0.17	0.16	0.14	0.11	0.10	0.08	0.06	0.05	0.04

Simulation with $a = b = 1, c = 2$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.20	0.20	0.21	0.22	0.23	0.24	0.27	0.29	0.33
y^e	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
y^p	0.40	0.38	0.36	0.32	0.30	0.28	0.24	0.22	0.20
p^a	0.40	0.39	0.38	0.37	0.36	0.35	0.34	0.33	0.33
p^e	0.83	0.78	0.73	0.63	0.58	0.53	0.43	0.38	0.33
p^p	0.80	0.76	0.72	0.64	0.60	0.56	0.48	0.44	0.40
V^a	0.10	0.10	0.11	0.12	0.12	0.13	0.14	0.15	0.17
V^e	0.00	0.02	0.03	0.07	0.08	0.10	0.13	0.15	0.17
V^p	0.00	0.02	0.04	0.06	0.08	0.08	0.10	0.10	0.10
U^a	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.01	0.00
U^e	0.17	0.15	0.13	0.10	0.08	0.07	0.03	0.02	0.00
U^p	0.16	0.14	0.13	0.10	0.09	0.08	0.06	0.05	0.04

Simulation with $a = b = 1, c = 3$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.14	0.15	0.15	0.16	0.17	0.18	0.20	0.22	0.25
y^e	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
y^p	0.29	0.27	0.26	0.23	0.21	0.20	0.17	0.16	0.14
p^a	0.42	0.42	0.42	0.40	0.40	0.39	0.38	0.38	0.37
p^e	0.88	0.83	0.78	0.68	0.63	0.58	0.48	0.43	0.38
p^p	0.86	0.81	0.77	0.69	0.64	0.60	0.51	0.47	0.43
V^a	0.07	0.07	0.08	0.08	0.09	0.09	0.11	0.11	0.12
V^e	0.00	0.01	0.03	0.05	0.06	0.08	0.10	0.11	0.13
V^p	0.00	0.01	0.03	0.05	0.05	0.06	0.07	0.07	0.07
U^a	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.00
U^e	0.13	0.11	0.10	0.08	0.06	0.05	0.03	0.01	0.00
U^p	0.12	0.11	0.10	0.08	0.07	0.06	0.04	0.04	0.03

Simulation with $a = b = 1, c = 5$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.09	0.09	0.10	0.11	0.11	0.12	0.14	0.15	0.17
y^e	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
y^p	0.18	0.17	0.16	0.15	0.14	0.13	0.11	0.10	0.09
p^a	0.45	0.44	0.44	0.44	0.43	0.42	0.42	0.42	0.42
p^e	0.92	0.87	0.82	0.72	0.67	0.62	0.52	0.47	0.42
p^p	0.91	0.86	0.82	0.73	0.68	0.64	0.55	0.50	0.45
V^a	0.05	0.05	0.05	0.05	0.06	0.06	0.07	0.08	0.08
V^e	0.00	0.01	0.02	0.03	0.04	0.05	0.07	0.08	0.08
V^p	0.00	0.01	0.02	0.03	0.03	0.04	0.04	0.05	0.05
U^a	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.00
U^e	0.08	0.08	0.07	0.05	0.04	0.03	0.02	0.01	0.00
U^p	0.08	0.07	0.07	0.05	0.05	0.04	0.03	0.03	0.02

Simulation with $a = b = 1, c = 30$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03
y^e	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
y^p	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02
p^a	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
p^e	0.98	0.93	0.88	0.78	0.73	0.68	0.58	0.53	0.48
p^p	0.98	0.93	0.89	0.79	0.74	0.69	0.59	0.54	0.49
V^a	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
V^e	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.02
V^p	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01
U^a	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U^e	0.02	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00
U^p	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00

3. $a = 1$, different b and c

Simulation with $a = 1, b = 0.5, c = 0.25$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	1.00	1.00	0.99	1.01	1.03	1.04	1.11	1.17	1.33
y^e	1.33	1.33	1.33	1.33	1.33	1.33	1.33	1.33	1.33
y^p	2.00	1.90	1.80	1.60	1.50	1.40	1.20	1.10	1.00
p^a	0.25	0.24	0.23	0.22	0.21	0.20	0.19	0.18	0.17
p^e	0.67	0.62	0.57	0.47	0.42	0.37	0.27	0.22	0.17
p^p	0.50	0.48	0.45	0.40	0.38	0.35	0.30	0.28	0.25
V^a	0.50	0.51	0.51	0.53	0.54	0.56	0.60	0.62	0.67
V^e	0.00	0.07	0.13	0.27	0.33	0.40	0.53	0.60	0.67
V^p	0.00	0.10	0.18	0.32	0.38	0.42	0.48	0.50	0.50
U^a	0.12	0.12	0.11	0.10	0.09	0.08	0.05	0.03	0.00
U^e	0.67	0.60	0.53	0.40	0.33	0.27	0.13	0.07	0.00
U^p	0.50	0.45	0.41	0.32	0.28	0.25	0.18	0.15	0.13

Simulation with $a = 1, b = c = 0.5$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.67	0.67	0.68	0.71	0.73	0.75	0.81	0.87	1.00
y^e	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y^p	1.33	1.27	1.20	1.07	1.00	0.93	0.80	0.73	0.67
p^a	0.33	0.32	0.31	0.30	0.29	0.28	0.26	0.25	0.25
p^e	0.75	0.70	0.65	0.55	0.50	0.45	0.35	0.30	0.25
p^p	0.67	0.63	0.60	0.53	0.50	0.47	0.40	0.37	0.33
V^a	0.33	0.34	0.35	0.37	0.38	0.40	0.44	0.46	0.50
V^e	0.00	0.05	0.10	0.20	0.25	0.30	0.40	0.45	0.50
V^p	0.00	0.06	0.12	0.21	0.25	0.28	0.32	0.33	0.33
U^a	0.11	0.10	0.10	0.09	0.08	0.07	0.05	0.03	0.00
U^e	0.50	0.45	0.40	0.30	0.25	0.20	0.10	0.05	0.00
U^p	0.44	0.40	0.36	0.28	0.25	0.22	0.16	0.13	0.11

Simulation with $a = 1, b = c = 5$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.09	0.10
y^e	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
y^p	0.13	0.13	0.12	0.11	0.10	0.09	0.08	0.07	0.07
p^a	0.33	0.32	0.31	0.30	0.29	0.28	0.26	0.25	0.25
p^e	0.75	0.70	0.65	0.55	0.50	0.45	0.35	0.30	0.25
p^p	0.67	0.63	0.60	0.53	0.50	0.47	0.40	0.37	0.33
V^a	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.05	0.05
V^e	0.00	0.01	0.01	0.02	0.03	0.03	0.04	0.05	0.05
V^p	0.00	0.01	0.01	0.02	0.03	0.03	0.03	0.03	0.03
U^a	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00
U^e	0.05	0.05	0.04	0.03	0.03	0.02	0.01	0.01	0.00
U^p	0.04	0.04	0.04	0.03	0.03	0.02	0.02	0.01	0.01

Simulation with $a = 1, b = c = 30$

	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .4$	$\gamma = .5$	$\gamma = .6$	$\gamma = .8$	$\gamma = .9$	$\gamma = 1$
y^a	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
y^e	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
y^p	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01
p^a	0.33	0.32	0.31	0.30	0.29	0.28	0.26	0.25	0.25
p^e	0.75	0.70	0.65	0.55	0.50	0.45	0.35	0.30	0.25
p^p	0.67	0.63	0.60	0.53	0.50	0.47	0.40	0.37	0.33
V^a	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
V^e	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01
V^p	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
U^a	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U^e	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00
U^p	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00

Chapter 2
Do Waiting Times Reduce Hospital Costs?

2.1 Introduction

Waiting times are a major policy issue in many OECD countries. Average waiting times range between four and eight months for common procedures like cataract and hip replacement.

There are normally two rationales for justifying the existence of waiting times. The first is that waiting times act as a rationing mechanism that helps to bring in equilibrium the demand and supply of health care (see Lindsay and Feigenbaum 1984, Martin and Smith 1999, Cullis, Jones and Propper 2000): in the absence of price rationing and if benefit from treatment is to some extent unobservable, waiting times may deter patients with small benefit from asking treatment. A second rationale is that waiting times reduce the cost of provision of elective surgery. When demand is stochastic, waiting times may reduce idle capacity, therefore inducing a more efficient use of resources (see Iversen 1993, Iversen 1997, Barros and Olivella 2005).

Elaborating on this second aspect, Iversen (1993) argues that for low waiting times, hospital cost should reduce with waiting times, as a consequence of lower excess capacity. However, there is a point over which higher waiting times increase costs. This may be due for example to higher costs of managing the waiting list. Iversen (1993) also cites the increase of resources that are needed for repeated examination of patients (since their

status might change during the course of the waiting), and of treatment costs and average length of stay resulting from high cancellations rates typically found in hospitals with high capacity utilization. There is therefore, at least theoretically, a level of waiting time which minimises total costs. Above this optimal level, increasing waiting times should increase hospital cost.

To date, most empirical studies have focused on the effect of waiting times on individual or patient costs. For instance, Globerman (1991) estimates the economic costs of waiting for different procedures in Canada. The calculations take account of direct and indirect (intangible costs such as pain and anxiety) costs for patients and suggest that in aggregate these are comparable to losses resulting from labour strikes and lockouts.

More recently Hughes and McGuire (2003) have studied the effect of demand uncertainty on hospital cost structures via the accumulation of reserve capacity. The authors estimate that the marginal cost of elective admissions corresponds to 45% of emergency admissions. Moreover, they estimate that the holding of idle capacity adds 5% to the cost of emergency admissions. This is consistent with previous studies which have focused on the cost of excess bed capacity (e.g. Gaynor and Anderson 1995, Keeler and Ying 1996). The implication is that waiting times, by restraining excess capacity, should reduce hospital costs.

The purpose of this paper is to empirically estimate the elasticity of hospital costs with respect to waiting times. Our primary objective is to test whether, as implied by Hughes and McGuire (2003), waiting times reduce hospital costs or whether, as proposed

by Iversen (1993), there is an optimal level of waiting beyond which costs might increase. We use a sample of 137 acute hospitals over the period 1998-2002 in the English NHS.

The cross-sectional and panel data results suggest that waiting times have no significant effect on hospital costs. In most specifications the effect of waiting times is non-linear, suggesting a U-shaped relationship between hospital costs and waiting times, which is consistent with Iversen's (1993) model. However, the coefficients are generally not statistically significant. This suggests that, if anything, waiting times have no effect on hospital costs.

The study is organised as follows. The next section presents the empirical specification and section 2.3 describes the data. The results are presented in section 2.4 and the conclusion in section 2.5.

2.2 Econometric model

We estimate three types of regressions: pooled OLS, panel fixed effects and panel random effects. The pooled OLS model is given by:

$$y_{it} = \alpha + \mathbf{x}'_{it}\beta + u_{it}, \tag{2.1}$$

where y_{it} is the cost of hospital i at year t, $\mathbf{x}_{it}$ is a vector of explanatory variables, α is the individual effect, β is a vector of parameters to be estimated and u_{it} is the idyosincratic error. The crucial assumption of the pooled model is that the individual effects are constant over time and common to all observations in the sample.

An alternative approach is to assume that individual effects are specific to each observation. This leads to the fixed effects model:

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\beta + \varepsilon_{it}. \tag{2.2}$$

The hospital-specific fixed-effects α_i capture individual unobserved heterogeneity. If there are individual fixed-effects and these are correlated with the explanatory variables, then pooled OLS estimates would be biased (see Cameron and Trivedi 2005). Finally, we also estimate random-effects models:

$$\begin{aligned} y_{it} &= \alpha_i + \mathbf{x}'_{it}\beta + \varepsilon_{it}, \\ \alpha_i &\backsim N(\alpha, \sigma^2_\alpha), \\ \varepsilon_{it} &\backsim N(0, \sigma^2_\varepsilon). \end{aligned} \tag{2.3}$$

In this formulation the individual effects are randomly iid distributed over the population. We compare the fixed- and random-effects models using a Hausman test, which tests for systematic differences in coefficients between the two models. The rejection of the null hypothesis of no systematic difference in coefficients gives an indication of the existence of fixed-effects, and suggests that the random-effects estimates are biased. Otherwise, it is safe to assume the random-effects estimation.

Model selection also uses a Breusch-Pagan test for testing the variance of the random-effects term. In this case, the null hypothesis $var(\alpha_i) = 0$ suggests that there are no random effects, providing support for the pooled OLS regression. Rejecting the null, however, suggests the presence of individual effects, which might be modelled either by fixed- or random-effects.

It might be argued that the relationship between costs and waiting times is endogenous. If a hospital has high costs, it is more likely to have longer waiting times. There are several channels through which this may happen.

First, more inefficient hospitals have higher costs, due for example to poor management: if higher inefficiency also implies higher inefficiency in the management of the waiting list, then inefficient hospitals may have both higher costs and higher waiting times (a positive correlation). If the researcher has no access to variables correlated with inefficiency, then the OLS estimates of Equation ((2.1)) will be biased upwards. We use at least two variables that might be correlated with inefficiency: length of stay and proportion of day cases. Keeping other factors constant, more inefficient providers have a higher length of stay and a smaller proportion of day cases.

Second, hospitals with higher quality might have a higher cost, and at the same time attract a higher number of patients, which leads to a higher waiting time (again, a positive correlation). We use at least two variables that might be correlated with quality: length of stay and (age and gender adjusted) re-admission rates. Keeping other factors constant, providers with higher quality should have a higher length of stay and smaller readmission rates.

If there is some residual unobserved efficiency and quality, the OLS might still be biased. However, by estimating a fixed effect model, all unobserved inefficiency and quality will be captured by the individual fixed effect, as long as quality and inefficiency are time invariant, which seems plausible over short intervals of time.

Moreover, we estimate regressions with both balanced and unbalanced samples. This is an important sensitivity analysis to demonstrate the robustness of the results and to control for selection effects.

2.3 Data

The sample comprises 137 English NHS acute hospitals observed annually between 1998/1999 and 2001/2002, making an unbalanced panel with 440 observations.

This study uses several indicators related to hospital performance, institutional framework and the population served. The data were collected from several sources, including the Department of Health (DoH), the Hospital Episode Statistics (HES), the National Health Service Information Authority (NHSIA) the Chartered Institute of Public Finance and Accountancy (CIPFA) and Dr Foster. All variables used in this paper are calculated in an annual basis.

The DoH provides several statistics that are generally relevant for other government departments in the areas of healthcare, workforce, public health and social care. The dependent variable in this study uses the measure of total hospital cost that is calculated annually by the DoH based on the analysis of total expenditures at hospital level (TFR3). Other variables provided by the DoH include the total number of outpatient attendances, the number of beds available and the rate of emergency readmission within 28 days. HES is the other major data source used in this study. HES is the data warehouse of the English NHS, routinely collecting statistical information on episodes of patients admitted to hospital care by NHS hospitals in England (www.hesonline.nhs.uk). HES provides several indicators of

hospital activity, including total inpatient spells, emergency admissions, day cases and average length of stay. Moreover, the indicator of mean wait times, our explanatory variable of interest, also comes from the HES database. The HRG index that we use to control for patient casemix composition was provided by the NHSIA. This has now been discontinued and the corresponding information is available from the NHS Connecting for Health service. Dr Foster, an independent organization that provides information on the quality of health services, provided information on the number of competitors within 20km radius for each hospital.

Our dependent variable is the total hospital cost, measured in thousands of Pounds Sterling and excluding capital costs. It was compiled from the Department of Health and was transformed into real values for 2002 using the GDP deflator provided by HM Treasury. Our measure of waiting times is the mean wait for elective admissions, which was provided by the HES. It measures the average number of days between the decision of being admitted to the waiting list and the actual admission for treatment.

Table 2.1 provides a description of the variables employed in the analysis and corresponding sources of data. We divide the explanatory variables in six groups. Hospital activity is measured by the total number of inpatient spells and the total number of outpatient attendances. Both variables are measured in 1,000 cases. A second group of variables captures the severity of cases treated by the hospital and the demand on resources. It includes emergency admissions as a proportion of total spells and a HRG casemix index based on reference costs.

Hospital costs also depend on the efficiency on the use of resources. We control for the number of day cases as a proportion of elective surgeries and the average length of stay. More efficient hospitals are expected to have a higher proportion of surgeries carried out on a day case basis, and a lower average length of stay. As usual in the literature on hospital costs (e.g. Vita 1990, Vitaliano 1987), the capital stock is proxied by the number of available beds.

The quality of services is proxied by the percentage of emergency readmissions within 28 days. Contrary to the severity indicators, both efficiency and quality indicators are to some extent under the control of the hospital. Finally, we also include the number of hospitals within a 20km radius as a measure of the degree of competition in the geographical market (see Siciliani and Martin 2007).

We do not include salaries because information is not readily available. Also, salaries are nationally agreed and therefore there is no variations in salary across hospitals.

Table 2.1. Description of variables

Variable name	Description	Source*
(a) Hospital cost		
totcost	Total hospital cost (£000) (2002 real values using Treasury GDP deflator)	DoH
(b) Waiting times		
meanwait	Mean wait in days	HES
(c) Measures of activity		
totspells	Total inpatient spells (000)	HES
totop	Total outpatient attendances (000)	DoH
(d) Case mix		
emergadm	Number of emergency admissions as % of total inpatient spells	HES
hrgindrc	HRG casemix index based on Reference Costs	NHSIA
(e) Efficiency on use of resources		
daycase	Number of day cases as % of electives	HES
alos	Average length of stay	HES
(f) Capital inputs		
avbeds	Number of available beds	DoH
(g) Quality of services		
readmisnpc	Emerg. readm. % within 28 days, all ages, age sex std	DoH
(h) Competition		
nhosp20km	Number of hospitals within 20 km radius	
(i) Dummy variables		
acute	Hospital is an acute only hospital	CIPFA
teaching	Hospital is a teaching hospital	CIPFA
specialist	Trust provides specialist services	CIPFA
london	Trust is in London	CIPFA

* DoH: Department of Health, HES: Hospital Episode Statistics, NHSIA: National Health Service Information Authority, CIPFA: The Chartered Institute of Public Finance and Accountancy

Table 2.2 presents some descriptive statistics. The average hospital in the sample has a total cost of just over £100 million per year and an average waiting time for elective surgery of 97 days. It provides around 51,200 inpatient spells and 206,200 oupatient attendances, and faces the competition of 5 other hospitals on a 20km radius. Around 34% of inpatient spells are originated as emergency attendances, and the average HRG casemix index is at 100. With respect to the efficiency of resource use, each hospital admits on average 49% of the elective patients as day cases, with an average length of stay of 6.4 days. The proportion of emergency readmission within 28 days is around 6%.

Table 2.2. Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
totcost	809	113,798.3	71,957.0	3,688.5	554,866.7
meanwait	809	96.8	35.9	2.0	457.0
totspells	809	51.2	29.7	0.2	200.0
totop	809	206.2	128.5	10.8	1,093.9
emergadm	809	33.7	10.0	0	82.6
hrgindrc	809	100.9	22.0	68.5	229.7
daycase	804	48.6	11.2	0	95.5
alos	809	6.4	4.7	0.6	46.2
readmisnpc	612	5.9	1.1	3.7	18.4
avbeds	809	678.2	365.3	4.0	2,838.1
nhosp20km	617	5.3	5.4	1	20
london	809	0.2	0.4	0	1
y1998	809	0.3	0.4	0	1
y1999	809	0.3	0.4	0	1
y2000	809	0.2	0.4	0	1
y2001	809	0.2	0.4	0	1

The data present some limitations in terms of information availability, especially for the quality and competition indicators, which reduce significantly the sample.

2.3.1 Skewness correction

A common issue in this type of study is the high skewness usually found in cost and health expenditures data. Table 2.3 shows that the mean of total cost (£ 113,798) is 23% higher than the median (£ 94,824), the skewness statistic is 2.0 and the kurtosis is 8.9. This suggests that skewness is an issue especially for total cost, but less so for mean waiting times. The log transformation reduces the extent of this problem, although not eliminating it completely. After the log transformation, the mean of total cost becomes equal to the median (11.5), the skewness statistic is reduced to -0.6 and the kurtosis to 5.0.

Table 2.3. Additional descriptive statistics of selected variables (in levels and logs)

	Levels				Logs			
Variable	Mean	Median	Skewness	Kurtosis	Mean	Median	Skewness	Kurtosis
totcost	113,798.3	94,824.3	2.0	8.9	11.5	11.5	- 0.6	5.0
meanwait	96.8	94.0	1.6	15.8	4.5	4.5	- 2.1	13.4
totspells	51.2	47.5	1.1	6.0	3.7	3.9	- 1.6	7.3
totop	206.2	183.9	1.7	9.1	5.1	5.2	- 1.0	4.9
emergadm	33.7	35.2	- 0.7	6.3	3.4	3.6	- 3.6	20.7
hrgindrc	100.9	95.0	3.1	13.4	4.6	4.6	2.3	9.3
daycase	48.6	49.1	- 0.9	7.5	3.8	3.9	- 7.3	77.7
alos	6.4	5.6	4.9	32.8	1.7	1.7	1.1	8.0
readmisnpc	5.9	5.8	3.0	32.7	1.8	1.8	0.6	6.5
avbeds	678.2	627.0	1.4	8.0	6.4	6.4	- 1.5	8.3
nhosp20km	5.3	3.0	1.4	4.1	1.2	1.1	0.3	1.8

With the exception of emergency admissions, readmissions and day cases, all the other continuous variables are included in the log scale, which reduces skewness and allows the interpretation of coefficients as elasticities. Emergency admissions, readmissions and day cases, however, are kept in levels. Since they are measured as percentages, the log transformation would greatly increase both skewness and kurtosis.

2.4 Results

The results of the regression analysis for pooled OLS and fixed effects are reported in tables 2.4 and 2.5. The dependent variable in both regressions is the log of total hospital cost (lntotcost) in real values of 2002. Most variables are in logs, so coefficients can be interpreted as elasticities.

In each set of results there are seven different specifications[9], which add regressors progressively in order to test the stability of results. The basic regression (column (1) of

[9] However, there are only six specifications for panel regressions because the competition indicator is time-invariant, which prevents fixed effects estimation. Similarly, notice that the HRG index and the London dummy are not included in the panel regressions for the same reason.

each table) includes mean waiting times (linear and quadratic effect)[10] and activity indicators (inpatient spells and outpatient attendances), and controls for the HRG index, London dummy and year. We then progressively add controls for capital stock (available beds (2)), demand on resources (emergency admissions (3)), efficiency on use of resources (daycases (4) and average length of stay (5)), quality of service (emergency readmissions (6)) and competition (number of competitors in a 20 km radius (7)). Given the limited coverage of our sample, the inclusion of the quality and competition indicators reduces significantly the number of observations.

We also estimated the regressions using a trans-log specification, which is a second-order Taylor approximation adding square terms for the activity indicators. However, since the square terms were not significant in this specification, we decided to exclude them.

The cross-section regressions were estimated using standard errors robust to both heteroscedasticity and the serial correlation among observations of the same hospital over the years. Thus the tables report both the total number of observations (*N*) and the number of independent observations or clusters (*N clusters*).

Table 2.4 reports pooled cross-section estimates using unbalanced samples. As additional regressors are added the sample size decreases, falling from 440 observations in the basic regression to 319 observations in the regression with all independent variables.

---

[10] The quadratic effect is entered as $\ln meanwait^2/2$. Therefore, the elasticity of total cost with respect to waiting times can be calculated as $\varepsilon_w^C = \partial C_{it}/\partial \log w_{it} = \beta_1 + \beta_2 \ln meanwait$. β_1 is the estimated coefficient for the linear component and β_2 for the quadratic. Waiting times are computed at the sample mean.

Table 2.4. Unbalanced pooled OLS regressions of total hospital cost

Dependent variable: log(totcost); Robust standard errors

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
lnmeanwait	- 1.40	-.93*	-.93*	-.76*	-.73*	- 0.28	- 0.28
lnmeanwait2	0.29	.21*	.21**	.17*	.16*	0.07	0.07
lntotspells	.55***	.3***	.27***	.28***	.34***	.62***	.65***
lntotop	.38***	.28***	.28***	.28***	.29***	.23***	.24***
lnhrgindrc	1.2***	.81***	.83***	.8***	.77***	.73***	.72***
lnavbeds		.4***	.43***	.42***	.35***	0.14	0.10
emergadm			-3.10E-03	-3.00E-03	-3.00E-03	-2.10E-03	-1.10E-03
daycase				-2.20E-03	-.0027*	-.0037**	-.0034*
lnalos					0.08	.25***	.26**
readmisnpc						1.70E-03	4.10E-03
lnnhosp20km							-2.10E-03
london	.19***	.18***	.17***	.16***	.16***	.15***	.15**
y1999	.13***	.13***	.13***	.13***	.13***	.14***	.15***
y2000	.19***	.17***	.17***	.17***	.21***	.28***	.28***
y2001	.14***	.14***	.15***	.15***	.15***	.15***	.14***
_cons	4.7**	4.5***	4.4***	4.3***	4.4***	3.9***	3.9***
$\bar{R}^2$	0.86	0.89	0.89	0.89	0.89	0.9	0.89
RESET	0.1	0.5	0.5	1	1.2	2	1.8
Joint significance†	1.4	2	2.1	1.7	1.6	0.3	0.6
N	440	440	440	439	439	384	319
N_clusters	137	137	137	137	137	109	88

Legend: * p<.1; ** p<.05; *** p<.01; † Test for joint significance of lnmeanwait and lnmeanwait^2

All the regressions have been estimated with both linear and quadratic effects for waiting times, which allows us to control for nonlinearities in the hospital cost response to waiting times. Two reasons guided the choice of this functional form.

First, it gives a direct test of Iversen's (1993) suggestion of a nonlinear effect of waiting times. More specifically, according to Iversen (1993), waiting times reduce total cost to begin with, but then due to management and repeated examinations problems, the effect is reversed, and total costs start to increase. According to this hypothesis, we should expect to find a negative coefficient for the level of waiting times and a positive coefficient for the quadratic component.

The second reason for the inclusion of a quadratic effect of waiting times is to make sure that there are no misspecification problems. In all cases the RESET test is not significant, which suggests that the functional form is correctly specified and that there are no omitted variables.

Let us now focus on the coefficients estimated by the regressions, starting with waiting times. The cross-sectional results suggest that waiting times have no significant effect on hospital costs. In all cross-section regressions the estimated effect of waiting times is not statistically significant, either jointly or separately. However, in all regressions the effect of waiting times is non-linear, suggesting a U-shaped relationship between hospital costs and waiting times, which is consistent with Iversen's (1993) model. The coefficient for the linear component is negative, while the quadratic component is positive. As discussed above, this implies that waiting times have an initial negative impact on costs. However, after some point the trend is reversed and waiting times start to increase costs. Therefore, in principle there is an optimal level of waiting times that minimises total costs.

Although the effect of waiting times in the basic regression (column (1) of Table 2.4) is not statistically significant, its estimated magnitude is substantial. Using the formula $\varepsilon_w^C = \beta_1 + \beta_2 \ln meanwait$, the elasticity of hospital cost with respect to waiting times at the sample average of waiting times is estimated at -0.07. Therefore, when taken at the sample average waiting times are expected to reduce hospital cost. However, this effect is expected to vanish if waiting times are increased. In fact, we can calculate the level of waiting times where the elasticity turns from negative to positive. This is the waiting times level that solves: $\beta_1 + \beta_2 \ln meanwait = 0$, which yields 125 days. That is, after a median waiting time of 125 days waiting times would start to increase hospital costs, which agrees with Iversen's (1993) model.

Other control variables display significant effects. As expected both inpatient and outpatient activity increase cost, as does the HRG index. On average, hospitals costs in

London are approximately 20% higher than in the rest of the country. Real costs increased significantly between 1998 and 1999, possibly due to nation-wide salary increase from 1999/2000 onwards.

Adding available beds (column (2)) affects the results, causing the coefficients of waiting times to become significant. However, the separate effects of waiting times are significant only at 10% and the joint effect is not statistically significant. The effect of available beds, our proxy for capital, is itself positive in most regressions where it is included.

The introduction of emergency admissions or day cases does not affect the results significantly (see columns 3 and 4). Emergency admissions have no effect on hospital costs. Day cases have a negative and significant coefficient, suggesting that hospitals with a higher proportion of elective admissions treated as day cases have lower costs. In column (5), we add average length of stay, which has a positive and significant effect on hospitals' costs, as expected. The effect of waiting times is not altered and the RESET test is still not significant.

Next we include readmission rates (column (6)), which impose a positive, but very small and not statistically significant effect on hospital cost. However, there is another important consequence. In sharp contrast with previous specifications, the effects of waiting times and available beds cease to be significant. This might be due to an omitted variable bias or to selection in the sample of hospitals with readmission information. Notice that readmission rates cause a sizable reduction in the sample. The RESET test is not significant, suggesting that there are no omitted variables.

Table 2.5. Unbalanced fixed effects regressions of total hospital cost

Dependent variable: log(totcost); Robust standard errors						
Variable	(1)	(2)	(3)	(4)	(5)	(6)
lnmeanwait	- 0.19	- 0.18	- 0.13	- 0.14	- 0.13	- 0.16
lnmeanwait2	0.06	0.06	0.04	0.04	0.04	0.04
lntotspells	.16*	0.13	.18**	.18**	.19**	0.02
lntotop	.12*	.11*	.11*	.11*	.11*	0.10
lnavbeds		.24***	.17**	.17**	.15**	.23***
emergadm			.0068**	.0065*	.0061*	3.00E-03
daycase				-8.20E-04	-1.00E-03	-1.10E-03
lnalos					0.05	0.05
readmisnpc						- 0.01
y1999	.13***	.13***	.13***	.13***	.13***	.13***
y2000	.15***	.14***	.14***	.14***	.16***	.17***
y2001	.21***	.2***	.19***	.19***	.19***	.2***
cons	10***	9***	8.9***	9***	9***	9.5***
R^2 within	0.67	0.68	0.69	0.69	0.69	0.69
R^2 between	0.71	0.82	0.77	0.77	0.78	0.86
R^2 overall	0.68	0.81	0.77	0.77	0.78	0.76
$corr(\alpha_i, Xb)$	0.64	0.69	0.63	0.65	0.66	0.71
σ	0.3	0.24	0.25	0.25	0.26	0.27
σ_u	0.29	0.23	0.25	0.24	0.25	0.27
σ_e	0.065	0.063	0.063	0.063	0.063	0.063
ρ	0.95	0.93	0.94	0.94	0.94	0.95
Hausman	68.5***	56***	62***	60.5***	59.5***	72.6***
Breusch-Pagan	234.2***	260.4***	249.5***	243.3***	236.7***	241.2***
Joint significance†	5.6***	2.7*	1.9	1.7	1.5	0.5
N	440	440	440	439	439	384
N clusters	137	137	137	137	137	109

Legend: * p<.1; ** p<.05; *** p<.01; † Test for joint significance of lnmeanwait and lnmeanwait^2

Finally we evaluate the effect of local competition from other hospitals (column (7)). The estimated effect of competition is negative, although not significant.

In addition to the pooled cross-sectional analysis we also estimate fixed- and random-effects panel regressions. Results from the fixed-effects estimations are reported in Table 2.5. Notice that the time-invariant regressors (namely, hospital type dummies, London dummy and number of competitors) are excluded from the fixed-effect regressions. Also, the two regressions where the samples reduce to just only one wave cannot be estimated either by fixed- or random effects. Therefore, there are only 6 fixed- and random-effects regressions.

The effect of waiting times estimated by fixed-effects is similar to the pooled OLS case, with negative sign for the linear component and positive for the quadratic. As ex-

pected, both total inpatient spells and outpatient attendances increase total hospital costs. Average beds and emergency admissions also increase costs and are statistically significant. Although not statistically significant, treating more elective surgeries as day cases and a higher readmission rate both reduce hospital costs, whilst higher average length of stay has the opposite effect.

Moreover, there are some interesting results from the comparison of the models. In all cases the Breusch-Pagan test rejects the hypothesis of null variance of the individual-effects, suggesting that pooled OLS estimates are biased. There are also indications that random-effects estimation is not appropriate. Individual-specific effects are much correlated with independent variables and in all cases the Hausman test rejects random effects model.

The fixed effects and the pooled OLS estimates have many similarities, especially in terms of the patterns of significance. The fixed effects coefficients tend to be of lower magnitude.

Table 2.6 reports the results of random effects estimations. We notice that in some of the random effects estimations the effect of waiting times is significant and in accordance with the hypothesis proposed by Iversen (1993). The linear coefficient of waiting times is negative, but the quadratic is positive, suggesting that increasing waiting times up to a certain level decreases costs, but past this level the effect is reversed.

However, the effects are significant at relatively low levels (10%). Tests for joint significance are also are also significant, but at low level. Moreover, as mentioned above, the Hausman test rejects the random effects specification in all cases. Therefore, although

the results from table 2.6 are indicative of a nonlinear effect of waiting times on cost, we cannot rule out that the true effect is as depicted by table 2.5.

Table 2.6. Unbalanced random effects regressions of total hospital cost

Dependent variable: log(totcost); Robust standard errors

Variable	(1)	(2)	(3)	(4)	(5)	(6)
lnmeanwait	- 0.72	-.55*	-.54*	-.5*	-.45*	- 0.24
lnmeanwait2	0.17	.13**	.13**	.12**	.11**	0.06
lntotspells	.36***	.18***	.19***	.21***	.24***	.35***
lntotop	.31***	.22***	.22***	.23***	.24***	.23***
lnavbeds		.45***	.45***	.43***	.39***	.35***
emergadm			4.60E-04	-6.90E-05	-5.00E-04	-1.70E-03
daycase				-.0024*	-.0028**	-.0041***
lnalos					0.08	.11**
readmisnpc						- 0.01
y1999	.13***	.13***	.12***	.13***	.13***	.14***
y2000	.14***	.13***	.13***	.14***	.17***	.19***
y2001	.18***	.18***	.18***	.19***	.18***	.18***
cons	9.8***	7.7***	7.7***	7.8***	7.6***	7***
R^2 within	0.62	0.66	0.66	0.66	0.66	0.65
R^2 between	0.77	0.85	0.85	0.86	0.86	0.9
R^2 overall	0.8	0.86	0.85	0.86	0.87	0.88
σ	0.18	0.16	0.16	0.15	0.15	0.13
σ_u	0.17	0.14	0.14	0.14	0.14	0.11
σ_e	0.065	0.063	0.063	0.063	0.063	0.063
ρ	0.87	0.84	0.84	0.83	0.83	0.75
Joint significance†	9.2**	6.4**	6.2**	5.1*	4.6*	3.8
N	440	440	440	439	439	384
N_clusters	137	137	137	137	137	109

Legend: * p<.1; ** p<.05; *** p<.01; † Test for joint significance of lnmeanwait and lnmeanwait^2

2.4.1 Regressions with balanced sample

The results above are based on the progressive inclusion of additional explanatory variables.

In some cases this affects the statistical significance of the coefficient of waiting times. As noted in the discussion of the results, there are two possible explanations: a) it might be the case that the regressions with a limited number of regressors suffer from omitted variable bias, or b) the inclusion of regressors leads to selection in the sample of observations.

In order to further explore this issue we re-estimate the regressions using a balanced sample. That is, all regressions in Table 2.7 are estimated using only the sample of 88 hos-

Table 2.7. Balanced pooled OLS regressions of total hospital cost

Dependent variable: log(totcost); Robust standard errors

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
lnmeanwait	- 0.29	- 0.45	- 0.44	- 0.45	- 0.27	- 0.28	- 0.28
lnmeanwait2	0.07	0.11	0.11	0.11	0.07	0.07	0.07
lntotspells	.67***	.45***	.43***	.43***	.65***	.65***	.65***
lntotop	.28***	.23***	.24***	.24***	.24***	.24***	.24***
lnhrgindrc	1.1***	.81***	.81***	.78***	.72***	.73***	.72***
lnavbeds		.3***	.32***	.32***	0.10	0.10	0.10
emergadm			-1.30E-03	-1.50E-03	-6.00E-04	-1.20E-03	-1.10E-03
daycase				-1.90E-03	-.0034*	-.0033*	-.0034*
lnalos					.26**	.25**	.26**
readmisnpc						4.10E-03	4.10E-03
lnnhosp20km							-2.10E-03
london	.2***	.19***	.18***	.17***	.15***	.15***	.15**
y1999	.14***	.15***	.15***	.15***	.15***	.15***	.15***
y2000	.18***	.17***	.17***	.18***	.28***	.28***	.28***
y2001	.13***	.14***	.14***	.15***	.14***	.14***	.14***
cons	2.9**	3.6***	3.6***	3.9***	3.9***	3.9***	3.9***
$\bar{R}^2$	0.88	0.89	0.89	0.89	0.89	0.89	0.89
RESET	0.7	0.8	0.8	1.4	1.9	1.8	1.8
Joint significance†	0.4	1.1	1.1	1	0.7	0.7	0.6
N	319	319	319	319	319	319	319
N_clusters	88	88	88	88	88	88	88

Legend: * p<.1; ** p<.05; *** p<.01; † Test for joint significance of lnmeanwait and lnmeanwait^2

pitals that we obtain when all regressors are included (column (7) of Table 2.4). Similarly, all regressions in Table 2.8 use the sample of column (6) of Table 2.5 and all regressions in Table 2.9 use the sample of column (6) of Table 2.6. Moreover, the balanced panel data regressions (Tables 2.8 and 2.9) include only hospitals with observations for all four years. This explains why the sample sizes are slightly smaller.

The regression results with balanced sample are qualitatively similar to the unbalanced. The main difference is that now the effect of waiting times ceases to be significant throughout.

Table 2.8. Balanced fixed effects regressions of total hospital cost

Dependent variable: log(totcost); Robust standard errors

Variable	(1)	(2)	(3)	(4)	(5)	(6)
lnmeanwait	0.67	0.42	0.37	0.35	0.29	0.32
lnmeanwait2	- 0.13	- 0.08	- 0.07	- 0.07	- 0.06	- 0.06
lntotspells	- 0.01	- 0.08	- 0.05	- 0.02	- 0.00	- 0.00
lntotop	.11*	0.10	0.11	0.11	.11*	.11*
lnavbeds		.3***	.26***	.25***	.22***	.22***
emergadm			2.90E-03	2.50E-03	2.20E-03	2.60E-03
daycase				-1.10E-03	-1.50E-03	-1.50E-03
lnalos					0.06	0.06
readmisnpc						-3.50E-03
y1999	.14***	.14***	.13***	.14***	.14***	.14***
y2000	.16***	.15***	.15***	.15***	.18***	.18***
y2001	.22***	.21***	.21***	.21***	.21***	.21***
cons	9.1***	8.2***	8.3***	8.3***	8.5***	8.5***
R^2 within	0.65	0.68	0.68	0.68	0.68	0.68
R^2 between	0.49	0.83	0.82	0.84	0.84	0.83
R^2 overall	0.27	0.68	0.69	0.71	0.7	0.69
$corr(\alpha_i, Xb)$	0.3	0.64	0.64	0.66	0.65	0.65
σ	0.32	0.26	0.26	0.25	0.26	0.26
σ_u	0.31	0.25	0.25	0.24	0.25	0.25
σ_e	0.07	0.067	0.067	0.067	0.067	0.067
ρ	0.95	0.93	0.93	0.93	0.93	0.93
Hausman	185.4***	171.5***	51.3***	170.5***	181.8***	195***
Breusch-Pagan	216.5***	215.7***	213.8***	213.9***	186***	183.6***
Joint significance†	3.9**	1.2	0.9	0.8	0.6	0.6
N	300	300	300	300	300	300
N_clusters	75	75	75	75	75	75

Legend: * p<.1; ** p<.05; *** p<.01; † Test for joint significance of lnmeanwait and lnmeanwait^2

Table 2.9. Balanced random effects regressions of total hospital cost

Dependent variable: log(totcost); Robust standard errors

Variable	(1)	(2)	(3)	(4)	(5)	(6)
lnmeanwait	- 0.32	- 0.48	- 0.51	- 0.46	- 0.50	- 0.51
lnmeanwait2	0.09	0.12	0.12	0.11	0.12	0.12
lntotspells	.49***	.22***	.23***	.29***	.37***	.37***
lntotop	.29***	.23***	.23***	.22***	.23***	.23***
lnavbeds		.44***	.42***	.4***	.32***	.32***
emergadm			1.60E-03	5.40E-04	3.70E-04	-6.10E-05
daycase				-.0035*	-.0043**	-.0043**
lnalos					0.12	0.12
readmisnpc						3.50E-03
y1999	.13***	.13***	.13***	.13***	.14***	.14***
y2000	.13***	.14***	.13***	.14***	.19***	.19***
y2001	.18***	.18***	.18***	.19***	.19***	.19***
cons	8.5***	7.4***	7.5***	7.6***	7.7***	7.7***
R^2 within	0.59	0.65	0.65	0.65	0.65	0.65
R^2 between	0.84	0.88	0.87	0.88	0.89	0.89
R^2 overall	0.81	0.85	0.85	0.86	0.87	0.87
σ	0.15	0.13	0.13	0.13	0.13	0.13
σ_u	0.13	0.12	0.12	0.12	0.11	0.11
σ_e	0.07	0.067	0.067	0.067	0.067	0.067
ρ	0.77	0.75	0.75	0.75	0.72	0.72
Joint significance†	4.1	3.1	2.8	2.5	2.4	2.6
N	300	300	300	300	300	300
N_clusters	75	75	75	75	75	75

Legend: * p<.1; ** p<.05; *** p<.01; † Test for joint significance of lnmeanwait and lnmeanwait^2

2.5 Conclusions

This chapter addresses an important question: What is the effect of waiting times on hospital costs? The motivation for this study comes from the theoretical literature on hospital regulation, where one can identify two different views. On the one hand, some authors (e.g. Gaynor and Anderson 1995, Keeler and Ying 1996, Hughes and McGuire 2003) argue that hospitals use waiting times to regulate excess capacity and that consequently waiting times should reduce hospital costs. On the other hand, Iversen (1993) argues that this negative effect exists only to a certain extent. Beyond a given point, however, the management of the waiting list and repeated examinations become too costly, and then waiting times start to increase total hospital costs.

In our view this is essentially an empirical question, which remains open until now. To answer this question we use a sample of 259 acute hospitals from the English NHS observed in the fours years between 1998 and 2001. Our primary purpose is to estimate the elasticity of hospital costs with respect to waiting times.

We estimated pooled OLS, fixed effects and random effects models for total hospital cost. In each case waiting times are entered in the regression in both linear and quadratic terms, which should assist in identifying the curvature of the cost function. The signs of the estimated coefficients are consistent with Iversen's (1993) model: the coefficient is negative for the linear effect and positive for the quadratic effect, suggesting a U-shaped relationship between hospital costs and waiting times. However, the coefficients are generally not statistically significant. It is possible to argue that to a certain extent the results

give some support for Iversen's (1993) model: although the coefficients are not significant, the signs are correctly estimated.

The results suggest that waiting times have no impact on hospital costs. This casts some doubt on the effectiveness of waiting times as a tool to control hospital costs. In the future more research would be desirable in order to further understand this issue.

Chapter 3
Preventive Care in Competitive Health Insurance Markets with Adverse Selection

3.1 Introduction

Economists have long acknowledged the importance of informational asymmetries in influencing individual behaviour. In a classic paper, Rothschild and Stiglitz (1976) describe how imperfect information, in the form of adverse selection, leads to distortions of insurance contracts. In particular, they show that, if individuals in an insurance market are heterogeneous with respect to unobservable risk, an equilibrium may fail to exist; if the equilibrium exists, it will necessarily separate the individuals according to risk type. In this case, low-risk individuals will be worse-off, while high-risk are no better-off, implying that a symmetric information equilibrium would be welfare-improving.

In the case of health insurance, most applications have focused on how insurers distort the provision of curative care in order to sort patients according to risk type. By contrast, preventive health services have received less attention. With the increase in pre-paid or prospective health insurance, however, this situation tends to change. Receiving a fixed premium per patient, insurers face increased incentives for cost-efficiency. That might lead to greater emphasis in prevention, contributing to reduce the need for future curative care (see Herring 2002, Miceli and Heffley 2002). This paper aims to extend the conventional

analysis of health insurance markets in order to consider the incentives for the provision of preventive care.

Following the distinction proposed by Ehrlich and Becker (1972), most authors (e.g. Phelps 1978, Kenkel 1994, Kenkel 2000, Barigozzi 2004, Welch 2004) separate preventive care into primary and secondary prevention.[11] The first category, also denoted self-protection by Ehrlich and Becker (1972), refers to activities that aim to reduce the probability of illness. Examples include the regular practice of physical exercises and vaccination campaigns. On the other hand, secondary prevention or self-insurance consists on efforts to attenuate the consequences of illness, without necessarily affecting the probability of occurrence. Examples related to health include screening for early cancer, high blood pressure, heart disease and dental plaque.

Phelps (1978) argues that the demand for prevention depends on its effect on the likely out-of-pocket medical expenditures, the work-loss associated with illness and, more importantly, on the potential health damage that can be avoided. However, as remarked by Breyer (1982), none of these effects can be accurately anticipated by the consumer or the insurer.

Ehrlich and Becker (1972) were the first to provide a formal model for prevention in the presence of market insurance. They show that market insurance decreases secondary prevention, in the sense that they can be seen as substitutes to redistribute income between alternative states-of-the-world. On the other hand, market insurance has two opposite effects on primary prevention: although primary prevention is discouraged by a lower mar-

---

[11] Kenkel (2000) also discusses tertiary prevention, which would aim to reduce the disability associated with a chronic illness.

ginal gain in the reduction of the difference of incomes between states, it is encouraged if the insurance premium is negatively affected by the reduction in the probability of loss. The authors argue that in general primary prevention can be considered a complement to market insurance. In particular, with quadratic utility function and probability of loss higher than 1/2, market insurance unambiguously increases optimal primary prevention.

Courbage and de Coulon (2004) and Courbage and Rey (2006) study the effect of prudence and risk aversion in the demand for prevention. Courbage and de Coulon (2004) conclude that private health insurance does not lead to less prevention (probability of exercising, regular check-ups and smoking) in the UK, although they employ only one cross-section of data.

The argument above implicitly assumes that the premium is actuarially fair, depending on the amount of effort to reduce the probability of loss. Other authors (e.g. Zweifel and Breyer 1997, Kenkel 2000, de Ven and Ellis 2000, Zweifel and Manning 2000, Newhouse 2002) have pointed out that unfair premiums, by breaking the link between self-protection and premium reduction, give rise to *ex-ante* moral hazard, that is reduction in the demand for self-protection. If the regulator cannot observe the self-protective effort, *ex-ante* moral hazard imposes an externality on other individuals, who need to pay higher premiums in order to compensate for the increase in the propensity to loss. However, as Kenkel (2000) points out, as long as curative care is unable to completely restore health, individuals have incentives to prevention, even with complete financial coverage and fair premiums.

Bond and Crocker (1991) describe the use of risk classification to inhibit the *ex ante* moral hazard. In their model individuals consume some goods that are correlated with the

propensity to loss, like smoking. If the consumption of the hazardous good is observable, risk rated premiums are shown to mitigate the ex ante moral hazard, and the first best can be achieved, even if the underlying propensity to loss is unobservable. Shavell (1979) reaches a similar conclusion, showing that full coverage is preferable if prevention is observable. However, if the insurer observes preventive effort imprecisely, full coverage is not optimal, and that introduces an additional risk for the insured.

Focusing on the US context, Heffley (1982) and Miceli and Heffley (2002) analyse the claim that Health Maintenance Organizations (HMOs) have a presumable "preventive orientation", that is, "because of the contractual obligation to provide medical care for a fixed premium, prepaid groups may have a stronger obligation to deliver preventive services that reduce the need for future treatment" (Miceli and Heffley 2002, p. 429). They compare consumer demand for prevention and provider's investment on capacity under three alternative systems: pure retrospective (fee-for-service), mixed and pure prospective (pre-paid). They show that the first two systems induce optimal prevention, whilst under pure capitation consumers either over or under consume prevention, and providers manipulate the level of capacity investment in order to control demand. A similar argument has been put forward by Herring (2002), who argues that enrolee turnover reduces the benefit from investments in prevention from the health plan's perspective. Using data from employer-sponsored health plans, Herring (2002) shows that employment-induced insurer turnover results in under-investment in prevention as measured by mammography utilization.

In recent years, a topic related to prevention that has been receiving a lot of interest is the study of screening, more precisely of genetic testing (see Tabarrok 1994, Strohmenger and Wambach 2000, Fagart and Fombaron 2003, Hoel and Iversen 2002, Hoel, Iversen, Nilssen and Vislie 2006, Hoy and Polborn 2000, Hoy, Orsi, Eisinger and Moatti 2003, Polborn, Hoy and Sadanand 2006). The successful sequencing of the human genome has brought renewed optimism for the treatment of the most diverse ailments, from cancer to countless genetic related disorders. It is believed that in the future doctors will be able to prescribe specially tailored treatments and medicines based on the examination of the genetic endowment of patients.

One of the earliest analyses in this area was provided by Tabarrok (1994), who argues that the increasing availability of genetic tests produces both benefits and costs. As mentioned above, genetic testing improves the ability to detect diseases at early stages, facilitating the targeting of treatments and potentially contributing to improve health. On the other hand, that creates additional classification risk. Those individuals that are identified with high risk of developing genetic diseases might end up virtually uninsurable, given the high premiums that would be charged. Several authors have explored the potential welfare effects of genetic testing, with particular focus on whether insurers should be allowed to use test results in insurance contracts. There is no general consensus about the optimal policy, since results depend on the informational and institutional framework of each application. Some authors suggest that it is preferable to allow insurers to use information from genetic tests (e.g. Hoel and Iversen 2002, Hoel et al. 2006), whilst others sustain that genetic

testing should be banned in order to improve social welfare (e.g. Hoy et al. 2003, Polborn et al. 2006).

Although focusing on a different context, Sappington and Lewis (1999) extend the implications of partially observable prevention by considering screening costs. They model the provision of health care when the likely treatment costs per patient cannot be perfectly anticipated. However, the health care provider can perform tests which increase the level of information, but also imply a positive cost. For a relatively high screening cost the purchaser is able to implement conventional risk adjustment. However, if the screening cost is low, it is likely that providers would engage in inefficient selection of patients (dumping). In the latter case the authors propose a system of subjective risk adjustment.

Two recent papers analyse how different payment systems affect investments on prevention. In Barros and Martinez-Giralt's (2003) model, prevention determines the referral rates between primary and secondary treatments centres. They show that under cost-reimbursement there is no incentive for prevention, neither by the primary care centre nor the hospital. However, integrated management of the two units of service, combined with prospective payments are shown to stimulate prevention, which in turn decreases referral rates, and increases the efficiency in the provision of services. Barigozzi (2004) explores the optimal reimbursement rule in two opposite cases, with prevention being either a complement or a substitute to curative care. The author cites the prevention and treatment of HIV/AIDS as examples of complementary goods, since in general the early detection implies higher costs of treatment. Conversely, screening for certain types of breast cancer can be considered a substitute to treatment, because early detection prevents the development

of more serious conditions. Barigozzi (2004) shows that the optimal rule will always encourage investments in treatment. On the other hand, prevention will be promoted if it is a substitute to curative care, but will be discouraged if the two types of care are complements.

Cropper (1977) discusses the effect of age on the demand for preventive services in a human capital framework (see Grossman 1972, Grossman 2000). The author shows that, if longevity is fixed, the incentives for prevention decrease as the individual ages, because the period for pay-off of investment decreases. However, if death is endogenous and the depreciation of capital stock increases with age, then prevention will also increase with age. In an empirical application, Kenkel (1994) estimates the determinants of breast and cervical cancer screening, showing that, although the risk of cancer increases with age, women use less preventive services as they get older. The author recognises the possibility of other explanations for this pattern, but sustains that it provides evidence for the first alternative of the relationship between age and prevention.

Byrne and Thompson's (2001) model explores inconsistencies in the intertemporal evaluation of screening and prevention. Their results suggest that patients are unlikely to comply with appropriate recommendations because they have myopic expectations with respect to the benefits. In an empirical study, Wu (2003) also shows that health status affects the demand for different types of preventive services. Individuals with poor health are more likely to be screened for relatively simple conditions, like cholesterol checks, but less likely to test for several types of cancer. The author suggests this difference is due to higher levels of anxiety among sicker patients, which prevents them from testing for serious illnesses.

The present paper focuses on the incentives for the provision of prevention in competitive health insurance markets. The insurer receives a fixed premium per patient, and in turn provides both curative care, which restores the health of sick patients, and preventive care, which decreases of probability of falling ill. Examples of preventive services include screening for asymptomatic diseases, vaccination campaigns and programs to promote lifestyle changes (e.g. increasing physical activities and smoking cessation). On the other hand, patients are divided into two classes, which are heterogeneous with respect to the probability illness and the efficiency of prevention in reducing it.

We analyse two cases, the first-best, when the patient type is observable, and the unregulated competitive market equilibrium with adverse selection. Following established results in the literature (e.g. Rothschild and Stiglitz 1976, Glazer and McGuire 2000), we show that under imperfect information low-risk patients are separated by receiving less than optimal curative care. Moreover, the main contribution of the paper is to demonstrate that the level of preventive care is also distorted, with the direction of distortion depending on the relative efficiency of prevention for each risk type. Low-risk patients receive lower (higher) marginal benefit from preventive care if prevention is relatively more (less) efficient for them.

The paper is organized as follows. The next section presents the basic model, describing the patient utility, the insurer objective, the social welfare function and the effect of prevention on the probability of illness. Section 3.3 characterises the first best, which corresponds to the equilibrium when the patient type is observable, while section 3.4 analy-

ses the unregulated market equilibrium, when only the average proportion of each risk type is observable. Section 3.5 concludes.

3.2 The model

There are two types of individuals, L and H, distinguished according to the probability of illness $p^i, i \in \{L, H\}$. The probability of illness is strictly positive, and H-types are assumed to face a higher risk than L-types:

$$p^H > p^L > 0. \tag{3.1}$$

Let $\theta_i, i = H, L$, with $\theta_H + \theta_L = 1$, denote the proportion of type i individuals in the population. The probability of illness is private information of the patient, so it is not observed by the insurer. However, the insurer knows the distribution of each type of individuals. It is then possible to calculate the average probability of illness $\bar{p}$:

$$\bar{p} = \theta_H p^H + \theta_L p^L. \tag{3.2}$$

For each enrolled patient the plan receives a premium r. In return the plan provides preventive care ϕ and, if the individual gets sick, delivers curative care m. In this specification ϕ and m indicate the monetary value of health care. The costs of providing preventive and curative care are denoted by $g(\phi)$ and $h(m)$, respectively. The insurance contract then takes the form $C(\phi, m, r)$.

Patients do not make any other payment beyond the premium, and we assume that all individuals in the population take part into the contract. Additionally, we assume that "each plan can offer only one contract, and that each consumer chooses only one contract"

(Glazer and McGuire 2000, p. 1059). Initially suppose the regulator charges the same premium $r \geq 0$ for all individuals, regardless of their risk level.

3.2.1 Patient's utility

Following Glazer and McGuire (2000), the utility of a healthy individual is normalised to zero: $U_h = 0$, where the subscript refers to the health status (h stands for healthy and s for sick). If the individual becomes ill, there is a utility loss D associated with the discomfort of the disease in terms of pain and suffering, and with the reduced ability to perform the normal activities.

On the other hand, receiving medical care m increases utility by $v = v(m)$, with $v_m > 0$ and $v_{mm} < 0$. We assume that, although the insurance contract provides coverage for the monetary loss, there is typically a significant health effect which cannot be fully recovered through treatment (see Kenkel 2000). That is, curative care at best restores the original level of health of the patient, which implies:

$$U_s = v(m) - D \leq 0, \ \forall \, m. \tag{3.3}$$

The expected utility the individual receives from a contract $C(\phi, m, r)$ is given by:

$$\begin{aligned} V(C) &= (1 - p(\phi))U_h + p(\phi)U_s \qquad (3.4) \\ V(C) &= p(\phi)\,(v(m) - D) - r, \end{aligned}$$

where the simplification is due to the normalisation of the utility when healthy.

The effect of prevention

Individuals do not accrue any direct utility from preventive care. Instead, they value preventive care only for its effect on the probability of illness. Prevention can be distinguished into two categories, according to its particular effect on health (Ehrlich and Becker 1972):

1. Primary prevention or self-protection, which reduces the probability of illness occurrence, and

2. Secondary prevention or self-insurance, which reduces the health consequences of the disease, conditional on its occurrence.

Our model focuses on the first type of prevention and assumes that preventive care does not affect the level of curative care that is needed in case of illness. Consequently, the utility from curative care can be considered the same for both H-types and L-types. Examples of primary prevention include both public provided interventions such as vaccination and also lifestyle decisions related to diet, exercise and smoking. One important aspect of primary prevention is that it might be not observable by the insurer and the regulator (Barigozzi 2004).

The principal assumption of the model is that receiving preventive care ϕ decreases the probability of becoming ill $p(\phi)$:

$$p = p(\phi), \text{with } p(0) \leq 1, p_\phi < 0, p_\phi(0) = -\infty, p_{\phi\phi} > 0, \tag{3.5}$$

where subscripts denote partial derivatives. The probability of illness is strictly convex and restricted to the interval $0 < p \leq 1$. The last assumption guarantees that an interior solution for optimal prevention exists.

The model implicitly assumes that the probability of illness is monotonically decreasing on prevention. This is made for computational convenience. However, other assumptions could also be conceived. For instance, consider the example of vaccination campaigns. It might be argued that, from a population point-of-view, the first unit of vaccination has no effect on the probability of illness. This probability only starts to decrease after a minimal critical number of people is immunised. However, provided that the probability of illness is bounded between 0 and 1, the main results of the model are independent of this assumption.

Prevention can have different effects on the probability of illness, according to the individual risk type. It is possible to assume that either preventive care is at least as efficient for L-types as it is for H-types ($-p_\phi^L \geq -p_\phi^H$), or that it is more efficient for H-types ($-p_\phi^L < -p_\phi^H$).[12] In the former case, for a given amount of prevention the reduction in the probability of illness is larger for the L-types than for H-types. The intution for this assumption is that L-types tend to be healthier, and would consequently have a better response to the provision of prevention. For example, assume the risk level is primarily determined by age, with older people being exposed to higher probability of illness. In this case, it is

---

[12] Another possible assumption would be to consider that the distribution of preventive care efficiency does not follow strictly the risk classes. This could be modelled, for example, by assuming a random distribution of preventive efficiency. Each risk type would then face an exogenous probability of having either high or low efficiency of prevention. However, the main result of the paper, namely that the direction of distortion depends on the relative efficiency of prevention would still come through. Therefore, we focus on the more restrictive but simpler assumption that the efficiency of prevention follows risk classes.

reasonable to suppose that the demand for prevention would reduce with age, either because of the lower effectiveness of prevention, or because of a smaller period for pay-off of health investment (see Cropper 1977, Kenkel 1994). This example is not in compliance with the model assumptions, however, because age is an observable characteristic. A more appropriate example relates to the unobservable probability of developing a certain type of cancer. It might be argued that individuals with higher genetic propensity to developing the disease might also be less responsive to interventions aimed at reducing the probability of illness. In this case, the efficiency of prevention depends on an unobservable characteristic that is also related to risk class.

The second assumption, although less intuitive, is also possible. An example is the impact of smoking cessation on the overall health of young and old individuals. Kenkel (2000) argues that the gain of quitting for a young individual can be considered relatively small compared to an older individual. The young individual is likely to have smoked for a shorter period of time, and consequently to have imposed less harm over his/her health. Therefore, the benefit from stop smoking may be higher for the older individual (high risk).

3.2.2 Insurer's profit

Let $g(\phi)$ (with $g_\phi > 0, g_{\phi\phi} > 0$) and $h(m)$ (with $h_m > 0, h_{mm} > 0$) denote the cost of providing preventive and curative care, respectively. The insurer's expected profit per patient is given by the difference between the total amount collected in premiums and the expected cost of treatment. Thus, the expected profit of contract $C(\phi, m, r)$ is given by:

$$\Pi(C) = r - g(\phi) - p(\phi)h(m). \tag{3.6}$$

3.2.3 Social welfare

Assume a utilitarian regulator which gives the same importance for the welfare of patients and insurers alike. The welfare of the society can be seen as the weighted sum of the welfare of patients and the profits of insurers, where the weights are given by the proportion of each type of patient in the population:

$$\begin{aligned} W &= \sum_{i=L}^{H} \theta_i \left[V_i(C_i) + \Pi_i(C_i) \right] \\ &= \theta_H \left[p^H(\phi_H) \left(v(m_H) - D - h(m_H) \right) - g(\phi_H) \right] + \\ &\quad \theta_L \left[p^L(\phi_L) \left(v(m_L) - D - h(m_L) \right) - g(\phi_L) \right]. \end{aligned} \tag{3.7}$$

Note that for each type of contract the insurance premium is simply a transfer from patients to the insurer. Therefore, it does not affect the social welfare.

3.3 The first-best

Let us first assume the patient type is publicly observed by the procurer and the provider. Therefore the procurer can set both preventive and curative care at the optimal levels. This defines the first-best or complete information setting.

The procurer objective is to maximize the social welfare, by choosing the optimal levels of preventive and curative care for each patient type:

$$\begin{aligned} \max_{m_H, m_L, \phi_H, \phi_L} W &= \sum_{i=L}^{H} \theta_i \left[V_i(C_i) + \Pi_i(C_i) \right] \\ &= \theta_H \left[p^H(\phi_H) \left(v(m_H) - D - h(m_H) \right) - g(\phi_H) \right] + \\ &\quad \theta_L \left[p^L(\phi_L) \left(v(m_L) - D - h(m_L) \right) - g(\phi_L) \right]. \end{aligned} \tag{3.8}$$

The first order conditions are given by:

$$m_H^* \quad : \quad v_m(m_H^*) = h_m(m_H^*) \tag{3.9}$$

$$m_L^* \quad : \quad v_m(m_L^*) = h_m(m_L^*) \tag{3.10}$$

$$\phi_H^* \quad : \quad p_\phi^H(\phi_H^*)\left(v(m_H^*) - D - h(m_H^*)\right) = g_\phi(\phi_H^*) \tag{3.11}$$

$$\phi_L^* \quad : \quad p_\phi^L(\phi_L^*)\left(v(m_L^*) - D - h(m_L^*)\right) = g_\phi(\phi_L^*). \tag{3.12}$$

where the superscript * denotes the first best. Recall that the expected utility terms $v(m_i^*) - D - h(m_i^*)$, $i = H, L$, are negative, given the assumption of incomplete recovery from health losses.

Proof. The first order conditions are given by:

$$m_H \quad : \quad \theta_H p^H(\phi_H)\left(v_m(m_H) - h_m(m_H)\right) = 0$$

$$m_L \quad : \quad \theta_L p^L(\phi_L)\left(v_m(m_L) - h_m(m_L)\right) = 0$$

$$\phi_H \quad : \quad \theta_H\left[p_\phi^H(\phi_H)\left(v(m_H) - D - h(m_H)\right) - g_\phi(\phi_H)\right] = 0$$

$$\phi_L \quad : \quad \theta_L\left[p_\phi^L(\phi_L)\left(v(m_L) - D - h(m_L)\right) - g_\phi(\phi_L)\right] = 0,$$

which may re-arranged in order to get the expressions ((3.9))-((3.12)).

The second order conditions for the H-types' contract are:

$$\theta_H p^H(\phi_H^*)\left(v_{mm}(m_H^*) - h_{mm}(m_H^*)\right) < 0$$

$$\theta_H\left[p_{\phi\phi}^H(\phi_H^*)\left(v(m_H^*) - D - h(m_H^*)\right) - g_{\phi\phi}(\phi_H^*)\right] < 0$$

$$\theta_H p^H(\phi_H^*)\left(v_{mm}(m_H^*) - h_{mm}(m_H^*)\right)\theta_H\left[p_{\phi\phi}^H(\phi_H^*)\left(v(m_H^*) - D - h(m_H^*)\right) - g_{\phi\phi}(\phi_H^*)\right] >$$

$$\left[\theta_H p_\phi^H(\phi_H^*)\left(v_m(m_H^*) - h_m(m_H^*)\right)\right]^2.$$

Note that, according to ((3.9)), the right-hand side of the last equation is equal to zero. Hence the second order conditions for H-types are always satisfied. Similarly the second order conditions for L-types are:

$$\theta_L p^L(\phi_L^*)\left(v_{mm}(m_L^*) - h_{mm}(m_L^*)\right) < 0$$

$$\theta_L\left[p^L_{\phi\phi}(\phi_L^*)\left(v(m_L^*) - D - h(m_L^*)\right) - g_{\phi\phi}(\phi_L^*)\right] < 0$$

$$\theta_L p^L(\phi_L^*)\left(v_{mm}(m_L^*) - h_{mm}(m_L^*)\right)\theta_L\left[p^L_{\phi\phi}(\phi_L^*)\left(v(m_L^*) - D - h(m_L^*)\right) - g_{\phi\phi}(\phi_L^*)\right] >$$

$$\left[\theta_L p^L_{\phi}(\phi_L^*)\left(v_m(m_L^*) - h_m(m_L^*)\right)\right]^2.$$

A similar argument holds in this case to guarantee that the second order conditions for L-types are also always satisfied. ■

According to (3.9) and (3.10), the optimal level of curative care equalizes the marginal utility to the marginal cost. Since both types of patients enjoy the same benefit from curative care, they receive the same amount of treatment $m_H^* = m_L^* = m^*$.

On the other hand, according to (3.11) and (3.12), the optimal level of preventive care for each type of patient is achieved when the social welfare gain equals the cost of prevention. Although prevention is provided at the same cost for both types of patients, their capacity to benefit is different. Consequently, each risk type receives a different level of prevention. L-types receive more preventive care if they enjoy greater benefits in terms of reduction in the probability of illness ($-p^L_\phi \geq -p^H_\phi \Rightarrow \phi_L^* \geq \phi_H^*$), whilst the opposite is true if H-types are more benefited ($-p^H_\phi > -p^L_\phi \Rightarrow \phi_H^* \geq \phi_L^*$).

The main features of the optimal arrangement can be described as:

Proposition 3 *Under perfect information about individual risk types, the optimal menu of competitive contracts is characterized as a separating equilibrium.*

1. *Both types of patients receive the efficient level of curative care ($m_H^* = m_L^* = m^*$).*

2. *The provision of preventive care depends on the efficiency of prevention for each type. L-types receive more (less) care than H-types ($\phi_L^* \geq \phi_H^*$) if prevention is relatively more (less) efficient for their type ($-p_\phi^L \geq -p_\phi^H$). That is, $\phi_L^* \geq \phi_H^*$ if $-p_\phi^L \geq -p_\phi^H$, but $\phi_H^* \geq \phi_L^*$ if $-p_\phi^H > -p_\phi^L$.*

3.4 The unregulated market equilibrium

We now turn to the unregulated market equilibrium when the regulator and the plan cannot observe the individual type, but they know the average proportion of risk types in the population ($\theta_i, i = H, L$). Each plan decides which contract to offer and each individual chooses his preferred plan.

In the resulting separating equilibrium high risk individuals are offered their first-best contract given by:

$$\max_{m_H, \phi_H} V_H(\phi_H, m_H, r_H) = p^H(\phi_H)\,(v(m_H) - D) - r_H \tag{3.13}$$

$$\text{Subject to } \Pi_H(\phi_H, m_H, r_H) = r_H - g(\phi_H) - p^H(\phi_H)h(m_H) \geq 0, \tag{3.14}$$

where (3.14) is the zero-profit constraint, which guarantees that the total amount collected in premiums is enough to cover the expenses in both preventive and curative care. The

corresponding first-order conditions are:

$$m_H^m \;:\; v_m(m_H^m) = h_m(m_H^m) \tag{3.15}$$

$$\phi_H^m \;:\; p_\phi^H(\phi_H^m)\,(v(m_H^m) - D - h(m_H^m)) = g_\phi(\phi_H^m), \tag{3.16}$$

where the superscript m denotes the outcome of the unregulated market equilibrium.

Proof. In a competitive market the constraint ((3.14)) is binding in equilibrium. Hence substituting $r_H - g(\phi_H) - p^H(\phi_H)h(m_H)$ into ((3.13)) allows us to write the objective function of this maximisation problem as:

$$\max_{m_H,\phi_H} \mathcal{L} = p^H(\phi_H)\,(v(m_H) - D - h(m_H)) - g(\phi_H),$$

with the corresponding first-order conditions:

$$m_H^m \;:\; p^H(\phi_H)\,(v_m(m_H) - h_m(m_H)) = 0$$

$$\phi_H^m \;:\; p_\phi^H(\phi_H)\,(v(m_H) - D - h(m_H)) - g_\phi(\phi_H) = 0.$$

These equations are re-arranged in order to get expressions ((3.15)) and ((3.16)). ■

Comparing these conditions with (3.9) and (3.11) shows that the H-types' contract is equivalent to the first best, and may be characterised by $C_H^m = (\phi_H^*, m^*, r_H^*)$. The optimal level of curative care equalizes the marginal utility to the marginal cost, while optimal prevention is determined when the marginal reduction in the probability of illness for H-types times the expected utility equals the marginal cost of prevention. Therefore, the market contract for this risk category does not distort any type of care. H-types receive the first best levels of both preventive and curative care, and are charged the actuarially fair premium $r_H^* = g(\phi_H^*) + p^H(\phi_H^*)h(m_H^*)$.

Let us now consider the contract for L-types. The insurer wants to separate this category of patients, and therefore offers a contract that is different from the first-best contract purchased by the H-types. In effect, the L-types contract maximises their utility provided that H-types are not attracted to it. This contract solves the following problem:

$$\max_{m_L,\phi_L} V_L(\phi_L, m_L, r_L) = p^L(\phi_L)\left(v(m_L) - D\right) - r_L \tag{3.17}$$

$$\text{Subject to} \qquad \Pi_L(\phi_L, m_L, r_L) = r_L - g(\phi_L) - p^L(\phi_L)h(m_L) \geq 0, \tag{3.18}$$

$$\text{and} \qquad \begin{array}{c} V_H(\phi_H^*, m^*, r_H^*) = p^H(\phi_H^*)\left(v(m^*) - D\right) - r_H^* \geq \\ V_H(\phi_L, m_L, r_L) = p^H(\phi_L)(v(m_L) - D) - r_L \end{array}, \tag{3.19}$$

where (3.18) is the zero-profit constraint and (3.19) is the self-selecting constraint, which guarantees that the contract $C_L = (\phi_L, m_L, r_L)$ does not attract H-types. The first-order conditions are:

$$\begin{array}{rcl} & & v_m(m_L^m) = \\ m_L^m & : & h_m(m_L^m) + \frac{\lambda}{1-\lambda}\frac{p^H(\phi_L^m) - p^L(\phi_L^m)}{p^L(\phi_L^m)} v_m(m_L^m) \end{array} \tag{3.20}$$

$$\begin{array}{rcl} & & p_\phi^L(\phi_L^m)\left(v(m_L^m) - D - h(m_L^m)\right) = \\ \phi_L^m & : & g_\phi(\phi_L^m) + \frac{\lambda}{1-\lambda}\left(p_\phi^H(\phi_L^m) - p_\phi^L(\phi_L^m)\right)\left(v(m_L^m) - D\right) \end{array}, \tag{3.21}$$

where λ is the Lagrange multiplier associated with the incentive compatibility constraint (3.19).

Proof. Here we solve the maximization problem of the unregulated market equilibrium for L-type individuals. The Lagrangean for this maximization problem is given by:

$$\begin{aligned} \mathcal{L} \;=\; & p^L(\phi_L^m)\left(v(m_L^m) - D - h(m_L^m)\right) - g(\phi_L^m) + \\ & \lambda\left[p^H(\phi_H^*)\left(v(m^*) - D\right) - r_H^* - p^H(\phi_L^m)\left(v(m_L^m) - D\right) + p^L(\phi_L^m)h(m_L^m) + g(\phi_L^m)\right]. \end{aligned}$$

The first-order condition with respect to curative care is:

$$m_L^m : p^L(\phi_L^m)\left(v_m(m_L^m) - h_m(m_L^m)\right) - \lambda\left(p^H(\phi_L^m)v_m(m_L^m) - p^L(\phi_L^m)h_m(m_L^m)\right) = 0.$$

Adding and subtracting $p^L(\phi_L^m)v_m(m_L^m)$ to the term in parenthesis allow us to rewrite:

$$\begin{gathered} p^L(\phi_L^m)\left(v_m(m_L^m)-h_m(m_L^m)\right)- \\ \lambda\left(p^L(\phi_L^m)v_m(m_L^m)-p^L(\phi_L^m)h_m(m_L^m)+p^H(\phi_L^m)v_m(m_L^m)-p^L(\phi_L^m)v_m(m_L^m)\right)=0 \end{gathered}$$

$$(1-\lambda)\,p^L(\phi_L^m)\left(v_m(m_L^m)-h_m(m_L^m)\right) \;=\; \lambda\left(p^H(\phi_L^m)-p^L(\phi_L^m)\right)v_m(m_L^m)$$

$$v_m(m_L^m) \;=\; h_m(m_L^m)+\frac{\lambda}{1-\lambda}\frac{p^H(\phi_L^m)-p^L(\phi_L^m)}{p^L(\phi_L^m)}v_m(m_L^m).$$

On the other hand, the first-order condition with respect to preventive care is:

$$\begin{array}{rl} \phi_L^m: & \begin{gathered} p_\phi^L(\phi_L^m)\left(v(m_L^m)-D-h(m_L^m)\right)-g_\phi(\phi_L^m)+ \\ \lambda\left[-p_\phi^H(\phi_L^m)\left(v(m_L^m)-D\right)+p_\phi^L(\phi_L^m)h(m_L^m)+g_\phi(\phi_L^m)\right]=0 \end{gathered} \end{array}$$

Similarly, adding and subtracting $p_\phi^L(\phi_L^m)\left(v(m_L^m)-D\right)$ to the term in square brackets yields:

$$\begin{gathered} p_\phi^L(\phi_L^m)\left(v(m_L^m)-D-h(m_L^m)\right)-g_\phi(\phi_L^m)+ \\ \lambda\left[\begin{array}{c}\left(p_\phi^L(\phi_L^m)-p_\phi^H(\phi_L^m)\right)\left(v(m_L^m)-D\right)- \\ p_\phi^L(\phi_L^m)\left(v(m_L^m)-D-h(m_L^m)\right)+g_\phi(\phi_L^m)\end{array}\right]=0 \end{gathered}$$

$$(1-\lambda)\left[p_\phi^L(\phi_L^m)\left(v(m_L^m)-D-h(m_L^m)\right)-g_\phi(\phi_L^m)\right]+$$

$$\lambda\left(p_\phi^L(\phi_L^m)-p_\phi^H(\phi_L^m)\right)\left(v(m_L^m)-D\right)=0$$

$$p_\phi^L(\phi_L^m)\left(v(m_L^m)-D-h(m_L^m)\right)=g_\phi(\phi_L^m)+\frac{\lambda}{1-\lambda}\left(p_\phi^H(\phi_L^m)-p_\phi^L(\phi_L^m)\right)\left(v(m_L^m)-D\right).$$

The corresponding fair premium for the unregulated contract for L-types is $r_L^m=g\left(\phi_L^m\right)+$ $p^L(\phi_L^m)h\left(m_L^m\right)$. ■

According to equation (3.20), the marginal utility of curative care for L-types is higher than the marginal cost. Therefore, the level of curative care provided to L-types in the market equilibrium is below the first-best. This conclusion is similar to Rothschild and Stiglitz's (1976), and is necessary to prevent that H-types choose the contract designed for L-types. Similarly, according to (3.21), the level preventive care for L-types is also

distorted in the market equilibrium. In this case, however, the direction of distortion is determined by the relative efficiency of prevention,as explained below.

Comparing the contracts for the two risk types, we conclude:

Proposition 4 *The unregulated market equilibrium, if it exists, is characterized as a separating equilibrium.*

1. *Under the contract $C_H^m = (\phi_H^*, m^*, r_H^*)$ H-types receive the first-best marginal benefit from both types of services, and are charged the corresponding actuarially fair premium.*

2. *L-types, on the other hand, are offered the contract $C_L^m = (\phi_L^m, m_L^m, r_L^m)$. Compared to the first best, this contract provides higher marginal benefit of curative care: $v_m(m_L^m) > v_m(m^*)$.*

3. *L-types receive lower marginal benefit from preventive care if prevention is relatively more efficient for them ($-p_\phi^L \geq -p_\phi^H$), but higher marginal benefit from preventive care if prevention is relatively more efficient for H-types ($-p_\phi^L < -p_\phi^H$).*

These results are in line with Glazer and McGuire (2000) and reflect the standard Rothschild-Stiglitz separating equilibrium. L-type individuals are separated from the H-types by receiving less than the efficient levels of curative care ($m_L^m < m_H^m = m^*$).

On the other hand, when prevention is more efficient for L-types, under normal circumstances they can be expected to receive more preventive care than socially desired ($\phi_L^m > \phi_L^*$). This outcome is justified by two reasons. First note that L-types receive less

than the optimal level of curative care. Consequently, they would prefer to reduce the probability of illness to avoid utility losses. Therefore, the excess provision of preventive care will make them better off because it will increase their expected utility by reducing the probability of illness. H-types are not attracted by this contract because preventive care is relatively inefficient for them. Consequently, they prefer a contract that supplies less preventive care ($\phi_H^m = \phi_H^*$) but, at the same time, charges a relatively lower premium.

Conversely, when prevention is relatively less efficient for L-types, under normal circumstances they can be expected to receive less preventive care than socially desired ($\phi_L^m < \phi_L^*$). In this case, L-types are offered lower provision of both types of services. Nevertheless, they are not attracted by the H-types' contract because that would imply a higher premium.

It remains to show formally that the omitted incentive compatibility constraint for L-types is satisfied, that is, to check that L-types have indeed no incentives to mimic H-types. First, note that the incentive constraint of H-types is binding:

$$\begin{aligned} V_H(\phi_H^*, m^*, r_H^*) &= V_H(\phi_L^m, m_L^m, r_L^m) \\ p^H(\phi_H^*)\,(v(m^*) - D) - r_H^* &= p^H(\phi_L^m)(v(m_L^m) - D) - r_L^m \\ p^H(\phi_L^m)(v(m_L^m) - D) - p^H(\phi_H^*)\,(v(m^*) - D) &= r_L^m - r_H^*. \end{aligned} \tag{3.22}$$

Now consider the incentive constraint of L-types:

$$\begin{aligned} V_L(\phi_L^m, m_L^m, r_L^m) &\geq V_L(\phi_H^*, m^*, r_H^*) \\ p^L(\phi_L^m)(v(m_L^m) - D) - r_L^m &\geq p^L(\phi_H^*)\,(v(m^*) - D) - r_H^* \\ p^L(\phi_L^m)(v(m_L^m) - D) - p^L(\phi_H^*)\,(v(m^*) - D) &\geq r_L^m - r_H^*. \end{aligned} \tag{3.23}$$

Hence, from (3.22) and (3.23), the incentive constraint for L-types is satisfied if:

$$\begin{aligned} p^L(\phi_L^m)(v(m_L^m)-D) - p^L(\phi_H^*)\,(v(m^*)-D) \;&\geq\; p^H(\phi_L^m)(v(m_L^m)-D) - \\ &\quad\; p^H(\phi_H^*)\,(v(m^*)-D) \end{aligned}$$

$$\left(p^H(\phi_H^*) - p^L(\phi_H^*)\right)(v(m^*)-D) \;\geq\; \left(p^H(\phi_L^m) - p^L(\phi_L^m)\right)(v(m_L^m)-D).$$

Taking into account that curative care never fully recovers the patient health ($v(m) - D \leq 0, \forall m$), this implies:

$$\frac{p^H(\phi_H^*) - p^L(\phi_H^*)}{p^H(\phi_L^m) - p^L(\phi_L^m)} \leq \frac{v(m_L^m) - D}{v(m^*) - D}. \tag{3.24}$$

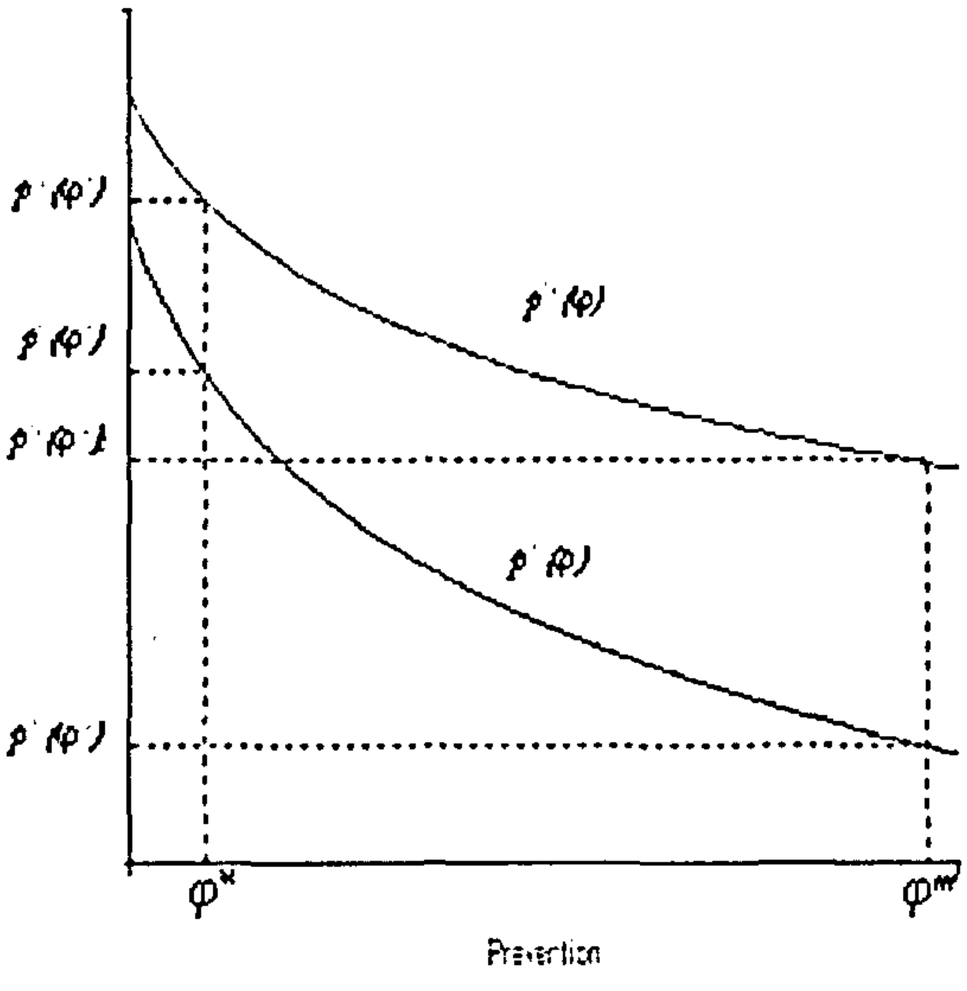
Since m^* maximises the patient utility ($0 \geq v(m^*) - D \geq v(m_L^m) - D$), the right hand side of this expression is always equal or greater than one ($\frac{v(m_L^m)-D}{v(m^*)-D} \geq 1$).

Figure 3.1 below shows the two alternative assumptions on the effect of prevention and compares the effect of prevention in the first-best and in the market equilibrium. Panel a shows the case where prevention is relatively more efficient for L-types ($-p_\phi^L \geq -p_\phi^H$). As prevention increases, the distance between the curves also increases, suggesting that the reduction in the probability of illness is bigger for L-types. Conversely, Panel b depicts the case where prevention is more efficient for H-types ($-p_\phi^L < -p_\phi^H$). The distance between the curves decreases as prevention increases. For all levels of prevention, we assume the probability of illness for H-types is always higher than for L-types, therefore the curves never cross.

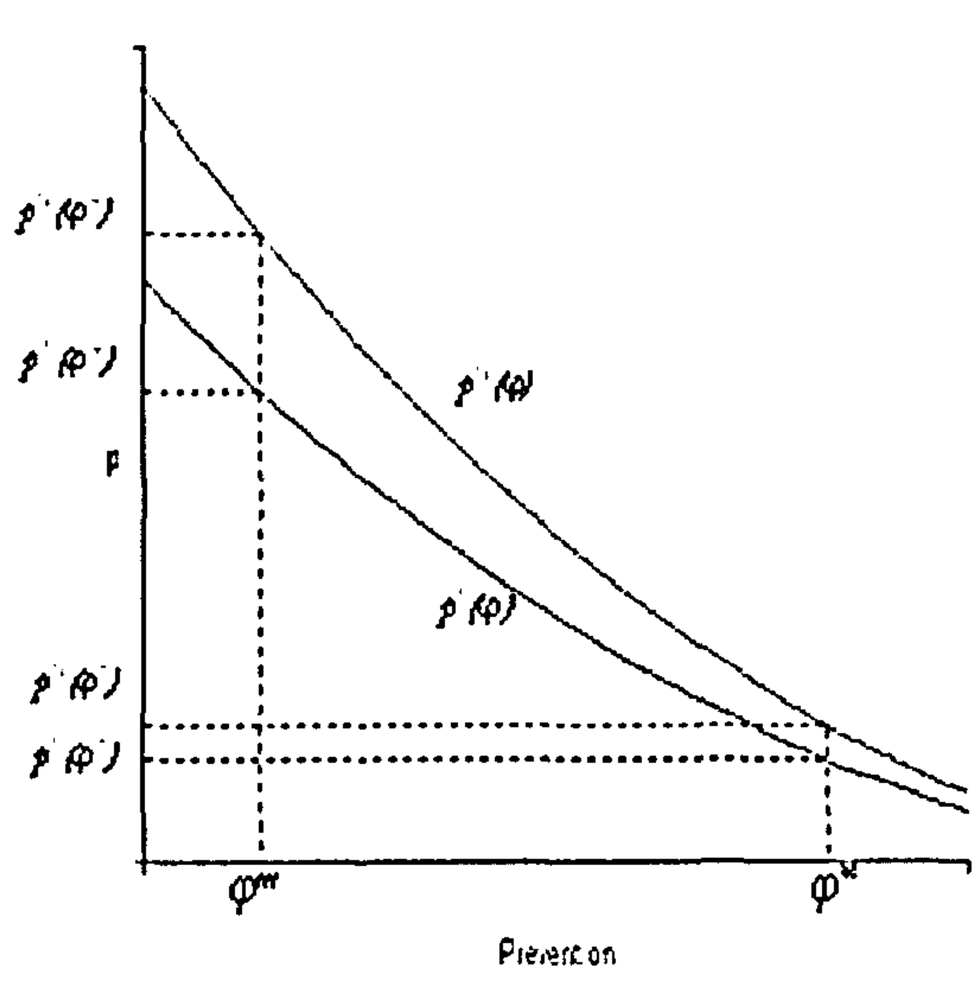
The ratio on the left hand side of equation (3.24) depends on the relative efficiency of prevention. When prevention is more efficient for L-types the market distortion will imply higher prevention for L-types compared to the first best ($\phi_L^m > \phi_H^*$). As depicted by Panel a

in Figure 3.1 below, in this case the difference between the probabilities of illness increases with prevention, therefore clearly we have $\frac{p^H(\phi_H^*)-p^L(\phi_H^*)}{p^H(\phi_L^m)-p^L(\phi_L^m)} \leq 1$.

Conversely, if prevention is more efficient for H-types the market equilibrium will provide L-types with lower prevention than socially optimal ($\phi_L^m < \phi_H^*$). Since the difference between the probabilities of illness decreases with prevention, in this case we will also have $\frac{p^H(\phi_H^*)-p^L(\phi_H^*)}{p^H(\phi_L^m)-p^L(\phi_L^m)} \leq 1$ (Panel b).



(a) Prevention is more efficient for L-types



(b) Prevention is more efficient for H-types

Fig. 3.1. Comparison of prevention in the first-best and the market equilibrium

Hence, the incentive compatibility constraint for L-types is always satisfied, regardless of the assumption about the relative efficiency of prevention, since $\frac{v(m_L^m)-D}{v(m^*)-D} \geq 1 \geq \frac{p^H(\phi_H^*)-p^L(\phi_H^*)}{p^H(\phi_L^m)-p^L(\phi_L^m)}$.

3.5 Conclusions

This paper has analysed the incentives for provision of curative and preventive care in a competitive health insurance market with adverse selection. Patients belong to two classes, which differ with respect to the probability illness and the efficiency of prevention in reducing it. We have analysed two informational frameworks: the first-best and the unregulated market with adverse selection, when patient type is unobservable.

Following established results in the literature (e.g. Rothschild and Stiglitz 1976, Glazer and McGuire 2000), under imperfect information low-risk patients are separated by receiving less than optimal curative care. Moreover, the main contribution of the paper is to demonstrate that the level of preventive care is also distorted, and that the direction of distortion depends on the relative efficiency of prevention for each risk type. Low-risk patients receive lower (respectively, higher) marginal benefit from preventive care if prevention is relatively more (less) efficient for them.

This result might have implications for the implementation of risk adjustment, in particular for the method of optimal risk adjustment proposed by Glazer and McGuire (see Glazer and McGuire 2000, Glazer and McGuire 2006). According to this method, the best way to avoid health plans distorting the amount and quality of services under capitation systems is to over- and underpay based on observable characteristics of individual enrolees.

However, this applies to curative care only. As the results of this analysis suggested, health plans face different incentives for the provision of curative and preventive care. It is desirable that optimal risk adjustment policies take this into account. Otherwise, the incentives for the provision of preventive care might in fact become even more distorted.

A potential empirical extension of this analysis might focus on policies that are currently in place among private health insurance regarding the access to preventive care. There are currently several examples of plans that provide financial benefits for enrolees that adopt healthy lifestyles. It would be interesting to test to what extent employers are taking that into account when providing the choice of health plans to employees and to measure the potential economic benefits that might be accrued from this.

Chapter 4
Socioeconomic inequalities in the burden of chronic diseases

4.1 Introduction

Recent publications from the World Health Organisation (WHO) and other international organizations (see Epping-Jordan et al. 2005, WHO 2005, World Bank 2005) highlight the increasing epidemiological burden and negative socioeconomic impact of chronic diseases worldwide. These publications indicate that contrary to widely held notions, chronic diseases affect the poor and are prevalent in low- and middle-income countries.

There are also suggestions that poverty predisposes individuals to chronic diseases. If these were so, chronic diseases may worsen socioeconomic inequality, and cause or aggravate poverty in individuals or households (Wagstaff 2002). Policy makers have been particularly targeted in the advocacies for intervention, as there are often elements of equity, poverty alleviation and social justice in many countries' political stewardship and health agenda. Wagstaff (2002) noted the increasing tendency for international agencies and countries to define intervention goals in terms of poverty reduction (Department for International Development 1997) as well as a broadening interpretation of the term poverty (World Bank 2001). One common trend among these reports is the lack of empirical evidence which clearly demonstrates a) the existence of socioeconomic inequality in the bur-

den of chronic diseases and b) that chronic diseases contribute to worsening socioeconomic inequality. This tends to weaken the evidence base for advocacy and intervention planning.

In this paper, we present an estimate of the socioeconomic related inequality in chronic diseases at the household level in Russia and Brazil, decomposing the estimated inequality into its contributing determinants. In addition, we show how the estimated inequalities have changed with changing prevalence of chronic diseases in the two countries over the period considered.

We show that in both countries poorer households face considerably higher probability of being affected by chronic diseases, and that socioeconomic inequality in chronic diseases has increased significantly in Russia during the period, but less so in Brazil. The concentration index in Russia starts at -0.021 in 2000, peaks at -0.046 in 2003, and is equal to -0.035 in 2004. In Brazil the concentration changes from -0.028 in 1998 to 0.003 in 2003.

The health inequity index, calculated by standardising the effect of demographic variables, is around -0.01 in Russia (with rising trend) and -0.06 in Brazil. Due to their indirect association with socioeconomic status, the standardising variables have opposite effects in each country, reducing the level of observed health inequality in Brazil, but increasing in Russia. This points out to the differential impact of recent economic changes on the socioeconomic inequality among demographic groups in each country.

The decomposition analysis highlights the importance of socioeconomic status, comorbidities and education to explain worsening health inequalities over the time. Moreover,

the bulk of this effect is due to changes in elasticities with respect to the determinants of chronic diseases, rather than changes in the concentration of such determinants.

The remainder of the paper is organised as follows. Section 4.2 reviews some of the evidence on the relation between health inequalities and socioeconomic status, emphasizing the role of chronic diseases risk factors (smoking, alcohol, high blood pressure, high cholesterol, low fruit and vegetable intake, overweight and physical inactivity). It also introduces the univariate and bivariate approaches to estimating health inequalities. Section 4.3 gives a detailed description of the methods to measure socioeconomic inequalities, focusing on the computation of concentration indices, the indirect standardisation using non-linear regression and the decomposition of changes in inequality. While the data is described in section 4.4, section 4.5 presents the results of the regression analysis, the indices for total inequality, the decomposition analysis and the decomposition of changes in inequalities (using both the Oaxaca decomposition and the total differential decomposition proposed by Wagstaff, van Doorslaer and Watanabe (2003)). Section 4.6 concludes.

4.2 Health inequalities and the poor

Much of the methodological and empirical work on measuring inequality in health has been carried out in Europe, and shows that inequalities in health (ill-health and access to care) almost always disfavour the poor. Overall, these studies suggest that health inequalities seem to be widening rather than narrowing (Wagstaff 2002) in both developing (see Victora, Vaughan, Barros, Silva and Tomasi 2000, Wagstaff et al. 2003) and developed countries (see Mackenbach and Kunst 1997, Mackenbach, Kunst, Cavelaars, Groenhof and

Geurts 1997, Pappas, Queen, Hadden and Fisher 1993, Schalick, Hadden, Pamuk, Navarro and Pappas 2000, Vega, Hollstein, Delgado, Perez, Carrasco and Marshall 2001, Houweling, Kunst, Borsboom and Mackenbach 2006). This situation may also be related to the skewed distribution of the proximate determinants which influence health outcomes at individual, household or community levels to disfavour the poor. These factors also vary widely between households (Wagstaff 2002) and among socioeconomic groups.

There are reasons to assume that this health inequality scenario is also relevant to chronic diseases. Risk factors associated with high incidence of chronic diseases tend to be concentrated among the poor and increasing in incidence in poor countries. For instance smoking and poor diet, tend to be concentrated among the lower socioeconomic groups in the United States of America and northern Europe. Incidence of chronic diseases and outcomes such as mortality rates are higher in the lower socioeconomic groups than the well-off which usually have better access to care than the poor in many societies (Wagstaff 2002). Higher income is associated with higher utilization of health services in developing countries (Castro-Leal, Demery and Mehra 1999). Human assets such as knowledge, literacy and education which influence the exposure to and the impact of chronic diseases, tend to be lower among the poor. These factors also reflect on the healthcare seeking behaviour of the less well off in the population. Perhaps the two most striking factors that might worsen inequality are the catastrophic health expenditure shocks and the tendency for missed income opportunities that follow a chronic disease. Calling to mind that chronic diseases present long term implications by nature, all these will tend to worsen existing socioeconomic and health inequalities. Decomposing the estimates of the inequality by the relative

contribution of these factors provides additional illumination to this process. This could be of relevance to planning interventions for chronic disease and evaluating such interventions.

There are two distinct approaches to estimating health inequality in the literature. First, the univariate approach (Wolfson and Rowe 2001) estimates the overall inequality in health (see Gakidou, Murray and Frenk 2000, Grand 1987). All inequality is measured irrespective of the characteristics of the units involved. This is typically done by constructing the Lorenz curve (Wagstaff and van Doorslaer 2004) which plots the cumulative proportion of individuals ranked by health on the x-axis against the cumulative proportion of health on the y-axis. Twice the area between it and the diagonal (equality line) and the curve equals the Gini coefficient, G.

The second approach (bivariate, according to Wolfson and Rowe (2001)) looks at a subset of health inequalities for instance occurring across the distribution of socioeconomic status (SES) (see Pamuk 1985, Schalick et al. 2000, Vagero and Erikson 1997, van Doorslaer et al. 1997, Wagstaff and van Doorslaer 2004) in contrast to ranking by health status. There is much debate around normative and ethical issues on which approach captures policymaker' or societal concerns and attempts have been made to bring these approaches together in a unifying methodology (Wagstaff and van Doorslaer 2004). Though this unification may help to better clarify the normative issues involved, our primary purpose is to explore empirical evidence of, and changes in socioeconomic inequality in chronic diseases. In doing so, we aim to explore evidence of the relationship between chronic diseases and poverty and social justice which are of interest to policy makers. In addition, we

aim to explore the determinants of SES inequality and how changes in these determinants are associated with the changes in inequality. We have therefore proceeded with the second approach because it directly link estimated inequality to socioeconomic status such as household income or consumption which levels could be indicative of poverty states.

4.3 Methods

4.3.1 Measuring SES inequalities in chronic disease

There are a number of ways in which relative inequalities in chronic diseases could be estimated. We use the concentration index (CI) approach which has been extensively applied in relevant literature (see Wagstaff, Paci and van Doorslaer 1991, van Doorslaer et al. 1997). The CI has been argued to be more appropriate than inequality indices derived from social welfare function if equity is defined with the social justice approach (Bommier and Stecklov 2002). Assuming we have a cardinal measure of health (utility) y_i, (and every one is ranked by a socioeconomic variable e.g. household consumption beginning from the lowest) a health concentration curve plots the cumulative proportion of the population against the cumulative proportion of health. Everyone enjoys the same health if the plot coincides with the diagonal. If it lies below the diagonal, inequalities in health exist in favour of the richer members of the society, proportionately to the degree that it lays away from the diagonal. The health concentration index, C, is defined as twice the area between the concentration curve and the diagonal, taking a value of zero if the curve coincides with the diagonal and positive (negative) values when it lies below (above) the diagonal.

C can be computed as (see Kakwani, Wagstaff and van Doorslaer 1997, Wagstaff et al. 1991):

$$C = \frac{2}{\mu} \sum_{i=1}^{N} y_i R_i - 1, \tag{4.1}$$

where μ is the mean health of the sample, N is the sample size, and R_i is the relative fractional rank of the ith individual in the SES distribution. Kakwani et al. (1997) showed that C might be computed by estimating β in the following convenient regression:

$$2\sigma_R^2 \left[\frac{y_i}{\mu}\right] = \alpha_2 + \beta_2 R_i + u_i, \tag{4.2}$$

where σ_R^2 is the variance of R_i and the estimator of β equals to:

$$\hat{\beta} = \frac{2}{\mu} \sum_{i=1}^{N} (y_i - \mu)\left(R_i - \frac{1}{2}\right). \tag{4.3}$$

As has been demonstrated by Wagstaff et al. (2003), the linear regression model is a straightforward way of decomposing inequalities to the contributions of various determinants of health:

$$y_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i. \tag{4.4}$$

Estimates from this regression model can then be used to decompose the concentration index as:

$$C = \sum_k \left(\frac{\beta_k \bar{x}_k}{\mu}\right) C_k + \frac{GC_\varepsilon}{\mu}, \tag{4.5}$$

where μ is the mean of y, $\bar{x}_k$ the mean of x_k, C_k the concentration index for x_k and GC_ε and is the generalized concentration index for ε_i. This shows that C can be viewed as being made of two components. The first being the deterministic or explained component is equal to the weighted sum of the concentration indices of the regressors where weights are simply the elasticities of y with respect to x_k. The second component (which can be computed

as a residual) is the unexplained component, which reflects the inequality in health that cannot be explained by systematic variation in the x_k across consumption groups. It is also regarded as a generalised concentration index for ϵ_i defined as:

$$GC_\epsilon = \frac{2}{n}\sum_{i=1}^{n}\epsilon_i R_i,$$

which is analogous to the Gini coefficient corresponding to the generalised Lorenz curve (Shorrocks 1983). This decomposition shows how each of the determinant's separate contribution to the explained consumption-related health inequity can be decomposed to its health elasticity and its consumption-related inequity (C_k), allowing further decomposition of each factor's contribution into these two terms (van Doorslaer and Jones 2003).

4.3.2 Indirect standardisation approach

A common problem with the computation of the concentration index refers to the presence of other factors, such as age and gender, which are likely to be associated with both SES and health. Although such standardising variables influence the distribution of health, in general their effect is not amenable to policy intervention. Therefore, their effect has to be neutralised in order to obtain a measure of the potentially avoidable health inequality. The relevant measure in this case is the inequity in the distribution of chronic diseases, which corresponds to the difference between the observed inequality and the inequality that one would observe if the standardising variables were proportionally distributed in the population. It is crucial to standardise the prevalence of chronic diseases since failing to account for this effect will cause the concentration index to "overstate the extent of avoidable or policy relevant income related inequality in health" (Gravelle 2003).

There are two possible methods for standardisation, direct and indirect. Following Wagstaff and van Doorslaer (2000) and O'Donnell, van Doorslaer, Wagstaff and Lindelow (2007), we choose to use the indirect standardisation approach. In this context, the indirect standardisation approach is preferable because it controls for non-confounding variables (non-standardising) which nonetheless are correlated with the confounding variables. Moreover, the indirect standardisation does not require the use of grouped data. This is more appropriate in this case because we want to keep the focus on households. If the direct standardisation approach were used instead, we would have to take into account the effect at the level of groups defined over the distribution of the socio-economic status indicator (O'Donnell et al. 2007, p. 61). Examples of other studies that also rely on the indirect standardisation method include Kakwani et al. (1997), van Doorslaer et al. (2000), van Doorslaer and Koolman (2004) and van Doorslaer, Koolman and Jones (2004).

Once we obtain the indirect standardised health indicator $\hat{y}_i$, it is possible to calculate its concentration index $\hat{C}$ and its standard error using the convenient regression method, in a similar way to observed health (see van Doorslaer et al. 2000, Wagstaff and van Doorslaer 2000). That is done through equations analogous to equations (4.2) and (4.3), where y_i is replaced by $\hat{y}_i$.

Finally, following Wagstaff and van Doorslaer (2000), the measure of health inequity or potentially avoidable health inequality HI_{WV} is given by the difference between the concentration indices for observed and standardised health indicators:

$$HI_{WV} = C - \hat{C} \tag{4.6}$$

Wagstaff and van Doorslaer (2000) argue that obtaining an estimate for the standard error for HI_{WV} is not trivial since C and $\hat{C}$ are not independently distributed. However, following Kakwani et al. (1997), they show the standard error for HI_{WV} can be estimated using the following convenient regression:

$$2\sigma_R^2 \left[\frac{y_i}{\mu} - \frac{\hat{y}_i}{\hat{\mu}}\right] = \alpha_3 + \beta_3 R_i + u_i, \tag{4.7}$$

where $\hat{y}_i$ and $\hat{\mu}$ stand for the indirectly standardised health indicator and its sample mean, respectively. As before, HI_{WV} and its standard error are given by the OLS estimation of β_3. Ultimately, that is the relevant measure of inequity in the prevalence of chronic diseases.

4.3.3 Non-linear regression models

In essence, the indirect standardised health indicator corresponds to the predicted values from a health equation. Early studies would include only the standardising variables among the controls, and then compute the fitted values from the regression. This approach, however, has been criticized on the basis that if there are other non-standardising variables relevant to explain health, the failure to control for them would give rise to omitted variables bias (Schokkaert and van de Voorde 2004).

Following Gravelle (2003), van Doorslaer and Koolman (2004) and van Doorslaer et al. (2004), we avoid this problem by distinguishing the explanatory variables in equation (4.4) into three groups $x = (x^r, x^s, x^p)$: the SES indicator x^r, the standardising variables (vector x^s) and the non-standardising, policy relevant variables (vector x^p). The indirectly standardised health indicator is then obtained as the fitted values of a regression of health

against these three groups, with the SES indicator and the policy relevant variables fixed at their sample means ($\bar{x}^r$ and $\bar{x}^p$ respectively).

Frequently, categorical variables measure health, in our case the incidence of chronic diseases is proxied by a binary indicator of whether at least one adult in the household reported chronic disease. The intrinsic nonlinear nature of the health regression implied in this case prevents estimation by OLS and calls for a linear approximation in the decomposition of the concentration index. Assuming a general nonlinear function G describes the health function, equation (4.4) can be restated as:

$$E(y_i|x_i) = G(\alpha + \sum_k \beta_k x_{ki}). \tag{4.8}$$

The indirectly standardised health indicator $\hat{y}$ is computed as the fitted values of this regression, setting SES and policy relevant variables at their sample means:

$$\hat{y}_i = G(\alpha + \sum_s \beta_s x_i^s + \bar{x}_i^r + \sum_p \beta_p \bar{x}_i^p). \tag{4.9}$$

The fact that we estimate a non-linear regression model has some implications for the decomposition of the concentration index, compared to the linear case given in equation (4.5). We follow van Doorslaer et al. (2004), who have proposed the use of a linear approximation to the decomposition in equation (4.5) based on the partial effects estimated from the non-linear model (4.9). The partial effects representation of the linear approximation is given by:

$$y_i = \alpha + \sum_k \beta_k^m x_{ki} + u_i, \tag{4.10}$$

where β_k^m are the partial effects from the non-linear model.[13]

---

[13] Partial effects are computed as the sample means of individual partial effects. For binary variables that is equivalent to the sample average of individual β_k^m for observations reporting that attribute. For continuous

According to van Doorslaer et al. (2004), some particular aspects should be taken into account when interpreting the results from the partial effects representation. First, recall the partial effect from a particular covariate estimated by a non-linear model is not constant over the range of the variable. Therefore, the standardisation procedure might not completely neutralise the effect of the standardising variables. In addition, contrary to the linear case, the linear approximation will not obtain an exact decomposition of the concentration index. That will be captured by the term u_i, which includes any approximation error from the partial effects representation.

4.3.4 Decomposing changes in inequalities

In explaining how the causes of changes in inequality are affected by changes in the determinants, we follow the approach put forward by Wagstaff et al. (2003). Using an Oaxaca-type decomposition method and denoting the elasticity of y with respect to x_k at time t by η_{kt}, the change in inequality (ΔC) over time can be computed as:

$$\Delta C = \sum_k \eta_{kt-1} \left(C_{kt} - C_{kt-1}\right) + \sum_k C_k \left(\eta_{kt} - \eta_{kt-1}\right) + \Delta \frac{GC_{\epsilon t}}{\mu_t}. \tag{4.11}$$

This approach allows disentangling the source of changes in health inequalities (C). It is possible to assess the "extent to which changes in health inequality are due to changes in inequality in the determinants of health, rather than to changes in their elasticities"(Wagstaff et al. 2003, p. 213).

---

variables, we calculate the individual partial effect as the change in the linear prediction of $\hat{y}$ around the observed value, holding other variables constant.

However, using the Oaxaca decomposition (4.11) it is not possible to disentangle the changes occurring within the elasticity (η_{kt}), whose components may change in different directions possibly having offsetting effects. For instance, a change in η_k may be due more to a change in β_k rather than a change in x_k or vice versa. In certain circumstances such as the evaluation of a programme, this offsetting effect may introduce errors in understanding the changes that have occurred.

As proposed by Wagstaff et al. (2003), a more detailed decomposition can be achieved by taking the total differential of equation (4.5) and allowing changes in the α, the β_k, the $\bar{x}_k$ and the C_k to affect C directly and indirectly through μ. The change in C (ΔC), is approximated by (see appendix of Wagstaff et al. (2003) for proof):

$$dC = \frac{dC}{d\alpha}d\alpha + \sum_k \frac{\bar{x}_k}{\mu}(C_k - C)\,d\beta_k + \sum_k \frac{\beta_k}{\mu}(C_k - C)\,d\bar{x}_k + \sum_k \frac{\beta_k \bar{x}_k}{\mu}dC_k + d\frac{GC_\epsilon}{\mu}. \tag{4.12}$$

This equation shows that change in the various components will produce different changes in the estimated inequality. For instance, since the probability of having at least one adult in the household with chronic diseases, being our y variable, is increasing in household ill-health, the average value is positive and C takes a negative value (that is, chronic disease is higher amongst the poor). In this case the first term in equation (4.12) $\frac{dC}{d\alpha} > 0$, so that a reduction in the prevalence of chronic diseases in the population results in C becoming more negative (worsening inequality). This indicate that a given reduction in the prevalence of chronic diseases represents a bigger proportional reduction for the better-off.

Compared to (4.11), (4.12) allows us to further decompose the changes in inequalities over time. For instance, as explained above it is possible to consider the effect on

inequalities in chronic diseases of changes in the prevalence of chronic diseases. Moreover, it is possible to determine whether changes in inequalities that result from changes in the elasticity of a particular explanatory variable are due either to: a) changes in the effect of that variable (β_k or the second term in (4.12)) or b) to changes in the average level of the variable ($\bar{x}_k$ or the third term in (4.12)). As with the Oaxaca decomposition, the total differential approach (4.12) also allows the evaluation of the impact of changes in the concentration index of explanatory variables and the residuals (fourth and fifth terms in (4.12), respectively).

4.4 Model, data and variable definitions

Our data consist of the latest five panels of the Russian Living Standards Measurement Surveys (LSMS) from 2000 to 2004, and the Brazilian National Household Survey (PNAD) for 1998 and 2003. Earlier panels of the Russian LSMS are somewhat inconsistent in the definition of some of the variables of interest. The focus is on inequalities in the probability of at least one individual reporting chronic disease (heart disease, hypertension and diabetes).

We have chosen to model this variable using a probit model as function of a vector of household variables (X_1), standardising variables (X_2) and fixed effects at the level of the community (X_3).

The household variables (X_1 vector) include controls for educational level of adults (average years of schooling among adults and proportion of adults with secondary and higher education), occupational status (employed, unemployed, economically inactive),

civil status of head of household[14], insurance status of household, number of overweighs, number of smokers and indicators of alcohol consumption among adults (risk factors associated with chronic diseases), and, in Brazil, the proportion of adults self-declared white. We also control for general health status, including the proportion of adults reporting non chronic (acute) illness in the period prior the interview, the proportion of adults with bad or very bad self-assessed health, and, in Brazil, the proportion of adults with at least one inpatient stay in the last 12 months and the proportion of adults with difficulty to perform at least one activity of daily life.

Our standardising variables (vector X_2) include essentially the age and gender composition of the household, which reflect the dependency ratio and ratio of male to female in household. Variables used include the proportion of men between 18 and 65 years, the proportion of women between 18 and 65 years, the proportion of men above 65 years, the proportion of women above 65 years and the average age of adults in the household. As explained above, the estimates of this non-linear regression are used to predict the standardised indicator of chronic diseases prevalence.

Finally, the community-level variables (vector X_3) included are dummies for urban-rural location and regional dummies. Regional dummies are an important control because the regions have markedly different levels of economic development. In both countries the reference category (excluded) is the richest region of the country, namely Moscow for Russia and the South-East region (where São Paulo and Rio de Janeiro are located) for

---

[14] The civil status of the household head is included to control for the fact that single-parent families can be expected to have a more fragile economic situation in many cases.

Brazil. Table 4.1 presents a detailed description of variables, including the codification of the regional dummies.

Table 4.1. Descriptive statistics

		Russia					Brazil	
		2000	2001	2002	2003	2004	1998	2003
ses	Log of household consumption (Russia) or income (Brazil) per equivalent adult	5.81	5.96	6.01	6.09	6.10	5.61	5.62
chronic	Dummy = 1 if at least one adult reported suffering from chronic disease (heart disease, hypertension or diabetes)	0.58	0.65	0.67	0.65	0.65	0.38	0.40
nonchronic	% adults reporting non chronic illness in the 30 days (Russia) or 2 weeks (Brazil) prior to the interview	0.43	0.44	0.43	0.43	0.42	0.07	0.08
bad	% adults with bad or very bad self assessed health	0.20	0.19	0.18	0.19	0.18	0.05	0.05
inpatient	% adults with at least one inpatient stay in the last 12 months						0.08	0.08
adl	% adults with difficulty to perform at least one activity of daily life						0.16	0.16
overweight	% overweight or obese adults	0.48	0.48	0.50	0.50	0.50		
smoke	% smoker adults	0.28	0.30	0.30	0.30	0.30		
alcohol	Average daily intake of alcohol among adults (in grams)	13.78	13.78	14.55	13.58	12.16		
kids	% children (below 18 years)	0.18	0.17	0.17	0.17	0.17	0.24	0.22
men	% men between 18 and 65 years	0.23	0.24	0.24	0.25	0.25	0.32	0.33
women	% women between 18 and 65 years	0.27	0.27	0.27	0.28	0.28	0.34	0.35
oldmen	% men above 65 years	0.09	0.09	0.09	0.08	0.07	0.04	0.04
oldwomen	% women above 65 years	0.23	0.23	0.23	0.23	0.23	0.05	0.06
age	Average age of adults in the household	47.22	47.25	47.25	46.95	46.94	40.14	40.63
secondary	% adults with secondary or higher education	0.63	0.65	0.66	0.67	0.68	0.33	0.41
education	Average years of schooling among adults	8.77	8.89	8.95	9.00	9.04	5.99	6.63
insured	% adults covered by private health insurance	0.72	0.77	0.78	0.78	0.79	0.20	0.20
inactive	% economically inactive adults	0.49	0.49	0.48	0.48	0.47	0.24	0.23
unemployed	% unemployed adults	0.04	0.03	0.03	0.03	0.03	0.08	0.09
white	% self-declared white adults						1.27	1.17
headmale	Gender of head of household	0.73	0.73	0.72	0.72	0.73	0.76	0.73
headsingle	Civil status of head of household	0.24	0.25	0.26	0.25	0.25	0.29	0.32
urban	Dummy = 1 for urban household	0.72	0.74	0.74	0.74	0.75	0.83	0.87
region1	Dummy = 1 for Siberia (Russia) or North region (Brazil)	0.20	0.18	0.17	0.19	0.19	0.07	0.10
region2	Dummy = 1 for Central (Russia) or North-East region (Brazil)	0.21	0.19	0.19	0.19	0.19	0.29	0.30
region3	Dummy = 1 for Volga (Russia) or South region (Brazil)	0.20	0.18	0.18	0.18	0.18	0.18	0.17
region4	Dummy = 1 for Caucasus (Russia) or Centre-West region (Brazil)	0.13	0.12	0.12	0.12	0.12	0.08	0.09
region5	Dummy = 1 for Ural (Russia) or Brasilia DF (Brazil)	0.15	0.14	0.14	0.14	0.14	0.03	0.03
region6	Dummy = 1 for Moscow (Russia) or South-East region (Brazil) - Reference region	0.05	0.13	0.14	0.12	0.12	0.35	0.31

Two variables are used as the key ranking indicator of socioeconomic status: household consumption in Russia and household income in Brazil. Of these two, household consumption has been argued to represent a better measure of living standards (Wagstaff and van Doorslaer 2004). Income ranking is problematic particularly for our data as a substantial part of income are not obtained in the formal sector and are difficult to clearly record. Income was self-reported in the LSMS surveys. According to O'Donnell et al. (2007) "income is an inferior measure, not only because of measurement challenges, but

also because for most households the fluctuation in income over time does not imply commensurate changes in living standards. In other words, if a household suffers a temporary negative income shock due to illness, but is able to maintain consumption through savings or insurance, it may be misleading to rank the household based on income or to express out-of-pocket payments as a share of income" (O'Donnell et al. 2007, p. 80).

This suggests that consumption is a more reliable indicator of socio-economic status and therefore should be a preferred choice over income. However, for the Brazilian survey we needed to use income as SES ranking variable since there was no information on expenditures. In both countries, the effect of inflation has been taken into account and monetary variables have been converted into international dollars of 2000.

The statistics presented in Table 4.1 provide some interesting insights into the recent evolution of the prevalence of chronic diseases and its determinants in Russia and Brazil. In both countries, it is possible to discern a tendency of increasing rates of chronic disease prevalence.

The average probability of having at least one individual reporting chronic in the household is considerably higher in Russia than in Brazil, as is the average proportion of individuals with bad or very bad self assessed health. A number of factors might contribute to explain this, including differences in health literacy, health seeking behaviours and access to diagnosis information. However, this was not further explored because the paper focuses mainly on inequalities in chronic diseases and not on explaining the differences in prevalence across the two countries.

Related to this point is the fact that on average the Russian sample is older. In Brazil children below 18 years form the third more populous group and the average age among adults is 40 years, whilst in Russia this position is occupied by the group of old women and the average age among adults is 47 years.

The Russian data provides a richer description of the evolution of risk factors associated with chronic diseases. Looking at the daily intake of alcohol among adults, we notice a tendency for decrease, in particular from 2002 onwards. The proportion of overweight and smokers initially increased, and then remained relatively stable, around 50% and 30% respectively.

Educational levels in Russia are markedly higher than in Brazil, with higher average years of schooling among adults, and higher proportion of adults with secondary and higher education. Although the proportion of inactive individuals in Russia has been falling, it remains significantly higher than in Brazil. Finally, the Brazilian sample is clearly more urban than the Russian. Also, notice that although the Brazilian capital (Brasília) is located in the Centre-West region, we enter it as a separate dummy because its socio-demographic profile is markedly distinct from the rest of the region.

4.5 Results

We proceed our analysis by first computing the static household inequality for chronic diseases and its determinants for each of the years in the panels. We use the convenient regression method, which allows us to estimate the standard error for the concentration indices (equation (4.2)). Second, we decompose the estimates of the concentration index for

chronic diseases into the contributions of the determinants (equation (4.5)) and compare these yearly estimates with the corresponding estimate of "prevalence of chronic diseases". We also compute the degree of health inequity, which compares the overall inequality in chronic with the inequality in standardised prevalence, according to equation ((4.6)). Finally, we compute the changes in inequality over the time for each country, decomposing it into the contribution of changes in the elasticities and changes in inequality of the determinants (equation (4.12)).

4.5.1 Regression analysis

As usual with health inequality studies (e.g. Wagstaff et al. 2003, van Doorslaer and Koolman 2004, van Doorslaer et al. 2004), the regression results presented here do not provide a behavioural description of the decision process determining the prevalence of chronic diseases. Since each equation uses only one year of data, we cannot rule out the possibility of unobserved heterogeneity and reverse causation between chronic diseases and the explanatory variables. This precludes any causal interpretation of the estimated parameters, which should be interpreted as simply reduced form equations for the determination of the probability of at least one individual with chronic diseases in the household.

This approach is valid since our main interest is to calculate a measure of the static inequality in each year. For our purposes, the most important result is the evidence of a clear negative relationship between the level of household income and the probability of chronic disease. The coefficient for socio-economic status is negative and statistically significant in all regressions, except for Brazil in 2003. Together with Figure 4.1, this

finding suggests households of lower SES tend to have a higher probability of chronic diseases. The regression result, however, is more robust in the sense that it controls for the potentially confounding effect of other covariates.

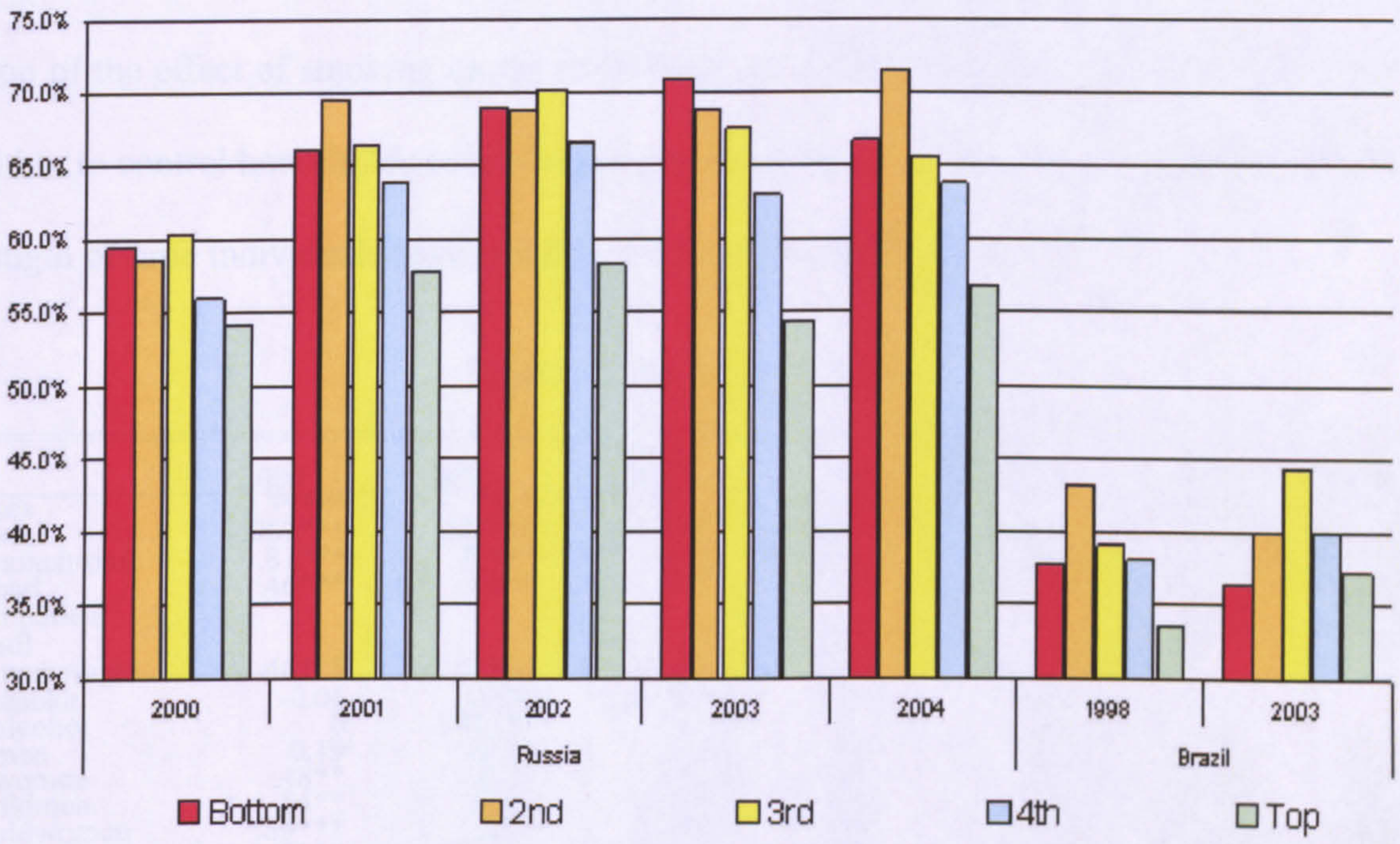


Fig. 4.1. Average probability of chronic diseases over SES quintiles

All the other variables have the expected effect. In particular, the morbidity indicators, such as the proportion of individuals reporting non-chronic illness, inpatient stays, ADL limitations and with bad or very bad health, all increase the probability of chronic

diseases. At least in Russia, where information was available, the effect of chronic disease risk factors is also apparent. The proportion of overweight individuals and the average intake of alcohol both tend to increase the probability of chronic diseases. The proportion of smokers, however, does not come out significant. This is possibly related to the fact that smoking has a cumulative harmful effect, which will only impact on individual health over longer periods of time than considered here. In order to obtain a more complete description of the effect of smoking on the probability of chronic diseases it would be necessary either to control households over a longer period of time or at least to take into account the length of time individuals have smoked (information not available).

Table 4.2. Probit regressions

	Russia					Brazil	
	2000	2001	2002	2003	2004	1998	2003
ses	-.75***	-.7***	-.56***	-.77***	-.67***	-.073**	0.02
ses2	.062***	.054***	.043***	.056***	.051***	0	0
nonchronic	.81***	.73***	.83***	.75***	.83***	.42***	.37***
bad	.46***	.6***	.57***	.61***	.41***	.56***	.5***
inpatient						.3***	.31***
adl						.74***	.69***
overweight	.46***	.47***	.51***	.57***	.55***		
smoke	-0.06	-0.04	0.03	-0.06	-0.1		
alcohol	0	.0014**	0	0	0		
men	0.19	-0.11	.39*	0.27	.41*	.57***	.51***
women	.48**	0.17	0.33	0.09	.5**	.53***	.56***
oldmen	.56**	-0.17	.65***	0.07	0.3	.95***	.96***
oldwomen	.89***	.39*	.68***	.39*	.77***	.83***	.86***
age	.065***	.055***	.06***	.043***	.065***	.12***	.13***
age2	-.00044***	-.00033***	-.00046***	0	-.00044***	-.0011***	-.0011***
secondary	-.24*	-0.13	-0.03	0.07	0.05	.38***	.41***
education	0.11	.27***	.14*	.3***	.25***	.026***	.027***
education2	0	-.013**	-0.01	-.017***	-.015***	-.0063***	-.0062***
insured	0.01	.25**	.22*	.23*	-0.05	.089***	.11***
inactive	0.05	-0.02	-0.03	-0.11	0.04	.49***	.47***
unemployed	0.28	0.23	0.07	-0.2	-0.12	.42***	.37***
white						.067***	.077***
headmale	-0.11	-0.05	-.27**	-.3**	-.25*	-.27***	-.26***
headsingle	-.41***	-.27**	-.51***	-.38***	-.36***	-.34***	-.34***
urban	0.03	0.02	-0.01	-0.02	-0.01	.093***	.067***
region1	-0.13	-0.04	0.01	.2***	0.11	0.03	-.1***
region2	0.01	0.07	-0.05	.21***	.23***	-0.01	-.032***
region3	-.18**	-0.02	-.15**	0.09	0.03	-.026*	-.064***
region4	-.22**	-0.13	-.17**	0.02	0.07	-0.02	0
region5	-0.12	0	0	0.12	.15*	.12***	0.01
cons	-1.7***	-1.7***	-1.5***	-1.1**	-1.8***	-3.6***	-4***
Pseudo R^2	22.00%	20.00%	21.00%	23.00%	23.00%	17.00%	18.00%
N	*3,754*	*4,126*	*4,326*	*4,382*	*4,346*	*86,976*	*103,746*

Prob. of at least one adult suffering from chronic disease in the household (heart disease, hypertension or diabetes);
Legend: * p<.1; ** p<.05; *** p<.01

4.5.2 Indices for total inequality

Let us now focus on the concentration indices for chronic diseases and the other explanatory variables. Theses indices were computed using the convenient regression method, according to equation (4.2). In each case, the relevant variable is transformed using its own sample mean and the variance of the SES fractional rank; the resulting variable is then regressed against the SES fractional rank, which allows us to recover the estimate of the concentration index and its standard error. Table 4.3 and Figures 4.2 and 4.3 report the results.

Table 4.3. Concentration indices of dependent and independent variables

	Russia					Brazil	
	2000	2001	2002	2003	2004	1998	2003
chronic	**-0.0214**	**-0.0286**	**-0.0291**	**-0.0462**	**-0.0349**	**-0.0276**	0.0028
ses	**0.0784**	**0.0768**	**0.0765**	**0.0752**	**0.0743**	**0.1023**	**0.0993**
ses2	**0.1334**	**0.1314**	**0.129**	**0.128**	**0.1269**	**0.2015**	**0.1945**
nonchronic	**-0.0381**	**-0.0262**	**-0.0424**	**-0.0358**	**-0.0312**	**-0.0841**	**-0.0609**
bad	**-0.1414**	**-0.1723**	**-0.1789**	**-0.2079**	**-0.1798**	**-0.238**	**-0.1944**
inpatient						**-0.0585**	**-0.0178**
adl						**-0.0986**	**-0.0504**
overweight	**0.0204**	0.0102	0.006	-0.0017	-0.0002		
smoke	**0.0265**	**0.0197**	**0.0295**	**0.0369**	**0.0404**		
alcohol	0.0375	**0.0562**	0.0159	0.0316	-0.0076		
men	**0.0361**	**0.046**	**0.0472**	**0.0506**	**0.0574**	**0.0485**	**0.0418**
women	**0.0727**	**0.074**	**0.0805**	**0.0879**	**0.0886**	**0.0409**	**0.0323**
oldmen	**-0.1103**	**-0.1286**	**-0.1406**	**-0.152**	**-0.201**	**0.0291**	**0.1329**
oldwomen	**-0.1475**	**-0.1643**	**-0.1655**	**-0.173**	**-0.1679**	**0.0589**	**0.1694**
age	**-0.034**	**-0.0389**	**-0.0425**	**-0.0427**	**-0.045**	**0.0124**	**0.0295**
age2	**-0.0672**	**-0.0769**	**-0.0841**	**-0.084**	**-0.0879**	**0.0222**	**0.0591**
secondary	**0.0743**	**0.0694**	**0.0755**	**0.0731**	**0.0659**	**0.3759**	**0.2788**
education	**0.0292**	**0.0275**	**0.03**	**0.0286**	**0.026**	**0.2416**	**0.1975**
education2	**0.0471**	**0.0452**	**0.0492**	**0.0474**	**0.0434**	**0.3914**	**0.331**
insured	**-0.0123**	**-0.0155**	**-0.0147**	**-0.0142**	**-0.0164**	**0.5068**	**0.5185**
inactive	**-0.104**	**-0.1023**	**-0.1256**	**-0.1285**	**-0.1382**	**-0.054**	**-0.0675**
unemployed	-0.0429	**-0.1211**	**-0.1272**	**-0.089**	**-0.1229**	**-0.2128**	**-0.2769**
white						**0.154**	**0.1644**
headmale	**0.0117**	**0.0151**	**0.0122**	**0.012**	**0.0121**	0.0001	**-0.0027**
headsingle	0.0191	0.0011	0.0124	0.0123	0.0049	**0.03**	**0.0571**
urban	**0.0483**	**0.055**	**0.0647**	**0.0674**	**0.0686**	**0.0815**	**0.0543**
region1	**0.053**	**0.0413**	0.0315	0.0153	0.0307	**-0.095**	**-0.1024**
region2	-0.0082	**-0.0403**	-0.0191	-0.0161	-0.0044	**-0.2681**	**-0.2691**
region3	**-0.1162**	**-0.1013**	**-0.1368**	**-0.1311**	**-0.1198**	**0.159**	**0.2103**
region4	**0.0525**	-0.0101	**-0.0537**	-0.0434	**-0.1234**	**-0.017**	**0.0231**
region5	**-0.0831**	**-0.0986**	**-0.095**	**-0.1262**	**-0.0788**	**0.3137**	**0.2731**

Note: Statistically significant indices in bold typeface (at $P<0.05$)

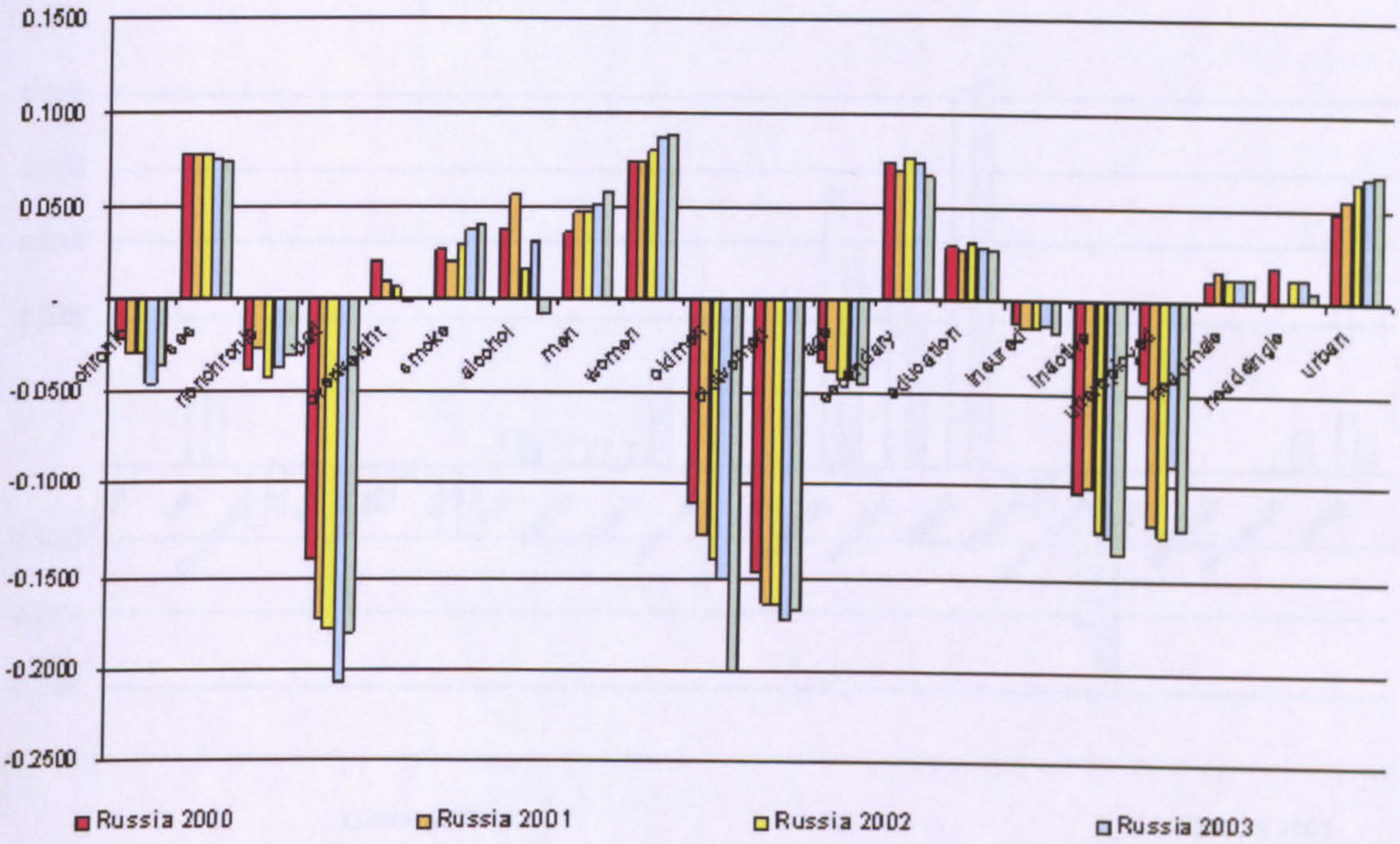


Fig. 4.2. Concentration indices for selected variables - Russia 2000-2004

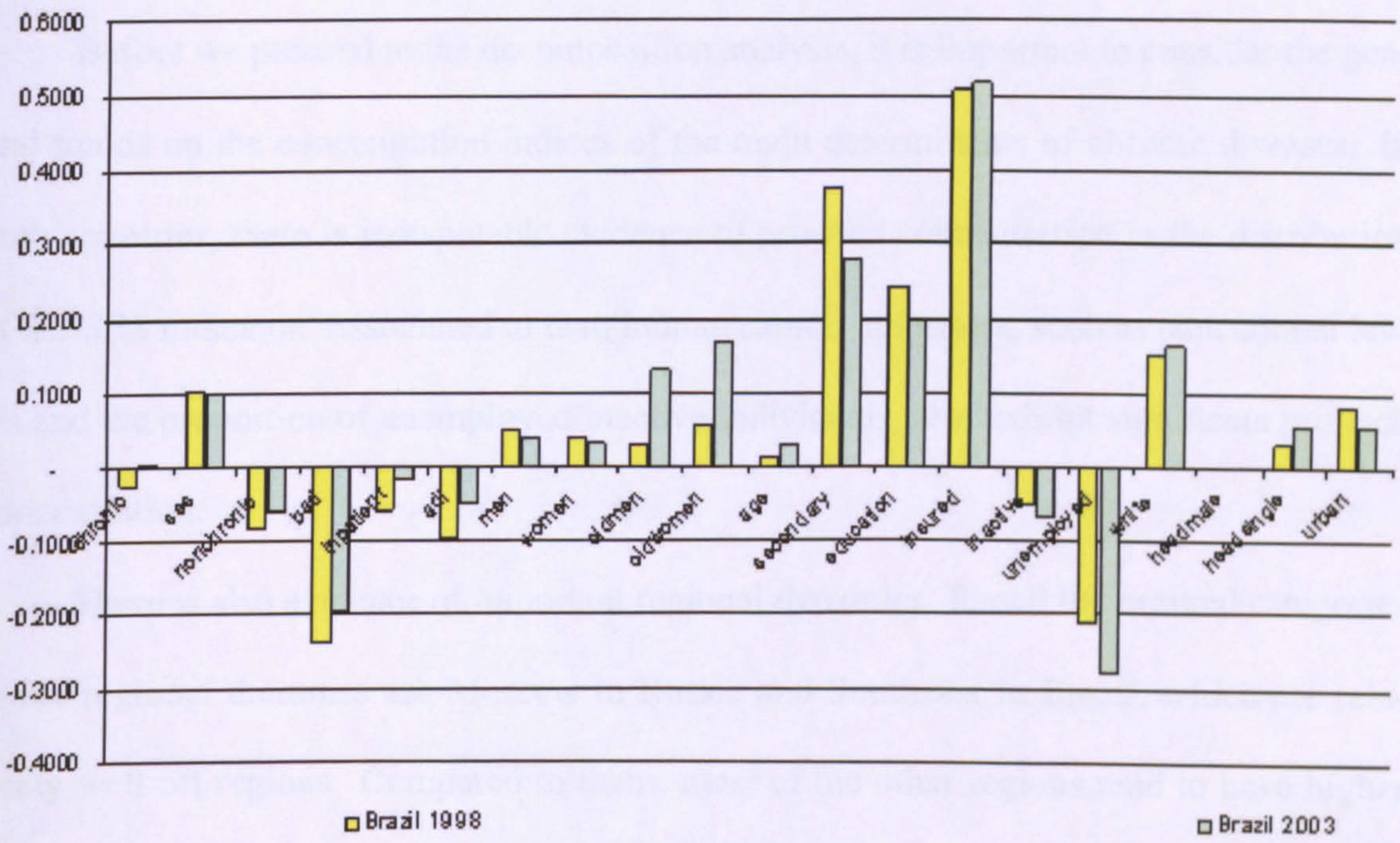


Fig. 4.3. Concentration indices for selected variables - Brazil 1998-2003

In all years, except in Brazil 2003, the concentration index for the probability of chronic diseases is negative and statistically significant, varying between -0.0214 (Russia 2000) and -0.0450 (Russia 2003). In Brazil 1998 it is about -0.0276. This general pattern of negative concentration indices suggests the probability of chronic diseases is significantly concentrated among households of lower SES. Importantly, there is also a tendency for increased concentration in Russia over the period analysed.

Before we proceed to the decomposition analysis, it is important to consider the general trends on the concentration indices of the main determinants of chronic diseases. In both countries, there is indisputable evidence of pro-rich concentration in the distribution of the SES indicator. Associated to that, human capital indicators, such as educational levels and the proportion of unemployed/inactive individuals, also exhibit significant pro-rich concentration.

There is also evidence of important regional dynamics. Recall the omitted categories in the regional dummies are Moscow in Russia and Southeast in Brazil, which are relatively well-off regions. Compared to them, most of the other regions tend to have higher concentration of poorer households. Similarly, urban areas are characterised by higher concentration of richer households.

All the morbidity indicators report negative and significant concentration indices, indicating concentration among households of lower SES. In the case of the proportion of individuals with bad or very bad self-reported health, the concentration indices are estimated at a considerably high level, around -0.20. With respect to risk factors, only the proportion of smokers has significant and positive concentration index.

The concentration index for health insurance is positive and considerably high in Brazil, but, somewhat surprisingly, negative in Russia. A possible explanation for this difference might be that the health insurance variable includes both mandatory and complementary coverage in Russia, but only complementary in Brazil.

Finally, there is an interesting difference about the effect of the standardising variables. In Brazil, household SES shows a tendency to increase with age. This fact suggests a probable indirect effect of income on the probability of chronic diseases through the positive association between income and average household age . In Russia, on the contrary, household SES presents a markedly negative association with the average age in the household. Thus, income is likely to exert an indirect negative effect on the probability of chronic diseases since richer households are disproportionately made up by younger individuals. Nevertheless, as noted above, the direct effect of SES on the probability of chronic diseases is negative in both countries.

4.5.3 Decomposition analysis

In this section, we discuss the results of the decomposition of the concentration index of chronic diseases, highlighting the most important factors to explain it. Due to the non-linear nature of the model explaining the probability of chronic diseases, the decomposition presented in Table 4.4 is a linear approximation based on a marginal effects representation (Equation (4.10)). That is, contrary to Wagstaff et al. (2003) and van Doorslaer and Koolman (2004), which decompose inequality using the elasticities recovered from linear re-

gressions, our analysis in based on marginal effects derived from the Probit model. Table 4.4 and Figure 4.4 present the results of the decomposition.

Table 4.4. Decomposition of inequality in the probability of chronic diseases

	Russia					Brazil	
	2000	2001	2002	2003	2004	1998	2003
C (actual)	-0.0214	-0.0286	-0.0291	-0.0462	-0.0349	-0.0276	0.0028
C (predicted)	-0.0246	-0.0246	-0.0271	-0.0439	-0.0338	-0.0249	0.0052
GC (resid)	0.0032	-0.004	-0.002	-0.0023	-0.0011	-0.0027	-0.0024
$HI_{WV} = C - \hat{C}$	0.0121	-0.0046	-0.0064	-0.0189	-0.0061	-0.0608	-0.0493
ses	**-0.1787**	**-0.1435**	**-0.11**	**-0.1562**	**-0.1319**	**-0.0344**	0.0102
ses2	**0.1881**	**0.148**	**0.1112**	**0.1484**	**0.1356**	0.0211	-0.0171
nonchronic	**-0.0071**	**-0.0039**	**-0.0065**	**-0.005**	**-0.0047**	**-0.0021**	**-0.0013**
bad	**-0.007**	**-0.0089**	**-0.0077**	**-0.0104**	**-0.0061**	**-0.0057**	**-0.0036**
inpatient						**-0.0012**	**-0.0003**
adl						**-0.0098**	**-0.0044**
overweight	0.0024	0.0012	0.0007	-0.0002	0.0001		
smoke	-0.0002	-0.0001	0	-0.0003	-0.0006		
alcohol	0.0001	**0.0005**	0	0	0		
men	0.0009	-0.0005	0.0017	0.0016	0.0025	**0.0073**	**0.0056**
women	**0.005**	0.0016	0.0028	0.0014	**0.0047**	**0.0061**	**0.005**
oldmen	**-0.0029**	0.0007	**-0.0032**	-0.0006	-0.0019	**0.0008**	**0.0039**
oldwomen	**-0.016**	-0.0071	**-0.0111**	-0.0073	**-0.0124**	**0.0022**	**0.0066**
age	**-0.0537**	**-0.045**	**-0.0506**	**-0.0379**	**-0.0584**	**0.0494**	**0.1194**
age2	**0.0378**	**0.0283**	**0.0398**	0.0171	**0.0401**	**-0.0355**	**-0.094**
secondary	-0.0054	-0.0027	-0.0003	0.0011	0.0011	**0.0379**	**0.0363**
education	0.0147	**0.0296**	0.0168	**0.0328**	**0.0247**	**0.0312**	**0.0283**
education2	-0.0016	**-0.0211**	-0.0108	**-0.0269**	**-0.0233**	**-0.1028**	**-0.0969**
insured	0.0001	**-0.0014**	-0.0011	-0.0008	0	**0.0074**	**0.0093**
inactive	-0.0013	0.0001	0.001	0.003	-0.001	**-0.0051**	**-0.0057**
unemployed	-0.0002	-0.0004	-0.0001	0.0002	0.0001	**-0.0058**	**-0.007**
white						**0.0107**	**0.0116**
headmale	-0.0005	-0.0003	**-0.001**	**-0.0011**	-0.001	**0**	**0.0004**
headsingle	**-0.001**	**0**	**-0.0006**	**-0.0005**	**-0.0002**	**-0.0023**	**-0.0047**
urban	0.0005	0.0004	-0.0001	-0.0005	-0.0001	**0.0051**	**0.0025**
region1	-0.0007	-0.0001	0	**0.0002**	0.0002	-0.0002	**0.0008**
region2	0	-0.0002	0.0001	**-0.0003**	**-0.0001**	0.0006	**0.0021**
region3	**0.0023**	0.0002	**0.0016**	-0.0008	-0.0003	-0.0006	**-0.0018**
region4	**-0.0008**	0.0001	**0.0004**	0	-0.0004	0	0
region5	0.0008	-0.0001	0.0001	-0.0008	-0.0007	**0.0008**	0

Note: Decomposition based on linear approximation using marginal effects from probit regression.
Significant contributions in bold typeface (at P<0.05)

The first thing to notice in Table 4.4 is that the marginal effects representation is at best an approximation, which intrinsically generates a residual. Therefore, the concentration index calculated using the convenient regression (C(actual), in the first line of Table 4.4) is necessarily different from the sum of the individual components of the decomposition (C(predicted)). The difference between the two is captured by the generalised concen-

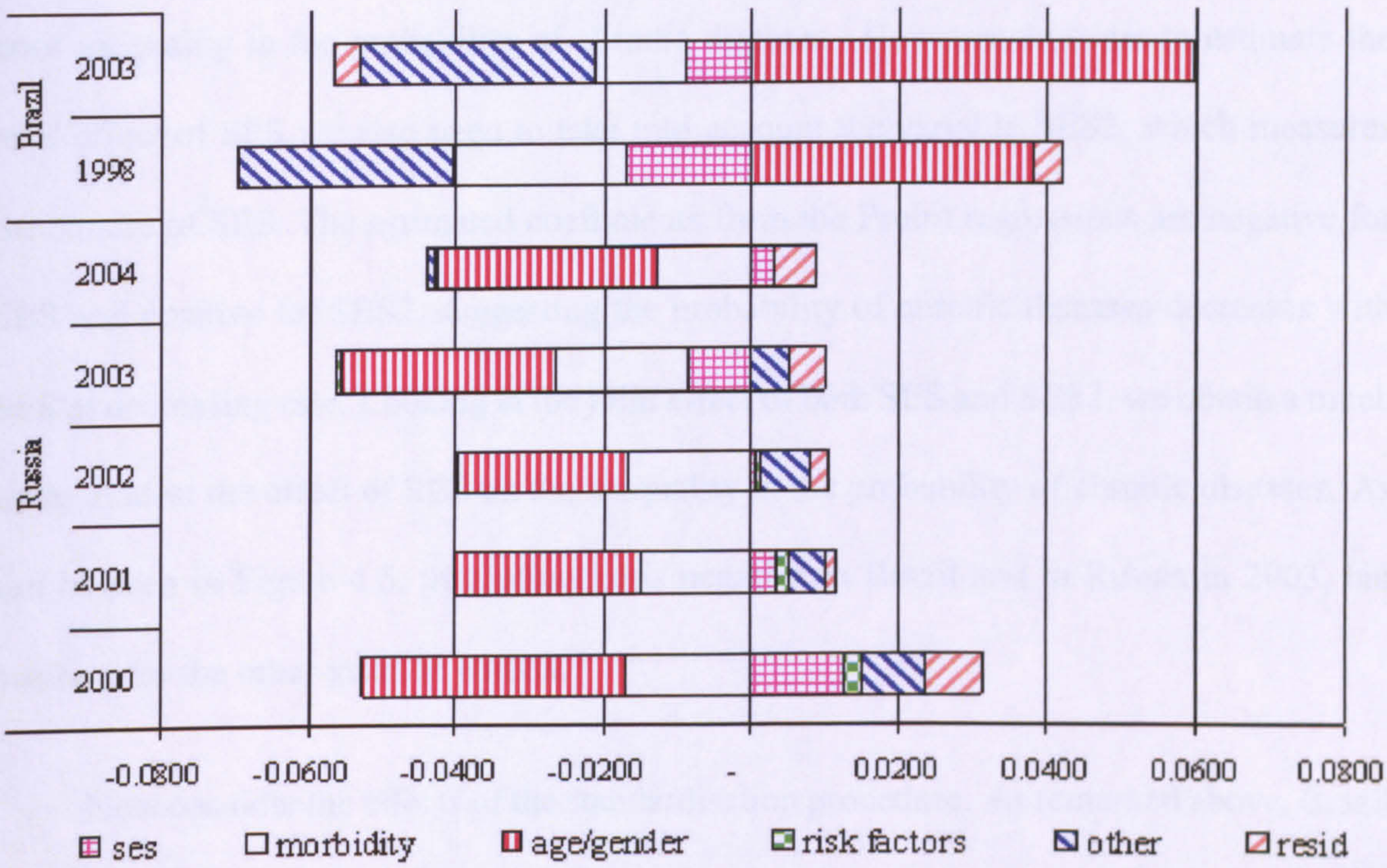


Fig. 4.4. Decomposition of inequality in the probability of chronic diseases

tration index of the residual (GC(resid)), which comprises both an estimation error and an approximation error (cf. van Doorslaer et al. 2004). In studies based on the linear decomposition, such as van Doorslaer and Koolman (2004) and Wagstaff et al. (2003), the actual and predicted concentration indices are exactly equal.

In almost all cases, the SES indicator contributes to a substantial increase in the pro-poor inequality in the probability of chronic diseases. However, in order to estimate the total effect of SES we also need to take into account the variable SES2, which measures the square of SES. The estimated coefficients from the Probit regressions are negative for SES and positive for SES2, suggesting the probability of chronic diseases decreases with SES at decreasing rate. Looking at the joint effect of both SES and SES2, we obtain a much more modest net effect of SES on the inequality in the probability of chronic diseases. As can be seen in Figure 4.5, this influence is negative in Brazil and in Russia in 2003, but positive for the other years in Russia.

Now consider the effects of the standardisation procedure. As remarked above, Brazil and Russia display a fundamental distinction with respect to the relationship between SES and the demographic composition of the households. In Brazil, household SES tends to increase with the average age among adults, whilst in Russia it tends to decrease. Consequently, when we look at the effect of standardising variables on the inequalities on chronic diseases, we find a significant positive contribution in Brazil, but a negative contribution in Russia.

The interpretation is that in Russia the richer households are also, in average, made up by younger individuals, which are less likely to be affected chronic disease relative to older

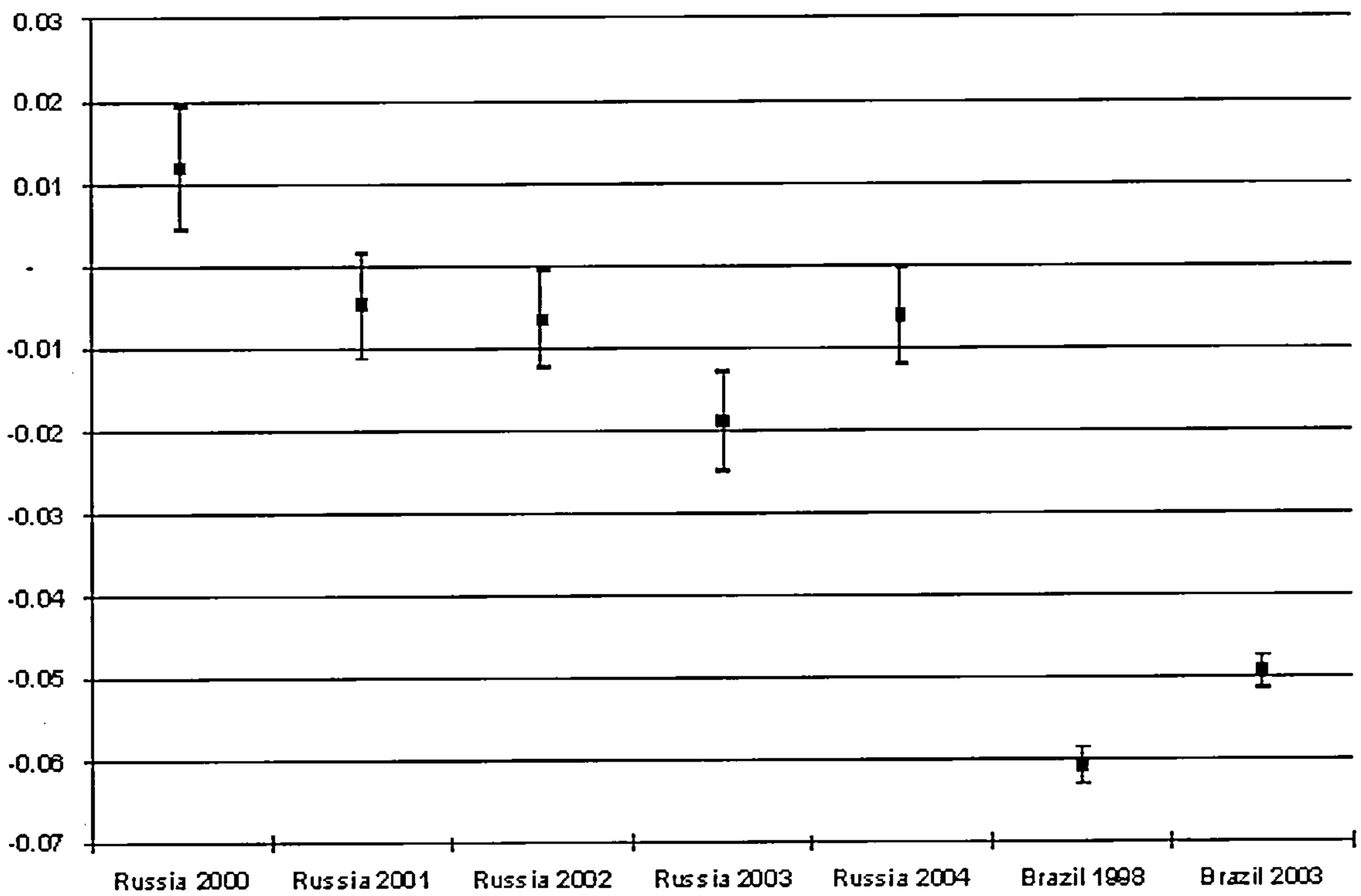


Fig. 4.5. Inequity indices for the probability of at least one adult with chronic illness (with 95% confidence intervals)

individuals. Therefore, relatively older households are associated with worsened inequality in the probability of chronic diseases. In Brazil, on the contrary, there is an offsetting effect resulting from the less clear-cut association between SES and demographic profile.

The fourth line in Table 4.4 reports the estimate of the inequity in the probability of chronic diseases (HI_{WV}) calculated according to the method proposed by Wagstaff and van Doorslaer (2000). This index measures the degree of socio-economic inequality in the probability of chronic diseases that would be observed if the standardising variables were uniformly distributed across the population. In other words, it gives the health inequity purged of age and gender influences, and potentially avoidable through policy intervention. Figure 4.5 gives a graphical representation of HI_{WV}, together with its standard error. Clearly, there is a significant degree of health inequity in all the years analysed in both countries, with households from lower SES having a disproportionate higher tendency to be affected by chronic illness. What is more, at least in Russia, health inequity displays a tendency to be worsening over time in the period considered.

Co-morbidities also exert a significant effect on inequalities in chronic diseases. In all cases, the decomposition analysis indicates negative and highly significant coefficients, which imply the current distribution of the burden of co-morbidities contributes to worsened pro-poor inequality in the probability of chronic diseases. Apart from the group of standardising variables, co-morbidities answer for the biggest share in explaining inequalities in the probabilities of chronic diseases. This information is very relevant for policy makers, since it gives indication of where efforts should be primarily concentrated in order to reduce inequalities in chronic diseases.

Risk factors have very limited influence in explaining inequalities in chronic diseases, and in most cases, their contribution is not statistically significant. Of the other non-standardising, policy relevant variables the most important are education (which reduces pro-poor inequality) and regional dummies. In Russia, in particular, there is evidence that the Central region, compared to Moscow, presents higher concentration of chronic diseases among households of lower SES, whilst the opposite is true for the Caucasus. In addition, in both countries the fact that the head of the household is single contributes to higher concentration of chronic diseases among poorer households.

4.5.4 Decomposition of changes in inequality

One of the most interesting applications of the method proposed by Wagstaff et al. (2003) is the ability to assess the relative importance of the different factors in explaining the evolution of inequality in the probability of chronic diseases over the time. Following Wagstaff et al. (2003), in this section we present the results of the analysis using both the Oaxaca decomposition (equation (4.11)) and the total differential approach (equation (4.12)).

Our aim is to unravel the main drivers of the changes in the concentration index of chronic diseases over the time in each country. Recall from Table 4.4 the concentration index in Russia starts around -0.021 in 2000, peaks at -0.046 in 2003 and in 2004 it is equal to -0.035, clearly describing a rising trend. In Brazil the concentration changes from -0.028 in 1998 to 0.003 in 2003.

Let us consider first the results from the Oaxaca decomposition. This method of decomposition consists on essentially separating the effect of each determinant into two components: changes in the elasticity of chronic disease with respect to the regressor, and changes in the sample means of the regressor itself. Table 4.5 presents the results.

Table 4.5. Oaxaca decomposition of changes in CI of probability of chronic diseases

	Russia								
	2001			2002			2003		
	$\Delta C \cdot \eta$	$\Delta \eta \cdot C$	*Total*	$\Delta C \cdot \eta$	$\Delta \eta \cdot C$	*Total*	$\Delta C \cdot \eta$	$\Delta \eta \cdot C$	*Total*
ses	0.0026	0.0325	0.0351	0.0007	0.0328	0.0336	0.0026	-0.0489	-0.0463
ses2	-0.0018	-0.0383	-0.0401	-0.0023	-0.0345	-0.0368	-0.0011	0.0383	0.0372
nonchronic	0.0019	0.0014	0.0033	-0.0025	-0.0001	-0.0026	0.0011	0.0004	0.0015
bad	-0.0016	-0.0003	-0.0019	-0.0002	0.0014	0.0012	-0.0014	-0.0013	-0.0027
inpatient									
adl									
overweight	-0.001	-0.0002	-0.0012	-0.0005	0	-0.0005	-0.001	0.0001	-0.0009
smoke	0	0.0001	0.0001	0	0.0001	0.0001	-0.0001	-0.0003	-0.0004
alcohol	0.0002	0.0003	0.0004	0	-0.0005	-0.0005	0	0	0
men	-0.0001	-0.0013	-0.0014	0	0.0022	0.0022	0.0001	-0.0003	-0.0001
women	0	-0.0034	-0.0034	0.0002	0.0009	0.0012	0.0001	-0.0015	-0.0014
oldmen	0.0001	0.0035	0.0036	-0.0003	-0.0036	-0.0039	0	0.0027	0.0027
oldwomen	-0.0007	0.0097	0.009	0	-0.004	-0.004	-0.0003	0.0041	0.0038
age	-0.0056	0.0143	0.0087	-0.0042	-0.0014	-0.0056	-0.0005	0.0132	0.0127
age2	0.0035	-0.013	-0.0094	0.0034	0.0081	0.0115	0.0001	-0.0228	-0.0227
secondary	0.0002	0.0026	0.0027	0	0.0024	0.0023	0	0.0015	0.0015
education	-0.0015	0.0163	0.0148	0.0013	-0.0141	-0.0128	-0.0015	0.0176	0.016
education2	0.0007	-0.0201	-0.0194	-0.0008	0.0111	0.0103	0.001	-0.0171	-0.0161
insured	-0.0003	-0.0011	-0.0014	0.0001	0.0001	0.0002	0	0.0002	0.0003
inactive	0	0.0014	0.0014	0.0002	0.0007	0.0009	0.0001	0.0019	0.002
unemployed	-0.0003	0.0001	-0.0002	0	0.0003	0.0003	-0.0001	0.0005	0.0004
white									
headmale	-0.0001	0.0003	0.0002	0.0002	-0.0009	-0.0007	0	-0.0001	-0.0001
headsingle	0.0006	0.0004	0.001	-0.0006	0	-0.0006	0	0.0002	0.0001
urban	0	-0.0001	-0.0001	0	-0.0005	-0.0005	0	-0.0003	-0.0003
region1	0	0.0005	0.0005	0	0.0001	0.0001	-0.0003	0.0005	0.0002
region2	-0.0002	0	-0.0002	-0.0001	0.0004	0.0003	0	-0.0003	-0.0003
region3	0	-0.0021	-0.0021	0.0004	0.001	0.0014	0	-0.0024	-0.0024
region4	0.0005	0.0004	0.0009	0.0004	0	0.0004	0	-0.0005	-0.0005
region5	0	-0.0008	-0.0008	0	0.0002	0.0002	-0.0002	-0.0007	-0.0009
resid			0.0317			0.4507			-1.562

Table 4.5 (continued)

	Russia 2004			Brazil 2003		
	$\Delta C \cdot \eta$	$\Delta \eta \cdot C$	*Total*	$\Delta C \cdot \eta$	$\Delta \eta \cdot C$	*Total*
ses	0.0015	0.0229	0.0243	-0.0003	0.0449	0.0446
ses2	-0.0011	-0.0117	-0.0128	0.0006	-0.0388	-0.0382
nonchronic	0.0007	-0.0003	0.0004	0.0005	0.0002	0.0007
bad	0.0009	0.0035	0.0044	0.0008	0.0014	0.0022
inpatient				0.0008	0	0.0008
adl				0.0043	0.0011	0.0053
overweight	0.0003	0	0.0003			
smoke	0	-0.0002	-0.0002			
alcohol	0	0	0			
men	0.0004	0.0005	0.0009	-0.0009	-0.0008	-0.0017
women	0	0.0033	0.0033	-0.0013	0.0002	-0.0011
oldmen	-0.0004	-0.0009	-0.0013	0.0031	0	0.0031
oldwomen	0.0004	-0.0054	-0.0051	0.0043	0.0001	0.0045
age	-0.0022	-0.0183	-0.0205	0.0692	0.0008	0.0701
age2	0.0014	0.0216	0.023	-0.0587	0.0001	-0.0585
secondary	-0.0001	0.0001	0	-0.0126	0.011	-0.0016
education	-0.0025	-0.0056	-0.0081	-0.0063	0.0034	-0.0029
education2	0.0021	0.0015	0.0036	0.0177	-0.0117	0.006
insured	0	0.0009	0.0009	0.0002	0.0017	0.0019
inactive	-0.0001	-0.0039	-0.004	-0.0011	0.0005	-0.0006
unemployed	0	-0.0001	-0.0001	-0.0016	0.0005	-0.0011
white				0.0007	0.0002	0.0009
headmale	-0.0001	0.0001	0	0.0004	0	0.0004
headsingle	0.0003	0	0.0003	-0.0022	-0.0001	-0.0023
urban	0	0.0003	0.0003	-0.0012	-0.0014	-0.0026
region1	0.0001	-0.0001	0	0.0001	0.0009	0.001
region2	0.0002	0	0.0002	0	0.0014	0.0014
region3	0	0.0005	0.0005	-0.0004	-0.0008	-0.0012
region4	-0.0003	-0.0001	-0.0004	0	0	0
region5	0.0004	-0.0003	0.0001	0	-0.0008	-0.0008
resid			-53.965			-1.9052

Considering the net contribution of SES (that is, taking into account both SES and SES2), we see it is has been an important contributor to worsened inequality in chronic diseases in Russia. In addition, education had a discernible impact to increase inequality in chronic diseases

In Brazil, SES and the standardising variables were the two most important factors to explain the decrease in pro-poor inequality. Nevertheless, the effect of the standardising variables was considerably more important, accounting for approximately half of the total change. This suggests that, if the impact of demographic composition had remained constant, the overall variation of inequality in the probability of chronic diseases would have been much less significant in Brazil.

Finally, we notice that in both countries the driving force behind these effects were changes in the elasticity of chronic diseases with respect to SES, since the concentration indices did not change substantially during the period. Although the concentration indices of both SES and education were approximately stable, the elasticities varied considerably.

The total differential method allows us to uncover one additional layer of the process, by decomposing the effect of changes in elasticities into changes in partial effects and changes in the means of the regressor, as displayed in Table 4.6. With respect to SES and education, the main changes have clearly affected the partial effects, not the sample means, which have been stable.

Table 4.6. Total differential decomposition of changes in CI of probability of chronic diseases

	Russia							
	2001				2002			
	β's	Means of x's	CI's	Total	β's	Means of x's	CI's	Total
ses	0.0234	-0.005	0.0026	0.021	0.0411	-0.0013	0.0007	0.0405
ses2	-0.0305	0.0076	-0.0018	-0.0247	-0.0389	0.0019	-0.0023	-0.0393
nonchronic	0	0	0.0019	0.0019	-0.0002	0	-0.0025	-0.0027
bad	-0.0015	0.0005	-0.0016	-0.0026	0.0008	0.0002	-0.0002	0.0007
inpatient								
adl								
overweight	0	0	-0.001	-0.0009	0.0001	0.0001	-0.0005	-0.0003
smoke	0.0002	0	0	0.0002	0.0003	0	0	0.0004
alcohol	0.0006	0	0.0002	0.0008	-0.0004	0	0	-0.0004
men	-0.0024	0	-0.0001	-0.0025	0.0036	0	0	0.0036
women	-0.004	0	0	-0.004	0.0015	0	0.0002	0.0017
oldmen	0.0029	0	0.0001	0.003	-0.0031	0.0001	-0.0003	-0.0033
oldwomen	0.0075	0	-0.0007	0.0068	-0.0034	-0.0001	0	-0.0035
age	0.0029	0	-0.0056	-0.0027	-0.0009	0	-0.0042	-0.0052
age2	-0.0066	0	0.0035	-0.003	0.0064	0	0.0034	0.0098
secondary	0.0028	-0.0001	0.0002	0.0029	0.0035	0	0	0.0034
education	0.0336	0.0008	-0.0015	0.0329	-0.0284	0.0002	0.0013	-0.0269
education2	-0.0315	-0.0007	0.0007	-0.0315	0.0183	-0.0001	-0.0008	0.0173
insured	0.001	0.0001	-0.0003	0.0008	-0.0001	0	0.0001	0.0001
inactive	0.0009	0	0	0.0009	0.0007	0	0.0002	0.0009
unemployed	0	0.0001	-0.0003	-0.0002	0.0003	0	0	0.0003
white								
headmale	0.0008	0	-0.0001	0.0007	-0.0025	0	0.0002	-0.0023
headsingle	0.0004	0	0.0006	0.001	-0.0011	0	-0.0006	-0.0017
urban	-0.0002	0	0	-0.0001	-0.0009	0	0	-0.0009
region1	0.0005	0	0	0.0005	0.0002	0	0	0.0002
region2	-0.0001	0	-0.0002	-0.0002	-0.0001	0	-0.0001	-0.0002
region3	-0.0011	0	0	-0.0011	0.0011	0	0.0004	0.0015
region4	0.0001	0	0.0005	0.0006	0	0	0.0004	0.0004
region5	-0.0006	0	0	-0.0006	0.0001	0	0	0.0001
resid				0.0317				0.4507

Table 4.6 (continued)

	Russia							
	2003				2004			
	β's	Means of x's	CI's	Total	β's	Means of x's	CI's	Total
ses	-0.0713	-0.0031	0.0026	-0.0717	0.0333	-0.0003	0.0015	0.0344
ses2	0.0447	0.0043	-0.0011	0.0479	-0.0154	0.0007	-0.0011	-0.0158
nonchronic	-0.0002	0	0.0011	0.0009	0	0	0.0007	0.0008
bad	-0.0008	-0.0002	-0.0014	-0.0024	0.0023	0.0001	0.0009	0.0033
inpatient								
adl								
overweight	0.0005	0	-0.001	-0.0005	0	0.0001	0.0003	0.0003
smoke	-0.0009	0	-0.0001	-0.001	-0.0003	0	0	-0.0004
alcohol	0	0	0	0	0	0	0	0
men	-0.0008	0.0001	0.0001	-0.0005	0.0009	0.0001	0.0004	0.0014
women	-0.0028	0	0.0001	-0.0027	0.0046	0	0	0.0046
oldmen	0.0019	0	0	0.0019	-0.001	0.0001	-0.0004	-0.0013
oldwomen	0.0032	0.0001	-0.0003	0.003	-0.0041	-0.0001	0.0004	-0.0039
age	-0.0013	0	-0.0005	-0.0018	-0.0045	0	-0.0022	-0.0067
age2	-0.0104	-0.0001	0.0001	-0.0104	0.0137	0	0.0014	0.0151
secondary	0.0024	0	0	0.0024	0.0002	0	-0.0001	0.0001
education	0.043	0.0005	-0.0015	0.0419	-0.0121	0.0003	-0.0025	-0.0143
education2	-0.0321	-0.0005	0.001	-0.0317	0.0028	-0.0003	0.0021	0.0046
insured	-0.0007	0	0	-0.0006	-0.0011	0	0	-0.0011
inactive	0.0012	0	0.0001	0.0013	-0.0031	0	-0.0001	-0.0032
unemployed	0.0001	0	-0.0001	0	-0.0002	0	0	-0.0001
white								
headmale	-0.0003	0	0	-0.0002	0.0004	0	-0.0001	0.0003
headsingle	0.0011	0	0	0.0011	0	0	0.0003	0.0002
urban	-0.0006	0	0	-0.0006	0.0005	0	0	0.0005
region1	0.0009	0.0001	-0.0003	0.0008	-0.0004	0	0.0001	-0.0003
region2	0.0007	0	0	0.0007	0	0	0.0002	0.0002
region3	-0.0015	0	0	-0.0015	0.0003	0	0	0.0004
region4	0	0	0	0	-0.0002	0	-0.0003	-0.0005
region5	-0.0006	0	-0.0002	-0.0007	-0.0001	0	0.0004	0.0003
resid				-1.562				-53.965

Table 4.6 (continued)

	Brazil 2003			
	β's	Means of x's	CI's	Total
ses	0.0413	0	-0.0003	0.041
ses2	-0.0363	0	0.0006	-0.0357
nonchronic	0.0002	-0.0001	0.0005	0.0006
bad	0.0004	0.0005	0.0008	0.0017
inpatient	0	0	0.0008	0.0008
adl	0.0004	0	0.0043	0.0047
overweight				
smoke				
alcohol				
men	-0.0006	0.0002	-0.0009	-0.0014
women	0.0002	0.0001	-0.0013	-0.001
oldmen	0	0.0003	0.0031	0.0034
oldwomen	0.0002	0.0004	0.0043	0.005
age	0.0041	0.0013	0.0692	0.0747
age2	-0.0004	-0.0023	-0.0587	-0.0614
secondary	0.0025	0.007	-0.0126	-0.0031
education	0.001	0.0027	-0.0063	-0.0026
education2	0.001	-0.0136	0.0177	0.0051
insured	0.0018	0.0002	0.0002	0.0022
inactive	0.0002	0.0002	-0.0011	-0.0007
unemployed	0.0009	-0.0004	-0.0016	-0.0012
white	0.0014	-0.001	0.0007	0.0012
headmale	0	0	0.0004	0.0003
headsingle	0.0001	-0.0004	-0.0022	-0.0025
urban	-0.0009	0.0001	-0.0012	-0.002
region1	0.0011	0.0003	0.0001	0.0014
region2	0.0014	0.0001	0	0.0015
region3	-0.0011	0.0001	-0.0004	-0.0014
region4	0	0	0	0
region5	-0.0007	0	0	-0.0007
resid				-1.9052

4.6 Conclusions

In this paper, we use data from the LSMS (2000-2004) and the PNAD (1998 and 2003) to quantify and analyse the determinants of the inequalities in chronic diseases in households in Russia and Brazil.

We focus our attention on three specific diseases: heart disease, hypertension and diabetes. Using the household as the unit of analysis, we estimate a Probit model for the probability of observing at least one adult suffering from one of these diseases. The results indicate SES measures play a significant role, and decrease, at decreasing rates, the probability of observing chronic illness in the household. That is, controlling for household

and community-level attributes, we show that chronic diseases are more likely to affect households from lower SES.

This initial finding is corroborated by the computation of inequality indices following the method proposed by Kakwani et al. (1997). We show that inequality in chronic diseases has risen in Russia during the period, with the poor bearing a disproportionately high share of the chronic diseases burden. Our estimates indicate the concentration index for chronic diseases in Russia starts around -0.021 in 2000, peaks at -0.046 in 2003 and in 2004 it is equal to -0.035, clearly describing a rising trend. In Brazil the concentration changes from -0.028 in 1998 to 0.003 in 2003.

The figures above, however, refer to the total SES inequality in chronic diseases, which includes the effect of standardising variables. However, the relevant measure of health equity should include only the fraction that can be modified through policy intervention. In order to estimate this policy amenable measure we neutralize the effect of standardising variables. We show that the inequity concentration index is negative in both countries, but considerably higher in Brazil. Additionally, in line with the overall concentration index, it has been rising in Russia.

The decomposition analysis (Wagstaff et al. 2003) reveals the important elements to explain inequality in chronic diseases. Taking into account that we use a non-linear regression model, our decomposition analysis implements a linear approximation based on a marginal effects representation (van Doorslaer et al. 2004). We find SES, co-morbidities and education make a significant contribution to worsening inequalities in chronic diseases.

Standardising variables have opposite effects in each country, increasing pro-poor inequalities in Russia, while decreasing them in Brazil.

Finally, we implement both the Oaxaca decomposition and the total differential decomposition, proposed by Wagstaff et al. (2003), in order to study the determinants of the evolution of inequality over the time. In line with van Doorslaer and Koolman (2004), the results in this case indicate that changes in inequality have been driven mainly by changes in the elasticities, rather than changes in the sample means of the determinants.

This analysis confirms the existence of significant levels of SES inequality in chronic diseases in Russia and Brazil. In both countries, poorer households face considerably higher probability of being affected by chronic diseases. The paper also provides evidence that inequalities in education and co-morbidities contribute to worse this scenario. This evidence suggests that efforts to relieve the burden of chronic diseases from poor households should aim to maximise the impact of externalities from other related policy areas. In that sense, policy makers should consider income redistribution as one instrument to be used in combination with others.

Chapter 5
The Economic Impact of Chronic Diseases: Evidence from Brazil, India and Russia

5.1 Introduction

Evidence suggests that the burden of chronic diseases and their risk factors is increasing worldwide. Recent projections from the WHO indicate that in 2005, 35 of the 58 million expected deaths worldwide will have been from chronic noncommunicable diseases (see Strong, Mathers, Leeder and Beaglehole 2005, WHO 2005). A projected 388 million people will die from chronic diseases in the next ten years and the majority of these deaths will occur in the most productive (economically active) age group if nothing is done (WHO 2005). An estimated 80% of predicted deaths will occur in the low- and middle-income countries. The economic impact of this pandemic of chronic diseases on countries, households and individuals is increasingly becoming a subject of considerable interest. Based on the WHO projections, countries stand to lose billions of dollars in national income over the next 10 years due to the impact of deaths from chronic diseases on labour supplies and national savings and poorer countries are relatively worse off (WHO 2005). Indications are that chronic disease burden may expose individuals and household to poverty just as poverty exposes people to greater risks of chronic diseases. This is despite a common misconception that chronic diseases affect mostly the wealthy and affluent. Emerging data from various sources indicate that the poor could be more

vulnerable to chronic diseases (eg. Bartley, Sacker, Firth and Fitzpatrick 1999, Bartley, Fitzpatrick, Firth and Marmot 2000, Chaturvedi, Jarrett, Shipley and Fuller 1998, Stronks, van de Mheen and Mackenbach 1998, White 2000, World Bank 2005) a situation that could worsen existing socio-economic inequalities in access to care in many countries.

The literature on the macroeconomic impact of disease discusses possible socio-economic channels through which chronic diseases may affect the individuals' household or national economic well being. However, empirical evidence is scarce and the mechanisms by which chronic diseases impact on the economy are not entirely clear at present. For instance, there is very little information on household or individuals' coping mechanisms in the face of expenditure from chronic diseases especially in low- and middle-income economies although the economies of such countries are typically volatile and health systems are frequently poorly developed in addition to inadequate social security or formal insurance systems. Therefore, it would be expected that the ability of households to respond to catastrophic healthcare expenditure (idiosyncratic) shocks (see Cochrane 1991, Kochar 1995, Morduch 1995) or to maintain non-health consumption levels over periods of economic challenge from chronic disease is likely to be limited given these conditions. Coping processes could be particularly consequential when household heads or breadwinners are affected.

This paper partly addresses a gap in information relating to how chronic diseases impact on the economy of and the coping mechanisms adopted by households. We focus our analysis on Brazil and India (being developing countries) and the Russian Federation where chronic diseases are projected to increase in prevalence (WHO 2005). These countries en-

joy relatively higher incomes than other but comparable countries and are experiencing epidemiological and economic transition, increased urbanization and/or changes in traditional lifestyles (Greenberg, Raymond and Leeder 2005). We use micro data from the Living Standards Measurement Study (LSMS) previously conducted by the World Bank in these countries for our analysis.

The results of the two-part model indicate that chronic diseases are significantly associated with increased demand for health care and reduced labour earnings in the Russia federation, Brazil and India. Each additional case of chronic disease in the household generate a conditional increase on health expenditures of 21%, 120% and 14% in Brazil, India, and Russia, respectively.

Since the potential to work is affected by chronic diseases, labour income reduces by 7.9% in Brazil and 4.3% in Russia. The analyses indicate that remittances to households increased by 3.1% in Brazil and 8.7% in Russia, and this may have provided some insurance against the expected reduction in non-health consumption, and partly against the increased household health expenditure. In Brazil, the net effect on non-health consumption is an increase of 3.9%, suggesting that households are able to insure non-health consumption against chronic diseases, possibly from remittances. In Russia, however, where unobserved heterogeneity is accounted for, there is a net reduction of non-health consumption of 3.2%, suggesting that chronic diseases in fact reduce household overall welfare.

The remainder of this paper is organized as follows. The next section reviews the literature on the economic consequences of diseases and describes the possible association between economic development, risk factors and chronic diseases. A brief description of

the health systems in Brazil, India and Russia, is then provided, emphasizing the growing importance of chronic diseases. Section 5.3 explains the methodology and the econometric models employed for this investigation, and Section 5.4 describes the data. Section 5.5 presents the results and Section 5.6 concludes. The Appendix 5.A provides an extension using Generalized Linear Models.

5.2 Evidence of economic consequences of illness

In recent times, there has been an increasing interest among health policy makers in the linkages between health and economic growth (see Howitt 2005, López-Casasnovas, Rivera and Currais 2005, van Zon and Muysken 2005). However, many of the empirical macro-economic studies on the links between health, diseases and the economy have focused on HIV/AIDS (eg. Anand, Pandav and Nath 1999, Bloom and Mahal 1997, Cuddington 1993, Cuddington and Hancock 1994, Hanson 1992), malaria (eg. Gallup and Sachs 2001, Sachs and Malaney 2002) and other communicable diseases. There is however very little exploration in contemporary literature of the impact of chronic disease on economic wellbeing and growth. Recent studies by Bhargava, Jamison, Lau and Murray (2001) and Gyimah-Brempong and Wilson (2001) merely provide anecdotal evidence of the indirect link between chronic diseases and economic well-being since the working-age segment of populations are most at risk of chronic diseases. These studies demonstrate the positive effect of adult survival rates, stock of and investment in, health human capital on GDP growth rates and per capita income growth rates. It may be that these linkages take root from the unit of economic production – at both individual and the household level. Economic pro-

duction in households frequently faces different risks, like job loss, weather uncertainty and crop failure and sudden debilitating chronic diseases may be among the most important risks for both rural and urban households (see Asfaw and von Braun 2004, Dercon and Krishnan 2000, Gertler and Gruber 2002, Sauerborn, Adams and Hein 1996, Schultz and Tansel 1997, Wagstaff 2005).

Individuals or households challenged by chronic diseases may inevitably face the double jeopardy of increased healthcare and self-care expenditure on one hand, and diminished productivity on the other. It is often left to the affected to invent private coping mechanisms (Sauerborn et al. 1996) especially where social security coverage is limited. As a means of coping, households may be compelled to increase supply of labour (though compromised) to markets, decrease consumption of other goods (intrahousehold compensation), or obtain transfers such as gifts, donations and other forms of remittances from other community members (interhousehold compensation). Using data from the 1997 LSMS in Côte d'Ivoire and Ghana, Schultz and Tansel (1997) showed that the number of days away from work resulting from illness reduce wages by at least 10%. Gertler and Gruber (2002) also showed that the level of consumption insurance enjoyed by Indonesian households is affected by the magnitude of the health shock. Wagstaff (2005) estimates the economic consequences of health shocks in Vietnam during the period 1993-1998 using a similar model as Gertler and Gruber's (2002), concluding that adverse health shocks (reduction in Body Mass Index (BMI)) decrease income earned from work and increase medical expenditure. In addition, although there is possibly an increase in unearned (transfer income) remittances, gifts and goods from relatives and friends, there is an undisputed

reduction in both food and non-food consumption, in which case consumption insurance failed. Shocks due to chronic disease may worsen inequalities in household-welfare distribution. For example, using data on adult nutrition in Ethiopia, Dercon and Krishnan (2000) showed that women are more intensely affected by adverse illness shocks and this effect is accentuated by the age and pre-marriage wealth differences between husbands and wives. It is therefore to be expected that the failure of coping mechanisms could threaten household recovery from shocks due to chronic diseases and that the risk of failure may be proportional to the number of individuals suffering chronic diseases in the household.

In this paper we use empirical data to investigate the evidence to how chronic diseases affect household healthcare expenditure, non-health consumption, labour (earned) income, and to demonstrate how households' unearned (transfer) income may provide some insurance against shocks from chronic diseases. Our a priori hypothesis is that because chronic diseases naturally occur throughout the course of a lifetime, they will adversely affect the health status of individuals to the detriment of the household's economic production. Our working questions are as follow: Do chronic diseases burden increase household health expenditure and to what extent? How do chronic diseases affect non-health expenditure? How do households fund increased household expenditure due to chronic diseases? Is there household productivity loss (indirect costs) with chronic diseases?

In 2004 the GDP per capital in Russia Brazil and India was $10,865, $8,140 and $1,830 (international dollars; source WHO (2006)). These countries are in different stages of epidemiological transition where chronic diseases are growing in importance. Deaths form chronic diseases are projected to increase by 22% in Brazil and 18% in India in the

next ten years (WHO 2005). Only around 59%, 45% and 25% of total health expenditure is funded from public sources in Russia, Brazil and India respectively and private sector funding through either private insurance or out-of-pocket spending is still a major source of funding for health care in these countries. Out-of-pocket payments represent 97%, 71% and 64% of the total private health expenditure in India, Russia and Brazil respectively (WHO 2006).

5.2.1 Chronic diseases in developing countries

The objective of this paper is to provide evidence on the effect of chronic diseases on household health care expenditures and income, and also to establish the extent of consumption insurance provided to non-health items. Two reasons determine the focus on chronic diseases.

The first is that chronic diseases, by their very nature, affect the long term health status of individuals, imposing substantial direct and indirect costs. For example, a victim of stroke will require continued medical attention after the event. Since the monetary costs, especially in developing countries, are not always fully covered by private insurance or the public health system, ultimately the burden of treating the patient is responsibility of the household. Also, it is not uncommon that the patient's ability to work and to take care of himself becomes compromised. Consequently, other members of the household are forced to either stop working in order to take care of the patient, or to increase their own supply of labor in order to generate additional resources. It is clear that there are many unanswered questions and it is relevant to study the consequences of chronic diseases, since they inflict

the same kind of adverse effect imposed by the catastrophic shocks documented by Gertler and Gruber (2002) and Wagstaff (2005).

The second motivation for the study of economic effects of chronic diseases is their increasing relevance for the populations of developing countries. The burden of disease in these countries has traditionally associated to communicable diseases. However, at the present moment several changes are taking place which are projected to increase the importance of chronic diseases. On the first place, the recent relative advances in the combat of communicable diseases and child mortality have contributed to the increase of life expectancy. As a consequence, the older populations are more vulnerable to the development of chronic diseases.

More importantly, however, is the current trend of change in traditional life styles, with the adoption of habits associated with risk factors for chronic diseases. Among such risks factors it is possible to include the ingestion of saturated fats and sugars, the increased use of tobacco and alcohol and the decreased levels of physical activity. Consequently, the incidence of chronic is projected to increase significantly in developing countries in the next decades (see Beaglehole and Yach 2003, Greenberg et al. 2005), with coronary heart disease and stroke being projected to become the "leading causes of death and loss of disability-adjusted life years by 2020" (Beaglehole and Yach 2003).

Beaglehole and Yach (2003) also remark that the adoption of risk factors for chronic diseases is not uniform among developing countries. Specifically, they show that countries where the control of communicable diseases has been more succesful, like Brazil and China, are also where risk factors for chronic diseases are currently more prevalent. This

process has been particularly intense in eastern and central Europe, the region that has registered the highest rates of mortality due to non-communicable diseases.

5.3 Methodology

The two-part model (2PM) was adopted for the analysis to address three problems that are typically encountered with household health and expenditure data:

- First, the usual nontrivial proportion of zero response in any given period,
- Second, the characteristically skewed distribution of expenditure data and
- Third, household health and expenditure data often suffer endogeneity either from missing data, reverse causality or simultaneity.

A number of modelling alternatives were considered prior to the selection of to the two-part model (see Blough, Madden and Hornbrook 1999, Cameron and Trivedi 2005, Jones 2000, Mullahy 1998). The two-part model distinguishes the decision process into two separate stages. For example, in the case of health expenditures, households first decide whether to incur health expenditure or not and later decide on the intensity of the expenditure. A binary choice model is specified for the first stage or censoring mechanism, and a linear regression model is specified for the outcome conditional on the outcome being observed on the second stage (Cameron and Trivedi 2005, p. 545).

The implicit assumption is that the two decisions are generated by separate probability processes (Pohlmeier and Ulrich 1995). We use the logit model to model the first stage

partly because it is more amenable to the panel data estimation planed for Russia among other reasons.

Let y_i denote one of the five continuous outcomes and x_i' the explanatory covariates. The probability of reporting the outcome of interest y_i is estimated as:

$$\Pr(y_i > 0|x) = \frac{\exp(x_i'\beta^{\text{logit}})}{1 + \exp(x_i'\beta^{\text{logit}})} = \Lambda(x_i'\beta^{\text{logit}}). \tag{5.1}$$

Where $\Lambda(\cdot)$ represents the logistic distribution, x_i is the $1Xk$ vector of covariates and β^{logit} is the $kX1$ vector of parameters to be estimated.

The second part is then estimated using ordinary least squares (OLS) regression on $\ln(y_i)$:

$$\ln(y_i) = z_i'\beta^{OLS} + \varepsilon^{OLS}, \varepsilon^{OLS} \sim N(0, \sigma_\varepsilon^2), \tag{5.2}$$

where z_i is the $1Xq$ vector of covariates and β^{OLS} is the $qX1$ vector of parameters to be estimated and ε^{OLS} is the error term. The vectors x_i and z_i are not restricted to contain the same set of regressors. We estimated the effect of chronic diseases on five different outcomes: (1) household healthcare expenditure, (2) non-healthcare expenditures, (3) earned income, (4) unearned income and (5) work loss for the head of the household. Two of the outcome variables: non-health expenditure and earned income that were reported by most households were however estimated using only OLS models following log transformation.

Although the effect of chronic diseases on each outcome is estimated separately, we cannot rule out the possibility that the outcomes are determined simultaneously. In this case it would be preferable to run a joint estimation of the system of equations determining the outcomes of interest. However, we follow other published studies which nonetheless estimate the effect separately (e.g. Gertler and Gruber 2002, Wagstaff 2005). The implicit

assumption in this modelling strategy is that the effects of chronic diseases on the different outcomes do not interact.

The main advantage of using the two-part model is to obtain estimates of the effect of explanatory variables on both the propensity to engage and on the conditional level of engagement. This can provide important insights on the decision process followed by households that might help the design of policy interventions. For instance, assume by hypothesis that in a given country chronic diseases affect the first stage of determination of health expenditures but not the second. That is, chronic diseases might increase the propensity to have health expenditures, but do not affect the conditional level of expenditures. If that is the case, then policy interventions should place additional effort on making sure that households do not incur on health expenditures in the first place, instead of trying to limit the overall level of expenditures. Of course, other effects with different implications are also possible, the important point being that the two-part model allows disentangling the effect on the two separate decisions.

The predictions from the two parts of the model were combined to recover estimates of the unconditional mean of the dependent variable in the models. Predictions on the original scale of y were based on Duan's smearing estimator, which prevents bias from the nonlinear transformation of the log residuals (see Duan 1983, Manning 1998, Manning and Mullahy 2001):

$$E(y_i > 0|x) = \boldsymbol{E}(\exp(x_i^0\beta^{OLS} + \frac{\sigma_\epsilon^2}{2})). \tag{5.3}$$

The Russian LSMS panel was unbalanced; therefore, both fixed and random effects models were estimated. This allowed for accounting for household unobserved heterogene-

ity (see Cameron and Trivedi 2005, Gertler and Gruber 2002, Wagstaff 2005). Similarly to the cross-sectional case, the following equation was implemented for the household-specific effects model:

$$\mathbf{Y}_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}. \tag{5.4}$$

Where $\mathbf{Y}_{it}$ represents the logarithm of any of the outcomes of interest for household i at period t, x_{it} is a vector of covariates, α_i represents unobservable household-specific effects, which are assumed to be time-invariant. Denote the averages of household outcome, covariates and errors over-the-time by $\bar{Y}_i = \frac{\sum_t \mathbf{Y}_{it}}{T}$, $\bar{x}_i = \frac{\sum_t x_{it}}{T}$ and $\bar{\varepsilon}_i = \frac{\sum_t \varepsilon_{it}}{T}$, respectively. The fixed effects models were implemented as:

$$Y_{it} - \bar{Y}_i = \alpha_i + x'_{it}\beta + \varepsilon_{it} - (\alpha_i + \bar{x}'_i\beta + \bar{\varepsilon}_i)\,. \tag{5.5}$$

If unobserver effects are assumed to be time-invariant, taking the difference between each observation and the household average eliminates α_i. Though estimates from Fixed Effects (FE) models are usually less biased, they were compared with the Random Effects (RE) models. The RE model ignores the over-time variance in the data, and uses only the variation between households to estimate the marginal effect of chronic diseases and is estimated as follows:

$$Y_{it} - \hat{\lambda}\bar{Y}_i = \left(x_{it} - \hat{\lambda}\bar{x}_i\right)'\beta + \left(1 - \hat{\lambda}\right)(\mu + \alpha_i) + \left(\varepsilon_{it} - \hat{\lambda}\bar{\varepsilon}_i\right). \tag{5.6}$$

Where μ is scaling parameter and $\hat{\lambda}$ provides a measure of the association between α_i and ε_{it} . Let σ^2_α denote the variance of the household-specific unobserved effec, and σ^2_e denote the variance of the idiosyncratic error. Then $\hat{\lambda}$ is a consistent estimator for:

$$\lambda = 1 - \frac{\sigma_e}{\sqrt{\sigma^2_e + T\sigma^2_\alpha}}. \tag{5.7}$$

In the absence of correlation between α_i and the regressors the RE estimator is more efficient than the FE. However, it is inconsistent if the true model is FE. We used Hausman test to compare all the parameters jointly and separately (Jones 2000). The RE model was assumed if the null hypothesis of no correlation between the disturbances and the regressors could be rejected.

Complementing our application of a 2PM, we employed Instrumental Variables (IV) estimator to further explore and control for endogeneity in order to obtain robust estimates. Respondents' place of residence and indicators related to the head of household (gender, age and month of birth) were considered possible candidates for controlling for suspected endogeneity for the cross-sectional data. For the Russian panel data, values of the dependent variables were additionally used as instruments. These instruments were expected to affect the health condition of the members of the household, whilst being uncorrelated with the error term in the main equation. The validity of the instruments was examined using the Hausman and Sargan's tests. All models were implemented using STATA software.

5.4 Data

The data for this study were obtained from the Life Standards Measurement Study (LSMS) household surveys for Brazil, India and Russia. The LSMSs are a series of surveys that the World Bank has performed in about 30 different countries since 1985. According to the World Bank, "the purpose of these surveys was to gather data that would allow the detailed study of household behavior and several aspects of living" (Grosh and Glewwe 1995).

The data collection for Brazil was carried out in 1996-1997 in the South-Easth and the North-East regions. The Indian study was carried out in 1997-1998 in south and eastern Uttar Pradesh and north and central Bihar, and covered only rural areas. The Brazilian and Indian surveys provide only one cross-section of data. The Russian data is a panel set that follows the same households over time. We used the eight most recent rounds (from 1997 to 2004) to avoid the variation in the definition of the variables across earlier rounds in Russia. These eight rounds have standardized variable definition but combine into an unbalanced panel. There were 4,940 household samples in the Brazil data, 2,249 in the Indian data and an average of 4,179 in the data from Russia. Table 5.1 provides descriptive statistics for the variables used in the analysis.

The LSMS sample design is based on the collection of data at the household level. For some variables, such as hours worked, the questions are also asked in an individual basis, so it is possible to disaggregate the observations up to the individual level. For most variables, however, such separation is not possible. This is particularly true for the variables on expenditures, which are the main interest for us. The majority of LSMS data was obtained for a 30-day period prior to the time of interview.

The questionnaires are composed by several modules, such as labour, education, migration and health, which cover different aspects of the living standards of the household and its individual members. The data collection is based on interviews performed at the households. We have access to information relative both to the household as a whole (e.g., income and expenditure) and to each of the other members in particular (e.g., age, sex, education, health status and hours worked for each member).

Five dependent variables were constructed to capture: household health expenditure, non-health expenditure, earned income, unearned income, and work days lost, to explore the possible dimensions to a household's economic adjustment to chronic diseases (see Gertler and Gruber 2002, Wagstaff 2005). All monetary variables are valued at 2000 international currency.

5.4.1 Dependent variables

Household health expenditure

About 56% of Brazil's and 62% of Russia's households reported incurring healthcare expenditure, compared to 98% of households in India. As discussed in section 5.2, this difference is related to the organisation of the health system in India, which relies much more on private out-of-pocket payments compared to the other two countries. The household healthcare expenditure was obtained by aggregating all household expenditures on healthcare goods and services. On average, healthcare expenditures ranged between 6%-10% of total household expenditures. They accounted for 6.4% of total expenditures in Brazil, 9.6% in India, and around 7.0% in Russia.

Household non-health expenditure

Households may shift non-health consumption to health consumption in the event of chronic diseases: in this situation, non-health expenditure may decrease even as health expenditure increases (i.e. the occurrence of consumption insurance against chronic diseases). All non-health expenditure was aggregated to a dependent variable (nonhealthexp).

Table 5.1. Descriptive statistics

Variable	Description	Brazil	India	Russia
earned	Labour income*	773.52	139.00	772.68
unearned	Non-labour income*	590.99	22.34	522.47
dunearned	Dummy unearned>0	0.54	0.21	0.82
healthexp	Health expenditures*	66.57	13.62	86.86
dhealthexp	Dummy healthexp>0	0.56	0.98	0.62
nonhealthexp	Non-health expenditures*	543.56	129.15	1,655.24
chronic	Individuals with chronic disease	0.50	0.10	0.32
nonchronic	Individuals with non chronic disease	0.55	0.56	0.89
bad	Individuals with bad health	0.11		0.38
physical	Individuals engage in physical activities	0.43		
smoke	Smokers			0.71
overweight	Overweight individuals	0.80		1.08
headactiv	Days of activity lost by head in last 30 days*	8.18	3.40	10.08
dheadactiv	Dummy headactiv>0	0.08	0.26	0.04
headwork	Dummy head works	0.72	1.00	0.44
headmale	Dummy head is male	0.77	0.96	0.74
single	Dummy head is single	0.12	0.03	0.25
kids	Children (below 18 years)	1.48	2.98	0.67
men	Male adults	1.01	1.56	0.73
women	Female adults	1.29	1.63	0.78
oldmen	Males above 65 years	0.15	0.26	0.17
oldwomen	Females above 65 years	0.19	0.20	0.42
age	Average age**	40.20	38.20	47.26
education	Average schooling years**	6.75	2.51	8.77
insured	Insured individuals	0.64		1.71
urban	Dummy urban area	0.79		
caste	Dummy middle or backward castes		0.80	
NE	Dummy NE region	0.50		
moscow	Dummy Moscow region			0.10
north	Dummy North region			0.07
central	Dummy Central region			0.20
volga	Dummy Volga region			0.19
caucasus	Dummy Caucasus region			0.12
ural	Dummy Ural region			0.15
siberia	Dummy Siberia region			0.19
Year 1997				0.13
Year 1998				0.13
Year 1999				0.13
Year 2000				0.14
Year 2001				0.15
Year 2002				0.16
Year 2003				0.16
Year 2004				0.16
N		4,938	2,249	27,487

*Include only outcome>0; **Among adults; Monetary values in international dollars of 2000.

Earned (labour) and unearned (transfer) incomes.

Changes in earned (labour) and unearned (transfer) incomes consequent upon chronic diseases were regressed against determinants to capture income impact of disease. This will also provide indications as to the strategies adopted by households to cope with the additional burden caused by chronic diseases. For Brazil and Russia, labour income accounts for around 60% of total income, while for India it accounted for around 86%.[15]

Indirect cost of chronic diseases

Assuming that household productivity loss is inevitable in the event of chronic disease, these indirect costs were estimated by measuring the impact on the head of the household's labour supply. This was proxied by the number of activity days lost to disease within the 30 days preceding the interview (headactiv). This information was available for Brazil and Russia though unavailable in the Indian data.

5.4.2 Independent variables

Crucial to this analysis are two explanatory variables: chronic and nonchronic, which indicate the number of individuals reported to have chronic and non-chronic (acute) diseases respectively in households. On average, there were 0.5 individuals reporting chronic disease per household in Brazil, 0.1 in India and 0.32 in Russia. Given the average number of adults per household of 2.63 in Brazil, 3.66 in India and 2.11 in Russia, these imply an overall household "prevalence" of reporting chronic diseases to be 19.0% in Brazil and

---

15 Some noted discrepancies between total income and total expenditures, especially for Russia may be due to underreporting of the income variables.

15.0% in Russia, but only 2.9% in India.[16] The prevalence of non-chronic (acute) diseases also varies between 21.0% in Brazil and 15.2% for India, and is slightly higher in Russia - around 42.2%. The variable bad, which refers to the number of individuals reporting bad health, was included to control for the fact that the perception of the severity of illness may influence reporting and help-seeking behaviour.

Other covariates included in the models were selected either on theoretical basis or have been shown in related studies to influence household income and expenditure. Demographic variables include household adults mean age (age); number of women (women); number of adults in the household; marital status of household heads (single); and caste in India data. All household variables were computed around number of adults (age 18 and above) in the household as a denominator, since chronic diseases are infrequent below adult age group.

Some chronic disease associated risk factors may also generate need for medical expenditure. For Brazil, these were physical (number of individuals that engage in physical activity), overweight (number of individuals with body mass index (BMI) above 25 units in the household), while overweight and smoke (number of individuals that smoke) were included in Russia. Information on risk factors was unavailable in Indian data. The variables overweight and obese were computed by taking into account the *weight* and *height* of individuals. With such information it is possible to compute the individual body mass

---

[16] The list of diseases varied among the surveys. For Brazil, they include heart problem, high blood pressure, diabetes, respiratory problems, digestive problems, gynaecological problems, prostate problems, allergy, cancer, bone/muscle/joint problems, neuro-psychiatric problems and high cholesterol. Included for India, are respiratory problems, heart problem, blood pressure, cataract, and permanent disability. For Russia, the chronic diseases included are diabetes, infarction and stroke.

index (BMI), by using the following formula $BMI = \frac{weight}{height^2}$. Individuals with a BMI between 25 and 30 units are classified as overweight, while those with BMI equal or greater than 30 are considered obese. Again, there is an evident difference between the countries, and we note a higher incidence of both overweight and obesity in Russia as compared to Brazil.

Accessibility to care reflects strongly on the healthcare expenditure relative to income. We broadly controlled for household geographical location by including rural or urban areas (as opposed to metropolitan areas). We assume rural dwellers have less access to chronic disease healthcare compared to urban dwellers. For Brazil, regional location was also controlled for because the Northeast region is markedly less developed compared to the Centre-South of the country. For Russia, the RE models include dummies for six areas (Moscow, Central, Volga, Caucasus, Ural and North/Siberia). The Indian survey covered only rural areas.

5.5 Results

In discussing the main results, the tests that were performed to assess the validity of the instruments and to determine the preferred estimation method (OLS or IV for Brazil and India; OLS-FE, IV-FE, OLS-RE or IV-RE for Russia) are highlighted. The complete set of results is reported in the appendix.

5.5.1 Health expenditures

The presence of chronic diseases in a household significantly increased the propensity (probability) and the magnitude of incurring health expenditure (Table 5.2). The marginal effects for the logit models were computed approximately as $\bar{y}\,(1-\bar{y})\,\beta$, $\bar{y}$ being sample mean healthcare expenditure and β the estimated coefficient (Cameron and Trivedi 2005). Although there was no effect on the probability of reporting health expenditures in India (98% of households in India reported health expenditures), each additional case of chronic disease in the household increased the probability of incurring health expenditure by 17.8% points in Brazil, and 5.9% points in Russia. The Hausman test for Russia rejected the hypothesis of no difference between FE and RE estimators, suggesting the existence of unobserved household fixed effects in the determination of the propensity to report health expenditure.

In the second part of the model, the Hausman test for Brazil and India did not show systematic differences between the OLS and IV estimates (see Hausman test IV for Brazil and India in Table 5.2). Although the instrument used for Brazil did not pass Sargan's test, we generally interpreted the result as evidence of no correlation between the endogenous regressors and the error term. Considering that IV estimates are potentially less efficient than OLS (Cameron and Trivedi 2005), we opted to assume the OLS estimates for Brazil and India. In the case of Russia, the Hausman IV test indicated significant differences between the OLS and IV estimators for the fixed effects models, providing evidence of the presence of endogeneity. Similarly, the Hausman FE test detected systematic differences

Table 5.2. Two part regressions for health expenditures

	Brazil		India		Russia RE	
Variable	*Logit*	*OLS*	*Logit*	*OLS*	*Logit*	*OLS*
chronic	.72***	.21***	1.2	1.2***	.36***	.15***
nonchronic	.84***	.12***	2.4***	.82***	1.1***	.27***
bad	0.08	.19***			.16***	.23***
lnincome	.14***	.2***	-0.12	.064**	.23***	.2***
smoke		0.01			-0.01	
physical	0.02	0.04				.072***
overweight	-0.02				.058**	-.059***
headmale	0	0.06	0.39	0.02	-.4***	0.04
single	-.32***	0.03	-1.8**	-.47**	-.64***	-.1**
kids	.15***	-0.01	0.09	.059***	.13***	.028*
men	-.15***	-0.04	0	-0.03	-.3***	-.096***
women	-0.03	0.03	0.06	.13***	-.26***	-.084***
oldmen	-0.03	-.13*	0.6	0.1	-.54***	-.086**
oldwomen	0.12	0.06	0.67	-.15*	-.35***	-.19***
age	-.025*	.017*	.16**	0.01	-.043***	-.0098*
age^2	.00034**	0	-0017*	0	00047***	.000099*
education	.2***	.098***	0.12	.19***	.15***	0.01
$education^2$	-.01***	0	-0.02	-0.01	-.0062*	.0041*
insured	.096**	.065***			.09***	.054***
urban	0.1	-.14**				
NE	.22***	-.44***				
caste			0.07	-0.03		
1998					.12**	.094**
1999					.44***	0.03
2000					.59***	0
2001					.73***	0.02
2002					.87***	0.04
2003					.76***	.17***
2004					.95***	.14***
moscow					0.1	.14***
central					.31***	-.11***
volga					.14**	-.19***
caucasus					.25***	.27***
ural					-0.08	-.2***
cons	-1.9***	.85***	-0.08	2.2***	-2.2***	3.1***
N	*4,805*	*2,695*	*1,920*	*1,886*	*26,986*	*16,787*
R^2	14.00%	26.00%	15.00%	31.00%		12.00%
R^2 within						5.90%
R^2 between						15.00%
Sargan test		22***		1.3		
Hausman test IV		1.61		2.12		76.6***
Hausman test FE		336.64***				79.1***
σ_m						0.62
σ_ϵ						1.2
ρ						22.00%

Methods: Logit and Logit-RE (Russia) in first part; OLS and OLS-RE (Russia) in second. Legend: * p<.1; ** p<.05; *** p<.01.

between the OLS-FE and OLS-RE estimators, which suggest the presence of unobservable household fixed effects. However, the null hypothesis of non-systematic differences in coefficients of the OLS and IV estimators (Hausman IV for RE column) could not be rejected, nor can the null for RE-IV and the FE-IV (Hausman FE for IV column). Similarly, for both Brazil and India, we opted for the OLS estimator with Random Effects, which provides greater efficiency when compared with the IV estimator.

Each additional case of chronic disease increased household health expenditure by 21% in Brazil and 14% in Russia. The estimates for many of the variables are consistent for Brazil and Russia. For instance, the estimated income elasticity of healthcare expenditure is between 0.12 and 0.22 for both countries. Similarly, the two other health indicators (nonchronic and bad) were also consistently associated with increased probability and increased level of health expenditures in all countries. The conditional effect of non-chronic conditions was between 12% and 28%, whilst for bad health it was between 18% and 29%.

Male and single civil status of household headship was consistently associated with lower probabilities of expenditure, and lower levels of expenditure (for India and Russia). This negative association may indicate the existence of access problems for households headed by male or single individuals. Average schooling in the household positively increased the probability of expenditures, as often reported elsewhere (eg. Cowell 2006, Fuchs 2004, Grossman 2000). Health insurance increases both the propensity and the average levels of health expenditure.

NE households in Brazil had higher probability of spending, but lower average levels of health expenditure (urban households also report lower conditional levels of health

expenditure). In Russia, households from Moscow and Caucasus have higher average levels of health expenditure when compared with households from the North/Siberia, while households from other regions have lower levels of healthcare expenditure. An increasing time-trend was also observed.

5.5.2 Non-health expenditures

Results show that non-health expenditure increased positively with the average level of education of the household (at least in Brazil), average age of adults, while the number of overweight or bad health state (bad) reflects negatively on non-health expenditure. Only the second-stage (linear regression of log non-health expenditures) is used in this case because all households report non-health expenditures. Table 5.3 shows no significant differences between the OLS and IV estimators for the Brazil and India models hence the OLS estimator was considered efficient and preferred. Chronic disease showed no significant impact on non-health expenditures in the India data, while an increase in the number of people with chronic disease is associated with approximately 4% of growth in non-health expenditure in Brazil. For Russia, the Hausman tests indicated the existence of household unobserved heterogeneity (since both RE estimators are rejected) and evidence of endogeneity. Consequently, the IV-FE method was preferred, since it allows for controlling for both effects. Adopting this estimator, an increase in the proportion of household member with chronic disease resulted in 3.2% reduction in households' non-health expenditures as was expected.

The countries considered in this study have fragile systems of social and healthcare services. This implies that households are frequently exposed to the economic conse-

Table 5.3. Regression for non-health expenditures

Variable	Brazil	India	Russia
chronic	.039***	0	-.017*
nonchronic	.037***	0	.014**
bad	-.047*		-.069***
lnincome	.42***	.06***	.4***
smoke	.057***		
physical	.041***		.067***
overweight			0
headmale	.15***	.14**	.068***
single	-.15***	-.39***	-.17***
kids	.049***	.079***	.093***
men	-0.01	.12***	0.03
women	.028*	.12***	.033**
oldmen	-.064**	.13***	.058***
oldwomen	-0.01	-.065**	-0.01
age	.023***	.037***	.012***
age^2	-0029***	-0041***	-.0002***
education	.067***	.092***	-0.02
$education^2$	0	0	.0032***
insured	.077***		0.01
urban	.089***		
NE	.054***		
caste		-0.04	
1998			-.1***
1999			-.35***
2000			-.24***
2001			-.17***
2002			-.18***
2003			-.14***
2004			0.03
moscow			.092***
central			-.11***
volga			-.2***
caucasus			-0.03
ural			-.22***
cons	2***	5.3***	5.1***
N	*4,796*	*1,920*	*26,986*
R^2	65.00%	54.00%	45.00%
R^2 overall			21.00%
R^2 within			58.00%
Sargan test	81***	1	259.82***
Hausman test IV	1.19	3.97	1117.98***
Hausman test FE			0.38
σ_m			0.64
σ_ε			0.63

Methods: OLS and IV-FE (Russia). Legend: * p<.1; ** p<.05; *** p<.01.

quences of illness. Chronic diseases can impact on non-health expenditures (e.g. further expenses with informal carers), which may or may not be covered by the public system. On the other hand, results from Russia suggest the existence of unobservable fixed effects and that chronic diseases reduce non health expenditures. Whilst this sheds some doubts on the result from Brazil, we cannot rule out the possibility that the definition of what is considered health/non-health expenditures differs between the samples.

Considering that non-health expenditure is a reliable indicator of the income level of the household, these observations agree with the human capital notion of health status especially in the developing countries, that good health exerts positive effects on earnings potential (Thomas and Strauss 1997). Finally, in Russia there was a clear negative time-trend, which reflect general economic conditions in Russia during the period.

5.5.3 Earned (labour) income

As shown in Table 5.4, each additional case of chronic disease decreases the household earned income at an approximate rate of -7.9% in Brazil, and -4.3% in Russia. This reflects the household indirect costs (reduced productivity) due to chronic diseases. For Brazil, the Hausman test did not reject the (more efficient) OLS estimator, and similarly for Russia, FE-OLS was preferred.

The number of individuals with perceived bad health also had a negative impact: -7.9% in Russia and -15% in Brazil. In the three countries, non-chronic (acute) diseases had a lower impact on income earnings compared to chronic illnesses. This suggests that the household economic capacity is more seriously affected by long-term health problems, as in the case of chronic diseases, than by short-term shocks. This argument has been put forward by Gertler and Gruber (2002), which, together with Wagstaff (2005), have reported similar effects on the impact of chronic and acute diseases.

Similarly as is the case of non-health expenditure, measures of physical health (over-weight) and education, which reflect the human capital endowments, consistently increased

Table 5.4. Regression for earned (labour) income

Variable	Brazil	India	Russia
chronic	-.079***	0.02	-.043**
nonchronic	-.06***	0.02	0
bad	-.15***		-.059***
lnincome			
smoke			0
physical	0.03		
overweight	.089***		.062***
headmale	.25***	0.09	.16***
single	-.15***	-0.15	-.25***
kids	.017*	.048***	.067***
men	.23***	.18***	.2***
women	.15***	.059*	.16***
oldmen	-.13**	0.08	0.01
oldwomen	-.15***	0	-0.01
age	.035***	0.02	.11***
age^2	-.00028***	-.00033**	-.0013***
education	.13***	-.081*	-.077**
$education^2$	0	.021***	.0044*
insured	.18***		.054***
urban	.42***		
NE	-.46***		
caste		0.05	
1998			-.16***
1999			-.26***
2000			-.12***
2001			-0.03
2002			.052*
2003			.11***
2004			.33***
moscow			
central			
volga			
caucasus			
ural			
cons	3***	6.6***	5.7***
N	*4,126*	*1,810*	*22,866*
R^2	55.00%	17.00%	31.00%
R^2 within			12.00%
R^2 between			40.00%
Sargan test	133***	3.5	484.08***
Hausman test IV	3.84	0.08	-187.94
Hausman test FE			-446.84
σ_m			1
σ_ε			0.94
ρ			54.00%

Methods: OLS and OLS-FE (Russia). Legend: * p<.1; ** p<.05; *** p<.01.

household earnings. The number of smokers in the household had no effect on labour income in Russia.

5.5.4 Unearned (transfer) income

The results clearly show that chronic diseases increased the probability of receiving transfers (gifts and other remittances) by 7.0% in Brazil and 4.1% in Russia as shown in Table 5.5. This suggests that households were able to adapt to chronic health shocks by relying on informal coping mechanisms that reflect the degree of risk sharing among households (Gertler and Gruber 2002). Apart from chronic diseases, households headed by males had a lower probability of receiving remittances, while the number of old people increased the probability. The results also indicate a tendency for low caste Indian households to receive lower remittances, both in terms of the probability of receiving and of the level of the remittances.

5.5.5 Productivity losses

Finally, results show that chronic diseases may reduce labour productivity (measured by the number of workdays lost by the head of household) and as shown in Table 5.6. There was an increase in the probability of losing workdays in households with increasing cases of chronic diseases in the Indian data. However, in the Brazilian and Russian data, non-chronic illnesses showed significant positive effects on the probability of work loss.[17] Though

---

[17] Given the limited number of observations, it was not possible to estimate the second part of the model for this outcome.

Table 5.5. Two part regressions for unearned income

	Brazil		India		Russia RE	
Variable	*Logit*	*OLS*	*Logit*	*OLS*	*Logit*	*OLS*
chronic	.33***	0.03	0.02	-0.02	.29***	.083***
nonchronic	0	-0.03	-0.05	-.18**	.21***	.063***
bad	-0.05	-0.07			0.01	-0.02
lnincome					.067*	-0.01
smoke						
physical	.2***	.055*				
overweight	0.03	.1***			-.12***	0.01
headmale	-1.3***	-.31***	-1.3***	-0.02	-.51***	-0.02
single	-0.03	-.15*	-0.08	0.65	0.01	-0.05
kids	-0.04	-.042**	0.02	-0.02	.46***	-.024**
men	.1*	.11***	0.09	0.1	-0.07	.045*
women	0.09	.1***	0.11	.52***	-0.01	.043*
oldmen	1.5***	.44***	0.16	.37**	1.1***	.51***
oldwomen	1.5***	0.03	-0.01	-.36*	1.9***	.42***
age	-.052**	.026***	-.07**	0.03	-.12***	-.015***
age^2	.0011***	0	.001***	0	.0013***	.00017***
education	.096***	.093***	.23*	-0.02	.39***	.04*
$education^2$	0	.0022*	-0.02	0.03	-.019***	0
insured	0.01	.11***			.14***	.023**
urban	0.03	.25***				
NE	0.03	-.14***				
caste			-.29**	-.54***		
1998					-.81***	.063**
1999					-.56***	-.3***
2000					-0.09	-.26***
2001					0.12	-.12***
2002					.14*	-.048*
2003					.21***	.045*
2004					.31***	.065*
moscow					.39***	.092**
central					.4***	-.1***
volga					.29***	-.23***
caucasus					.22**	0.02
ural					.21***	-.21***
cons	0.23	3***	0.36	2.9***	1.5***	7***
N	4,938	*2,673*	2,247	*481*	27,438	*22,573*
R^2	23.00%	32.00%	3.80%	29.00%		13.00%
R^2 within						5.50%
R^2 between						17.00%
Sargan test		118***		6*		
Hausman test IV		8.61**		0.06		-85.11
Hausman test FE					88.13***	111.45***
σ_m						0.62
σ_ϵ						0.92
ρ						31.00%

Methods: Logit and Logit-FE (Russia) in first part; OLS and OLS-RE (Russia) in second. Legend: * p<.1; ** p<.05; *** p<.01.

Table 5.6. Regression for work days lost

Variable	Brazil	India	Russia FE
chronic	0.11	2.8***	-0.01
nonchronic	1.3***	3***	.98***
bad	0.21		0.13
lnincome	-0.07	-0.07	0.05
smoke			-0.11
physical	-0.13		
overweight	-0.03		-0.08
headmale	-.78***	-0.43	0.15
single	-0.02	0.65	-0.02
kids	0.01	0	0.01
men	-.34***	-.22**	-.51***
women	-.56***	-.38***	-.31**
oldmen	-0.28	.35*	-.96***
oldwomen	-0.2	-0.09	-1.2***
age	0	0.01	.084**
age^2	0	0	-.0011**
education	-0.04	-0.17	0.4
$education^2$	0	0.02	-0.02
insured	0.01		.12*
urban	-0.12		
NE	0.11		
caste		0.1	
1998			-.26**
1999			-0.2
2000			-.54***
2001			-.44***
2002			-.44***
2003			-.55***
2004			-.71***
moscow			
central			
volga			
caucasus			
ural			
cons	-1.1*	-1.9*	
N	4,805	1,925	4,596
Pseudo R^2	19.00%	42.00%	12.00%
Hausman test IV			
Hausman test FE			41.5**
σ_m			
σ_ϵ			
ρ			

Methods: Logit and Logit-FE (Russia). Legend: * p<.1; ** p<.05; *** p<.01.

this finding may seem counterintuitive, it should not be seen as completely unexpected, as work days lost may be more sensitive to non-chronic than chronic diseases: early retirement through disabilities would obviate labour participation at the time of interview. We would argue that productivity losses from chronic diseases could be better captured in data on early retirement from active labour than workdays lost within a period of 30 days as was captured in the data. Overall, the productivity loss model of our analysis had an ill fit, which prevented more rigorous analysis. One possible improvement would be to consider the effects at the individual, rather than household level.

Table 5.7. Marginal and conditional effects of chronic diseases

	Probability of outcome			Conditional level of outcome			Unconditional effect		
	Brazil	India	Russia	Brazil	India	Russia	Brazil	India	Russia
Health expenditures	17.8%*** (55.8%)	2.10% (98.2%)	5.9%*** (62.5%)	21%***	120%***	14%***	3.7%*** ($66.48)	2.50% ($13.62)	0.8%*** ($89.94)
Non-health expenditures				3.9%***	-	-3.2%*	3.9%*** ($542.85)	- ($129.14)	-3.2%* ($1620.26)
Earned income				-7.9%***	2%	-4.8%***	-7.9%*** ($772.51)	2% ($138.99)	-4.8%*** ($812.75)
Unearned income	8.2%*** (54.1%)	0.30% (21.4%)	4.1%*** (82.7%)	3.10%	-2%	8.7%***	8.20% ($590.23)	0% ($22.34)	4.1%*** ($542.05)

Average values given in parentheses; Monetary values in 2000 international dollars per 30-day period; Legend: * p<.1; ** p<.05; *** p<.01.

5.5.6 The effects of chronic disease in the household

A summary of the effects of chronic diseases on the evaluated outcomes are presented in Table 5.7. Each additional case of chronic disease increased the probability of incurring health expenditure. On average chronic diseases exposed households to 30% higher probability of incurring health expenditure relative to the benchmark average of 56% in Brazil and 10% higher than the 62% benchmark in Russia. Transfers and remittances to Russian and Brazilian households increased with chronic diseases. The probability of receiving transfers from other households in Russia and Brazil also increased with the occurrence of chronic diseases.

We show the annualized impact of chronic disease on households in Table 5.7. Chronic diseases significantly increased health expenditure to an annual $965 and $1,042 (2000 international currency) in Brazil and Russia respectively. If the 2003 per capita total expenditure estimates of $597 in Brazil and $ 551 in Russia (international dollars; source WHO (2006)) are converted to per-household estimates (by multiplying by the average individuals per household), the increase represents 61% and 90% total expenditure per household. Annual household non-health expenditure decreased considerably in Russia by $19.862 and increased by $6.523 (3.9%) in Brazil. No significant increase was observed in India.

Results indicate that the presence of a chronic disease in a household decreased labour income by 4.8% ($9,272) in Russia and by 7.9% ($9,282) in Brazil. The possibility for underreporting income should however be taken into account in interpreting these results. The reduction of labour income seemed to be at least partially offset by an increase in transfers and remittances associated with the presence of chronic diseases as observed in the three countries.

Overall, chronic diseases increased household healthcare expenditures and reduced earned income. Although there was an increase in transfers and remittances, the increase in health expenditures and the decrease in labour income tended to dominate, and consequently imposed a reduction in the consumption of non-health related items. The net effect in Russia was a reduction of 3.2% on the level of non-health expenditures[18], suggesting that chronic diseases in fact reduced household overall welfare.

In Figure 5.1 are displayed the impacts of the number of persons with chronic diseases in the average household on the outcomes of interest. Healthcare expenditure and remittances increased proportionately to the number of chronic diseases in households in a similar manner in the three countries regardless of the presence of health insurance schemes in Brazil and Russia. Income fell gradually in Russian and Brazil to the third chronic disease in a household. The observed rise in income in the fourth and fifth is mainly due to small number of households and outliers in this category. It would be expected that chronic disease related rise in health expenditure is insured by reduction in non-health expenditure.

---

[18] The estimated positive impact on household non-health expenditure in Brazil may be due to household unobserved heterogeneity which was significant in the in the Russian estimates, but could not be accounted for in the Brazilian data.

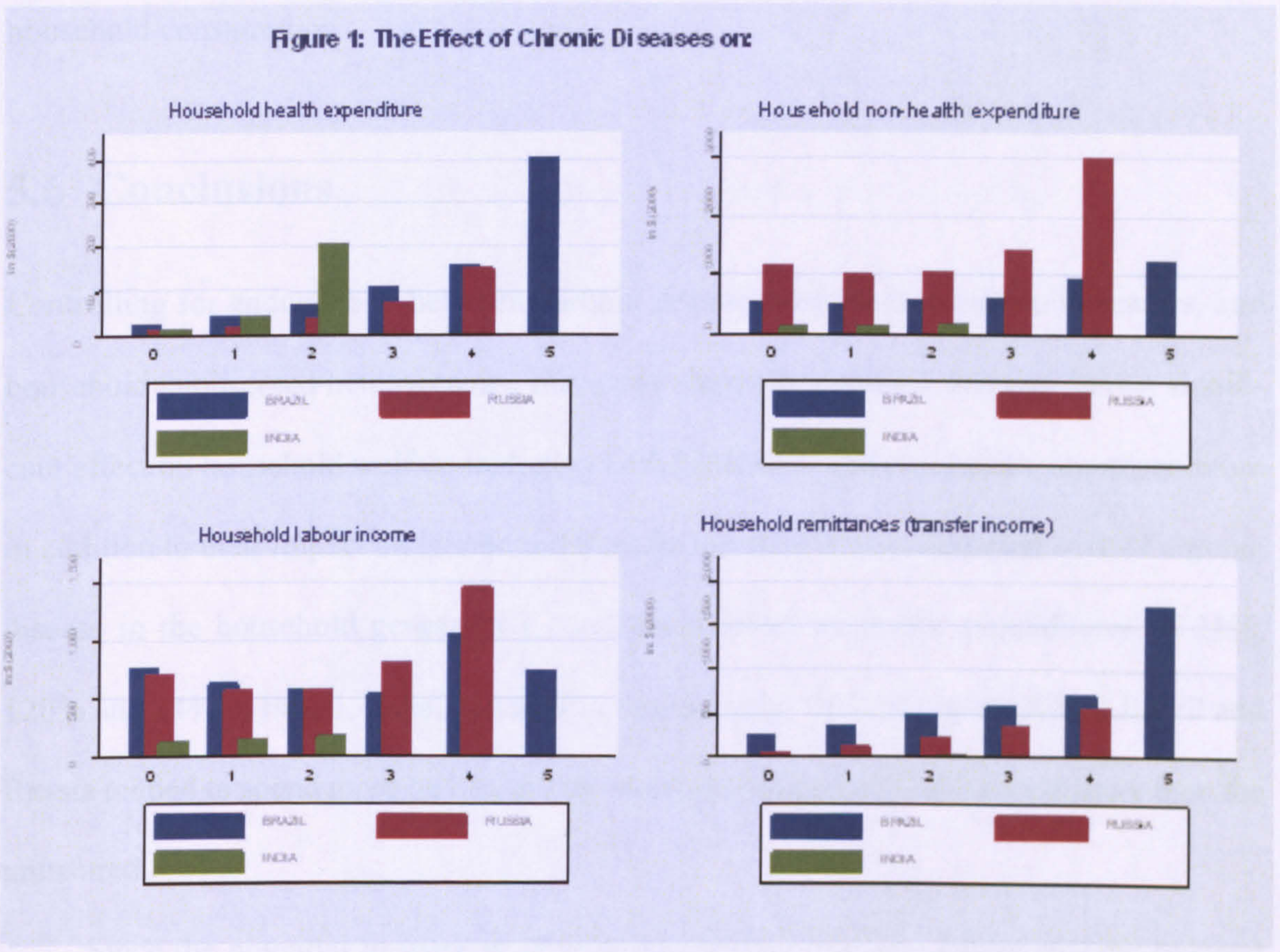


Fig. 5.1. The effect of chronic diseases

However, non-health expenditure gradually increased with increasing number of chronic diseases. This may be due to: (1) the increase in remittances to household with chronic diseases, whereby household health expenditure shocks; (2) increased self-care needs following chronic disease and; (3) it may be that households tend to protect against threats to household consumption.

5.6 Conclusions

Controlling for endogeneity between chronic diseases and socioeconomic indicators, and household unobserved heterogeneity, this study shows that chronic diseases have a significant effect on household welfare including both healthcare and non-healthcare expenditure in addition to their impact on labour and transfer incomes. Each additional case of chronic disease in the household generated a conditional effect on health expenditures of 21%, 120% and 14% in Brazil, India, and Russia, respectively. Insured households in Brazil and Russia tended to spend more on health care and incurred higher health expenditure than the uninsured.

Since the potential to work is affected by chronic diseases, labour income reduced by 7.9% in Brazil and 4.8% in Russia. On the other hand, households seemed to have received some compensation in the form of remittances, which increased by 3.1% in Brazil and 8.7% in Russia. However, the final effect of chronic diseases was an increase of 3.9% of non-health consumption in Brazil and a reduction of 3.2% in Russia. Overall, there is a negative net effect of chronic diseases on the household economy.

The prevalence of chronic diseases is projected to increase over the coming years especially in low- and middle-income countries (WHO 2005). Given that formal social support systems in many such countries are limited in coverage, households' socioeconomic wellbeing and indeed grass-roots development may be hindered by the burden of chronic diseases without appropriate intervention. This study provides empirical evidence for the theoretical links between chronic diseases, socioeconomic inequality and poverty. It has some policy implications for the treatment and control of chronic diseases in the countries included in the analysis. For instance, there is justification for reviewing the financing mechanisms to address chronic diseases particularly if the economic impact on individuals and households is to be mitigated. This is because there is a limit to what end informal financing could aid households in coping with the socioeconomic burden associated with chronic diseases.

The results also indicate that socio-economic status also influences the impact of chronic diseases on household welfare. Households headed by women or single parents reported consistently lower levels of labour income and higher levels of health expenditures. Higher average level of education in households is associated with higher earnings and consequently higher expenses. Households that are more educated had both higher labour and unearned (transfer) incomes, suggesting the importance of social capital indicators in determining access to assistance from other households. These suggest that the formulation of health financing policies should take into account country-specific inequalities that influence the impact of chronic diseases.

A number of limitations are noted. We have used self-reported data from the LSMS series of surveys although it was not possible to fully account for systematic misreporting of errors, which are probable with self-reported income and expenditure data. We assumed that the rigorous and detailed data collection process employed in the LSMSs might have palliated this problem. Except for Russia where an eight-year panel was available, we have used only cross-sections for both Brazil and India. Panel data would have been preferable for all countries were they available. Nevertheless, the conclusions are generally similar among all three countries.

5.A Appendix

In order to deal with the skewness of the data, we estimated two part models (2PM) with log transformation in the second part, which reduced the skewness of the cumulative distribution and stabilized the variance over the observations. The major difficult associated with this kind of model arises when one tries to retransform the dependent variable to its original scale. As several authors have remarked (see Duan 1983, Manning 1998, Mullahy 1998, Manning and Mullahy 2001, Deb and Burgess 2003), the retransformation from the log-scale to the original scale is affected by the variance of the log-scale. In our case, since the data did not present heteroscedasticity in the transformed scale, it was possible to apply Duan's smearing parameter (5.3).

An alternative method would be to employ generalized linear models (GLM) (see McCullagh and Nelder 1989). The advantages of this approach are that it imposes "minimal assumptions and obviates the need to transform the data; rather it represents a reparameterization of the model that retains the original scale (...) of the response" (Blough et al. 1999, p. 154). Although the use of GLM was not strictly necessary in this application, in this appendix we discuss how to implement them and present the results from the estimation. Overall these results are qualitatively similar to the estimation using the log transformation presented in the main text.

5.A.1 Two part model using GLM

Assume that y represents health care expenditures and x is the vector of covariates that one believes to affect y; 2PM make use of fundamental probability to decompose the determi-

nation of health care expenditures into two components:

$$E[y \mid \mathbf{x}] = P[y > 0 \mid \mathbf{x}] \cdot E[y \mid y > 0, \mathbf{x}]. \tag{5.8}$$

In this specification, one first estimates the probability of incurring expenses and then computes the expected disbursement for those with positive levels of expenditures. The overall expected expenditure is given by the product of the two estimates.

The first part, also referred to as the hurdle component, models the probability of incurring in any expenditure during a given period of time. It may be estimated using any standard model of binary choice, like the Probit and Logit models.

By its turn, the second part of the model considers only those observations with strictly positive expenditures. One cannot proceed by simply applying OLS to this subset of observations, since doing so would result in counter-intuitive estimates. In other words, in this case if one merely applies the OLS method (that by construction is defined to the whole of real numbers) the estimated values of expenditure might turn out to be negative, which is logically impossible. Another shortcoming of OLS is that, as it would not account for the possibility of heteroscedasticity in the distribution of expenditures, the estimates would be biased.

One possible and very popular alternative consists on imposing a transformation on the dependent variable, such as taking the logarithm of the expenditures. This transformation smoothes the distribution of expenditures, inducing it to approximate a Normal distribution. In this case, the estimated model would be:

$$E[\log(y_i) \mid y_i > 0, \mathbf{x}] = x_i\beta. \tag{5.9}$$

In principle, one could then retransform the estimated values fitted by Eq. ((5.9)) in order to recover the estimates for y in the original scale. However, There is an additional complication associated with that retransformation. As shown by Mullahy (1998, p. 252), according to Jensen's inequality, if one simply takes the exponential of the fitted values on the logarithmic scale:

$$E\left[\exp(\log(y_i)) \mid y_i > 0, \mathbf{x}\right] = E\left[y_i \mid y_i > 0, \mathbf{x}\right] > \exp(E\left[\log(y_i) \mid y_i > 0, \mathbf{x}\right]). \quad (5.10)$$

This implies a retransformation error that calls for the use of an additional component to compensate for it. Duan (1983) proposed the use of a smearing estimator that would get round the retransformation error represented in Eq. ((5.10)). As suggested by Deb and Burgess (2003, p. 5), the smearing factor ($\rho(x_i)$) may be computed in the following manner:

$$\rho(x_i) = \frac{1}{N}\sum_{i=1}^{N} \exp\left[\log(y_i) - x_i'\hat{\beta}\right]. \quad (5.11)$$

Hence, $\rho(x_i)$ depends on the deviations on the logarithmic scale. It is then used to recover the estimates for health expenditures in the original scale

$$\hat{y}_i = \exp(x_i'\hat{\beta}) \cdot \rho(x_i). \quad (5.12)$$

Also according to Deb and Burgess (2003, p. 5), however, the smearing estimator is "not robust to heteroscedasticity in the transformed scale".

5.A.2 Generalized linear models

GLMs are an extension of the classical linear models (see McCullagh and Nelder 1989, Blough et al. 1999). Therefore, following McCullagh and Nelder (1989), in order to sim-

plify the exposition we start with the structure of the classical linear models and subsequently we generalize it to the case of GLM.

Consider that the realization of a random variable Y_i is given by a vector of observations $y_i \sim N(\mu_i, \sigma_i^2)$, where the index i denotes the individual observations. The realization of y_i is governed by two components: the systematic and the random components (Nelder and Wedderburn 1972). The systematic component corresponds to the variation that is explained by covariate variables and is given by μ_i, while the random component is given by the residual variation that cannot be explained by the covariates. Thus μ_i may be defined as a function of the p covariates $x_{i1}, x_{i2}, \ldots x_{ip}$:

$$E(y_i) = \mu_i = \sum_{1}^{p} \beta_j x_{ij}. \tag{5.13}$$

The classical linear model is thus characterized by the following configuration, where the index i is suppressed by resorting to the matrix notation:

$$\begin{aligned} y &\sim N(\mu, \sigma^2) \\ E(y) &= \mu \\ \mu &= x\beta \end{aligned} \tag{5.14}$$

5.A.3 The generalization

GLMs extend the classical linear models by relaxing the assumptions behind ((5.14)). Thus, they are based on three components (see McCullagh and Nelder 1989, Blough et al. 1999):

1. The random component.

The independent variable is assumed to follow a probability distribution from the exponential family (see Table 5.8 below for list of possible distributions). This implies that there is a relationship between the variance and the mean and this relationship is given by the variance function:

$$var(y_i) = \sigma_i^2 = \phi V(\mu_i). \tag{5.15}$$

Thus, the specification of a distributional family implies a given relationship between the variance and the mean. For instance, "the variance function for the Gaussian distribution is 1, because variance and mean are independent, whereas for the skewed Gamma and Inverse Gaussian distributions the variance functions are quadratic ($var(y_i) \propto \mu_i^2$) and cubic ($var(y_i) \propto \mu_i^3$), respectively" (Barber and Thompson 2004, p. 198.).

2. The systematic component.

It is assumed the existence of a linear predictor η_i based on the covariates $x_{i1}, x_{i2}, \ldots x_{ip}$:

$$\eta_i = \sum_1^p \beta_j x_{ij}. \tag{5.16}$$

3. The link function $g(\mu_i)$.

The random and systematic components are assumed to be related by a linear function $g(\mu_i)$. This function, which is usually called link function, in fact determines how the mean of the independent variable (μ_i) is related to the linear predictor (η_i)

$$g(\mu_i) = \eta_i. \tag{5.17}$$

It is clear that the classical linear model is a special case of GLM where the probability distribution of the response variable (component 1) is Normal or Gaussian ($y_i \sim$

$N(\mu_i, \sigma_i^2)$) and the link function (component 3) is the identity (see Nelder and Wedderburn 1972, McCullagh and Nelder 1989).

Table 5.8 below (from Blough et al. (1999)) shows possible exponential distribution families, their natural link functions and the associated variance function:

Table 5.8. Exponential families of distributions

Distribution	Natural link function	Variance function
Gaussian	μ	1
Bernoulli	$\log\left(\frac{\mu}{1-\mu}\right)$	$\mu(1-\mu)$
Binomial	$\log\left(\frac{\mu}{1-\mu}\right)$	$\frac{\mu(1-\mu)}{n}$
Poisson	$\log(\mu)$	μ
Gamma	$\frac{1}{\mu}$	μ^2
Inverse Gaussian	$\frac{1}{\mu^2}$	μ^3
Quasi	$g(\mu)$	$V(\mu)$

Source: (Blough et al. 1999, p. 158.)

5.A.4 Estimated equations

In order to estimate the two part model households are divided into two groups, according to the existence or not of health expenditures. The first part of the model uses observations from both groups to model the probability of incurring in any health expenditure in the surveyed period. The second part uses only the observations with strictly positive expenditures and analyses the factors influencing the level of health care expenditures for households which reported any expenditures.

The anaysis of the determinants of the probability of reporting any expenditure is implemented as a Logit model given by equation ((5.1)). The second part can be estimated either using equation ((5.2)) or with a GLM model as follows:

$$g(\mu) = x\beta^{GLM} + \epsilon^{GLM}, \mu \sim E, \tag{5.18}$$

where $g(\cdot)$ is the link function relating the raw-scale mean of the dependent variable (μ) and the linear prediction from the regressors. We assume μ follows an unkown distribution from the exponential family (E). Both the link function and the probability distribution of the mean are selected according to the criteria specified below.

The set of variables in the first and second parts need not be the same, as is the case if one believes the two outcomes are driven by distinct data generating processes. In our case, however, we assume the two parts are affected by the same set of variables, although we do not restrict beforehand the effect of a given variable to be the same in both parts.

5.A.5 GLM model selection

Two crucial elements in the specification of the GLM model are the choice of the appropriate link function and the family of probability distribution. The link function is determined using a Box-Cox test, which selects from several transformations which one is most likely to render the dependant approximately normal. The general form of the Box-Cox transformation is given by:

$$T(y) = \frac{y^{\lambda} - 1}{\lambda}. \tag{5.19}$$

The dependent variable is transformed according to ((5.19)) using several values of the parameter λ. The resulting variable is then tested for normality, and the optimal value of λ is the one which maximizes the likelihood of normality for the transformed variable. Additionally, given integer values of λ are equivalent to familiar transformations. For instance, $\lambda = -1$ corresponds to the inverse transformation ($T(y) = \frac{1}{y}$), $\lambda = 0$ corresponds

to the logarithmic transformation ($T(y) = \ln(y)$) and $\lambda = 2$ corresponds to the quadratic transformation ($T(y) = y^2$).

On the other hand, the family of probability distribution is chosen using the algorithm proposed by Manning and Mullahy (2001). The Manning-Mullahy algorithm is based on the fact that each distribution family implies an specific relationship between the variance and mean ((5.15)), and uses the residuals from a candidate GLM in order to determine which distribution family is more appropriate.

Assume that the function ϕ in ((5.15)) is a power function of the form:

$$var(y_i) = \mu_i^{\gamma} \tag{5.20}$$

The question then becomes to determine the value of the parameter γ. In order to do that, we estimate the raw-scale residuals of the candidate GLM model. We then apply a Park test to determine the value of γ, which in turn implies a given family of probability distribution, according to Table 5.8. That is, if the variance is directly proportional to the mean ($\gamma = 1$) one should use the Poisson distribution, if it is proportional to the square of the mean ($\gamma = 2$) one should use the Gamma distribution, and if it is proportional to the cube of the mean ($\gamma = 3$) one should use the Inverse Gaussian distribution.

5.A.6 Results

Table 5.9 below reports the estimates for the parameter γ from equation ((5.20)), which are used to base the choice of the GLM family. In the great majority of cases γ is significantly estimated to be around 2, supporting the use of the Gamma distribution. Additionally, in

all cases we used the log link and, in the panel models for Russia, the regressions were estimated with exchangeable within-group correlation structure.

With respect to the regression results, the GLM estimated effects in general agree with the OLS models. More specifically, the GLM regressions show that chronic diseases a) increase health expenditures; b) increase non-health expenditures in Brazil, but have no effect in the two other countries; c) reduce labour income in Brazil and Russia, and d) have no statistically significant effect on remittances in any of the three countries.

Table 5.9. GLM Family test

Estimated γ in var(y)=E(y)$^{\gamma}$			
	Brazil	India	Russia
Health expenditures	2.1***	2***	1.9***
Non health expenditures	2***	2***	1.9***
Earned income	2***	2.2***	1.4***
Unearned income	2.3***	2***	3.1***
Work day loss	1.7***	3***	2***

Legend: * p<.1; ** p<.05; *** p<.01.

Table 5.10. GLM for health expenditures

Log link and gamma distribution			
Variable	Brazil	India	Russia‡
chronic	.16***	.99***	.15***
nonchronic	.16***	.91***	.24***
bad	.28***		.21***
lnincome	.19***	.079*	.2***
physical	0.01		
smoke			- 0.04
overweight	0.08		.067***
headmale	0.08	0.11	0.11
single	- 0.06	- 0.51	0.06
kids	- 0.02	.046*	- 0.01
men	- 0.05	- 0.05	-.084*
women	0.03	0.11	- 0.04
oldmen	-.27*	0.04	- 0.03
oldwomen	- 0.07	- 0.08	-.24***
age	0.02	0.03	.013*
age^2	- 0.00	- 0.00	-.00014**
education	.16***	.32***	0.05
education^2	-.0041*	- 0.02	0.00
insured	.079*		0.02
urban	-.2*		
NE	-.47***		
caste		- 0.09	
round7			0.00
round8			- 0.06
round9			- 0.03
round10			0.01
round11			0.04
round12			.14**
round13			.11*
moscow			.17***
central			-.23***
volga			-.21***
caucasus			.24***
ural			-.32***
cons	1.1**	- 0.23	1.9***
N	*2,695*	*1,886*	*19,647*
Pseudo R^{2†}	5.8%	13.1%	4.3%
AIC	26,626	12,468	
Deviance	3,360	3,175	32,601

Legend: * p<.1; ** p<.05; *** p<.01; † Computed as the square of the correlation between observed and predicted values for the dependent variable;

‡ Panel model with exchangeable within-group correlation structure;

Table 5.11. GLM for non-health expenditures

Log link and gamma distribution			
Variable	Brazil	India	Russia‡
chronic	.042**	- 0.02	0.02
nonchronic	.051***	0.00	- 0.01
bad	-.091**		- 0.07
lnincome	.43***	.058***	.34***
physical	.056***		
smoke			- 0.09
overweight	.037**		0.07
headmale	.15***	.11*	0.13
single	- 0.07	-.38***	- 0.20
kids	.05***	.074***	0.14
men	- 0.02	.12***	0.04
women	0.01	.13***	- 0.01
oldmen	-.097**	.11***	0.01
oldwomen	0.00	-.059*	- 0.07
age	.02***	.039***	- 0.00
age^2	-.00026***	-.00042***	- 0.00
education	.064***	.097***	- 0.00
$education^2$	- 0.00	0.00	0.00
insured	.078***		0.03
urban	.066*		
NE	0.02		
caste		-.068**	
round7			- 0.05
round8			0.03
round9			- 0.26
round10			- 0.26
round11			- 0.26
round12			- 0.19
round13			- 0.22
moscow			.54**
central			- 0.16
volga			- 0.24
caucasus			- 0.03
ural			- 0.28
cons	2.2***	2.7***	5.1***
N	*4,796*	*1,920*	*31,328*
Pseudo $R^{2\dagger}$	33.5%	44.0%	
AIC	66,782	22,113	
Deviance	1,810	377	23,101

Legend: * p<.1; ** p<.05; *** p<.01; † Computed as the square of the correlation between observed and predicted values for the dependent variable;

‡ Panel model with exchangeable within-group correlation structure;

Table 5.12. GLM for earned (labour) income

Log link and gamma distribution			
Variable	Brazil	India	Russia‡
chronic	-.092***	- 0.04	-.042***
nonchronic	-.041*	0.03	- 0.00
bad	-.2***		-.075***
physical	.06**		
smoke			- 0.02
overweight	.075***		.072***
headmale	.21***	0.12	.2***
single	-.14*	- 0.19	-.23***
kids	0.02	.035***	.022**
men	.26***	.17***	.17***
women	.15***	.065*	.11***
oldmen	- 0.06	0.09	.063**
oldwomen	-.19***	- 0.00	- 0.00
age	.029**	.026*	.081***
age^2	- 0.00	-.00036**	-.0011***
education	.16***	- 0.06	-.15***
education^2	0.00	.023***	.012***
insured	.17***		.038***
urban	.23***		
NE	-.36***		
caste		0.10	
round7			-.068***
round8			-.26***
round9			-.096***
round10			- 0.00
round11			.09***
round12			.14***
round13			.24***
moscow			.3***
central			-.24***
volga			-.38***
caucasus			-.45***
ural			-.35***
cons	3.4***	3.7***	5.3***
N	*4,126*	*1,810*	*26,522*
Pseudo R^{2†}	18.3%	16.1%	18.0%
AIC	59,234	21,267	
Deviance	2,932	1,028	27,250

Legend: * p<.1; ** p<.05; *** p<.01; † Computed as the square of the correlation between observed and predicted values for the dependent variable;

‡ Panel model with exchangeable within-group correlation structure;

Table 5.13. GLM for unearned income

Log link and gamma distribution			
Variable	Brazil	India	Russia‡
chronic	- 0.02	- 0.02	0.06
nonchronic	0.05	- 0.14	0.05
bad	- 0.15		- 0.06
physical	0.09		
smoke			-.097***
overweight	.12***		0.05
headmale	- 0.01	- 0.10	0.13
single	- 0.05	0.77	- 0.05
kids	- 0.03	- 0.04	- 0.04
men	.1*	0.12	0.10
women	.11*	.47***	0.03
oldmen	.23**	.63***	.28***
oldwomen	- 0.03	- 0.35	0.07
age	0.01	.089**	- 0.02
age^2	0.00	-.00084*	0.00
education	.16***	.41**	0.02
education^2	- 0.00	- 0.03	0.00
insured	.14***		0.02
urban	.16*		
NE	-.26***		
caste		- 0.26	
round7			0.03
round8			-.33***
round9			-.25***
round10			-.2**
round11			- 0.04
round12			0.14
round13			0.13
moscow			0.10
central			-.23***
volga			-.29***
caucasus			.19**
ural			-.35***
cons	3.6***	- 0.97	6.2***
N	*2,673*	*481*	*26,320*
Pseudo R^{2†}	12.3%	13.9%	0.7%
AIC	37,335	3,577	
Deviance	3,323	878	33,720

Legend: * p<.1; ** p<.05; *** p<.01; † Computed as the square of the correlation between observed and predicted values for the dependent variable;

‡ Panel model with exchangeable within-group correlation structure;

Chapter 6
Conclusion

This thesis has set out to analyse two classes of questions in health economics. The first part has focused on a range of issues related to the provision of health care and the regulation of providers and insurers. The results of this analysis provide some important extensions of the literature.

Chapter 1 has compared the process of negotiation between purchasers and providers of health care under three scenarios: the purchaser sets the price, but activity is bargained between purchaser and provider: *activity* bargaining; the price is bargained between purchaser and provider, but activity is chosen unilaterally by the provider: *price* bargaining; and price and activity are bargained simultaneously between purchaser and provider: *efficient* bargaining. We have compared the scenarios and the effect of changing the bargaining balance in terms of four outcomes: prices, activity, the welfare of purchaser and the welfare of the provider. In the cases without closed-form solutions, we have performed numerical simulations of the equilibrium outcomes.

The comparison of the scenarios provides interesting policy implications. We show that if the purchaser has low bargaining power, *efficient* bargaining leads to higher price, lower activity and lower purchaser's utility, compared to *price* bargaining. Consequently, the first policy implication is that purchasers with low bargaining power might be better off if restricted to bargaining on prices only, and not on price and activity.

The second result of the analysis is that if purchasers can set prices (*activity* bargaining), net consumer welfare is highest. This result holds for any level of bargaining power of the purchaser. The analysis therefore supports policies such as "payment by results" in the UK, where prices are fixed by the purchaser or the regulator.

One less intuitive result concerns the effect of changing from *efficient* or *price* bargaining to *activity* bargaining. In the first case the results suggest an unambiguous decrease in activity, while in the second case the effect depends on the bargaining power of the purchaser. More precisely, activity decreases if the bargaining power of the purchaser is low, but increases in the opposite case.

Among possible extensions that might contribute to take matters further, Chapter 1 proposes some interesting hypotheses which would deserve empirical examination. In particular, a useful exercise would be to estimate the balance of bargaining power in health care markets. Although there has been some effort in this direction (e.g. Melnick et al. 1992, Dor and Watson 1995, Propper 1996, Brooks et al. 1997), most analyses are restricted to the US context and aim only indirectly at estimating the bargaining power. This information might be useful for governments to decide whether to encourage purchasers to bargain on prices only or on price and activity simultaneously. Another possible empirical extension would focus on the effect of bargaining power and negotiation framework on the level of activity. A particularly interesting case would be to test whether policies such as "payments by results" in the UK, which might be identified as an example of *activity* bargaining, are likely to increase or decrease activity compared to previous policies.

Chapter 2 addresses an important question: What is the effect of waiting times on hospital costs? It is possible to identify in the literature some suggestions that waiting times, by regulating excess demand, should reduce hospital costs (e.g. Gaynor and Anderson 1995, Keeler and Ying 1996, Hughes and McGuire 2003). However, Iversen (1993) argues that this negative effect exists only to a certain extent. Beyond a given point, however, the management of the waiting list and repeated examinations become too costly, and then waiting times start to increase total hospital costs.

In this chapter we use a sample of 259 acute hospitals over the period 1998-2002 in the English NHS to estimate the elasticity of hospital costs with respect to waiting times. We have estimated pooled OLS, fixed effects and random effects models for total hospital cost. In each case waiting times are entered in the regression in both linear and quadratic terms, which should assist in identifying the curvature of the cost function. The signs of the estimated coefficients are consistent with Iversen's (1993) model: the coefficient is negative for the linear effect and positive for the quadratic effect, suggesting a U-shaped relationship between hospital costs and waiting times. However, the coefficients are generally not statistically significant. It is possible to argue that to a certain extent the results give some support for Iversen's (1993) model: although the coefficients are not significant, the signs are correctly estimated.

The results suggest that waiting times have no impact on hospital costs. This casts some doubt on the effectiveness of waiting times as a tool to control hospital costs. In the future more research would be desirable in order to further understand this issue.

Chapter 3 provides an analysis of the incentives for the provision of health care in a competitive health insurance market. This study extends the model of health insurance with adverse selection proposed by Rothschild and Stiglitz (1976) by considering the effect on both curative and preventive care. We assume that prevention affects the probability of illness. Patients belong to two groups, which have different probability of illness and different levels of efficiency of prevention.

The results suggest that, in addition to distortions on curative care, the level of prevention is also distorted. The direction of distortion depends on the relative efficiency of prevention for each risk type: low-risk patients receive lower (respectively, higher) marginal benefit from preventive care if prevention is relatively more (less) efficient for them.

This result might have implications for the implementation of optimal risk adjustment schemes (see Glazer and McGuire 2000, Glazer and McGuire 2006). According to this method, the best way to avoid health plans distorting the amount and quality of services under capitation systems is to over- and underpay based on observable characteristics of individual enrolees. However, this applies to curative care only. As the results of this analysis suggested, health plans face different incentives for the provision of curative and preventive care. It is desirable that optimal risk adjustment policies take this into account. Otherwise, the incentives for the provision of preventive care might in fact become even more distorted.

A potential empirical extension of this analysis might focus on policies that are currently in place among private health insurance regarding the access to preventive care, perhaps focusing on the effect on lifestyle choices. There are currently several examples of

plans that provide financial benefits for enrolees that adopt healthy lifestyles. It would be interesting to test to what extent employers are taking that into account when providing the choice of health plans to employees and to measure the potential economic benefits that might be accrued from this.

The second part of the thesis has focused on the economics of chronic diseases, particularly in developing countries. Until recently, a widespread view has sustained that chronic diseases are a problem restricted to rich countries. According to this view, the real issue to be tackled in developing countries would concern communicable diseases.

In recent years, however, new evidence has started to challenge this view. International studies have drawn attention to the fact that chronic diseases are becoming increasingly prevalent in developing countries (see WHO 2005, World Bank 2005). This transition has been associated with several elements which are characteristic of developing nations, such as the rapid increase in urbanisation, changes in traditional lifestyles and the lack of information and access to good nutrition. Chapters 4 and 5 contribute to this debate by presenting two strands of evidence.

Chapter 4 starts by analysing the association between socioeconomic inequality and chronic diseases. We quantify and analyse the determinants of the inequalities in the prevalence of heart disease, hypertension and diabetes in two countries, Brazil (1998 and 2003) and Russia (2000 to 2004). The results of the analysis suggest that socioeconomic status is a significant determinant of the prevalence of chronic diseases and that households from lower standing bear a disproportionately high share of the chronic diseases burden. The computation of concentration indices shows that i) inequality in chronic diseases has

risen in Russia during the period and ii) the inequity concentration index is negative in both countries, but considerably higher in Brazil.

We then set out to analyse the determinants of the socioeconomic inequality in chronic diseases. We apply three different methods, as proposed by Wagstaff et al. (2003): decomposition analysis, the Oaxaca decomposition and the total differential decomposition. We find that socioeconomic status, co-morbidities and education make a significant contribution to worsening inequalities in chronic diseases. Standardising variables have opposite effects in each country, increasing pro-poor inequalities in Russia, while decreasing them in Brazil. Moreover, the results indicate that changes in inequality have been driven mainly by changes in the elasticities, rather than changes in the sample means of the determinants.

The main conclusion of the analysis is that socioeconomic inequality is not the only determinant of inequality in chronic diseases. Inequalities in education and co-morbidities also contribute to worsening the inequality in the prevalence of chronic diseases. Therefore, efforts to relieve the burden of chronic diseases from poor households should focus not only on income redistribution, but should also aim to maximise the impact of externalities from other related policy areas.

The main limitation of this analysis is that it is based on cross-sectional methods. Possible extensions of this study could incorporate longitudinal methods, which are more appropriate to deal with unobserved heterogeneity and could potentially provide more robust results.

The study of the economic effects of chronic diseases is taken further in Chapter 5. This chapter provides some additional evidence which might help to disentangle the

relationship between the chronic diseases and the socioeconomic status of the household. The analysis estimates the effect of chronic diseases on household health expenditures, non-health expenditures, labour income and transfers income in Brazil (1996), India (1997) and Russia (1997 to 2004).

The results suggest that chronic diseases have a significant effect on household welfare. In particular, chronic diseases are shown to imply higher health expenditures, lower labour productivity and labour income. On the other hand, this negative impact is partially offset by an increase in the amount of remittances from other households The net effect, at least in Russia where we were able to account for unobserved heterogeneity using longitudinal data, was a significant reduction on non-health consumption, which is generally considered a measure of household welfare.

This analysis could be improved if we had access to more complete and uniform data for the three countries. Although for Russia an eight-year panel was available, we have used only cross-sections for both Brazil and India. Panel data would have been preferable for all countries were they available. Nevertheless, the conclusions are generally similar among all three countries.

The prevalence of chronic diseases in projected to increase over the coming years especially in low- and middle-income countries (WHO 2005). Given that formal social support systems in many such countries are limited in coverage, households' socioeconomic wellbeing and indeed grass-roots development may be hindered by the burden of chronic diseases without appropriate intervention.

Overall, the analysis contained in Chapters 4 and 5 provide empirical evidence for the theoretical links between chronic diseases, socioeconomic inequality and poverty. It has some policy implications for the treatment and control of chronic diseases in the countries included in the analysis. For instance, there is justification for reviewing the financing mechanisms to address chronic diseases. This is because there is a limit to what end informal financing could aid households in coping with the socioeconomic burden associated with chronic diseases.

Bibliography

Anand, K., Pandav, C. and Nath, L.: 1999, Impact of HIV/AIDS on the national economy of India, *Health Policy* **47**(3), 195–205.

Aronsson, T., Lofgren, K.-G. and Wikstrom, M.: 1993, Monopoly union versus efficient bargaining: Wage and employment determination in the Swedish construction sector, *European Journal of Political Economy* **9**, 357–370.

Asfaw, A. and von Braun, J.: 2004, Is consumption insured against illness? Evidence on vulnerability of households to health shocks in rural Ethiopia, *Economic development and cultural change* **53**(1), 115–129.

Barber, J. and Thompson, S.: 2004, Multiple regression of cost data: Use of generalized linear models, *Journal of Health Services Research and Policy* **9**(4), 197–204.

Barigozzi, F.: 2004, Reimbursing preventive care, *The Geneva Papers on Risk and Insurance Theory* **29**(2), 165–186.

Barros, P. and Olivella, P.: 2005, Waiting lists and patient selection, *Journal of Economics and Management Strategy* **14**(3), 623–646.

Barros, P. P. and Martinez-Giralt, X.: 2000, Selecting negotiation processes with health care providers, *Technical report*, Unitat de Fonaments de l'Anàlisi Econòmica and Institut d'Anàlisi Econòmica. Available at http://ideas.repec.org/p/aub/autbar/467.00.html.

Barros, P. P. and Martinez-Giralt, X.: 2003, Preventive health care and payment systems, *Topics in Economic Analysis and Policy* **3**(1), 1–21.

Barros, P. P. and Martinez-Giralt, X.: 2005a, Negotiation advantages of professional associations in health care, *International Journal of Health Care Finance and Economics* **5**(2), 191–204.

Barros, P. P. and Martinez-Giralt, X.: 2005b, Bargaining and idle public sector capacity in health care, *Economics Bulletin* **9**(5), 1–8.

Barros, P. P. and Martinez-Giralt, X.: 2006, Models of negotiation and bargaining in health care, *in* A. Jones (ed.), *The Elgar Companion to Health Economics*, Edward Elgar Publishing Limited, Cheltenham, chapter 21.

Bartley, M., Fitzpatrick, R., Firth, D. and Marmot, M.: 2000, Social distribution of cardiovascular disease risk factors: Change among men in england 1984-1993, *Journal of Epidemiology and Community Health* **54**(11), 806–814.

Bartley, M., Sacker, A., Firth, D. and Fitzpatrick, R.: 1999, Understanding social variation in cardiovascular risk factors in women and men: The advantage of theoretically based measures, *Social Science and Medicine* **49**(6), 831–845.

Beaglehole, R. and Yach, D.: 2003, Globalisation and the prevention and control of non-communicable disease: The neglected chronic diseases of adults, *The Lancet* **362**(9397), 903–908.

Bhargava, A., Jamison, D. T., Lau, L. J. and Murray, C. J. L.: 2001, Modeling the effects of health on economic growth, *Journal of Health Economics* **20**(3), 423–440.

Bloom, D. and Mahal, A.: 1997, Does the AIDS epidemic threaten economic growth?, *Journal of Econometrics* **77**(1), 105–124.

Blough, D., Madden, C. and Hornbrook, M.: 1999, Modeling risk using generalized linear models, *Journal of Health Economics* **18**(2), 153–171.

Bommier, A. and Stecklov, G.: 2002, Defining health inequality: Why Rawls succeeds where social welfare theory fails, *Journal of Health Economics* **21**(3), 497–513.

Bond, E. and Crocker, K.: 1991, Smoking, skydiving, and knitting: The endogenous categorization of risks in insurance markets with asymmetric information, *Journal of Political Economy* **99**(1), 177–200.

Breyer, F.: 1982, Rational purchase of medical care and differential insurance coverage for diagnostic services, *Journal of Health Economics* **1**(2), 147–156.

Brooks, J., Dor, A. and Wong, H.: 1997, Hospital-insurer bargaining: An empirical investigation of appendectomy pricing, *Journal of Health Economics* **16**, 417–434.

Bulkley, G. and Myles, G.: 1997, Bargaining over effort, *European Journal of Political Economy* **13**, 375–384.

Byrne, M. and Thompson, P.: 2001, Screening and preventable illness, *Journal of Health Economics* **20**(6), 1077–1088.

Cameron, A. C. and Trivedi, P.: 2005, *Microeconometrics - Methods and applications*, Cambridge University Press, New York.

Castro-Leal, F., Demery, L. and Mehra, K.: 1999, Public social spending in africa: Do the poor benefit?, *World Bank Research Observer* **14**(1), 49–72.

Chalkley, M. and Malcomson, J.: 1998a, Contracting for health services when patient demand does not reflect quality, *Journal of Health Economics* **17**, 1–19.

Chalkley, M. and Malcomson, J.: 1998b, Contracting for health services with unmonitored quality, *The Economic Journal* **108**, 1093–1110.

Chaturvedi, N., Jarrett, J., Shipley, M. and Fuller, J.: 1998, Socioeconomic gradient in morbidity and mortality in people with diabetes: Cohort study findings from the whitehall study and the who multinational study of vascular disease in diabetes, *British Medical Journal* **316**(7125), 100–105.

Clark, D.: 1995, Priority setting in health care: An axiomatic bargaining approach, *Journal of Health Economics* **14**, 345–360.

Cochrane, J.: 1991, A simple test of consumption insurance, *The Journal of Political Economy* **99**(5), 957–976.

Courbage, C. and de Coulon, A.: 2004, Prevention and private health insurance in the UK, *The Geneva Papers on Risk and Insurance* **29**(4), 719–727.

Courbage, C. and Rey, B.: 2006, Prudence and optimal prevention for health risks, *Health Economics* **15**(12), 1323–1327.

Cowell, A. J.: 2006, The relationship between education and health behavior: Some empirical evidence, *Health Economics* **15**(2), 125–146.

Cropper, M.: 1977, Health, investment in health, and occupational choice, *Journal of Political Economy* **85**(6), 1273–1294.

Cuddington, J.: 1993, Modeling the macroeconomic effects of AIDS, with an application to Tanzania, *World Bank Economic Review* **7**(2), 173–189.

Cuddington, J. and Hancock, J.: 1994, Assessing the impact of AIDS on the growth path of the Malawian economy, *Journal of Development Economics* **43**(2), 363–368.

Cullis, J., Jones, P. and Propper, C.: 2000, Waiting lists and medical treatment: Analysis and policies, *in* A. Culyer and J. Newhouse (eds), *Handbook of Health Economics*, Elsevier, Amsterdam, chapter 23.

de Fraja, G.: 2000, Contracts for health care and asymmetric information, *Journal of Health Economics* **14**(5), 663–677.

de Ven, W. V. and Ellis, R. P.: 2000, Risk adjustment in competitive health plan markets, *in* A. Culyer and J. Newhouse (eds), *Handbook of Health Economics*, Elsevier, Amsterdam, chapter 14.

Deb, P. and Burgess, J.: 2003, A quasi-experimental comparison of econometric models for health care expenditures, *Technical report*, Hunter College: Department of Economics. Available at http://ideas.repec.org/p/htr/hcecon/212.html.

Department for International Development: 1997, Eliminating world poverty: A challenge for the 21st century, *Technical report*, The Stationary Office, London. White Paper on International Development.

Dercon, S. and Krishnan, P.: 2000, In sickness and in health: Risk sharing within households in rural Ethiopia, *Journal of Political Economy* **108**(4), 688–727.

Dor, A. and Watson, H.: 1995, The hospital-physician interaction in US hospitals: Evolving payment schemes and their incentives, *European Economic Review* **39**, 795–802.

Duan, N.: 1983, Smearing estimate: A nonparametric retransformation method, *Journal of the American Statistical Association* **78**(383), 605–610.

Ehrlich, I. and Becker, G.: 1972, Market insurance, self-insurance, and self-protection, *Journal of Political Economy* **80**(4), 623–649.

Ellis, R. and McGuire, T.: 1990, Optimal payment systems for health services, *Journal of Health Economics* **9**(4), 375–396.

Epping-Jordan, J., Galea, G., Tukuitonga, C. and Beaglehole, R.: 2005, Preventing chronic diseases: Taking stepwise action, *The Lancet* **366**, 1667–1671.

Fagart, M.-C. and Fombaron, N.: 2003, Value of information and prevention in insurance markets, *Technical report*, LEI/CREST and THEMA Université de Paris X. Available at http://www.crest.fr/pageperso/fagart/fagartfombaron.pdf.

Figueras, J., McKee, M., Mossialos, E. and Saltman, R.: 2005, *Purchasing to improve health systems performance*, European Observatory on Health Systems and Policies Series, Open University Press, UK.

Fuchs, V.: 2004, Reflections on the socio-economic correlates of health, *Journal of Health Economics* **23**(4), 629–636.

Gakidou, E., Murray, C. and Frenk, J.: 2000, Defining and measuring health inequality: An approach based on the distribution of health expectancy, *Bulletin of the World Health Organization* **78**(1), 42–54.

Gal-Or, E.: 1997, Exclusionary equilibria in health care markets, *Journal of Economics and Management Strategy* **6**(1), 5–43.

Gal-Or, E.: 1999a, The profitability of vertical mergers between hospitals and physician practices, *Journal of Health Economics* **18**(5), 621–652.

Gal-Or, E.: 1999b, Mergers and exclusionary practices in health care markets, *Journal of Economics and Management Strategy* **8**(3), 315–350.

Gallup, J. and Sachs, J.: 2001, The economic burden of malaria, *American Journal of Tropical Medicine and Hygiene* **64**(1), 85–96.

Gaynor, M. and Anderson, G.: 1995, Uncertain demand, the structure of hospital costs, and the cost of empty hospital beds, *Journal of Health Economics* **14**(3), 291–317.

Gertler, P. and Gruber, J.: 2002, Insuring consumption against illness, *The American Economic Review* **92**(1), 51–70.

Glazer, J. and McGuire, T.: 2000, Optimal risk adjustment in markets with adverse selection: An application to managed care, *The American Economic Review* **90**(4), 1055–1071.

Glazer, J. and McGuire, T.: 2006, Optimal risk adjustment, *in* A. Jones (ed.), *The Elgar Companion to Health Economics*, Edward Elgar Publishing Limited, Cheltenham, chapter 26.

Globerman, S.: 1991, A policy analysis of hospital waiting lists, *Journal of Policy Analysis and Management* **10**(2), 247–262.

Grand, J. L.: 1987, Inequalities in health: Some international comparisons, *European Economic Review* **31**(1-2), 182–191.

Gravelle, H.: 2003, Measuring income related inequality in health: Standardisation and the partial concentration index, *Health Economics* **12**(10), 803–819.

Greenberg, H., Raymond, S. and Leeder, S.: 2005, Cardiovascular disease and global health: Threat and opportunity, *Health Affairs* **Web Exclusive**.

Grosh, M. and Glewwe, P.: 1995, A guide to living standards measurement study surveys and their data sets, *Living Standards Measurement Study Working Papers* **WP 120**.

Grossman, M.: 1972, On the concept of health capital and the demand for health, *Journal of Political Economy* **80**(2), 223–255.

Grossman, M.: 2000, The human capital model, *in* A. Culyer and J. Newhouse (eds), *Handbook of Health Economics*, Elsevier, Amsterdam, chapter 7.

Gyimah-Brempong, K. and Wilson, M.: 2001, Health human capital and economic growth in Sub-Saharan Africa and OECD countries, *The Quartely Review of Economics and Finance* **44**(2), 296–320.

Hanson, K.: 1992, AIDS: What does economics have to offer?, *Health Policy and Planning* **7**(4), 315–328.

Heffley, D.: 1982, Allocating health expenditures to treatment and prevention, *Journal of Health Economics* **1**(3), 265–290.

Herring, B.: 2002, Sub-optimal coverage of preventive care due to market-level turnover among private insurers: Evidence from mammography, *Yale University* .

Hoel, M. and Iversen, T.: 2002, Genetic testing when there is a mix of compulsory and voluntary health insurance, *Journal of Health Economics* **21**(2), 253–270.

Hoel, M., Iversen, T., Nilssen, T. and Vislie, J.: 2006, Genetic testing in competitive insurance markets with repulsion from chance: A welfare analysis, *Journal of Health Economics* **25**(5), 847–860.

Houweling, T., Kunst, A., Borsboom, G. and Mackenbach, J.: 2006, Mortality inequalities in times of economic growth: Time trends in socioeconomic and regional inequalities in under 5 mortality in indonesia, 1982-1997, *Journal of Epidemiology and Community Health* **60**(1), 62–68.

Howitt, P.: 2005, Health, human capital, and economic growth: A Schumpeterian perspective, *in* G. López-Casasnovas, B. Rivera and L. Currais (eds), *Health and Economic Growth: Findings and Policy Implications*, MIT Press, Cambridge, MA.

Hoy, M., Orsi, F., Eisinger, F. and Moatti, J. P.: 2003, The impact of genetic testing on healthcare insurance, *Geneva Papers on Risk and Insurance - Issues and Practice* **28**(2), 203–221.

Hoy, M. and Polborn, M.: 2000, The value of genetic information in the life insurance market, *Journal of Public Economics* **78**(3), 235–252.

Hughes, D. and McGuire, A.: 2003, Stochastic demand, production responses and hospital costs, *Journal of Health Economics* **22**(6), 999–1010.

Iversen, T.: 1993, A theory of hospital waiting lists, *Journal of Health Economics* **12**(1), 55–71.

Iversen, T.: 1997, The effect of a private sector on the waiting time in a national health service, *Journal of Health Economics* **16**(4), 381–396.

Jones, A.: 2000, Health econometrics, *in* A. Culyer and J. Newhouse (eds), *Handbook of Health Economics*, Elsevier, Amsterdam, chapter 6.

Kakwani, N., Wagstaff, A. and van Doorslaer, E.: 1997, Socioeconomic inequalities in health: Measurement, computation, and statistical inference, *Journal of Econometrics* **77**(1), 87–103.

Kalai, E.: 1977, Nonsymmetric Nash solutions and replications of 2-person bargaining, *International Journal of Game Theory* **6**, 129–133.

Keeler, T. and Ying, J.: 1996, Hospital costs and excess bed capacity: A statistical analysis, *The Review of Economics and Statistics* **78**(3), 470–481.

Kenkel, D.: 1994, The demand for preventive medical care, *Applied Economics* **26**(4), 313–325.

Kenkel, D.: 2000, Prevention, *in* A. Culyer and J. Newhouse (eds), *Handbook of Health Economics*, Elsevier, Amsterdam, chapter 31.

Kochar, A.: 1995, Explaining household vulnerability to idiosyncratic income shocks, *The American Economic Review* **85**(2), 159–164.

Le Grand, J., Mossialos, E. and Saltman, R.: 1999, *Health care and cost containment in the EU*, Ashgate, UK.

Lindsay, C. and Feigenbaum, B.: 1984, Rationing by waiting lists, *The American Economic Review* **74**(3), 404–417.

López-Casasnovas, G., Rivera, B. and Currais, L.: 2005, Introduction: The role health plays in economic growth, *in* G. López-Casasnovas, B. Rivera and L. Currais (eds),

Health and Economic Growth: Findings and Policy Implications, MIT Press, Cambridge, MA.

Ma, A.: 1994, Health care payment systems: Cost and quality incentives, *Journal of Economics and Management Strategy* **3**(1), 93–112.

Mackenbach, J. and Kunst, A.: 1997, Measuring the magnitude of socio-economic inequalities in health: An overview of available measures illustrated with two examples from europe, *Social Science and Medicine* **44**(6), 757–771.

Mackenbach, J., Kunst, A., Cavelaars, A., Groenhof, F. and Geurts, J.: 1997, Socioeconomic inequalities in morbidity and mortality in Western Europe. The EU working group on socioeconomic inequalities in health, *The Lancet* **349**(9066), 1655–1659.

Manning, A.: 1987, An integration of trade union models in a sequential bargaining framework, *The Economic Journal* **97**, 121–139.

Manning, W.: 1998, The logged dependent variable, heteroscedasticity and the retransformation problem, *Journal of Health Economics* **17**(3), 283–295.

Manning, W. and Mullahy, J.: 2001, Estimating log models: To transform or not to transform, *Journal of Health Economics* **20**(4), 461–494.

Martin, S. and Smith, P.: 1999, Rationing by waiting lists: An empirical investigation, *Journal of Public Economics* **71**(1), 141–164.

McCullagh, P. and Nelder, J.: 1989, *Generalized linear models*, Chapman and Hall, London.

McDonald, I. and Solow, R.: 1981, Wage bargaining and employment, *The American Economic Review* **71(5)**(5), 896–908.

Melnick, G., Zwanziger, J., Bamezai, A. and Pattison, R.: 1992, The effects of market structure and bargaining position on hospital prices, *Journal of Health Economics* **11**(3), 217–233.

Miceli, T. and Heffley, D.: 2002, Do HMOs encourage prevention? An analysis of alternative health care plans, *Contemporary Economic Policy* **20**(4), 429–439.

Morduch, J.: 1995, Income smoothing and consumption smoothing, *Journal of Economic Perspectives* **9**(3), 103–114.

Mougeot, M. and Naegelen, F.: 2005, Hospital price regulation and expenditure cap policy, *Journal of Health Economics* **24**(1), 55–72.

Mullahy, J.: 1998, Much ado about two: Reconsidering retransformation and the two-part model in health econometrics, *Journal of Health Economics* **17**(3), 247–281.

Muthoo, A.: 1999, *Bargaining Theory with Application*, Cambridge University Press, Cambridge.

Nash, J.: 1950, The bargaining problem, *Econometrica* **18**, 155–162.

Nash, J.: 1953, Two-person cooperative games, *Econometrica* **21**, 128–140.

Nelder, J. and Wedderburn, R.: 1972, Generalized linear models, *Journal of the Royal Statistical Society, A* **135**(3), 370–384.

Newhouse, J.: 2002, *Pricing the priceless - A health care conundrum*, MIT Press, Cambridge.

O'Donnell, O., van Doorslaer, E., Wagstaff, A. and Lindelow, M.: 2007, *Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation*, The World Bank, Washington, D.C.

Osborne, M. and Rubinstein, A.: 1990, *Bargaining and Markets*, Academic Press, Inc., San Diego.

Oswald, A.: 1985, The economic theory of trade unions: An introductory survey, *Scandinavian Journal of Economics* **87(2)**, 160–193.

Pamuk, E.: 1985, Social class inequality in mortality from 1921 to 1972 in England and Wales, *Populational Studies (Cambridge)* **39**(1), 17–31.

Pappas, G., Queen, S., Hadden, W. and Fisher, G.: 1993, The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986, *New England Journal of Medicine* **329**(2), 103–109.

Pecorino, P.: 2002, Should the US allow prescription drug reimports from Canada?, *Journal of Health Economics* **21**, 699–708.

Phelps, C.: 1978, Illness prevention and medical insurance, *The Journal of Human Resources* **13, Supplement: NBER Conference on the Economics of Physician and Patient Behavior**, 183–207.

Pohlmeier, W. and Ulrich, V.: 1995, An econometric model of the two-part decision process in the demand for health, *The Journal of Human Resources* **30**(1), 339–361.

Polborn, M., Hoy, M. and Sadanand, A.: 2006, Advantageous effects of regulatory adverse selection in the life insurance market, *The Economic Journal* **116**(508), 327–354.

Propper, C.: 1996, Market structure and prices: The responses of hospitals in the UK National Health Service to competition, *Journal of Public Economics* **61**(3), 307–335.

Rothschild, M. and Stiglitz, J.: 1976, Equilibrium in competitive insurance markets: An essay on the economics of imperfect information, *Quarterly Journal of Economics* **90**(4), 629–649.

Sachs, J. and Malaney, P.: 2002, The economic and social burden of malaria, *Nature* **415**(6872), 680–685.

Sampson, A.: 1993, Bargaining over effort and the monitoring role of unions, *European Journal of Political Economy* **9**, 371–381.

Sappington, D. and Lewis, T.: 1999, Using subjective risk adjusting to prevent patient dumping in the health care industry, *Journal of Economics and Management Strategy* **8**(3), 351–382.

Sauerborn, R., Adams, A. and Hein, M.: 1996, Household strategies to cope with the economic costs of illness, *Social Science and Medicine* **43**(3), 291–301.

Schalick, L., Hadden, W., Pamuk, E., Navarro, V. and Pappas, G.: 2000, The widening gap in death rates among income groups in the United States from 1967 to 1986, *International Journal of Health Services* **30**(1), 13–26.

Schokkaert, E. and van de Voorde, C.: 2004, Risk selection and the specification of the conventional risk adjustment formula, *Journal of Health Economics* **23**(6), 1237Ű 1259.

Schultz, T. P. and Tansel, A.: 1997, Wage and labor supply effects of illness in Cote d'Ivore and Ghana: Instrumental variables estimates for days disabled, *Journal of Development Economics* **53**(2), 251–286.

Shavell, S.: 1979, On moral hazard and insurance, *The Quarterly Journal of Economics* **93**(4), 541–562.

Shorrocks, A.: 1983, Ranking income distributions, *Economica* **50**(197), 1–17.

Siciliani, L. and Martin, S.: 2007, An empirical analysis of the impact of choice on waiting times, *Health Economics* **16**(8), 763–779.

Strohmenger, R. and Wambach, A.: 2000, Adverse selection and categorical discrimination in the health insurance markets: The effects of genetic tests, *Journal of Health Economics* **19**(2), 197–218.

Strong, K., Mathers, C., Leeder, S. and Beaglehole, R.: 2005, Preventing chronic diseases: How many lives can we save?, *The Lancet* **366**(9496), 1578–1582.

Stronks, K., van de Mheen, H. and Mackenbach, J.: 1998, A higher prevalence of health problems in low income groups: Does it reflect relative deprivation?, *Journal of Epidemiology and Community Health* **52**(9), 548–557.

Tabarrok, A.: 1994, Genetic testing: An economic and contractarian analysis, *Journal of Health Economics* **13**(1), 75–91.

Thomas, D. and Strauss, J.: 1997, Health and wages: Evidence on men and women in urban Brazil, *Journal of Econometrics* **77**(1), 159–185.

Vagero, D. and Erikson, R.: 1997, Socioeconomic inequalities in morbidity and mortality in western europe, *The Lancet* **350**(9076), 516; author reply 517–518.

van Doorslaer, E. and Jones, A.: 2003, Inequalities in self-reported health: Validation of a new approach to measurement, *Journal of Health Economics* **11**(4), 61–87.

van Doorslaer, E. and Koolman, X.: 2004, Explaining the differences in income-related health inequalities accross European countries, *Health Economics* **13**(7), 609–628.

van Doorslaer, E., Koolman, X. and Jones, A.: 2004, Explaining income-related inequalities in doctor utilisation in Europe, *Health Economics* **13**(7), 629–647.

van Doorslaer et al., E.: 1997, Income-related inequalities in health: Some international comparisons, *Journal of Health Economics* **16**(1), 93–112.

van Doorslaer et al., E.: 2000, Equity in the delivery of health care in Europe and the US, *Journal of Health Economics* **19**(5), 553–583.

van Zon, A. and Muysken, J.: 2005, Health as a principal determinant of economic growth, *in* G. López-Casasnovas, B. Rivera and L. Currais (eds), *Health and Economic Growth: Findings and Policy Implications*, MIT Press, Cambridge, MA.

Vega, J., Hollstein, R., Delgado, I., Perez, K., Carrasco, S. and Marshall, G.: 2001, Chile: Socioeconomic differentials and mortality in a middle-income nation, *in* T. Evans, M. Whitehead, F. Diderichsen, A. Bhuiya and M. Wirth (eds), *Challenging Inequalities in Health: From Ethics to Action*, Oxford University Press, Oeford.

Victora, C. G., Vaughan, J. P., Barros, F. C., Silva, A. C. and Tomasi, E.: 2000, Explaining trends in inequities: Evidence from Brazilian child health studies, *The Lancet* **356**(9235), 1093–1098.

Vita, M.: 1990, Exploring hospital production relationships with flexible functional forms, *Journal of Health Economics* **9**(1), 1109–1119.

Vitaliano, D.: 1987, On the estimation of hospital cost functions, *Journal of Health Economics* **6**(4), 305–318.

Wagstaff, A.: 2002, Poverty and health sector inequalities, *Bulletin of the World Health Organization* **80**(2), 97–105.

Wagstaff, A.: 2005, The economic consequences of health shocks, *World Bank Policy Research Working Paper* **3644**.

Wagstaff, A., Paci, P. and van Doorslaer, E.: 1991, On the measurement of inequalities in health, *Social Science and Medicine* **33**(5), 545–557.

Wagstaff, A. and van Doorslaer, E.: 2000, Measuring and testing for inequity in the delivery of health care, *Journal of Human Resources* **35**(4), 716–733.

Wagstaff, A. and van Doorslaer, E.: 2004, Overall versus socioeconomic health inequality: A measurement framework and two empirical illustrations, *Health Economics* **13**(3), 297–301.

Wagstaff, A., van Doorslaer, E. and Watanabe, N.: 2003, On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam, *Journal of Econometrics* **112**(1), 207–223.

Welch, H. G.: 2004, *Should I be Tested for Cancer? Maybe Not and Here is Why*, University of California Press, Berkeley.

White, M.: 2000, Researching the changing social pattern of cardiovascular diseases, *Journal of Epidemiology and Community Health* **54**(11), 804–805.

WHO: 2005, *Preventing Chronic Diseases - A Vital Investment*, World Health Organization, Geneva.

WHO: 2006, Working together for health, *in* WHO (ed.), *World Health Report 2006*, World Health Organization, Geneva.

Wolfson, M. and Rowe, G.: 2001, On measuring inequalities in health, *Bulletin of the World Health Organization* **79**(6), 553–560.

World Bank: 2001, *World Development Report 2000/2001: Attacking Poverty*, Oxford University Press, New York.

World Bank: 2005, *Dying too young: Addressing premature mortality and ill health due to non communicable diseases and injuries in the Russian Federation*, World Bank, Washington, DC.

Wright, D.: 2004, The drug bargaining game: Pharmaceutical regulation in Australia, *Journal of Health Economics* **23**(4), 785–813.

Wu, S.: 2003, Sickness and preventive medical behavior, *Journal of Health Economics* **22**(4), 675–689.

Zweifel, P. and Breyer, F.: 1997, *Health Economics*, Oxford University Press, New York.

Zweifel, P. and Manning, W.: 2000, Moral hazard and consumer incentives in health care, *in* A. Culyer and J. Newhouse (eds), *Handbook of Health Economics*, Elsevier, Amsterdam, chapter 8.