Social Insurance, Organization and Hospital Care

Rehn, Eric

2009

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Social Insurance, Organization and Hospital Care

Eric Rehn

Lund Economic Studies Number 152

Distributed by the Department of Economics
Lund University
P.O. Box 7082
S-220 07 Lund
Sweden
Telephone: +46 (0)46 222 00 00
Fax: +46 (0)46 222 41 18

Printed in Sweden by
KFS i Lund AB, Lund, 2009

This is dedicated to
the woman I love,
wherever she stands
I give my dedication

Acknowledgements

Having no real interests (except sports) I've always done a little bit of this and a little bit of that - a little bit of political science and a little bit of law, a little bit of forestry and a big chunk of economics. My thesis is a reflection of this - a little economics about this and a little economics about that. Writing this thesis has been a strange experience; the infuriating and constant lack of knowledge, the great feeling when something is finished, the frustration over the fact that things are never finished, the conflict between details and the broader picture, the discrepancy between the initial idea and the final result, and the excitement of actually making progress. At times it has been thoroughly boring, to use my father's words when I've described what I'm doing, and other times it has been inspiring and fun. For the latter times I thank the following persons:

Professor Fredrik Andersson - who has put up with my ideas and bad spelling for almost nine years now. Without Fredrik's support, ideas and patience its not implausible that this would be, in some sense, an acknowledgement without a thesis. If you find something interesting in this thesis it almost certainly can be traced back to a discussion with Fredrik.

Professor Carl Hampus Lyttkens - who has found my ideas interesting and always been willing to contribute with facts, additional ideas and improvements. Without Carl Hampus encouragement many things would have been different - no conference presentations, no badminton tournaments, and no money to support my research in the near future.

Professor Lars Söderström - who has shown me around in the real world, outside my theoretical models, by letting me be a part of his projects. Moreover, he has been willing to listen to my ramblings about everything and nothing during our travels to Stockholm, Umeå and Cyprus - for that I'm especially thankful.

Tommy Andersson - for teaching me everything about "Microeconomics C" (once again) and for advice about everything from my research to child care and heavy metal (notably Tommy is heavy metal while I'm just a "poser" wearing a studded belt but no long hair). Fredrik Gallo - for helping me with integrals, functions, and Deus Ex - where he did not help me as much as kill me. Andreas Bergh - for being able to clarify. If you wonder about something just ask Andreas and he will clarify it.

Robin Rander - for helping me with matlab (among many other things) and for a fantastic "Guitar Hero" session in the seminar room. Ola Andersson - for the bitterness, the bitters and the bargaining models. Therese Nilsson (or is it Pernilla Johansson?) - for the parties, the hallandic genome, and for talking when I'm quiet... (which does not happen that often). Pernilla Johansson (or is it Therese Nilsson?) - for the parties, the competitive spirit, the altruism and for repeatedly beating me on the Squash court. Mårten Wallette - for crashing into "my" office, taking me to Pickwick, Bishop's Arms (Lund) and Bishop's Arms (Umeå) - it's a shame that we never could visit Mucky Duck together. Martin Nordin - for showing me how to make a sandwich.

All the PhD-candidates of 2003 - for sharing my suffering during the first years... My lunch companions at EC1 - for intriguing questions about my lunch, spirited conversations about fruit and for tremendous listening skills! The NEK Football team - for its defensive stoutness - I never liked losing. Everyone at the Department of Economics - for the hellos and the "how are you"s, and for help with administrative as well as research related questions.

Oddvar Kaarbøe - for being a great final seminar opponent. The Jan Wallander and Tom Hedelius Foundation - for financial support. Seminar and conference participants at the following venues: Department of Economics Lund University, Department of Economics Uppsala University, 6th World Congress on Health Economics, Copenhagen & Lund, Health Economics Study Group Conference in Aberdeen, First World Meeting of the Public Choice Society Amsterdam. Jaya Reddy - for correcting my english.

Igul IBK - for being my non-economic haven. Jonne - for knowing that he will lose when it counts. Ferry - for letting me carry his golf clubs. Kalle - for being much smarter than me. Leif - Leif is Leif. Fast - thank you for the music. Björn - for knowing when to quit and get a real job. Johan - if you can find him you can't find a better man. Marcus - for the anecdotes, the openness and the "smörgåstårtor".

My penultimate "thank you" goes to... Maggie - for putting up with me (I'm sort of a brute when you get to know me) and my music (my wife does not...), for being such a good friend, and for buying me candy when needed. It wouldn't have been the same without Mags.

Finally, I'd like to thank my family - "you didn't have to love me but you did, and I thank you!" The usual disclaimer applies....

"I used to be such a sweet, sweet thing [just ask my mother]
Till they got a hold of me [the economists]
I opened doors for little old ladies [it just don't seem rational]
I helped the blind to see [I've stop talking to non-economists]
I got no friends 'cause they read the papers [or at least this acknowledgement]
They can't be seen with me [have they ever?]
And I'm gettin' real shot down [cf. the defense]
And I'm feelin' mean [wouldn't you?]
No more Mister Nice Guy [embracing soft, non-economic, values]
No more Mister Clean [Herr Rehn]"

Contents

Chapter 1

Introduction

This thesis consist of three chapters, besides this introductory chapter, dealing with such disparate topics as social insurance, organization and hospital care. The common ground lies in the author's interest in incentives and organization in the public sector, a theme that all three chapters touch upon to some extent. This introduction provides some background, in terms of related literature, and short summaries of the three chapters.

1.1 Social Insurance

Social insurance policies provide income support to individuals who temporarily or permanently are unable to sustain a sufficient income from working. This could be due to age, unemployment, accidents or sickness. When unemployed the individual typically is unable to work due to low demand for her particular skills and when ageing there might be both legal and physical restrictions on the individual's ability to work. Accidents and/or sickness might disable the individual from working. The inability to work, irrespective of reason (although reversing one's age might be difficult), may either be permanent or temporary. For each reason, and depending on the longevity of the disability, social insurance policies provide support (to eligible individuals). In universal social insurance systems unemployment insurance provides support to the temporarily unemployed (permanently unemployed usually end up in social assistance programs), the retired receive pensions, the temporarily sick are covered by a sickness insurance, and individuals injured at work are covered by workers-compensation-plans. Permanently sick and permanently disabled, due to accidents (work related as well as other accidents), generally receive support from a disability insurance. Chapter 2 of this thesis primarily deals with the design of disability insurance policies, but in many cases the results are also applicable to other social insurance policies. The remainder of this section provides some background for chapter 2.

Disability Insurance - Classification Errors and Labor Supply

Disability insurance, in general, partly covers income losses due to permanent or long-term sickness or disability. Beneficiaries of the disability insurance have to fulfill a number of eligibility requirements, e.g. being an employee or of a certain age, having contributed to the insurance for a certain time (e.g. by being employed for at least one year) and, foremost, the eligible individuals have to exhibit a certain degree of disability. These requirements differ between countries (see e.g. Söderström et. al. (2006) or de Jong (2003)). Most of these requirements are easily checked, e.g. an individuals age, work status and contribution to the insurance. The obvious exception is the individual's degree of disability; determining the degree to which an individual is able to work, or even if disabled or not, is generally difficult.

All individuals applying for benefits from disability insurance go through a screening process. To become eligible for benefits the individual must be judged to be disabled according to the definition of disability in the disability insurance. The definition of disability, as well as the requirements above, varies from country to country, but in general entails both medical and employment-related requirements. Obviously, an individual's disability will depend on the definition of disability, but also on other features of the economy e.g. the flexibility in wages and the demand for certain skills - all in all, disability is a relative concept (Söderström & Rehn, 2008). The fact that disability is relative makes the screening process difficult and subjective, and typically the process produces two types of classification errors: rejection errors (also known as type-I errors) and award errors (type-II errors). Rejection errors are made when truly disabled individuals are denied disability benefits, while award errors are the opposite i.e. when truly able individuals receive disability benefits.

Although the existence of classification errors is widely acknowledged there are, to my knowledge, only a couple of academic studies dealing with the size of these errors, Nagi (1969) and Benitez-Silva et. al. (2004). Both these studies deal with the American disability insurance (Social Security Disability Insurance - SSDI). Nagi (1969) uses an objective expert group to evaluate applicants' disability status after the classification was made by administrators of the insurance – the expert group was unaware of the outcome of the classification. The expert group's classification differed from that of the administrators in 30 percent of the cases. Furthermore, both the award error and the rejection error were around 20 percent. More than thirty years later Benitez-Silva et. al. (2004) reach similar results using self-reported health status of disability applicants. They find that the award error is approximately 20 percent and the rejection error might be as high as 60 percent.

The existence of classification errors creates problems for social insurance systems; obviously social welfare would be enhanced if only the truly disabled, and all of them, were beneficiaries to disability insurance. If this was the case, the number of truly poor individuals (who are not able to work, receive no disability benefits and live on welfare) in the economy would be reduced and the cost of monitoring beneficiaries would vanish. Moreover the replacement rate, i.e. benefit divided by previous income, could be increased if the screening process did not entail classification errors - ruling out excessive use of the insurance i.e. moral hazard.

In spite of the very limited research on the size of classification errors there exists extensive research that is indicative of their existence. This research mainly focuses on the effects of disability benefits on labor supply. The reasoning is the following: if only the truly disabled have disability benefits (apply for benefits) then there would be no labor supply effects (application effects) of altered benefit levels (cf. Kreider, 1998). Alternatively, as noted by Leonard (1986), large disability benefit elasticities might be consistent with award errors in the disability screening process.

The bulk of empirical evidence suggests that labor supply is affected by disability benefits, but the size of this effect is uncertain. Reviews of the literature suggest that the elasticity of labor force non-participation with respect to benefit levels ranges from 0.06 to 0.93 (Danzon, 1993, Bound & Burkhauser, 1999), implying that a 10 percent increase in the disability benefits might give a 9.3 percent decrease in labor supply. Besides being indicative of problems with the screening process, the labor supply literature suggests that individuals respond to monetary incentives.

Disability Insurance - Theoretical Issues

Insurance policies are generally vulnerable to moral hazard (i.e. excessive use) and disability insurances are no exception, as reflected in the labor supply effects presented above. The imperfect screening process gives opportunities for excessive use of disability insurance, which is plainly expressed by Diamond & Mirrlees (1978) in the introduction to their seminal paper on optimal disability insurance: "...neither private insurance nor public arrangements distinguish fully between low income by choice and low income by necessity. Full insurance might defeat itself through moral hazard." (Diamond & Mirrlees, 1978:295). They then study the optimal benefit structure for the insurance in an economy of identical individuals in all aspects but ability to work - the individuals are either able or disabled. They conclude that the optimal benefit is the greatest benefit compatible with a positive work effort from the able. This implies that if the disability benefits are sufficiently unattractive, the able will choose to work. They also formulate a condition for the presence of moral hazard, which has been used in several articles (e.g. Whinston, 1983; Diamond & Sheshinski (1995); chapter 2 in this thesis)[1]. They contend that moral hazard is present if equalizing the utility for work-

[1] Whinston (1983) extends the Diamond & Mirrlees-model by introducing *ex ante* differences between individuals - the probability of becoming disabled differs - this enables an analysis of the interaction between moral hazard and adverse selection in disability insurances. Diamond & Sheshinski (1995) will be returned to below.

ers and non-workers imply that the marginal utility of extra consumption is higher for non-workers.

The problem with the optimal insurance policy proposed by Diamond & Mirrlees is that it is costly for the targeted group in the disability insurance; the disabled will get less of their income loss replaced if the benefits are made unattractive for the able (cf. Parsons 1991)[2]. That is, the replacement rate will generally be low, implying a low efficiency in achieving the goal for the insurance policy - income support to individuals who cannot sustain a sufficient work income due to disability.

The optimal disability insurance in Diamond & Mirrlees (1978) relies on self-revelation (i.e. any information from a screening process, if there is one, is ignored), i.e. that the benefit structure is such that the able reveal that they are able and thus work. However, the replacement rate of the insurance could be increased if the information gained in the screening process is used to differentiate between individuals - place them in different hypothetical groups (give them different "tags") depending on their individual characteristics (cf. Akerlof, 1978; Parsons, 1996). This practice has in the literature become known as tagging, starting with Akerlof (1978). The obvious problem with tagging, as illustrated by the empirical literature above, is that it is imperfect. Tagging produces both type-I and type-II errors; thus, when constructing a disability insurance policy the insurance administration must heed these errors. Hence, theoretical models of disability insurances must incorporate these errors.

Both Diamond & Sheshinski (1995) and Parsons (1996) provide theoretical models with two types of individuals, able and disabled, and a screening process that produces a imperfect tagging. The dividing line between these two models is that Diamond & Sheshinski adopt a three-price structure that does not include a specific salary to tagged workers while Parsons includes such a salary. A three-price structure is common in many welfare systems, e.g. salaries to workers, disability benefits to tagged non-workers and social assistance to untagged non-workers, but this ignores some of the informa-

[2]Parsons (1991) primarily investigates self-screening mechanisms in disability insurances.

tion available to the insurance administration. When it is realized that the screening process produces an imperfect tagging, it becomes clear that some of the tagged non-workers could actually work. Knowing this the insurance administration could include a fourth price in the welfare system, a salary paid to tagged workers, to provide work incentives for this group. As already noted, this is done by Parsons (1996), who shows that including a specific salary for tagged workers improves social welfare. Salanié (2002) shows that Parsons' results are very general as long as leisure is a normal good.

Imperfect Tagging Revisited - Moving Beyond the Two-Type-Economy

Chapter 2 of this thesis revisits the analysis in Parsons (1996) and moves beyond the two-type-economy by introducing, first, a third type (partially disabled) and, subsequently, a continuum of types. These models are here called the three-type-model and the continuous model, respectively. The three-type-model stems from the recognition that many disability insurances have more than one level of disability. The Swedish disability insurance has four levels of disability: 25%, 50%, 75%, 100%; hence a second tag (a partial disability tag) is introduced to match the introduction of a third type. The continuous model recognizes the fact that disability (ability) is continuously distributed over the population and is not a discrete variable.

Throughout the analysis it is assumed that workers and non-workers have different utility of consumption. There are several reasons to assume state dependent utility functions: e.g. that non-workers may have more time than workers to enjoy their consumption, implying a higher utility of a certain consumption or, on the other hand, that the non-workers' health status may not allow them to enjoy their consumption as much as the, presumably healthy, workers.[3] In chapter 2 it is assumed that for a given consumption, c, individuals have greater utility of c when not working than when working. This is

[3]Concerning state dependent utility functions: in the health insurance literature it is often assumed that $U(X) > V(X)$, where $U(X)$ is the utility of consuming X when in good health and $V(X)$ is the utility when ill (cf. Viscusi & Evans, 1990)

done to accentuate the fact that non-workers need not have a bad health status even if disabled, i.e. individuals may be unable to work (given the rules of the insurance policy etc.) without their health status being obstructive to living an otherwise normal life (or obstructive to consumption). The assumption also implies that able individuals would choose (if they could choose) not to work if they had the same income when not working, a reasonable assumption when leisure is a normal good.

The basic results from Parsons (1996) are in essence transferable to the three-type-model and the continuous model in chapter 2. That is, individuals should be given incentives to work in line with their ability and this leaves room for being more generous towards the targeted group(s). The composition of these incentives, however, might be somewhat different in the continuous model compared to the discrete models. The continuous model allows for any relation between the income of tagged workers and untagged workers, whereas the income of tagged workers is always greater in the discrete models.

The consumption allocations for the different groups (i.e. salaries and benefits respectively) in the three-type model cannot be fully ranked, making the results less straight forward than in Parsons' model. It is however obvious that mechanism behind the results is similar, two logarithmic examples highlight this fact.

The tagged in the continuous model may be given incentives to work by a lower cut-off level i.e. they may stop working at a lower disability level, and it cannot be ruled out that the cut-off level for the tagged is greater than or equal to the cut-off level of the untagged. This is an unappealing result in the sense that the targeted group for the disability insurance (the tagged) is treated with less or equal generosity to the untagged in terms of cut-off level. However, it is found that the premium for tagged workers is greater than the premium for untagged workers in this case, possibly outweighing the less generous cut-off level.

1.2 Organization

The question of how to organize an economic transaction is an eternal question. Should a specific transaction be carried out in a firm, in the market or in some hybrid form of organization? This kind of questions often end up in the domain of transaction cost economics (e.g. Klein *et. al.* (1978), Williamson (1985)). An alternative, but related, way to analyze the organization of economic transactions is the property rights approach to organization (hereinafter PRA) (Grossman & Hart (1986), Hart & Moore (1990), Hart (1995)).[4]

The property rights approach to organization has not achieved the same impact, and broad familiarity, as transaction cost theory, but has all the same inspired a non-trivial amount of research and critique. This section is primarily devoted to familiarizing the reader with the model and the critique of the model, but also summarizes the findings in chapter 3 which adapts the PRA to trilateral transactions.

The Property Rights Approach to Organization

While transaction cost theory suggests that integration reduces opportunistic behavior, when contracting is incomplete, it does not explain how this happens. Starting from this contention the seminal work by Grossman & Hart (1986) and Hart & Moore (1990) proposes a theory of how to organize firms (determine the boundaries of firms) in a world with incomplete contracts. This theory focuses on the importance of asset ownership for the *relationship-specific investments* made in a transaction (trade relationship). Relationship-specific investments are investments that are more valuable in a particular transaction than outside the transaction. That is, relationship-specific investments are investments by party B in the relationship with party S, and vice versa, that increase the mutual dependence and the coordination between the parties - ensuring a more rewarding cooperation between B and S; i.e. increasing the value of the relationship. These investments are

[4] Also known as the Grossman-Hart-Moore model.

non-contractible and the level of investments typically increases with asset ownership, because asset ownership gives the owner control of contingencies not specified in the contract and hence a stronger bargaining position (the owner becomes less vulnerable to withdrawal from trade by a party that does not invest or invests less). Organizational form in this theory is defined by the distribution of assets; e.g. integration means that one party owns all assets used in the transaction. Notably, the PRA focuses on organizational issues and abstracts from demand (consumer side) issues that might affect the benefits of a certain organization. Next the PRA is briefly discussed to familiarize the reader with its typical assumptions and results.

The discussion here, as in later chapters, is based on Hart (1995:chapter 2) who presents a simple and tractable version of the PRA. The model in Grossman & Hart (1986) is similar albeit more elaborate e.g. when formalizing the residual control rights. The Hart & Moore (1990) model, on the other hand, is in some sense the most sophisticated version of the model, allowing for multiple (instead of just two) agents and changing the bargaining solution, and will be discussed in further detail below.

The most common representation of the PRA is a two-period model with two parties (bilateral trade); in fact most of the literature assumes a bilateral trade relationship with some exceptions e.g. Hart & Moore (1990), Bolton & Whinston (1993) and chapter 3 of this thesis. The bilateral trade envisioned may have the following form: let the two parties be B (buyer) and S (seller), planning to trade with each other. In this trade B produces the final good and S produces an input to this production, for which B and S use one physical asset each: b and s. Before they trade with each other, they make realtionship-specific investments, typically in their human capital. These investments increase the value of trade for both parties. However, if B and S fail to agree on trade, they may withdraw from this specific transaction and buy respectively sell the input on the open market (this is called *no-trade* and the bargaining interpretation of this disagreement point is discussed extensively below).

There is no uncertainty about costs and benefits, and no asymmetric information in this model. Moreover, the parties can make correct calculations about the expected return of any action and have an unlimited wealth - ensuring that the parties can buy any asset i.e. that wealth constraints do not hinder an efficient allocation of assets. However, there is *ex ante* uncertainty about the quality of the input - its characteristics cannot be contracted on in contingent manner. Figure 1.1 describes the timing of the model; notably assets are already allocated, i.e. the organizational structure for the transaction is decided when the investments are made in period 0. In period 1 the parties trade with each other, the uncertainty about input quality is resolved, and the parties bargain, heeding that both have the option of no-trade, over the division of the surplus.

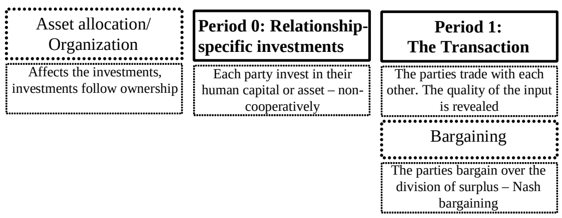

Figure 1.1: The timing of the PRA-model

Both Grossman & Hart (1986) and Hart (1995) apply the Nash bargaining solution (NBS) (good textbook references are Osborne & Rubinstein (1990) and Muthoo (1999)) to the bargaining problem. Hence they assume cooperative (even if NBS is the outcome in some non-cooperative bargaining structures as well) and *ex post* efficient bargaining. This efficiency implies costless renegotiations i.e. any contract may be renegotiated in the model at any time (until an agreement is reached and the transaction is carried out).

A *period*-0-*contract*, e.g. establishing the basic conditions for the transaction, cannot specify the relevant characteristics (the quality) of the input in a verifiable manner, nor can it specify all possible future contingencies - hence it is an incomplete contracting model. If complete contracting was possible the parties could choose their investments to maximize their joint *ex post* surplus. Instead of making first-best investments, i.e. maximize their joint *ex post* surplus, B and S will choose second-best investments, which typically imply underinvestments (Grossman & Hart, 1986, Hart & Moore, 1990, Hart, 1995).[5] Notably, investment are observable to both parties, but not verifiable to outsiders (not enforceable)(cf. Hart 1995).

When it comes to organization, the interaction between asset ownership and relationship-specific investments determines the organizational structure. It is assumed that all investments are beneficial for the investing party irrespective if she trades with the other party or not. However, the benefit is greater if the investing parties trade with each other (cf. Hart, 1995). The investments are typically made in human capital and connected to the use of the assets. Moreover it is assumed that human capital cannot be transferred between the parties.

Asset ownership gives the asset owner control over contingencies concerning assets not specified in the contract (residual control rights), implying e.g. that agents not involved in the transaction may be excluded from using the assets. Residual control rights are important given the uncertainty in the model, stemming from the presence of unforeseen contingencies and the *ex ante* uncertainty about input quality. In general, asset ownership creates greater incentives for investments, because the increased control makes the asset owner less vulnerable to a withdrawal from trade by the other party. Hence the incentives for investments crucially depend on asset allocation, which enables a comparative analysis of the organizational structure. The best organizational structure in the comparative analysis is the structure that supports the greatest relationship-specific investments, and thus creates the greatest surplus.

Some results from Hart (1995) may serve as examples of this comparative analysis:

- If the assets b and s are independent, then non-integration is optimal.
- If the assets b and s are complementary, then some form of integration

[5] Grossman & Hart (1986) does not rule out overinvestments in the second-best. In Hart & Moore (1990) and Hart (1995), however, overinvestments are ruled out by assuming that the marginal incentive to invest is increasing in the number of assets (and agents) involved in the trade. See assumption 6 in Hart & Moore (1990) and assumptions 2.2 and 2.3 in Hart (1995).

(e.g. B owning both b and s) is optimal.

- If B's (S's) human capital is essential[6] for the transaction, then B (S) should own both assets.
- Joint ownership is suboptimal when assets are complementary and the parties make investments in human capital.
- If B's (S's) investment is relatively unproductive, contributes relatively less to the total value of the transaction, then S (B) should own all assets.

Using the same basic assumptions, e.g. about incomplete contracts, timing and information, Hart & Moore (1990) analyze the control structure (the ownership structure) of coalitions of agents in a multi-agent and multi-asset setting. These coalitions may be thought of as cooperative projects, some form of production, involving many agents. However, the analysis abstracts from the actual structure of this production as it just assumes that if a coalition is formed (in period 1) it will choose an efficient allocation of assets. That is, it is not specified who produces the inputs and who produces the final good and how they are connected (except in the application to many assets where there are many upstream firms and one downstream firm (cf. Bolton & Whinston (1993)). These coalitions create a surplus that is divided between the agents according to their Shapley value, which gives each agent his expected contribution to a coalition. Hence Hart & Moore (1990) take a cooperative approach to their bargaining problem and assume that the *ex post* bargaining outcome is efficient (costless renegotiation). The value of a coalition typically depends (as the value of trade) on the investments that the agents make, these investments (as above) are assumed to be complementary.

As in Hart (1995) and in chapter 3 of this thesis there is inherent under-investment in the model since the full gains from the investment will not be

[6]Meaning that this human capital pivotal for the transaction such that the other party's (with non-essential human capital) investment incentives are muted in its absence. Muted in the sense that the party will make the same investment irrespective of asset ownership. Hence access to this human capital i.e. being part of the transaction, in a sense, becomes more important than the residual control rights stemming from asset ownership. See more about this in chapters 3 and 4.

reaped by the investor (since the created surplus is divided over all agents in the bargaining). This setup yields several results e.g.:

- If only one agent, i, has an investment then i should own all assets.
- Joint ownership with each party having veto power is ruled out i.e. never optimal[7] and, additionally, no more than one agent should have veto power over an asset.
- If an asset is idiosyncratic, i.e. irrelevant to all other agents, to i then i should own it.
- If i is indispensable to an asset then i should own this asset.
- If two (or more) assets are complementary they should be owned or controlled (be in the same coalition) together.

These results are similar to the results presented above. Hart & Moore (1990) also introduce a new line of results since their setup enables an analysis of employee incentives (if one accepts the idea that employees bargain over the surplus of their production).[8] In this setup an employee/worker is someone who works on an asset without owning it. They find that employee incentives (for investments in human capital specific to the asset) change when the ownership structure changes - in particular they find that an employee will invest more if the owner of the asset that she works with acquires more assets.

Related Literature and Critique of the PRA

Starting with the critique it is obvious that the PRA has limitations and this is also acknowledged by Hart & Moore in a recent article. Hart & Moore (2008) state that although literature related to the PRA "... has generated

[7] Specifically, joint ownership (with each party having veto-power) is suboptimal due to the complementarity between investments (in human capital) (supermodularity). This also implies that ownership by non-investing outsiders is never desirable. See Holmström (1999) for a discussion of this issue.

[8] Grossman & Hart (1986) and the basic model in Hart (1995) focus on management incentives.

some useful insights about firm boundaries, it has some shortcomings". In particular they mention:

1. that it is incredible that non-contractible *ex ante* investments, although important, single-handedly determine the organizational structure. Moreover such investments are hard to measure empirically, making it difficult to test the empirical validity of the PRA (cf. Whinston, 2003).

2. that the approach is unsuitable for analyzing the internal organization of firms, e.g. allocation of authority and employee behavior.

3. that the approach has some foundational weaknesses; there is a discrepancy between the assumption of bounded rationality when contracting (incomplete contracts due to transaction costs) and the assumption of full rationality in all other aspects. That is, why are otherwise rational individuals not able to devise a way to deal with the uncertainty in contracting?[9]

In response to these shortcomings Hart & Moore (2008) suggest a framework with contracts as *reference points* i.e. the agents evaluate the outcome of trade with reference to the contract they agreed upon in period 0. Moreover this framework also considers *ex post* inefficiencies - the agents may exert less effort (in unverifiable dimensions) if they feel that they have received less than the contract promised. Hart & Moore (2008) argue that this framework sheds light on some of the issues not dealt with in the PRA-literature.

An often cited critique of the model is found in Holmström (1999), who claims that while PRA aims at providing a framework for understanding the boundaries of a firm it fails to do so. The reasons is that it fails to explain how different activities are distributed across firms. In his view the model deals with individual ownership of assets if taken literally, and if the individuals are interpreted as firms it implies that the activities of the firm

[9]This point was made by Maskin & Tirole (1999b), see also Hart & Moore (1999) where they respond to this critique and argue that Maskin & Tirole's results are less damaging to the incomplete contracts theory than it might seem. This discussion, although interesting, is too intricate to be brief about and therefore not included here.

are exogenously given. Thus, Holmström believes that the theory serves to explain how assets are distributed across firms for a given set of activities, but not to explain how activities are distributed across firms and it is the activities that determine the boundaries of the firm. Moreover, Holmström asserts that the PRA, by focusing on hold-ups, overlooks a "... great variety of instruments that can be used to influence employee incentives". Hence, as already noted, the model is unsuitable for explaining employee behavior within firms. Finally, Holmström notes that the model "unfortunately" is vulnerable to changes in bargaining assumptions. This notion is discussed in detail in the next subsection.

Threat points, Outside options and the Robustness of the PRA

The *no-trade payoffs* (specifically that they are increasing in asset ownership) are vital for the conclusions drawn from the PRA. No-trade is the *disagreement point*, in a general sense,[10] in the negotiation over the input price and as such its characteristics (e.g. is it a definite disagreement or could there be an agreement in a later period?) are important for the results of the model. Two contemporaneous articles Chiu (1998) and De Meza & Lockwood (1998) show that results in the PRA are not invariant to changes in bargaining assumptions. In particular they underline that the interpretation of no-trade affects the results.

In bargaining theory, in general, a difference is made between *inside options* (also known as disagreement, threat or *status quo* points) and *outside options* (see e.g. Muthoo, 1999). An inside option is the payoff an agent receives up to the time an agreement is reached. An outside option, on the other hand, is the payoff an agent receives when the negotiation irretrievably breaks down (cf. De Meza & Lockwood, 1998, Chiu, 1998, Muthoo, 1999). Even though different bargaining solutions are used in Grossman & Hart (1986) / Hart (1995) and in Hart & Moore (1990) the interpretation of no-trade, even if it is not clearly stated, is the same in both settings, namely

[10] That is, I allow myself to use *disagreement point* as a general expression for all types fallback positions from the bargaining. In the bargaining literature disagreement points are often synonymous with *threat points*.

that no-trade can be interpreted as an inside or, if you like, a *status quo* option (cf. De Meza & Lockwood, 1998, Chiu, 1998).

Chiu (1998) as well as De Meza & Lockwood (1998) use the *outside option principle* to show that interpreting no-trade as an outside option, instead of an inside option, changes the results of the PRA. The outside option principle states that if the outside option is binding, i.e. greater than the payoff the agent would get without an outside option, then the agent receives this payoff. However, if the outside option is non-binding then the agent receives the payoff she would receive if she had no outside option - the outside option has no effect on the negotiation (to read more about the outside option principle see e.g. Osborne and Rubinstein, 1990, Muthoo, 1999).

De Meza & Lockwood use *Rubinstein's alternating-offers* model (Rubinstein, 1982) with outside options for both agents to model the bargaining. Chiu, on the other hand, proposes a different bargaining solution consistent with a range of bargaining scenarios, among them the alternating-offers model (Chiu, 1998, section V). In both articles it is found that the loss of asset ownership does not necessarily imply (weakly) lower investments (as in the PRA). Instead it is found that losing assets may in certain circumstances (a binding outside option) increase the investments by the agent losing the asset. This finding is obviously at odds with the findings in PRA, but both articles present circumstances when the PRA-result (that investments increase with asset ownership) also holds in their setup. Chiu contends that the PRA-result holds in a situation with repeated trade, non-binding outside options under any ownership structure (which could be interpreted as non-existing outside options) and, importantly, inside options that increase in value with asset ownership. In the De Meza & Lockwood setup the PRA-result holds if it is assumed that the total return to investments is weakly increasing in the number of assets owned (De Meza & Lockwood, 1998, assumption 5) and the acquiring agent's outside option is binding with probability one already before the transfer of the asset.

It seems that the PRA, given the studies discussed above, is robust as long as no-trade is interpreted as inside options and there are no outside options. Is this interpretation of no-trade restrictive for applicability of the

PRA? Chiu (1998) contends that it is restrictive because people often seek outside options to strengthen their bargaining position.[11] De Meza & Lockwood states that the restrictiveness depends on the context analyzed; if the analysis deals with a situation where there is a natural inside option, e.g. when the agents are already trading with other parties and will continue to do so throughout the negotiation, then it is obviously not restrictive. An example of this can be found in chapter 4 of this thesis i.e. within hospitals. Clearly, a radiology department trades with a number of different medical departments at the same time and the medical departments may consult external radiologists without ruling out future trade between them. Hence, in the negotiation between the radiology department and e.g. a surgery department both have inside options or threat points. That is, the radiology department may threaten to focus on other medical departments (i.e. put the surgery department's requests at bottom of the stack) and the surgery department can threaten to use external radiologists. Other examples outside the hospital setting are easily constructed - entrepreneurs with alternative employments that are bargaining over a common project or upstream firms and downstream firms that continue to trade on a spot-market for inputs until they reach an agreement.

Related literature

Bolton & Whinston (1993) study multilateral trade relationships with, in its simplest form, one upstream supplier and two downstream producers of final products - where the two downstream producers compete for the input that creates supply assurance concerns. In this setup they analyze the effects of different integration structures and the determinants of socially efficient organizational structures. In contrast to Hart & Moore (1990), who also look at a multilateral setting, the investments are substitutes in their setting. Among other things, they find that both downstream firms underinvest (compared to first-best) under non-integration, while integration leads

[11] Furthermore, Chiu (1998) (as well as Rajan & Zingles (1998), but in a somewhat different setting) question the validity of the assumption that the value of the inside-options increases with asset ownership.

to over-investments by the integrated downstream firm.

Segal & Whinston (2000) analyze the use of *exclusive contract*, i.e. prohibiting trade with another party, to protect non-verifiable investments in a model similar to the PRA. They find that if investments are fully relationship-specific then exclusivity does not matter, i.e. does not increase the investment level, while the opposite applies for investments that are valuable outside the relationship. Moreover, it is also found that exclusivity, which can be interpreted as an insurance against hold-ups of the same type as ownership in PRA, may in fact discourage investments (a similar result to the results of De Meza & Lockwood (1998) and Chiu (1998)). Furthermore it is found that if investments are substitutes, and not complements as in the PRA-setup, then it may be optimal to give ownership to a non-investing party. Notably, Rajan & Zingales (1998) reach the same conclusion.

Rajan & Zingales (1998) suggest that *access to critical resources* may be better in explaining the internal organization and boundaries of the firm than the control (power) arising from asset ownership. This approach differs from PRA in a number of ways, for example. it defines a firm as more than its assets, there is no need for enforceable property rights, and obviously asset ownership is not the only source of power within the firm ("soft" assets such as talent and ideas also matter). It is also argued that control (security) provided by asset ownership might "... breed complacence" and thus provide weaker incentives for investments than access to a critical resource (because the value of access is critically dependent on the investment while ownership has a value *per se*). Furthermore, they show that the results of the PRA, rather unsurprisingly, change when the assumptions of the model are altered.

In an extension of the PRA, Matouschek (2004) introduces *ex post* inefficiency (and abstracts from *ex ante* investment inefficiency). This inefficiency stems from asymmetric information between the agents i.e. each agent has private information. Matouschek shows that the inefficiency depends on the ownership structure and finds the optimal (inefficiency minimizing) ownership structure in different circumstances. In particular he finds that ownership structures that reduce the total no-trade payoff (i.e. the sum of no-trade payoffs for both agents in bilateral trade) increase the probability of efficient trade. Interestingly, Matouschek shows that joint ownership might be optimal, in contrast to the PRA where joint ownership is suboptimal given investments in human capital, when the expected gains from trade are large.[12] If the expected gains are small, on the other hand, then non-integration or integration is optimal depending on asset characteristics. Similarly to the PRA, non-integration is optimal when assets are *non-synergistic* (in essence independent) and integration is optimal when assets are *synergistic* (complementary).

Trilateral Trade and Asset Allocation

PRA has, in spite of the presented critique, one advantage over most alternative models - its tractability. The simple, yet formal, structure of the model facilitates the analysis of organizations and produces useful insights, even if they are limited to certain circumstances (e.g. inside options and managerial incentives). Chapter 3 of this thesis takes the most tractable version of the PRA, the one presented in Hart (1995), and extends it to a trilateral trade relationship.

Trilateral trade is a transaction where one downstream party produces the final good and two upstream parties supply inputs for this production. One of the upstream suppliers also supplies the other upstream supplier with an input. It might simplify intuition to think about the following situation: a downstream producer of mobile phones, D, needs both microchips and memory cards (that contain microchips) to produce the mobile phones. Let U_1 be the producer of microchips and U_2 be the producer of memory cards. Now trilateral trade, as depicted here, is the situation where U_1 supplies microchips, significantly different types of microchips, to D and U_2, while U_2 supplies D with memory cards.

Chapter 3 analyzes the optimal organizational structure for the trilateral

[12] The notion that joint ownership might be optimal in certain circumstanses may also be found elsewhere e.g. if the value of no-trade falls with increased investment (Rajan & Zingales, 1998), if the *ex post* bargaining is modelled by alternating-offers bargaining (De Meza & Lockwood, 1998, Chiu, 1998), if the parties have the option to sell their share (Maskin & Tirole, 1999a)

trade under different assumptions about asset and human capital. In general it is found that there is a tendency towards partial integration in the trilateral model, while full integration is more of an exception. This result suggests a more general finding i.e. that a downstream firm, in many instances, should not treat all of its suppliers the same way (integrate all or integrate none). Instead, it is beneficial to integrate one party and let the other party remain an independent contractor, implying that the latter party's investment incentive generally is greater than it would be under full integration. Which party to integrate depends on the characteristics of assets, human capital and investments; generally the party that is least sensitive, indicated by unchanged incentives for investments, to losing control over its asset may be integrated.

Moreover, chapter 3 discusses the relative productivity of investments and the effects of changing the starting point of the analysis. The latter is an issue that has received little attention in earlier expositions of the PRA, but is important both for the model's intuitive appeal and the conclusions drawn from it. Finally, chapter 3 discusses the bargaining assumptions of the model and provides a microeconomic foundation for the chosen bargaining model. The setup for the bargaining is the following: all parties bargaining bilaterally (i.e. an agreement between two parties cannot be conditioned on the participation of the third party) and simultaneous with rational expectations about the outcome in the other negotiations. The symmetric Nash bargaining solution is used to determine the outcome of this bargaining.

1.3 Hospital Care

Hospitals are the visible signs of the health care system (and represent the bulk of health care expenditure), and are often impressive building complexes that are easily noted in the skyline of the cities. Within these buildings a wide variety of interdependent transactions are carried out to provide specialized care. Hospitals are at the centre of attention of citizens as well as politicians and are often seen, in the public debate, as inefficient and overly bureaucratic organizations (at least in public health care systems). In spite of this the organization of hospitals in general, and especially the organization of public hospitals, has rendered little interest among economists. Chapter 4 of this thesis addresses the organization of public hospitals by applying the property rights approach to organization to the problem. This approach is new to the hospital literature.

Typical Characteristics of Hospitals and Hospital Care

Hospital output typically consists of two ingredients: hospital care (the actual treatment) and defensive capacity (cf. Zweifel & Breyer, 1997). Defensive capacity is a readiness to satisfy an option demand i.e. spare resources in terms of both equipment and personnel that can be used to meet (unforeseen) upswings in demand e.g. in case of major accidents or major epidemics.

The major part of hospital output is obviously hospital care, and the production of hospital care is typically a joint production carried out by a number of interdependent medical departments (cardiology, orthopedic dep. etc.), support services (e.g. laboratories and radiology) and, to some extent, hotel services (cooking, cleaning etc.). In the joint production of hospital care every patient "receives customized attention" consisting of inputs from different parts of the hospital - both from different medical departments and from different support services (cf. Harris, 1977). The customized attention that each patient receives requires decision-making to be decentralized. That is, that patient care decisions are made by the treating hospital departments because they have information about the patient that the hospital management does not have (asymmetric information).

Physicians (medical departments) will, in most instances, have an informational advantage over both hospital managements and patients (see e.g. Arrow, 1963). Moreover, the irreducible uncertainty about treatment outcomes makes it difficult for the hospital management to assess the departments' actions *ex post* (cf. McGuire, 2000). This implies that the hospital management cannot contract, in an enforceable manner, the hospital departments to take certain actions e.g. to ensure good cooperation in the joint production of hospital care. This has implications, or at least should have, for the organization of hospitals; the organization of the joint production in itself must be such that it e.g. ensures a high level of cooperation and coordination.

Vertical integration is the most common organizational form for hospitals (e.g. Coles & Hesterly, 1998, Söderström & Lundbäck, 2002) possibly in response to the need for cooperation and coordination, but there are also other possible explanations e.g. inertia and/or a need for political control (public hospitals) - the need for political control spurs a preference for hierarchic public organizations (cf. Williamson, 1999).

Economic Research and Hospitals

Much of the economic research on hospitals has focused on ownership forms (and efficiency), optimal scale of hospitals or reimbursement, while little has been done in terms of actual modelling of hospitals and hospital organization. Most of this research, explicitly or implicitly, studies the US health care system. This is especially true for the "ownership literature", reviewed in Sloan (2000), which compares the behavior and performance of for-profit and (private) not-for-profit hospitals (and to a lesser extent public hospitals that seem to have a clearly different orientation to their private counterparts e.g. providing more uncompensated care). The general finding in this literature is "...that for-profit and private not-for-profit are far more alike than different..." both in terms of efficiency and quality (Sloan, 2000). It has also been suggested that the quality of public hospitals, in the US system, is lower than their private counterparts (recent research on German hospitals reach the opposite conclusion (Herr, 2008)). These results will typically depend on the institutional structure of the health care market and are hence likely to be country-specific.

The institutional structure of the health care system, e.g. competition, funding and hospital organization, will also affect the optimal scale of hospitals, but it seems likely that the research results on optimal scale are less country-specfic than the ownership results. The general tendency towards vertically integrated hospitals and hospital mergers holds across countries and seems to be driven by the same rationale, i.e. that larger hospitals are more efficient due to better use of resources. Economic research on this topic is supportive of this contention, but with the proviso that this only holds up to a certain size above which the efficiency falls due to increased costs of management etc. Posnett (2002) reviews the literature trying to pinpoint this "certain size" and finds that the optimal size of hospitals lies somewhere in the range 100-300 beds. Notably, the average size in Sweden is above 450 beds (cf. Söderström & Lundbäck, 2002). However, economies of scale should only be one of the determinants for hospital organization. Even if 450 beds are above the most efficient scale, there are other considerations, such as interdependency between different hospital services or control of quality/production, that might warrant a vertically integrated structure entailing 450 beds. Coles & Hesterly (1998) (starting from the contention that transaction cost are central to understanding vertical integration) find that asset-specificity (assets or investments that are specific to a certain transaction) is more important for explaining vertical integration than economies of scale. Moreover they find that hospitals typically will integrate core services, even if transaction costs are low, to ensure control over quality. These findings have obvious connections to the findings in chapter 4: that the need for control over transactions (residual control rights) and specific investments creates a preference for an integrated structure.

Another strand of literature where it is commonplace to talk about hospitals is the "reimbursement literature" dealing with the reimbursement of health care providers e.g. hospitals. This is a quite large literature focusing on how the reimbursement affects provider behavior e.g. in terms of "skimping", "cream skimming" and "dumping". Typical examples of this literature are Ellis & McGuire (1986) and Ma (1994). The effects of different reimbursement schemes on hospital organization, if there are such effects, are not dealt with in this literature.

The economic modelling of hospitals and hospital organization is sparse with Newhouse (1970) being the most obvious exception. Newhouse (1970) develops an informal model of not-for-profit hospitals based on the contention that not-for-profit hospitals maximize quantity and quality subject to a zero profit constraint. He finds that not-for-profit hospitals will achieve least-cost production, but have a bias against low quality care (care that some consumers may demand). The latter, together with high barriers to entry is thought to hinder economic efficiency on the hospital market.

The economic research on the internal organization of hospitals seems, as already noted, very limited. Harris (1977) discusses a conceptual model of US hospitals and, in doing this, provides some general insights into hospitals and hospital care with implications for hospital organization; e.g. that hospital care is a joint production, that decision-making is decentralized and that there is a need for a defensive capacity. He also discusses the internal market within hospitals i.e. that hospitals consist of an array of suppliers and demanders - laboratories supplying test results to medical departments, medical departments demanding treatments from other medical departments and so on. Hospitals in essence consist of a range of specialists performing tasks that are complementary to each other, creating an internal market of supply and demand, and a need for coordination and cooperation among these tasks. These contentions will of course affect the optimal organization of hospitals. Chapter 4 of this thesis tries to incorporate some of these insights into a formal model of public hospitals.

Public Hospitals - Incentives and Organization

Chapter 4 of this thesis proposes a novel way to analyze the organization of public hospitals (and to a lesser extent hospitals in general). The property rights approach to organization (PRA) is put forward as conducive to the analysis of hospital organization, and especially so for public hospitals. The PRA (presented above and extended in chapter 3) is suggested to capture many features of hospital care, most importantly the importance of a well-functioning joint production. The PRA also misses out on some features (e.g. asymmetric information). The conduciveness of the PRA is discussed quite extensively in chapter 4 before moving on to the analysis.

The analysis explores issues concerning the privatization and integration of public hospital services and produces a number of results. Most interestingly, it is found, when studying a transaction involving a medical department (also acting as public principal) and a support service, that integration should be opted for both when constructing new treatment units as well as when considering privatization of the support service. Both results are foremost driven by the realization that the medical department's human capital is essential for production of hospital care. This is intuitively appealing, and somewhat trivial; without the cooperation with medical departments most support services would experience difficulties in giving patients suitable treatment. Furthermore, this suggests that public ownership of all assets in the transaction is the best option as long as the public principal's human capital is essential. Obviously, if the public principal is not a medical department, as assumed in chapter 4, assuming that its human capital is essential for hospital care would be more questionable - an interesting topic for future research.

The analysis of a transaction between a medical department and a hotel service reveals that the organization of the transaction depends on the characteristics of investments. In certain circumstances integration should be opted for, while in other circumstances the best option is non-integration; and in yet other circumstances joint ownership could be the solution. The fundamental lesson concerning this transaction is that being dependent on access to the other party's assets and/or investment lowers the incentive for investments unless some residual control rights over assets can be granted - either through integration or joint ownership. Specific investments in physical capital and complementary investments create such dependence.

If the analysis is extended to trilateral transactions it is found that privatization, i.e. disintegrating an integrated structure, is a Pareto improvement only in very special circumstances. However, it is also found that when constructing new treatment units then full integration, in most instances, is not an improvement over non-integration or partial integration. Thus trilateral transactions that are already integrated should remain integrated and new treatment units should either be non-integrated or partially integrated.

Generally, the predictions of the model are supportive of medical departments (with essential human capital) owning support services' assets, but not other medical departments' assets (unless the other medical departments are indifferent over ownership).

Bibliography

Akerlof G A, 1978, "The Economics of Tagging as Applied to Optimal Income Tax, Welfare Programs, and Manpower Planning, American Economic Review 68:8-19

Arrow, K J., 1963, "Uncertainty and the Welfare Economics of Medical Care", American Economic Review, 53:941-973

Benitez-Silva H, Buchinsky M & Rust J, 2004, "How large are the classification errors in the social security disability award process", NBER Working paper 10219, http://www.nber.org/papers/w10219

Bolton, P & Whinston, M D, 1993, "Incomplete Contracts, Vertical Integration and Supply Assurance", Review of Economic Studies 60:121-148

Bound J & Burkhauser R V, 1999, "Economic analysis of transfer programs targeted on people with disabilities", i Ashenfelter O & Card D (eds.), *Handbook of labor economics vol 3*, Amsterdam: North Holland, 3417-3528.

Chiu, S, 1998, "Noncooperative Bargaining, Hostages, and Optimal Asset Ownership", American Economic Review 88:882-901

Coles, J W & Hesterly, W S, 1998, "Transaction costs, quality, and economies of scale: examining contracting choices in the hospital industry", Journal of Corporate Finance, 4:321-345

Danzon P M, 1993, "The economic implications of public disability insurance in the United States", Journal of Labor Economics 11:170-200.

De Meza, D & Lockwood, B, 1998, "Does Asset Ownership Always Motivate Managers? Outside Options and the Property Right Theory of the Firm, Quarterly Journal of Economics 113:361-386

de Jong P R, 2003, “New directions in disability (benefit) policy: The Dutch experience”, in Gould R & Laitinen-Kuikka S (eds.), *Current trends in disability pensions in Europe*, Helsingfors: Finnish Centre for Pensions.

Diamond P A & Mirrlees J A, 1978, “A Model of Social Insurance with Variable Retirement”, Journal of Public Economics 10:295-336

Diamond P A & Sheshinski E, 1995, “Economic Aspects of Optimal Disability Benefits”, Journal of Public Economics 57:1-23

Ellis, R P & McGuire, T G, 1986, "Provider Behavior under Prospective Reimbursement - Cost Sharing and Supply", Journal of Health Economics, 5:129-151

Grossman, S J & Hart, O, 1986, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", Journal of Political Economy 94:691-719

Harris, J E, 1977, "The Internal Organization of Hospitals: Some Economic Implications", Bell Journal of Economics, 8:467-482

Hart, O & Moore, J, 1990, "Property Rights and the Nature of the Firm", Journal of Political Economy 98:1119-1158

Hart, O, 1995, *Firms, Contracts and Financial Structure*, Claredon Lectures in Economics, Claredon Press, Oxford

Hart, O & Moore, J, 1999, "Foundations of Incomplete Contracts", Review of Economic Studies, 66:115-138

Hart, O & Moore, J, 2008, "Contracts as Reference Points", Quarterly Journal of Economics CXXIII:1-48

Herr, A, 2008, "Cost and technical efficiency of German hospitals: does ownership matter?", Health Economics, 17:1057-1071

Holmström, B, 1999, "The Firm as a Subeconomy", Journal of Law, Economics, & Organization 15:74-102

Klein, B, Crawford, R G & Alchian, A, 1978, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", Journal of Law and Economics, 21:297-326

Kreider B, 1998, “Workers’ applications to social insurance programs when earnings and eligibility are uncertain”, Journal of Labor Economics 16:848-877

Leonard J S, 1986, “Labor supply incentives and disincentives for disabled persons”, in Berkowitz M & Hill M A (eds.), *Disability and the labor market: Economic problems, policies and programs*, Ithaca, New York: ILR press, 64-94

Ma, A, 1994,"Health Care Payment Systems: Cost and Quality Incentives", Journal of Economics & Management Strategy, 3:93-112

Maskin, E & Tirole J, 1999a, "Two Remarks on the Property-Rights Literature", Review of Economic Studies 66:139-149

Maskin, E & Tirole J, 1999b, "Unforeseen Contingencies and Incomplete Contracts", Review of Economic Studies, 66:83-114

Matoushek, N, 2004, "Ex post Inefficiencies in a Property Rights Theory of the Firm", Journal of Law, Economics & Organization 20:125-147

McGuire, T G, 2000, "Physician Agency", in *Handbook of Health Economics*, vol:1A, Culyer A J & Newhouse J P eds, 462-536

Muthoo, A, 1999, *Bargaining Theory with Applications*, Cambridge University Press, Cambridge

Nagi S Z, 1969, “Disability and rehabilitation: Legal, clinical and self-concepts and measurement”, Ohio State University Press

Newhouse, J P, 1970, "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital", The American Economic Review, 60:64-74

Osborne, M J & Rubinstein, A, 1990, *Bargaining and Markets*, Academic Press, San Diego

Parsons D O, 1991, “Self-Screening in Targeted Public Transfer Programs”, Journal of Political Economy 99:859-876

Parsons D O, 1996, “Imperfect ‘tagging’ in social insurance programs”, Journal of Public Economics 62:183-207

Posnett, J, 2002, "Are Bigger Hospitals Better?", in *Hospitals in a Changing Europe*, McKee M & Healy J eds., Open University Press, Buckingham, 100-118

Rajan, R G & Zingales, L, 1998, "Power in a Theory of the Firm", Quarterly Journal of Economics 113:387-432

Rubinstein, A, 1982, "Perfect Equilibrium in a Bargaining Model", Econometrica, 50:97-109

Salanié B, 2002, "Optimal Demogrants with Imperfect Tagging", Economic Letters 75:319-324

Segal, I R & Whinston, M D, 2000, "Exclusive Contracts and Protection of Investments", RAND Journal of Economics 31:603-633

Sloan, F, 2000, "Not-for-profit ownership and hospital behavior" in Handbook of Health Economics, vol.1, Culyer A J & Newhouse J P eds, Elsevier, Amsterdam, 1141-1174

Söderström, L & Lundbäck, M, 2002, *Stor och liten - Om vårdeffektiva sjukhus*, Reformera Nummer 15, Reforminstitutet, Bromma

Söderström L, Ekbladh M & Rehn E, 2006, ”Vad säger ekonomerna? En forskningsöversikt” Samtal om Socialförsäkringen Nr 9, Socialförsäkringsutredningen, Stockholm.

Söderström L & Rehn E, 2008, "Arbetsförmåga - en fråga om efterfrågade färdigheter" in Lotta Vahlne Westerhäll (ed.), *Arbets(o)förmåga - ur ett mångdisiplinärt perspektiv*, Santerus Förlag, Stockholm

Viscusi, W K & Evans, W N, 1990, "Utility Functions That Depend on Health Status: Estimates and Economic Implications", American Economic Review, 80:353-374

Whinston, M D, 1983, “Moral Hazard, Adverse Selection, and the Optimal Provision of Social Insurance”, Journal of Public Economics 22:49-71

Whinston, M D, 2003, "On the Transaction Cost Determinants of Vertical Integration", Journal of Law, Economics & Organization 19:1-23

Williamson, O E, 1985, *The Economic Institutions of Capitalism*, Free Press, New York

Williamson, O E, 1999, "Public and Private Bureaucracies: A Transaction Cost Economics Perspective", Journal of Law, Economics and Organization, 15:306-342

Zweifel, P & Breyer, F, 1997, *Health Economics*, Oxford University Press, New York

Chapter 2

Imperfect Tagging Revisited - Moving Beyond the Two-Type Economy

Abstract: Chapter 2 moves beyond the disability insurance with two types, able and disabled, in Parsons (1996). This is done in two ways: by introducing a third type, the partially disabled, and by allowing for a continuum of types. It is assumed throughout that the individuals have state-dependent utility of consumption. The results are in essence consistent with Parsons: individuals should be given incentives to work in line with their ability and this leaves room for being more generous towards the targeted group(s). The continuous model, however, opens up for using a broader range of incentives than in the discrete models.

Keywords: disability insurance, imperfect tagging, partial disability, continuous disability

JEL classification: H21; H53

2.1 Introduction

Disability, be it partial or full, lowers individuals' ability to sustain a sufficient income. Disability insurance aims to provide income support to these individuals. Like all insurance policies a disability insurance is vulnerable to excessive use - moral hazard. In their seminal paper Diamond & Mirrlees (1978) define the basic moral hazard condition and show that in a two-type-economy, i.e. individuals are either able or disabled, an optimal disability insurance can be constructed taking moral hazard into account. This policy induces all able individuals to work if the income from working is sufficiently more generous than the benefit offered to non-workers. This implies a great strain on the targeted group for the disability insurance - the disabled, who cannot work. The benefits will generally be low, implying a low replacement rate. To improve the efficiency (the replacement rate) of disability insurance, tagging/classification of individuals is both used in practice and suggested in theory.

Tagging is the practice of assigning a tag to individuals based on their characteristics, e.g. able and disabled. Akerlof (1978) shows that differentiating between individuals through tagging may improve efficiency. The screening process (tagging process) in his model never mistakes an able individual for a disabled one (however not all disabled necessarily receive a tag), implying that the tagged disabled will be compensated fully for their loss of income. Empirical studies and casual observation suggest that mistakes are made and that the classification of individuals is less accurate than Akerlof envisions. The screening process produces two types of errors, often labeled type-I error and type-II error. Type-I errors arise when the truly disabled are not tagged, while type-II errors are the opposite i.e. able individuals receiving a disability tag. Empirical studies show that the classification errors are substantial; in an early study of the American disability insurance Nagi (1969) finds that the type-I error as well as the type-II error is 20%. Benitez-Silva et. al. (2004) uses self-reported ability to evaluate the classification errors and finds similar results for the type-II error but concludes that the type-I error might be as large as 60%.

Imperfect tagging in disability insurance in a theoretical setting has been studied by e.g. Diamond & Sheshinski (1995) and Parsons (1996). Both articles assume that some individuals are misclassified by the screening process. Diamond & Sheshinski's two-type model recognizes the presence of both type-I and type-II errors but does not model a specific salary for tagged workers.

Parsons, on the other hand, differentiate the salaries for tagged and untagged workers in the optimal program. He shows, in a two-type model, where individuals are either able or disabled, that it is optimal to provide a higher consumption allocation (i.e. salary for workers, benefit for non-workers) to individuals with a disability tag. Salanié (2002) shows that Parsons results are very general as long as leisure is a normal good.

This chapter tries to move beyond the two-type model of imperfect tagging by first introducing a third type, the partially disabled, and thereafter constructing a model with a continuum of types (similar to yet distinct from the continuous model in Diamond & Sheshinski (1995)). Throughout, it is assumed that workers and non-workers have different utilities of consumption i.e. that utility is state-dependent.[1]

The chapter is organized as follows: section 2.2 familiarizes the reader with Parsons' two-type model and it is found that the ranking of incomes (consumption allocations) becomes incomplete when utility is assumed to be state dependent. Section 2.3 analyzes a three-type model (with two tags) where individuals are able, disabled or partially disabled. Here it is found that the main intuition from the two-type model still holds, although it is less straight forward, i.e. it is optimal to reward tagged individuals for working in line with their ability instead. Section 2.4 deals with a continuous version of the imperfect tagging model and it is found that, apart from the fact that the complexity of the issue increases, tagged individuals is treated more generously than untagged individuals either in terms of higher income or by not working at lower disability levels or both. Section 2.5 concludes the chapter.

[1] This is not done by Parsons (1996) who assumes state independent utilities, but this assumption can also be found in Diamond & Sheshinski (1995).

2.2 Imperfect Tagging - Two Types and One Tag

Consider an economy with two types of individuals: able and disabled. The proportion of able individuals is ℓ^A, accordingly the proportion of disabled is $\ell^D = 1 - \ell^A$. Individuals in this model are either fully able to work (able) or not able to work at all (disabled). θ is the degree of disability which is dichotomous in the two-type model, 0 (able) or 1 (disabled). All individuals in the economy are identical in all aspects but the degree of disability. All working individuals have the same marginal product and the marginal product is normalized to 1.

A disability insurance is designed to cover the loss of income (at least partially) and a screening process decides whether an individual is eligible for disability benefits. The screening process is exogenous to the model and assigns a disability "tag" to an individual with probability p_θ. The tagging is imperfect in that it, with positive probability, fails to assign disability tags to truly disabled individuals and, with positive probability, assigns disability tags to able individuals.

2.2.1 The screening process

Able individuals receive a disability tag (T) with probability p_0 and no tag (NT) with probability $\varphi_0 = 1 - p_0$. The disabled, however, receive a tag with probability p_1 and no tag with probability $\varphi_1 = 1 - p_1$. The probability of getting a disability tag is greater for the truly disabled than for the able, that is $p_1 > p_0 > 0$ and thus $\varphi_0 > \varphi_1 > 0$. Thus, the type-1 error is $\varphi_1$and the type-2 error is p_0.

2.2.2 The utility functions

Able individuals who are working, W, have the following utility of consumption:

$$U_W = u\left(c\right) - D_\theta$$

where $D_\theta > 0$ is the disutility of working depending on the degree of disability; for the able we have $\theta = 0$, thus $U_f^A = u\left(c\right) - D_0$. Furthermore, the utility of consumption when not working, N, is given by:

$$U_N = v\left(c\right)$$

The utility functions $u(c)$ and $v(c)$ are concave and increasing in consumption, and the marginal utilities go from ∞ to 0 as c goes from 0 to ∞. Work and consumption is preferable to no work and no consumption: $u(c) - D_0 > v(0)$. That is, able individuals will work in the absence of a disability insurance. Work is unpleasant such that; $u(c) - D_0 < v(c)$, *all* c. It is assumed that if $u\left(c\right) - D_0 = v(\tilde{c})$ it follows that $u'\left(c\right) < v'(\tilde{c})$. This is the moral hazard condition introduced by Diamond & Mirrlees (1978). The moral hazard condition states that equating the utilities between non-workers and workers will render a marginal utility higher for non-workers. This assumption is needed to characterize the relation between the utility functions $u\left(c\right)$ and $v\left(c\right)$. Notably, the moral hazard condition is satisfied in the models presented here and also for $u(c) = v(c)$.

2.2.3 Characterizing individuals and policy instruments

Given the two-sided classification error in the screening process, the individuals are characterized by their work status, their ability and their tag status and may thus be divided in to 8 groups

$$\begin{gathered}[ability, work, tag] : \left[A, W, NT\right], \left[A, W, T\right], \left[A, N, NT\right], \left[A, N, T\right], \\ \left[D, W, NT\right], \left[D, W, T\right], \left[D, N, NT\right], \left[D, N, T\right].\end{gathered}$$

Since it is impossible for the disabled to work, the two groups $[D, W, NT]$ and $[D, W, T]$ are not realized in the model. The disability insurance administration, performing the screening process and deciding on benefit levels, knows that there are six groups of individuals in the economy. However, it cannot fully distinguish between able and disabled individuals, i.e. disability is imperfectly observed. Work- and tag-status are observed by the insurance administration. Since the screening process is exogenous to the model, the insurance administration's only available policy instrument is the benefit levels, here modelled as the different consumption allocations to the distinguishable groups in the economy: c_W, c_W^T, $c_N$and c_N^T, where the superscript is the tag status and the subscript is the work status.[2]

2.2.4 The optimization problem

Assume that all able individuals work in the optimal program (cf. Parsons, 1996). The policy objective is to maximize the expected social welfare by choosing the benefit level (the consumption vector). In this case the expected social welfare is given by:

$$(1-p_0)\ell^A(u\left(c_W\right)-D_0)+p_0\ell^A(u\left(c_W^T\right)-D_0)+ \\ +(1-p_1)(1-\ell^A)v\left(c_N\right)+p_1(1-\ell^A)v\left(c_N^T\right)$$

In maximizing expected social welfare the insurance administration is constrained by the resources available in the economy and by the work constraints. The resource constraint is:

$$(1-p_0)\ell^A c_W+p_0\ell^A c_W^T+(1-p_1)(1-\ell^A)c_N+p_1(1-\ell^A)c_N^T\leq M$$

where M is the resources in the economy. It is assumed that all working individuals have the same marginal product and this is normalized to 1; thus $M=\ell^A\times 1$. Note that individuals have no reason to forgo consumption in

[2] That is, the insurance administration uses the information it possesses, about work- and tag-status to construct the policy instrument.

this model and the resource constraint therefore holds with equality (Parsons, 1996). The insurance administration is also constrained by the work constraints of the able individuals. If the able are to work irrespective of their tag, the following constraints need to be fulfilled:

$$u\left(c_W\right) - D_0 \geq v\left(c_N\right)$$

for untagged individuals, and

$$u\left(c_W^T\right) - D_0 \geq v\left(c_N^T\right)$$

for tagged individuals.

Thus, the insurance administration solves the following maximization problem:

$$\max_{c_W, c_W^T, c_N, c_N^T} (1-p_0)\ell^A(u\left(c_W\right) - D_0) + p_0\ell^A(u\left(c_W^T\right) - D_0) + \\ +(1-p_1)(1-\ell^A)v\left(c_N\right) + p_1(1-\ell^A)v\left(c_N^T\right) \tag{2.1}$$

subject to the resource constraint:

$$(1-p_0)\ell^A c_W + p_0\ell^A c_W^T + (1-p_1)(1-\ell^A)c_N + p_1(1-\ell^A)c_N^T = M \tag{2.2}$$

and the work constraints:

$$u\left(c_W\right) - D_0 \geq v\left(c_N\right) \tag{2.3}$$

$$u\left(c_W^T\right) - D_0 \geq v\left(c_N^T\right) \tag{2.4}$$

The first order conditions, where λ_j for $j = i, ii, iii$ are the Lagrange multipliers, for this maximization are the following: :

$$\frac{\partial L}{\partial C_f} = (1-p_0)\ell^A u'\left(c_W\right) - \lambda_i(1-p_0)\ell^A + \lambda_{ii}u'\left(c_W\right) = 0 \tag{2.5}$$

$$\frac{\partial L}{\partial C_d} = (1-p_1)(1-\ell^A)v'(c_N) - \lambda_i(1-p_1)(1-\ell^A) - \lambda_{ii}v'(c_N) = 0 \quad (2.6)$$

$$\frac{\partial L}{\partial C_f^d} = p_0\ell^A u'\left(c_W^T\right) - \lambda_i p_0\ell^A + \lambda_{iii}u'\left(c_W^T\right) = 0 \quad (2.7)$$

$$\frac{\partial L}{\partial C_d^d} = p_1)(1-\ell^A)v'\left(c_N^T\right) - \lambda_i p_1(1-\ell^A) - \lambda_{iii}v'\left(c_N^T\right) = 0 \quad (2.8)$$

Since the resource constraint is binding we know that $\lambda_i > 0$, and it can be shown that $\lambda_{ii} > 0$ and $\lambda_{iii} > 0$ (see appendix), i.e. the work constraints bind at the optimum.

2.2.5 The consumption allocations

How are the different consumption allocations related to each other? First, in the optimal program both work constraints are binding implying that:

$$u(c_W) - v(c_N) = u\left(c_W^T\right) - v\left(c_N^T\right) \quad (2.9)$$

Moreover the first order conditions yield the following redistribution principle (cf. Parsons, 1996):

$$\omega\frac{1}{u'(c_W)} + (1-\omega)\frac{1}{v'(c_N)} = \omega^T\frac{1}{u'(c_W^T)} + (1-\omega^T)\frac{1}{v'(c_N^T)} \quad (2.10)$$

where $\omega = \frac{(1-p_0)\ell^A}{(1-p_1)(1-\ell^A)+(1-p_0)\ell^A}, \omega^T = \frac{p_0\ell^A}{p_1(1-\ell^A)+p_0\ell^A}$. Both these weights are positive (by the screening mechanism) and less than or equal to one. Concerning the ranking of the consumption allocations, the following proposition is made:

Proposition 2.1:

$$c_W^T > c_W \lesseqgtr c_N^T > c_N$$

Proof. First determine the ranking within the tag-states. The binding work constraints and the assumptions about the utility functions give that $c_W > c_N$ and $c_W^T > c_N^T$. Next, note that $p_1 > p_0$ implying $\omega > \omega^T$. Knowing this, now assume that $c_W = c_W^T$, which implies that $c_N = c_N^T$ since the work constraints are binding. Thus $\frac{1}{u'(c_W)} = \frac{1}{u'(c_W^T)}$ and $\frac{1}{v'(c_N)} = \frac{1}{v'(c_N^T)}$, and the left-hand side of (2.10) is greater than the right-hand side. That is

$$\omega \frac{1}{u'(c_W)} + (1-\omega)\frac{1}{v'(c_N)} > \omega^D \frac{1}{u'(c_W^T)} + (1-\omega^D)\frac{1}{v'(c_N^T)}$$

However in the optimal program the redistribution principle is satisfied with equality, thus the redistribution principle requires that c_W is reduced relative to c_W^T (since LHS is an increasing function of c_W and RHS an increasing function of c_W^T) . Thus $c_W < c_W^T$ in the optimal program, implying that $c_N < c_N^T$, since the work constraints are binding and both $v(c)$ and $u(c)$ are increasing functions. ■

Concerning the ranking of c_N^T and c_W it may be concluded that $u'(c_W) < v'(c_N^T)$ in the optimal program, since the smallest element in a weighted average cannot exceed or equal the greatest element of another weighted average if the averages are equal and have positive weights, that is $\frac{1}{u'(c_W)} > \frac{1}{v'(c_N^T)}$. Notably $u(c_W) - D_0 < v\left(c_N^T\right)$ since $c_N^T > c_N$. Assuming $v'(c) = u'(c)$ all c would ensure that the consumption allocations could be fully ranked ($c_N^T < c_W$), and this is implicitly assumed in the examples below, but not assumed in the general model.

It is not obvious that tagged non-workers should have a lower income (consumption) than untagged workers in this model. It might be optimal, from a utilitarian point of view, to increase the consumption of tagged non-workers above the level of the untagged workers if the marginal utility of consumption for the former group is high enough. However, this setup follows Parsons' results in providing a premium for being tagged and working, and also in keeping income for untagged non-workers low (e.g. low social assistance level).

2.2.6 Logarithmic example

To illustrate the model a logarithmic example is constructed. Assume that $U_f^A = u(c) - D_0 = \ln c - D_0$ and $U_d^i = v(c) = \ln 2c$ for $i = A, D$. This specification of the model is solved for c_W, c_W^T, $c_N$and c_N^T.

Figure 2.1 shows the consumption allocations assigned to the different groups in the optimal insurance program when the probability of the able getting a disability tag (p_0) varies. This is done under the following assumptions: $p_1 = 0.8$, $D_0 = 0.5$ and $M = \ell^A = 0.8.$[3]. Notably, the required difference in consumption between working and not working is greater than in Parsons' example. For $p_0 = 0.2$, the consumption allocations are $\left[c_W, c_N, c_W^T, c_N^T\right] = [0.83, 0.25, 1.23, 0.37]$ in this example compared to $[0.819, 0.497, 0.996, 0.604]$ in Parsons' example. Thus, this model requires greater reward for workers, as expected. Note that when $p_0 > 0.8$ this violates the assumption of the model, i.e. $p_1 > p_0$, explaining the "exponential" growth of the consumption allocations for the non-tagged groups.

Moreover when the assumption that $p_0 > p_1$ is violated, the tagged and untagged switch roles. This can be seen in figure 1 where not having a tag is a better signal of disability than having a tag when $p_0 > p_1$and thus the consumption allocations are greater for the untagged groups in the economy in this range. Clearly, when p_1 is high and p_0 low, i.e. the screening process is very reliable, the model approaches full insurance and it is optimal to compensate the few able and tagged individuals substantially if they choose to work in spite of the tag.

[3]These assumptions follow the assumptions made by Parsons (1996) and figure 1 thus resembles Parsons' figure 2 (p.197). The difference between the figures stems from the specification of the utility functions, that is the higher utility of consumption for non-workers.

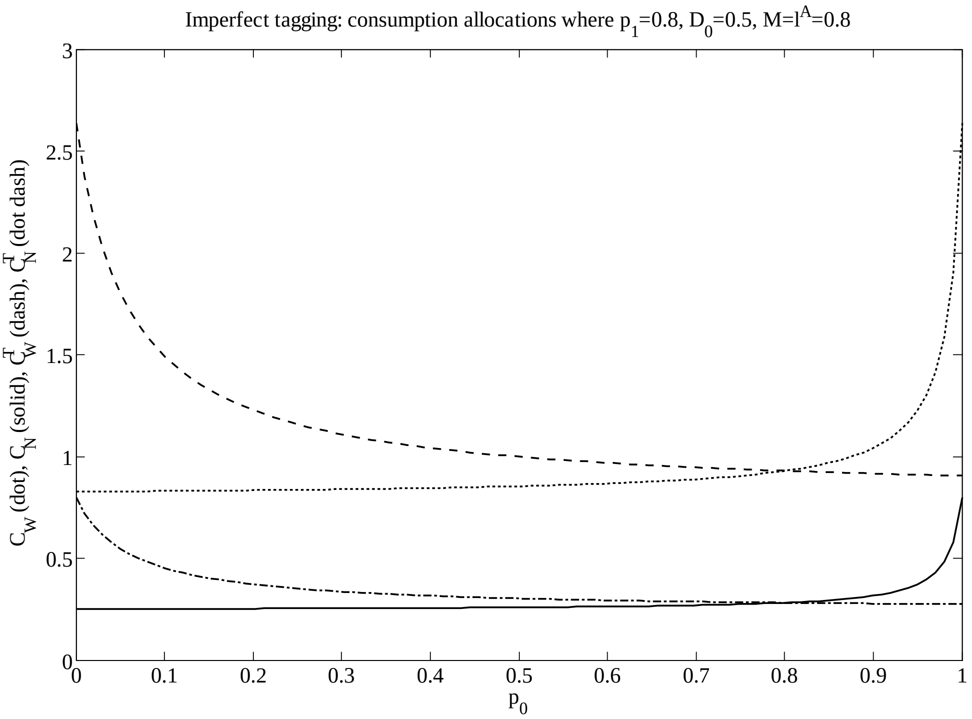

Figure 2.1: Logarithmic example imperfect tagging

2.3 Imperfect Tagging - Three Types and Two Tags

Now consider an economy with three types of individuals: able, partially disabled and disabled, whose population weights are ℓ^A, ℓ^P, and $\ell^D = 1 - (\ell^A + \ell^P)$, respectively. These population weights are exogenous to the model. The three types differ in their disutility of working, which is increasing in the degree of disability, $0 \leq \theta \leq 1$. All individuals in the economy are identical in all aspects but the degree of disability. All working individuals have the same marginal product normalized to 1. The disabled are unable to work. Furthermore, all degrees of partial disability are, for simplicity, treated as one type. Assume that partial disability is limiting such that the partially disabled may only work part-time or not work at all. The able, however, may work full-time, part-time or not at all. Able individuals have $\theta = 0$, partially disabled $\theta = \bar{\theta}$ ($\bar{\theta} \in (0, 1)$), and disabled $\theta = 1$. Apart from the the addition of a third type this model is similar to the two-type model - with appropriate modifications to accommodate for the third type.

2.3.1 The screening process

Able individuals receive a disability tag (T) with probability p_0, partial disability tag (P) with probability π_0 and no tag (NT) with probability $\varphi_0 = 1 - \pi_0 - p_0$. The disabled receive a disability tag with probability p_1, partial disability tag with probability π_1 and no tag with probability $\varphi_1 = 1 - \pi_1 - p_1$. For the partially disabled these probabilities are given by $p_{\bar{\theta}}, \pi_{\bar{\theta}}$ and $\varphi_{\bar{\theta}} = 1 - \pi_{\bar{\theta}} - p_{\bar{\theta}}$ respectively. The probability of getting a disability tag is greater for the truly disabled than for the partially disabled, which in turn is greater than the probability for the able, that is $p_1 > p_{\bar{\theta}} > p_0 > 0$. Furthermore; $\pi_{\bar{\theta}} > \pi_1 > \pi_0 > 0$. These two conditions state that the screening mechanism is more capable of separating the able from the partially disabled and the disabled than the latter two from each other - this since the able have the smallest probabilities of being tagged in both instances. It is also assumed that $0 < \varphi_1 < \varphi_{\bar{\theta}} < \varphi_0$, implying that $p_1 - p_{\bar{\theta}} > \pi_{\bar{\theta}} - \pi_1$, which may be interpreted as follows; the screening mechanism will relatively more often tag a disabled as a partially disabled than tag a partially disabled as a disabled. Furthermore, it is assumed that the probabilities within each type are ranked as follows:

$$\begin{aligned} p_1 &> \pi_1 > \varphi_1 \text{ for the disabled} \\ \pi_{\bar{\theta}} &> p_{\bar{\theta}} > \varphi_{\bar{\theta}} \text{ for the partially disabled} \\ \varphi_0 &> \pi_0 > p_0 \text{ for the able} \end{aligned}$$

Thus the probability of getting the "right" tag is greatest for each type, followed by the probability of getting the tag closest to the individual's actual status. The assumption for the partially disabled implicitly says that the partially disabled are closer to being disabled than able. The imperfect tagging is more complex in the three-type-economy than in the two-type-economy. The number of possible erroneous judgements in the screening process is increased from two to six: able getting T-tag (p_0), able getting P-tag (π_0), partially disabled getting T-tag ($p_{\bar{\theta}}$), partially disabled getting no tag ($\varphi_{\bar{\theta}}$), disabled getting P-tag (π_1), and disabled getting no tag (φ_1).

2.3.2 The utility functions

The individuals' utility depend on their work-status, ability and of course the consumption they are allocated. Able and partially disabled have different disutility of working, and partially disabled have a greater disutility of working $D_{\bar{\theta}}$ than able individuals, i.e. $D_{\bar{\theta}} > D_0$. Remember, the disutility of working is increasing in degree of disability. The disutility of working, for both types, is reduced by working part time.

Able individuals who are working full-time, W, have the following utility of consumption:

$$U_W = u(c) - D_0$$

Able individuals may also choose to work part-time, p, instead of full-time.

If they do so the utility of consumption is given by:

$$U_p = u\left(c\right) - kD_0$$

where k measures the reduction in disutility from working part-time, $0 < k \leq 1$. k can be interpreted as the extent of the part-time work, for example working 50 %.Thus it is assumed that the disutility of working part-time may equal the disutility of working full-time, but not equal zero for able individuals. Utility of consumption for the partially disabled (partial disability is denoted by the superscript P) working part-time is:

$$U_p^P = u\left(c\right) - kD_{\bar{\theta}}$$

For all types the utility of consumption when not working, N, is given by:

$$U_N = v\left(c\right)$$

As in the two-type-economy $u(c)$ and $v(c)$ are both concave and increasing. $u'(c)$ goes from ∞ to 0 as c goes from 0 to ∞. This also holds for $v'(c)$. For all $c > 0$ and $\theta < 1$ it is assumed that $u(c) - D_\theta > v(0)$ implying that $u(c) - kD_\theta > v(0)$, that is work and consumption is preferred to no work and no consumption by both the able and the partially disabled. Since work is unpleasant, it entails a disutility, thus $u(c) - kD_\theta < v(c)$, implying that $u(c) - D_\theta < v(c)$ for all c. The moral hazard condition from the two-type model is extended to suit the three-type model:[4]

$$u(\hat{c}) - kD_\theta = v(\tilde{c}) \Rightarrow u'(\hat{c}) < v'(\tilde{c})$$

That is, when the utility of working (full-time as able or part-time as partially disabled) is equal to the utility of not working and receiving disability benefits then the marginal utility of extra consumption is higher for not working.

[4]Notably this implies that $u(c) - D_\theta = v(\tilde{c}) \Rightarrow u'(c) < v'(\tilde{c})$ through the concavity of $u(\bullet)$

2.3.3 Characterizing individuals and policy instruments

Once again individuals are characterized by their work-status, their ability and their tag-status and, as mentioned above, partially disabled cannot work full-time and disabled individuals cannot work at all. This implies that there are 18 different groups in the economy (this can easily be verified by the reader) and thus the number of policy instruments is extended to match the increase of possible states (compared to the two-type-economy). The insurance policy now consists of nine consumption allocations:

$$c_W, c_p, c_N, c_W^P, c_p^P, c_N^P, c_W^T, c_p^T, c_N^T$$

where the superscript is the tag-status and the subscript is the work-status.

2.3.4 Optimality and the optimization problem

The insurance administration uses the policy instruments to maximize social welfare. It is assumed that all able individuals work full-time and all partially disabled individuals work part-time in the optimal program. The insurance administration optimizes the following social welfare function:

$$\begin{gathered} SWF = \varphi_0 \ell^A (u(c_W) - D_0) + \pi_0 \ell^A (u(c_W^P) - D_0) + p_0 \ell^A (u(c_W^T) - D_0) + \\ + \varphi_{\bar\theta} \ell^P (u(c_p) - kD_{\bar\theta}) + \pi_{\bar\theta} \ell^P (u(c_p^P) - kD_{\bar\theta}) + p_{\bar\theta} \ell^P (u(c_p^T) - kD_{\bar\theta}) + \\ + \varphi_1 \ell^D v(c_N) + \pi_1 \ell^D v\left(c_N^P\right) + p_1 \ell^D v\left(c_N^T\right) \end{gathered}$$

The insurance administration is constrained in its choice of consumption allocations by a resource constraint:

$$\begin{gathered} \varphi_0 \ell^A c_W + \pi_0 \ell^A c_W^P + p_0 \ell^A c_W^T + \varphi_{\bar\theta} \ell^P c_p + \pi_{\bar\theta} \ell^P c_p^P + p_{\bar\theta} \ell^P c_p^T + \\ + \varphi_1 \ell^D c_N + \pi_1 \ell^D c_N^P + p_1 \ell^D c_N^T = M \end{gathered}$$

Where M could be equal to the production in the economy (remember that each able individual has a marginal product equal to one and assume that the marginal product of partially disabled is k), i.e. $M = \ell^A * 1 + \ell^D * k$. As in the two-type-economy the resource constraint is binding. The insurance administration also has to ensure that able individuals will work full-time and that partially disabled will work part-time in the optimal program. As discussed earlier, the able have three options concerning work status irrespective of tag status. To induce them to work full-time in all tag states, the following constraints need to be fulfilled:

For able with no tag:

$$\begin{aligned} u(c_W) - D_0 &\geq u(c_p) - kD_0 \\ u(c_W) - D_0 &\geq v(c_N) \end{aligned}$$

For able with a P-tag:

$$\begin{aligned} u(c_W^P) - D_0 &\geq u(c_p^P) - kD_0 \\ u(c_W^P) - D_0 &\geq v(c_N^P) \end{aligned}$$

For able with a T-tag:

$$\begin{aligned} u(c_W^T) - D_0 &\geq u(c_p^T) - kD_0 \\ u(c_W^T) - D_0 &\geq v(c_N^T) \end{aligned}$$

These work constraints ensure that able individuals will choose to work full-time over working part-time and not working at all. The options open to the partially disabled are limited since they cannot work full-time. For the partially disabled the (part-time) work constraints are:

For partially disabled with no tag:

$$u(c_p) - kD_{\bar{\theta}} \geq v(c_N)$$

For partially disabled with a P-tag:

$$u(c_p^P) - kD_{\bar{\theta}} \geq v(c_N^P)$$

For partially disabled with a T-tag:

$$u(c_p^T) - kD_{\bar{\theta}} \geq v(c_N^T)$$

The part-time work constraints ensure that the partially disabled will choose to work part-time over not working at all, irrespective of tag status. Disabled individuals cannot work and thus there is no need for constraints on their behavior. Before setting up the optimization problem it will be helpful to simplify notation, let: $a = \varphi_0 \ell^A, b = \pi_0 \ell^A, c = p_0 \ell^A, d = \varphi_{\bar{\theta}} \ell^P, e = \pi_{\bar{\theta}} \ell^P, f = p_{\bar{\theta}} \ell^P, g = \varphi_1 \ell^D, h = \pi_1 \ell^D, i = p_1 \ell^D$. The optimization problem thus becomes:

$$\max_{c_W, c_p, c_N, c_W^P, c_p^P, c_N^P, c_W^T, c_p^T, c_N^T} SWF \tag{2.11}$$

subject to

$$ac_W + bc_W^P + cc_W^T + d\ell^D c_p + ec_p^P + fc_p^T + gc_N + hc_N^P + ic_N^T = M \tag{r1}$$

$$\begin{aligned} u(c_W) - D_0 &\geq u(c_p) - kD_0 && \text{(r2)} \\ u(c_W) - D_0 &\geq v(c_N) && \text{(r3)} \end{aligned}$$

$$\begin{aligned} u(c_W^P) - D_0 &\geq u(c_p^P) - kD_0 && \text{(r4)} \\ u(c_W^P) - D_0 &\geq v(c_N^P) && \text{(r5)} \end{aligned}$$

$$\begin{aligned} u(c_W^T) - D_0 &\geq u(c_p^T) - kD_0 && \text{(r6)} \\ u(c_W^T) - D_0 &\geq v(c_N^T) && \text{(r7)} \end{aligned}$$

$$u(c_p) - kD_{\bar{\theta}} \geq v(c_N) \tag{r8}$$

$$u(c_p^P) - kD_{\bar{\theta}} \geq v(c_N^P) \tag{r9}$$

$$u(c_p^T) - kD_{\bar{\theta}} \geq v(c_N^T) \tag{r10}$$

Obviously, not all work constraints are slack or binding at the same time. Moreover, it is straightforward that constraints r3, r5 and r7 are fulfilled with strict inequality if the other constraints are satisfied with equality. It can also be shown that the last case is a solution candidate to the optimization problem (see appendix). Thus r3, r5 and r7 are redundant and other constraints are binding in the optimal program.

Now, let λ_j for $j = 1, ..., 10$ be the Lagrange multipliers of the optimization problem with $\lambda_j = 0$ for $j = 3, 5, 7$ and $\lambda_i > 0$, for $i = 1, 2, 4, 6, 8, 9, 10$. Then the first order conditions are the following:

$\frac{\partial \mathcal{L}}{c_W} = au'(c_W) - \lambda_1 a + \lambda_2 u'(c_W) = 0$
$\frac{\partial \mathcal{L}}{c_p} = du'(c_p) - \lambda_1 d - \lambda_2 u'(c_p) + \lambda_8 u'(c_p) = 0$
$\frac{\partial \mathcal{L}}{c_N} = gv'(c_N) - \lambda_1 g - \lambda_8 v'(c_N) = 0$
$\frac{\partial \mathcal{L}}{c_W^P} = bu'(c_W^P) - \lambda_1 b + \lambda_4 u'(c_W^P) = 0$
$\frac{\partial \mathcal{L}}{c_p^P} = eu'(c_p^P) - \lambda_1 e - \lambda_4 u'(c_p^P) + \lambda_9 u'(c_p^P) = 0$
$\frac{\partial \mathcal{L}}{c_N^P} = hv'(c_N^P) - \lambda_1 h - \lambda_9 v'(c_N^P) = 0$
$\frac{\partial \mathcal{L}}{c_W^T} = cu'(c_W^T) - \lambda_1 c + \lambda_6 u'(c_W^T) = 0$
$\frac{\partial \mathcal{L}}{c_p^T} = fu'(c_p^T) - \lambda_1 f - \lambda_6 u'(c_p^T) + \lambda_{10} u'(c_p^T) = 0$
$\frac{\partial \mathcal{L}}{c_N^T} = iv'(c_N^T) - \lambda_1 i - \lambda_{10} v'(c_N^T) = 0$

2.3.5 Redistribution principle and consumption allocations

Now, note that the work constraints give the following ranking of consumption allocations within each tag-status group:

$c_W > c_p > c_N$, for not tagged

$c_W^P > c_p^P > c_N^P$, for P-tagged

$c_W^T > c_p^T > c_N^T$, for T-tagged

This is not surprising given the setup of the model. However, the relation between consumption allocations across tag-status is less obvious and thus more interesting . As in the two-type-economy a redistribution principle can be elaborated from the first order conditions. The redistribution principle describes the redistribution among untagged, P-tagged and T-tagged individuals, and states that the weighted average of the inverse marginal utilities is equalized across tag-status.

$$\omega_a \frac{1}{u'(c_W)} + \omega_g \frac{1}{v'(c_N)} + (1 - \omega_a - \omega_g) \frac{1}{u'(c_p)} =$$
$$= \omega_b \frac{1}{u'(c_W^P)} + \omega_h \frac{1}{v'(c_N^P)} + (1 - \omega_b - \omega_h) \frac{1}{u'(c_p^P)} =$$
$$= \omega_c \frac{1}{u'(c_W^T)} + \omega_i \frac{1}{v'(c_N^T)} + (1 - \omega_c - \omega_i) \frac{1}{u'(c_p^T)}$$

where

$$\begin{aligned}
\omega_a &= \frac{\varphi_0 \ell^A}{\varphi_0 \ell^A + \varphi_{\bar{\theta}} \ell^P + \varphi_1 \ell^D}, \omega_g = \frac{\varphi_1 \ell^D}{\varphi_0 \ell^A + \varphi_{\bar{\theta}} \ell^P + \varphi_1 \ell^D}, \\
\omega_b &= \frac{\pi_0 \ell^A}{\pi_0 \ell^A + \pi_{\bar{\theta}} \ell^P + \pi_1 \ell^D}, \omega_h = \frac{\pi_1 \ell^D}{\pi_0 \ell^A + \pi_{\bar{\theta}} \ell^P + \pi_1 \ell^D}, \\
\omega_c &= \frac{p_0 \ell^A}{p_0 \ell^A + p_{\bar{\theta}} \ell^P + p_1 \ell^D}, \omega_i = \frac{p_1 \ell^D}{p_0 \ell^A + p_{\bar{\theta}} \ell^P + p_1 \ell^D}
\end{aligned}$$

Ranking of consumption allocations across tag-status

Could it be optimal to ignore the individuals' tag-status and give all workers the same consumption allocation? Formally this is represented by setting $c_W = c_W^P = c_W^T$, implying, through the work constraints that $c_p = c_p^P = c_p^T$, which in turn implies $c_N = c_N^P = c_N^T$. Under these conditions the redistribution principle becomes (see appendix):

$$\omega_a \frac{1}{u'(c_W)} + \omega_g \frac{1}{v'(c_N)} + (1 - \omega_a - \omega_g) \frac{1}{u'(c_p)} > \tag{A}$$

$$\omega_b \frac{1}{u'(c_W^P)} + \omega_h \frac{1}{v'(c_N^P)} + (1 - \omega_b - \omega_h) \frac{1}{u'(c_p^P)} > \tag{B}$$

$$\omega_c \frac{1}{u'(c_W^T)} + \omega_i \frac{1}{v'(c_N^T)} + (1 - \omega_c - \omega_i) \frac{1}{u'(c_p^T)} \tag{C}$$

However, the redistribution principle is fulfilled with equality in the optimal program; thus it cannot be optimal to ignore the information that the tag-status gives. Note that A is an increasing function in c_W since $\frac{1}{u'(c_W)}$ is increasing in c_W , and that B (C) is increasing in c_W^P (c_W^T) by the same reasoning. Therefore c_W needs to be lowered compared to both c_W^P and c_W^T, and c_W^P needs to be lowered compared to c_W^T to achieve equality, implying that $c_W < c_W^P < c_W^T$ in the optimal program. Now, the binding work constraints yield that $c_p < c_p^P < c_p^T$ and $c_N < c_N^P < c_N^T$. Notably, the information contained in the screening process is used at the optimum. That is, individuals with the same work-status but different tag-status will receive different consumption allocations. Thus, the ranking between tag-status groups for individuals with the same work status is:

$c_W^T > c_W^P > c_W$, for full-time workers

$c_p^T > c_p^P > c_p$, for part-time workers

$c_N^T > c_N^P > c_N$, for non-workers

Combining this with the ranking within each tag-status group, a partial ranking may be achieved. Obviously c_W^T is the greatest consumption allocation and c_N the smallest, in line with the results from the two-type model. Table 1 contains the partial ranking.

	c_W	c_p	c_N	c_W^P	c_p^P	c_N^P	c_W^T	c_p^T	c_N^T
c_W	$=$	$>$	$>$	$<$	$\lesseqqgtr$	$\lesseqqgtr$	$<$	$\lesseqqgtr$	$\lesseqqgtr$
c_p	$<$	$=$	$>$	$<$	$<$	$\lesseqqgtr$	$<$	$<$	$\lesseqqgtr$
c_N	$<$	$<$	$=$	$<$	$<$	$<$	$<$	$<$	$<$
c_W^P	$>$	$>$	$>$	$=$	$>$	$>$	$<$	$\lesseqqgtr$	$\lesseqqgtr$
c_p^P	$\lesseqqgtr$	$>$	$>$	$<$	$=$	$>$	$<$	$<$	$\lesseqqgtr$
c_N^P	$\lesseqqgtr$	$\lesseqqgtr$	$>$	$<$	$<$	$=$	$<$	$<$	$<$
c_W^T	$>$	$>$	$>$	$>$	$>$	$>$	$=$	$>$	$>$
c_p^T	$\lesseqqgtr$	$>$	$>$	$\lesseqqgtr$	$>$	$>$	$<$	$=$	$>$
c_N^T	$\lesseqqgtr$	$\lesseqqgtr$	$>$	$\lesseqqgtr$	$\lesseqqgtr$	$>$	$<$	$<$	$=$

Table 1: the partial ranking

The inequality signs show how the consumption allocations

in the vertical column are related to the cons. allocation in the top row.

The assumption $v'(c) = u'(c)$ for all c gives us a complete ranking in the two-type case, and imposing the same assumption for the three-type model ensures that $c_W > c_N^T$, $c_W^P > c_N^T$ and that $c_W > c_W^P$. This follows from the fact that the smallest element of a weighted average cannot exceed or equal the greatest element of another weighted average if equality is to hold.[5] Thus the assumption improves the ranking but does not make it complete for any weights between zero and one. The ranking of consumption allocations given to part-time workers, compared to full-time workers of different tag status, is undetermined in this case and will depend on the disutility of working for the groups as can be seen in the logarithmic example below. The ranking between part-time workers and non-workers of different tag-status is also undetermined and will depend (similar to the two-type model) on the marginal utility of consumption. If non-workers have a very high marginal utility of consumption, it might be socially optimal to give them a higher consumption than part-time workers (from a utilitarian point of view).

[5] For positive weights and as in this case $0 < \omega_j < 1$, $j = a, b, c, d, e, f, g, h, i$

2.3.6 Logarithmic example

To illustrate the three-type model it is once again assumed that $u(c) = \ln c$ and $v(c) = \ln 2c$. The somewhat rigorous assumptions about the screening process limit the opportunities for comparisons, but some tendencies may be found in the examples below. Especially the effect of work disutility on the consumption allocations can be easily depicted. Figure 2.2 plots the consumption allocations over D_0 (the disutility of working for able individuals) under the following assumptions concerning the screening process and the population weights: $p_0 = 0.1$, $p_1 = 0.7$, $p_\theta = 0.2$, $\pi_0 = 0.2$, $\pi_1 = 0.2$, $\pi_\theta = 0.7$, $\ell^P = 0.1$, $\ell^D = 0.1$, $\ell^A = 0.8$, $D_\theta = 1$. Furthermore it is assumed that $k = 0.5$, i.e. the part-time workers work 50 percent.

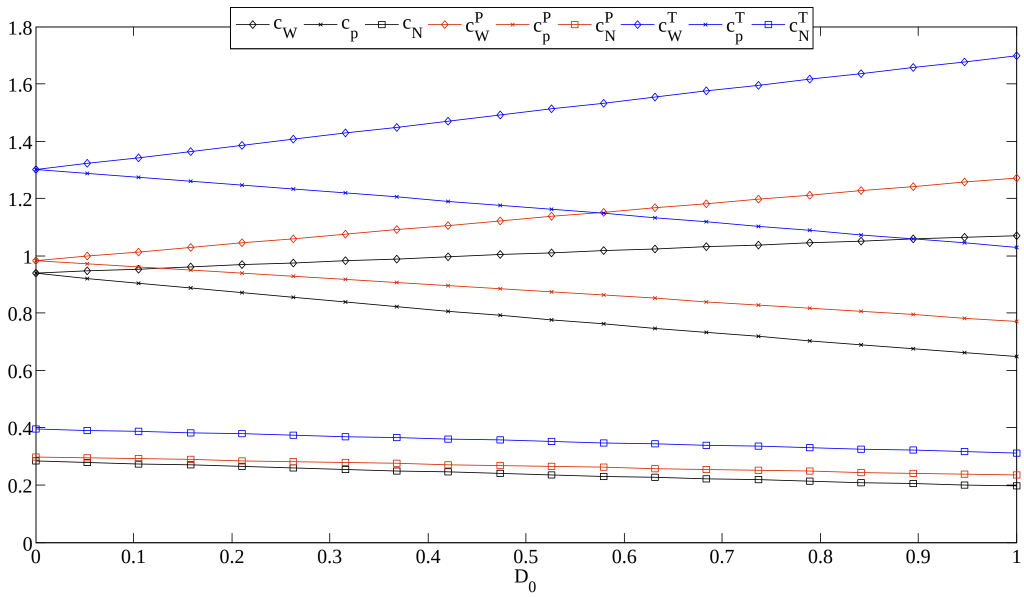

Figure 2.2: The three-type economy: logarithmic example I

Notably the consumption allocations for non-workers are significantly lower than for both part-time and full-time workers, while the ranking between part-time workers and full-time workers of different tag-status varies with D_0. The income for T-tagged part-time workers (c_p^T), for example, is greater than the income for both P-tagged and untagged full-time workers over a wide range of D_0. However, as D_0 grows difference in income between part-time and full-time workers within the same tag-status grows - to ensure that full-time work is an attractive alternative. When the disutility of working is sufficiently great, above 0.9, then the consumption allocations are ranked in a familiar way i.e.:

$$c_W^T > c_W^P > c_W > c_p^T > c_p^P > c_p > c_N^T > c_N^P > c_N$$

similar to the logarithmic example in the two-type model.

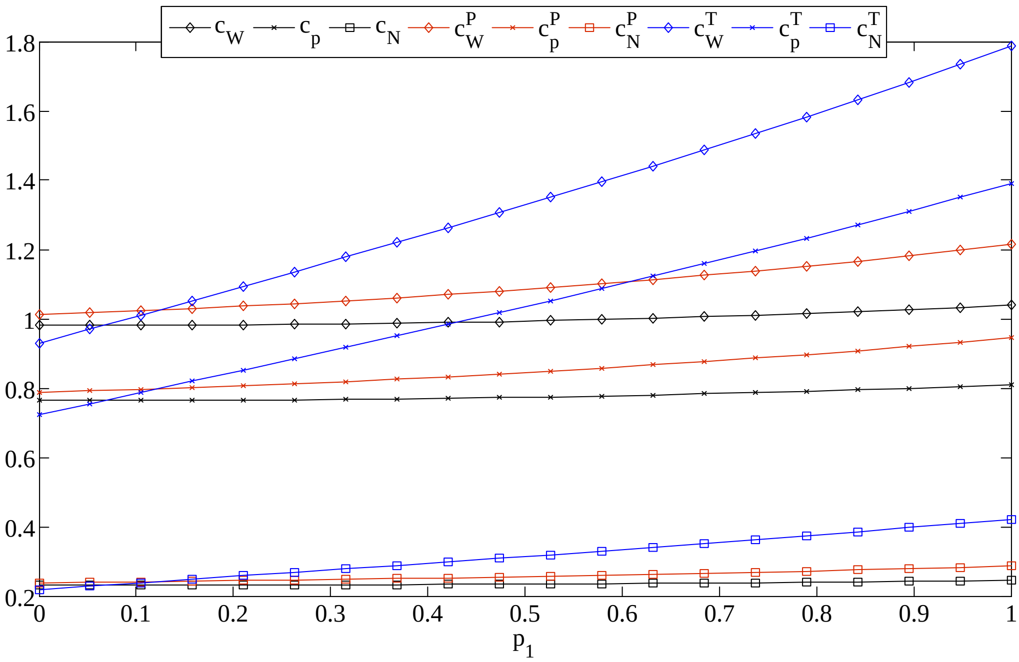

Figure 2.3: The three-type economy: logarithmic example II

As already noted, the assumptions about the screening process limit the amount of reasonable examples, but figure 2.3 plots the consumption allocations over p_1. It shows that the coverage of the income loss for the disabled improves as the probability that the disabled receive a disability tag becomes higher. A similar effect can be found for the partially disabled as the probability of assigning a P-tag to the partially disabled increases (not presented here). It is apparent that similar forces are at work in the three-type model to those in the two-type model, but the results are less straightforward and obvious - a natural effect of introducing a third type. The logarithmic examples underline this finding.

2.4 Imperfect Tagging - A Continuum of Types and One Tag

In this section the economy consists of a continuum of types i.e. all degrees of disability are represented. However, it is assumed that the disability insurance once again, as in section 2.2, only has one disability level, disabled, and thus only one tag (T). The continuous model in this section is similar to the continuous model in Diamond & Sheshinski (1995), but relaxes the restriction that the salary is the same for tagged and untagged workers and assumes that all individuals are applicants for the disability insurance (i.e. applying for a tag is not voluntary in the model presented here). The former implies that the continuous model presented here opens up for differentiating salary over tag-status (as is done by Parsons (1996)).

2.4.1 The Distribution of Disability and the Screening Process

Disability, θ, is distributed continuously over the population. Let $F(\theta)$ be the distribution of disability with the density $f(\theta)$ and let θ go from 0 to 1. The fully able (hereinafter called *able*) individuals have $\theta = 0$ and completely disabled (*disabled*) have $\theta = 1$. Assume that the disabled cannot work and that all other individuals may choose not to work. As in the previous sections the individuals go through a screening process and are either judged to be able or disabled. Disabled individuals are eligible for disability benefits. The screening is imperfect and individuals are judged disabled (able) with probability $p(\theta)$ $(1 - p(\theta))$ where $p(\theta)$ is monotonically increasing in θ. If

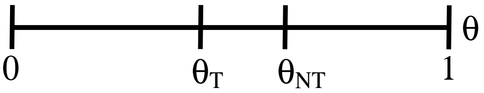

Figure 2.4: Possible relation θ_T and θ_{NT}

judged disabled the individual receives a tag T; otherwise she receives no tag NT.

Furthermore, let the cut-off levels - above this level no one works - for tagged and untagged individuals be θ_T and θ_{NT}, respectively. The work constraints, introduced below, ensure that untagged individuals with $\theta \in [\theta_{NT}, 1)$, and tagged individuals with $\theta \in [\theta_T, 1)$, do not work. That is, above these levels the disutility of working is such that it is better to receive disability benefits (if tagged) or social assistance (if untagged). Given the usual relation between disability and social assistance benefits, with the former being greater, and the results in the previous sections it is likely that $\theta_{NT} > \theta_T$ (see figure (2.4)). Diamond & Sheshinski (1995) show that this holds in their model, but they have a restriction on their model that is not present here, i.e. that $c_W = c_W^T$, implying that the cut-off for tagged will be lower than the cut-off for untagged as long as $p(\theta)$ is increasing in θ - the targeted group (the tagged) is treated more generously in this program (which in the end is the purpose of a disability insurance). Allowing for a premium for tagged workers opens up for alternative relations between the cut-off levels (i.e. $\theta_{NT} \leq \theta_T$), but $\theta_{NT} > \theta_T$ remains the most appealing option given the purpose of introducing a disability insurance.

2.4.2 The Utility Functions

Individuals who are working, W, have the following utility of consumption:

$$U_W = u(c) - D(\theta)$$

where $D(\theta) > 0$ is the disutility of working depending on the degree of disability. Assume that $D(\theta)$ is increasing in θ such that $D(\theta) \to \infty$ when $\theta \to 1$. Furthermore, the utility of consumption when not working, N, is given by:

$$U_N = v(c)$$

Similar to the previous sections, it is assumed that the utility functions $u(c)$ and $v(c)$ are concave and increasing in consumption and that the marginal utilities go from ∞ to 0 as c goes from 0 to ∞. Moreover, work is unpleasant such that; $u(c) - D(\theta) < v(c)$, for all c and all θ. It is assumed that if $u(c) - D(\theta) = v(\tilde{c})$, for all θ, it follows that $u'(c) < v'(\tilde{c})$ (the moral hazard condition). Furthermore assume that if $u(c) - D(\theta) = v(\tilde{c})$, $\forall\theta \Rightarrow u'(c) < v'(\tilde{c})$ and $u'(c+x) < v'(\tilde{c}+x)$ $\forall x \geq 0$.

2.4.3 Maximizing social welfare

Assume that all individuals with disability below θ_{NT} and θ_T, respectively, work in the optimal program. Social welfare in the continuous model is thus given by :

$$\int_0^{\theta_{NT}} \left[(1-p(\theta))\left(u(c_W) - D(\theta)\right)\right] dF(\theta) + \int_{\theta_{NT}}^{1} \left[(1-p(\theta))\left(v(c_N)\right)\right] dF(\theta) + \\ + \int_0^{\theta_T} \left[p(\theta)\left(u\left(c_W^T\right) - D(\theta)\right)\right] dF(\theta) + \int_{\theta_T}^{1} \left[p(\theta)\left(v\left(c_N^T\right)\right)\right] dF(\theta) \tag{2.12}$$

and the corresponding resource constraint is:

$$\begin{aligned} &\int_0^{\theta_{NT}} \left[(1-p(\theta))(c_W-1)\right] dF(\theta) + \int_{\theta_{NT}}^1 \left[(1-p(\theta)) c_N\right] dF(\theta) + \\ &+\int_0^{\theta_T} \left[p(\theta)\left(c_W^T-1\right)\right] dF(\theta) + \int_{\theta_T}^1 \left[p(\theta) c_N^T\right] dF(\theta) \leq M \end{aligned} \tag{2.13}$$

Finally, the work constraints are the following:

$$u(c_W) - D(\theta) \geq v(c_N) \text{ for } 0 \leq \theta \leq \theta_{NT} \tag{2.14}$$

$$u\left(c_W^T\right) - D(\theta) \geq v\left(c_N^T\right) \text{ for } 0 \leq \theta \leq \theta_T \tag{2.15}$$

Now note that because $D(\theta)$ is an increasing function, it is sufficient to check that the work constraints are fulfilled for the upper limits of θ. Following the same reasoning as in the discrete model, it can be shown that the work constraints are fulfilled with equality, for θ equal to θ_{NT} and θ_T respectively, in the optimal program. The resource constraint obviously binds - no reason to forgo consumption in the model.

The insurance administration wants to maximize social welfare by choosing consumption allocations to the different groups and is restricted in its choice by the resource constraint and the work constraints.

The first order conditions, where μ_j, $j = i, ii, iii$ are the Lagrange multipliers, are the following (note that $dF(\theta) = f(\theta)\, d\theta$):

$$
\begin{gathered}
u'\left(c_W\right)\int\limits_0^{\theta_{NT}}\left(1-p\left(\theta\right)\right)f\left(\theta\right)d\theta+ \\
-\mu_i\int\limits_0^{\theta_{NT}}\left(1-p\left(\theta\right)\right)f\left(\theta\right)d\theta+\mu_{ii}u'\left(c_W\right)=0
\end{gathered}
\tag{2.16}
$$

$$
\begin{gathered}
v'\left(c_N\right)\int\limits_{\theta_{NT}}^{1}\left(1-p\left(\theta\right)\right)f\left(\theta\right)d\theta+ \\
-\mu_i\int\limits_{\theta_{NT}}^{1}\left(1-p\left(\theta\right)\right)f\left(\theta\right)d\theta-\mu_{ii}v'\left(c_N\right)=0
\end{gathered}
\tag{2.17}
$$

$$
\begin{gathered}
u'\left(c_W^T\right)\int\limits_0^{\theta_T}p\left(\theta\right)f\left(\theta\right)d\theta+ \\
-\mu_i\int\limits_0^{\theta_T}p\left(\theta\right)f\left(\theta\right)d\theta+\mu_{iii}u'\left(c_W^T\right)=0
\end{gathered}
\tag{2.18}
$$

$$
\begin{gathered}
v'\left(c_N^T\right)\int\limits_{\theta_T}^{1}p\left(\theta\right)f\left(\theta\right)d\theta+ \\
-\mu_i\int\limits_{\theta_T}^{1}p\left(\theta\right)f\left(\theta\right)d\theta-\mu_{iii}v'\left(c_N^T\right)=0
\end{gathered}
\tag{2.19}
$$

The redistribution principle in the continuous model is established using the first-order conditions (see appendix):

$$
\Omega\frac{1}{u'\left(c_W\right)}+\left(1-\Omega\right)\frac{1}{v'\left(c_N\right)}=\Omega_T\frac{1}{u'\left(c_W^T\right)}+\left(1-\Omega_T\right)\frac{1}{v'\left(c_N^T\right)}
\tag{2.20}
$$

Where

$$\Omega = \frac{\int\limits_0^{\theta_{NT}} (1-p(\theta))f(\theta)d\theta}{\int\limits_0^1 (1-p(\theta))f(\theta)d\theta}$$

and

$$\Omega_T = \frac{\int\limits_0^{\theta_T} p\left(\theta\right) f\left(\theta\right) d\theta}{\int\limits_0^1 p\left(\theta\right) f\left(\theta\right) d\theta}$$

To be able to say something about the relative size of Ω and Ω_T, a relation between θ_{NT} and θ_T has to be established. Notably θ_i, where $i = T, NT$, may be ranked in three different ways namely: $\theta_T = \theta_{NT}$, $\theta_T > \theta_{NT}$, and $\theta_T < \theta_{NT}$, and $\theta_{NT} \geq \theta_T$ yields the following proposition:

Proposition 2.2: $\Omega > \Omega_T$ *when* $\theta_{NT} \geq \theta_T$

Proof. The proof proceeds in four steps. First note that:

$$\Omega = \frac{E\left[\left(1 - p\left(\theta\right)\right)|\theta \leq \theta_{NT}\right] \int\limits_0^{\theta_{NT}} f\left(\theta\right) d\theta}{E\left[\left(1 - p\left(\theta\right)\right)\right] \int\limits_0^1 f\left(\theta\right) d\theta}$$

$$\Omega_T = \frac{E\left[p\left(\theta\right)|\theta \leq \theta_T\right] \int\limits_0^{\theta_T} f\left(\theta\right) d\theta}{E\left[p\left(\theta\right)\right] \int\limits_0^1 f\left(\theta\right) d\theta}$$

Second:

$$\begin{aligned} E\left[\left(1-p\left(\theta\right)\right)|\theta \leq \theta_{NT}\right] &> E\left[\left(1-p\left(\theta\right)\right)\right] \\ &\Rightarrow \frac{E\left[\left(1-p\left(\theta\right)\right)|\theta \leq \theta_{NT}\right]}{E\left[\left(1-p\left(\theta\right)\right)\right]} > 1 \end{aligned}$$

since $(1-p(\theta))$ is greatest at the beginning of the interval. Third:

$$\begin{aligned} E\left[p\left(\theta\right)|\theta \leq \theta_{T}\right] &< E\left[p\left(\theta\right)\right] \\ &\Rightarrow \frac{E\left[p\left(\theta\right)|\theta \leq \theta_{T}\right]}{E\left[p\left(\theta\right)\right]} < 1 \end{aligned}$$

since $p(\theta)$ is greatest at the end of the interval. Fourth:

$$\int_0^{\theta_{NT}} f\left(\theta\right) d\theta \geq \int_0^{\theta_T} f\left(\theta\right) d\theta$$

since $\theta_{NT} \geq \theta_T$, this concludes the proof. ■

Notably $\theta_{NT} < \theta_T$, implies that the relation between Ω and Ω_T cannot be determined with certainty and that any relation is possible ($\Omega \gtreqless \Omega_T$) because the last part of the proof of proposition 2 no longer holds.

2.4.4 Consumption Allocations in the Continuous Model

This section discusses the ranking consumption allocations with the redistribution principle, relation (2.20), as the starting point. As already shown, the relation between the cut-off levels affects the relative size of the weights in (2.20). The main case with $\theta_T < \theta_{NT}$ is analyzed first, and thereafter the other possibilities are discussed.

The main case: θ_{NT} greater than θ_T

Proposition 2.3: *Given $\theta_{NT} > \theta_T$ the consumption allocations may be ranked in three different ways: 1) $c_W^T > c_W \lesseqgtr c_N^T > c_N$, 2) $c_W > c_W^T > c_N^T > c_N$ and 3) $c_W = c_W^T > c_N^T > c_N$*

Proof. The proof follows the same pattern as the proof in section 2. Notably, it is sufficient to exclude that $c_N^T \leq c_N$ to cover all possible rankings. Hence, given the binding work constraints, $\Omega > \Omega_T$, $\theta_{NT} > \theta_T$ and that the redistribution principle is fulfilled with equality in optimum it is obvious that the rankings 1 - 3 cannot be excluded (notably the ranking in 1 is incomplete for the same reasons as in the discrete model). Hence this proof focuses on showing that c_N is never greater than or equal to c_N^T. First, assume $c_N^T = c_N$, implying that $c_W > c_W^T$ (through the binding work constraints). This implies that the left-hand-side (LHS) of (2.20) is greater than the right-hand-side (RHS) since the weight (Ω) on the greatest element (c_W) is greater than the weight (Ω_T) on the second greatest element of the weighted average. In optimum LHS=RHS and thus c_N cannot be equal to c_N^T in optimum. A similar reasoning gives that c_N cannot exceed c_N^T. ■

At this stage something needs to be said about the interpretation and how reasonable these rankings are in terms of the model. Starting with the former, the continuous model, in contrast to the discrete model, opens up for using incentives other than consumptions to induce the tagged individuals to work. By lowering θ_T, i.e. the disability level below which all tagged individuals work, and thereby lowering the disutility of working for the tagged workers, the premium for tagged workers may be eliminated (ranking 3). It may even be optimal to give a "premium" to untagged workers (ranking 2). In these cases no workers (individuals with sufficiently low disutility of working) would apply for a tag, given that it is voluntary to apply for a tag (as in Diamond & Sheshinski (1995)). The continuous model also allows for pure monetary incentives to be used in the same way as in the discrete model (ranking 1). Notably, tagged non-workers always get a premium over untagged non-workers when $\theta_{NT} > \theta_T$, which is a reasonable feature of a disability insurance: the targeted group is made better off by the introduction of a disability insurance, and it is reasonable to assume that this premium reaches, at least a part of, this group since the probability of getting a tag is increasing in disability.

The cut-off levels θ_{NT} and θ_T are determined endogenously in the model and, by solving the optimization problem above for θ_{NT} and θ_T, a relationship between the marginal disutilities $D'(\theta_{NT})$ and $D'(\theta_T)$ may be established (see appendix, relation (A15)). This relation could serve as a robustness check of the rankings, but unfortunately it may only be concluded that none of the rankings give an obvious contradiction of, or are obviously in line with (A15). It might also be the case that ranking 1 holds in a greater number of circumstances because it allows for $(c_W - c_N) \lesseqqgtr \left(c_W^T - c_N^T\right)$ (see appendix).

The Other Possibilities: $\theta_{NT} \leq \theta_T$

A general observation when $\theta_{NT} \leq \theta_T$ is that the work premium is greater for the tagged than for the untagged workers (this is obviously not the case in ranking 2 and 3 above where $(c_W - c_N) > \left(c_W^T - c_N^T\right)$, but might be the case for ranking 1).

Proposition 2.4: *With $\theta_T \geq \theta_{NT}$ the premium for tagged workers is greater than the premium for untagged workers, i.e. $c_W^T - c_N^T > c_W - c_N$*

Proof. $\theta_T \geq \theta_{NT} \Rightarrow D(\theta_T) \geq D(\theta_{NT})$, implying that $u(c_W) - v(c_N) \leq u\left(c_W^T\right) - v\left(c_N^T\right) \Rightarrow v\left(c_N^T\right) - v(c_N) \leq u\left(c_W^T\right) - u(c_W)$ which may be rewritten as:

$$\int_{c_N}^{c_N^T} v'(z)\,dz \leq \int_{c_W}^{c_W^T} u'(z)\,dz$$

Now assume that $c_W^T - c_W = c_N^T - c_N = \Delta$ and note that this implies that:

$$\int_{c_N}^{c_N+\Delta} v'(z)\,dz > \int_{c_W}^{c_W+\Delta} u'(z)\,dz$$

since it is assumed that $u'(c+x) < v'(\tilde{c}+x)$ when $u(c) - D(\theta) = v(\tilde{c})$. Then, it is obvious that $\int_{c_N}^{c_N^T} v'(z)\,dz \leq \int_{c_W}^{c_W^T} u'(z)\,dz$ requires that $c_W^T - c_W > c_N^T - c_N \Leftrightarrow c_W^T - c_N^T > c_W - c_N$. ■

It is likely that when tagged individuals cannot be induced to work by a lower cut-off level than the untagged then the pure monetary incentives to work have to be stronger, explaining the fact that $c_W^T - c_W > c_N^T - c_N$.

θ_{NT} equals θ_T When $\theta_{NT} = \theta_T$ the binding work constraints imply that $u(c_W) - v(c_N) = u\left(c_W^T\right) - v\left(c_N^T\right)$, making this case similar to the discrete case in section 2.2, and to underline this point the consumptions allocations are ranked in the same way as in the discrete model.

Proposition 2.5: *With $\theta_T = \theta_{NT}$ the ranking of the consumption allocations is the same as in the discrete case with two types: $c_W^T > c_W \lesseqqgtr c_N^T > c_N$.*
Proof. The proof of the ranking follows the proof in section 2.2. ■

It is difficult to find clear-cut examples of when $\theta_T = \theta_{NT}$ is optimal, at least without changing the assumptions of the model. For example, assuming that $p(\theta)$ is constant over θ, i.e. eliminating the tag, would be helpful, but obviously contrary to the spirit of the model. Showing that $\theta_T = \theta_{NT}$ leads to a contradiction, e.g. of (A15), is infeasible and thus this case cannot be excluded in spite of the lack of clear-cut examples.

θ_{NT} smaller than θ_T When $\theta_{NT} < \theta_T$ the model opens up for a large number of cases stemming from the fact that the relation between Ω and Ω_T cannot be determined. In most of these cases, however, the results imply unappealing intuition. In all cases but one the tagged non-workers are not given a premium making the disability insurance unappealing for the targeted group - the disabled i.e. those who cannot work. Moreover, in all but one these cases this is combined with a premium for tagged workers, making it appealing for workers but unappealing for non-workers to be tagged (if tagging was voluntary only potential workers would apply). Hence, instead of going through all these unattractive cases, the focus here is on the case that seems realistic, namely that $c_W^T > c_W \lesseqqgtr c_N^T > c_N$.

Proposition 2.6: *If $\theta_{NT} < \theta_T$ and if this implies that $\Omega > \Omega_T$, then the consumption allocations may be ranked as follows: $c_W^T > c_W \lesseqqgtr c_N^T > c_N$*
Tentative proof. [6] The first part follows a similar pattern to the previous proofs, namely:

Note that $\theta_{NT} < \theta_T$ implies $D(\theta_{NT}) < D(\theta_T) \Rightarrow u(c_W) - v(c_N) < u\left(c_W^T\right) - v\left(c_N^T\right)$ and hence $v\left(c_N^T\right) - v(c_N) < u\left(c_W^T\right) - u(c_W)$. Now assume that $c_N^T > c_N$ implying that $v\left(c_N^T\right) - v(c_N) > 0$ and thus that $u\left(c_W^T\right) - u(c_W) > 0 \Rightarrow c_W^T > c_W$. Now consider the redistribution principle (2.20) and

[6] Notably, this "proof" also holds for $c_N^T \leq c_N$, which has been ruled out as unattractive above; hence the labelling "tentative proof".

note that it may be fulfilled with equality as long as $\frac{\Omega}{u'(c_W)} > \frac{\Omega_T}{u'\left(c_W^T\right)}$, which cannot be ruled out as long as $\Omega > \Omega_T$. The relationship between $c_W \lesseqgtr c_N^T$ cannot be determined for the same reasons as in the discrete model. ■

As for $\theta_{NT} = \theta_T$, it is hard to find examples that pinpoint when it is optimal to have $\theta_{NT} < \theta_T$. Obviously, it is first of all required that the number of tagged individuals is greater than the number of untagged individuals in the interval between θ_{NT} and θ_T, and it is likely that $f(\theta)$ needs to be increasing or constant over θ and that once again assuming that $p(\theta)$ is constant over θ would be helpful (see appendix for the effects of these assumptions on (A15)).

2.5 Concluding remarks

This chapter moves beyond the two-type-economy in Parsons (1996). This is done in two ways; by introducing a third type and by allowing for a continuum of types. Obviously, both these alterations complicate matters, but in essence the results remain the same, especially for the three-type model. Individuals should be given incentives to work in line with their ability and this leaves room for being more generous towards the targeted group(s). Still, the composition of these incentives might be somewhat different in the continuous model compared to the discrete models. The continuous model allows for any relation between the incomes of tagged workers and untagged workers, whereas the income of tagged workers is always higher in the discrete models. The tagged in the continuous model may also be given incentives to work by a lower cut-off level; i.e. they may stop working at a lower disability level.

In the continuous model it cannot be ruled out that the cut-off level for the tagged is greater than or equal to the cut-off level of the untagged. This may be seen as unappealing in the sense that the targeted group for the disability insurance (the tagged) is treated with less or equal generosity compared to the untagged in terms of cut-off level. However, it is found that the premium for tagged workers is greater than the premium for untagged workers in this case, possibly outweighing the less generous cut-off level.

There are several questions left for future research, ranging from further digging into the mechanisms of the continuous model, e.g. investigating the effect of an increase in resources and possible cross subsidies between the tagged and untagged, to more policy-oriented questions e.g. whether it is better to have a disability insurance allowing for several degrees of disability or a dichotomous disability insurance, or how the screening probabilities are affected by stricter or more lenient rules, and how this is manifested in the optimal program. Notably, rules, in most countries, are more common as policy instruments than incentives.

2.A Appendix

Two-Type Model

Complementary slackness

The complementary slackness conditions require that;

$\lambda_{ii} \geq 0$, $\lambda_{ii}\left[v\left(c_N\right)-u\left(c_W\right)+D_0\right]=0$ that is $\lambda_{ii}=0$ if $v\left(c_N\right)-u\left(c_W\right)+D_0<0$

$\lambda_{iii} \geq 0$, $\lambda_{iii}\left[v\left(c_N^T\right)-u\left(c_W^T\right)+D_0\right]=0$ that is $\lambda_{iii}=0$ if $v\left(c_N^T\right)-u\left(c_W^T\right)+D_0<0$

1. $\lambda_{ii}=\lambda_{iii}=0$ implies:

$$\begin{aligned} v\left(c_N\right)-u\left(c_W\right)+D_0 &< 0 \\ v\left(c_N^T\right)-u\left(c_W^T\right)+D_0 &< 0 \end{aligned}$$

implying that:

$$\begin{aligned} u'\left(c_W\right) &< v'(c_N) \\ u'\left(c_W^T\right) &< v'(c_N^T) \end{aligned}$$

since $u\left(c\right)-D_0=v(\tilde{c}) \Rightarrow u'\left(c\right)<v'(\tilde{c})$ and $u(c)-D_0<v(c), \forall\ c$

However, the first order conditions yield that:

$$u'\left(c_W\right)=v'(c_N)=u'\left(c_W^T\right)=v'(c_N^T)$$

$\Rightarrow$ contradiction, no solution candidate.

2. $\lambda_{ii}>0, \lambda_{iii}=0$ implies:

$$\begin{aligned} v\left(c_N\right)-u\left(c_W\right)+D_0 &= 0 \\ v\left(c_N^T\right)-u\left(c_W^T\right)+D_0 &< 0 \end{aligned}$$

implying that:

$$\begin{aligned} u'\left(c_W\right) &< v'(c_N) \\ u'\left(c_W^T\right) &< v'(c_N^T) \end{aligned}$$

but the first order conditions give that:

$$\begin{aligned} u'\left(c_W\right) &< v'(c_N) \\ u'\left(c_W^T\right) &= v'(c_N^T) \end{aligned}$$

$\Rightarrow$contradiction, no solution candidate.

3. $\lambda_{ii} = 0, \lambda_{iii} > 0$ implies:

$$\begin{aligned} v\left(c_N\right) - u\left(c_W\right) + D_0 &< 0 \\ v\left(c_N^T\right) - u\left(c_W^T\right) + D_0 &= 0 \end{aligned}$$

again this implies that:

$$\begin{aligned} u'\left(c_W\right) &< v'(c_N) \\ u'\left(c_W^T\right) &< v'(c_N^T) \end{aligned}$$

while the first order conditions imply that

$$\begin{aligned} u'\left(c_W\right) &= v'(c_N) \\ u'\left(c_W^T\right) &< v'(c_N^T) \end{aligned}$$

$\Rightarrow$contradiction, no solution candidate.

4. $\lambda_{ii} > 0, \lambda_{iii} > 0$ implies:

$$\begin{aligned} v\left(c_N\right) - u\left(c_W\right) + D_0 &= 0 \\ v\left(c_N^T\right) - u\left(c_W^T\right) + D_0 &= 0 \end{aligned}$$

implying:

$$\begin{aligned} u'(c_W) &< v'(c_N) \\ u'\left(c_W^T\right) &< v'(c_N^T) \end{aligned}$$

while the first order conditions imply that

$$\begin{aligned} u'(c_W) &< v'(c_N) \\ u'\left(c_W^T\right) &< v'(c_N^T) \end{aligned}$$

$\Rightarrow$ no contradiction, a solution candidate.

Three-Type Model

Complementary slackness

1. Assume that all work constraints are slack i.e. $\lambda_i = 0$ for $i = 2, 3, ..., 10$, which e.g. implies that $u'(c_W) = u'(c_p)$ and $u(c_W) - D_0 > u(c_p) - kD_0$. It is known that $D_0 \geq kD_0$ for all $D_0 > 0$ implying that $u(c_W) > u(c_p)$ which in turn implies that $c_W > c_p$ and thus $u'(c_W) < u'(c_p)$ - contradicting the assumption that $\lambda_i = 0$ for $i = 2, 3, ..., 10$.

2. Assume that all work constraints are binding in the optimal program, i.e. $\lambda_i > 0$ for $i = 2, 3, ..., 10$. Constraints r2, r3 and r8 cannot be fulfilled with equality at the same time. If this were the case $u(c_W) - D_0 = u(c_p) - kD_0$, $u(c_W) - D_0 = v(c_N)$, $u(c_p) - kD_{\bar{\theta}} = v(c_N)$ implying that $u(c_p) - kD_{\bar{\theta}} = u(c_p) - kD_0$. Since it is assumed that $D_{\bar{\theta}} > D_0$ it is obvious that $kD_{\bar{\theta}} > kD_0$, implying that $u(c_p) - kD_{\bar{\theta}} < u(c_p) - kD_0$ and contradicting the assumption that constraints r2,r3 and r8 are fulfilled with equality at the same time. The equivalent reasoning holds for constraints r4, r5 and r9 and constraints r6, r7 and r10.

3. Assume that constraint r3, r5 and r7 are slack and the other constraints are binding, i.e.$\lambda_i > 0$, for $i = 1, 2, 4, 6, 8, 9, 10$ and $\lambda_j = 0$ for $j = 3, 5, 7$. Why would these constraints be slack? Consider constraint

r3. If $u(c_p) - kD_{\bar{\theta}} \geq v(c_N)$ (r8) then $u(c_W) - D_0 > v(c_N)$ (r3) since it is required that $u(c_W) - D_0 \geq u(c_p) - kD_0$ and it is known that $u(c_p) - kD_{\bar{\theta}} < u(c_p) - kD_0$. With the equivalent reasoning it is obvious that constraints r5 and r7 are also slack. Do the other constraints bind under these conditions? Assume that $\lambda_2 = 0$, which implies (through the FOC:s) that $u'(c_W) > u'(c_p)$ and thus that $u(c_W) < u(c_p) \Leftrightarrow u(c_W) - D_0 < u(c_p) - kD_0$ since $D_0 \geq kD_0$, but $\lambda_2 = 0$ implies that $u(c_W) - D_0 > u(c_p) - kD_0$ - a contradiction arises. Thus it must be the case that $\lambda_2 > 0$, i.e. constraint r2 binds. The same reasoning may be applied to constraints r4 and r6. Now assume that $\lambda_8 = 0$ then the FOC:s give that $u'(c_p) > v'(c_N)$ when $u(c_p) - kD_{\bar{\theta}} > v(c_N)$, but the moral hazard condition gives that $u'(c_p) < v'(c_N)$ when $u(c_p) - kD_{\bar{\theta}} = v(c_N)$, and $v(c) > u(c)$ for all c; thus, it must be the case that $u'(c_p) < v'(c_N)$ when $u(c_p) - kD_{\bar{\theta}} > v(c_N)$ - once again a contradiction is reached. Thus $\lambda_8 > 0$, and the same reasoning holds for r9 and r10. A solution candidate for the optimization problem is now found.

Redistribution principle and consumption allocations

For increased comparability assume that $\ell^A > \ell^P = \ell^D = \ell$. First consider the weights in the redistribution principle: ω_a, ω_b, ω_c, ω_g, ω_h, ω_i, $\omega_d = (1 - \omega_a - \omega_g)$, $\omega_e = (1 - \omega_b - \omega_h)$, $\omega_f = (1 - \omega_c - \omega_i)$.

$$\begin{aligned}
\omega_a &= \frac{\varphi_0 \ell^A}{\varphi_0 \ell^A + \varphi_{\bar{\theta}} \ell + \varphi_1 \ell}, \omega_g = \frac{\varphi_1 \ell}{\varphi_0 \ell^A + \varphi_{\bar{\theta}} \ell + \varphi_1 \ell}, \\
\omega_b &= \frac{\pi_0 \ell^A}{\pi_0 \ell^A + \pi_{\bar{\theta}} \ell + \pi_1 \ell}, \omega_h = \frac{\pi_1 \ell}{\pi_0 \ell^A + \pi_{\bar{\theta}} \ell + \pi_1 \ell}, \\
\omega_c &= \frac{p_0 \ell^A}{p_0 \ell^A + p_{\bar{\theta}} \ell + p_1 \ell}, \omega_i = \frac{p_1 \ell}{p_0 \ell^A + p_{\bar{\theta}} \ell + p_1 \ell}
\end{aligned}$$

All weights are positive and may be partially ranked as follows, given the assumption about the screening process: $\omega_a > \omega_b > \omega_c$, $\omega_i > \omega_h > \omega_g$, $\omega_e > \omega_f > \omega_d$, and furthermore it may be shown that $\omega_a > \omega_d > \omega_g$, $\omega_e > \omega_h$, $\omega_i > \omega_f$.

It is known that $c_W > c_p > c_N$, $c_W^P > c_p^P > c_N^P$, $c_W^T > c_p^T > c_N^T$. Now, assume that $c_W = c_W^P = c_W^T$ implying that $c_p = c_p^P = c_p^T$ which in turn implies that $c_N = c_N^P = c_N^T$. Thus$\frac{1}{u'(c_W)} = \frac{1}{u'(c_W^P)} = \frac{1}{u'(c_W^T)} \equiv \alpha$, $\frac{1}{u'(c_p)} = \frac{1}{u'(c_p^P)} = \frac{1}{u'(c_p^T)} \equiv \beta$ and $\frac{1}{v'(c_N)} = \frac{1}{v'(c_N^P)} = \frac{1}{v'(c_N^T)} \equiv \gamma$. The redistribution principle may now be written (note that the relation between the groups is not specified):

$$\omega_a\alpha + \omega_g\gamma + \omega_d\beta \lesseqgtr \omega_b\alpha + \omega_h\gamma + \omega_e\beta \lesseqgtr \omega_c\alpha + \omega_i\gamma + \omega_f\beta$$

In the optimal program the redistribution principle is fulfilled with equality. Is this the case when the consumption allocations are equalized over work-status? First note that $\alpha > \beta > \gamma$, and that the left-hand side can be rewritten as:

$$\begin{aligned}
\omega_a\alpha + \omega_g\gamma + \omega_d\beta &= \\
(1 - \omega_g - \omega_N)\,\alpha + \omega_g\gamma + \omega_d\beta &= \\
\alpha - \alpha\omega_g - \alpha\omega_N + \omega_g\gamma + \omega_d\beta &= \\
&\alpha + \omega_g\,(\gamma - \alpha) + \omega_d\,(\beta - \alpha)
\end{aligned}$$

The middle expression of the redistribution principle is:

$$\begin{aligned}
\omega_b\alpha + \omega_h\gamma + \omega_e\beta &= \\
(1 - \omega_h - \omega_e)\,\alpha + \omega_h\gamma + \omega_e\beta &= \\
\alpha - \alpha\omega_h - \alpha\omega_e + \omega_h\gamma + \omega_e\beta &= \\
&\alpha + \omega_h(\gamma - \alpha) + \omega_e\,(\beta - \alpha)
\end{aligned}$$

and the right-hand side is:

$$\begin{aligned}
\omega_c\alpha + \omega_i\gamma + \omega_f\beta &= \\
(1 - \omega_i - \omega_W)\,\alpha + \omega_i\gamma + \omega_f\beta &= \\
\alpha - \alpha\omega_i - \alpha\omega_W + \omega_i\gamma + \omega_f\beta &= \\
&\alpha + \omega_i\,(\gamma - \alpha) + \omega_f\,(\beta - \alpha)
\end{aligned}$$

Thus:

$$\alpha+\omega_g(\gamma-\alpha)+\omega_d(\beta-\alpha) \lesseqqgtr \alpha+\omega_h(\gamma-\alpha)+\omega_e(\beta-\alpha) \lesseqqgtr \alpha+\omega_i(\gamma-\alpha)+\omega_f(\beta-\alpha)$$

implying

$$\omega_g(\gamma-\alpha)+\omega d(\beta-\alpha) \lesseqqgtr \omega_h(\gamma-\alpha)+\omega_e(\beta-\alpha) \lesseqqgtr \omega_i(\gamma-\alpha)+\omega_f(\beta-\alpha)$$

and $\omega_i > \omega_h > \omega_g \Rightarrow \omega_g(\gamma-\alpha) > \omega_h(\gamma-\alpha) > \omega_i(\gamma-\alpha)$ since $(\gamma-\alpha)$ is a negative number; furthermore $\omega_e > \omega_f > \omega_d \Rightarrow \omega_d(\beta-\alpha) > \omega_f(\beta-\alpha) > \omega_e(\beta-\alpha)$ since $(\beta-\alpha)$ is a negative number. Hence it can be concluded that

$$\omega_a\alpha+\omega_g\gamma+\omega_d\beta > \omega_b\alpha+\omega_h\gamma+\omega_e\beta$$

and

$$\omega_a\alpha+\omega_g\gamma+\omega_d\beta > \omega_c\alpha+\omega_i\gamma+\omega_f\beta$$

What about the relation between the middle and the right-hand side expressions?

Rewrite $\omega_b\alpha+\omega_h\gamma+\omega_e\beta$ as:

$$\begin{aligned}\omega_b\alpha+\omega_h\gamma+(1-\omega_b-\omega_h)\beta &= \\ &\beta+\omega_b(\alpha-\beta)+\omega_h(\gamma-\beta)\end{aligned}$$

and $\omega_c\alpha+\omega_i\gamma+\omega_f\beta$ as:

$$\begin{aligned}\omega_c\alpha+\omega_i\gamma+(1-\omega_c-\omega_i)\beta &= \\ &\beta+\omega_c(\alpha-\beta)+\omega_i(\gamma-\beta)\end{aligned}$$

Thus, comparing the middle expression and the right-hand side it is found that $\omega_b(\alpha-\beta)+\omega_h(\gamma-\beta) > \omega_c(\alpha-\beta)+\omega_i(\gamma-\beta)$ since $\omega_b(\alpha-\beta) > \omega_c(\alpha-\beta)$ ($\omega_b > \omega_c$ and $(\alpha-\beta)$ is positive) and $\omega_h(\gamma-\beta) > \omega_i(\gamma-\beta)$ ($\omega_i > \omega_h$ and $(\gamma-\beta)$ is negative). Hence,

$$\omega_a\alpha+\omega_g\gamma+\omega_d\beta > \omega_b\alpha+\omega_h\gamma+\omega_e\beta > \omega_c\alpha+\omega_i\gamma+\omega_f\beta$$

when the consumption allocations are the same for the same work status. This situation is not optimal since the redistribution principle is not fulfilled with equality.

Continuous Model

Redistribution Principle

Using (2.16) and (2.17) it is found that:

$$\mu_{ii} = \frac{\mu_i \int_0^{\theta_{NT}} (1-p(\theta)) f(\theta) d\theta - u'(c_W) \int_0^{\theta_{NT}} (1-p(\theta)) f(\theta) d\theta}{u'(c_W)} \tag{A1}$$

$$\mu_{ii} = \frac{v'(c_N) \int_{\theta_{NT}}^{1} (1-p(\theta)) f(\theta) d\theta - \mu_i \int_{\theta_{NT}}^{1} (1-p(\theta)) f(\theta) d\theta}{v'(c_N)} \tag{A2}$$

Set (A1) =(A2) and solve for μ_i:

$$\mu_i = \frac{v'(c_N) u'(c_W) \int_0^1 (1-p(\theta)) f(\theta) d\theta}{v'(c_N) \int_0^{\theta_{NT}} (1-p(\theta)) f(\theta) d\theta + u'(c_W) \int_{\theta_{NT}}^{1} (1-p(\theta)) f(\theta) d\theta} \tag{A3}$$

Using (2.18) and (2.19) it is found that:

$$\mu_{iii} = \frac{\mu_i \int_0^{\theta_T} p(\theta) f(\theta) d\theta - u'\left(c_W^T\right) \int_0^{\theta_T} p(\theta) f(\theta) d\theta}{u'\left(c_W^T\right)} \tag{A4}$$

$$\mu_{iii} = \frac{v'\left(c_N^T\right) \int_{\theta_T}^{1} p(\theta) f(\theta) d\theta - \mu_i \int_{\theta_T}^{1} p(\theta) f(\theta) d\theta}{v'\left(c_N^T\right)} \tag{A5}$$

Set (A4) =(A5) and solve for μ_i:

$$\mu_i = \frac{v'\left(c_N^T\right) u'\left(c_W^T\right) \int\limits_0^1 p\left(\theta\right) f\left(\theta\right) d\theta}{v'\left(c_N^T\right) \int\limits_0^{\theta_T} p\left(\theta\right) f\left(\theta\right) d\theta + u'\left(c_W^T\right) \int\limits_{\theta_T}^1 p\left(\theta\right) f\left(\theta\right) d\theta} \tag{A6}$$

Redistribution principle: set (A3) =(A6) and simplify:

$$\begin{aligned}
&\frac{1}{u'\left(c_W\right)} \frac{\int\limits_0^{\theta_{NT}} \left(1 - p\left(\theta\right)\right) f\left(\theta\right) d\theta}{\int\limits_0^1 \left(1 - p\left(\theta\right)\right) f\left(\theta\right) d\theta} + \frac{1}{v'\left(c_N\right)} \frac{\int\limits_{\theta_{NT}}^1 \left(1 - p\left(\theta\right)\right) f\left(\theta\right) d\theta}{\int\limits_0^1 \left(1 - p\left(\theta\right)\right) f\left(\theta\right) d\theta} = \\
&= \frac{1}{u'\left(c_W^T\right)} \frac{\int\limits_0^{\theta_T} p\left(\theta\right) f\left(\theta\right) d\theta}{\int\limits_0^1 p\left(\theta\right) f\left(\theta\right) d\theta} + \frac{1}{v'\left(c_N^T\right)} \frac{\int\limits_{\theta_T}^1 p\left(\theta\right) f\left(\theta\right) d\theta}{\int\limits_0^1 p\left(\theta\right) f\left(\theta\right) d\theta}
\end{aligned} \tag{A7}$$

The relationship between $D'\left(\theta_{NT}\right)$ and $D'\left(\theta_T\right)$

Maximizing the Lagrangian for the continuous model over θ_{NT} and θ_T yields the following first order conditions:

$$\begin{gathered}
u\left(c_W\right)\left(1 - p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right) - D\left(\theta_{NT}\right)\left(1 - p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right) + \\
-v\left(c_N\right)\left(1 - p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right) + \\
-\mu_i c_W \left(1 - p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right) + \mu_i c_N \left(1 - p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right) + \\
+\mu_i \left(1 - p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right) - \mu_{ii} D'\left(\theta_{NT}\right) = 0
\end{gathered} \tag{A8}$$

$$
\begin{gathered}
u\left(c_W^T\right) p\left(\theta_T\right) f\left(\theta_T\right) - D\left(\theta_T\right) p\left(\theta_T\right) f\left(\theta_T\right) - v\left(c_N^T\right) p\left(\theta_T\right) f\left(\theta_T\right) + \\
-\mu_i c_W^T p\left(\theta_T\right) f\left(\theta_T\right) + \mu_i c_N^T p\left(\theta_T\right) f\left(\theta_T\right) + \mu_i p\left(\theta_T\right) f\left(\theta_T\right) + \\
-\mu_{iii} D'\left(\theta_T\right) = 0
\end{gathered} \tag{A9}
$$

From (A8) it can be concluded that:

$$
\mu_i = \frac{u\left(c_W\right) - D\left(\theta_{NT}\right) - v\left(c_N\right)}{\left(c_W - 1 - c_N\right)} - \frac{\mu_{ii} D'\left(\theta_{NT}\right)}{\left(1 - p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right)\left(c_W - 1 - c_N\right)} \tag{A10}
$$

and (A9) gives that:

$$
\mu_i = \frac{u\left(c_W^T\right) - D\left(\theta_T\right) - v\left(c_N^T\right)}{\left(c_W^T - 1 - c_N^T\right)} - \frac{\mu_{iii} D'\left(\theta_T\right)}{p\left(\theta_T\right) f\left(\theta_T\right)\left(c_W^T - 1 - c_N^T\right)} \tag{A11}
$$

The work constraints are binding in the optimal program, implying that $u\left(c_W\right) - D\left(\theta_{NT}\right) - v\left(c_N\right) = 0$ and $u\left(c_W^T\right) - D\left(\theta_T\right) - v\left(c_N^T\right) = 0$. Hence in optimum the relationship between θ_{NT} and θ_T is determined by the following equality:

$$
\frac{\mu_{ii} D'\left(\theta_{NT}\right)}{\left(1 - p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right)\left(c_W - 1 - c_N\right)} = \frac{\mu_{iii} D'\left(\theta_T\right)}{p\left(\theta_T\right) f\left(\theta_T\right)\left(c_W^T - 1 - c_N^T\right)} \tag{A12}
$$

Determining μ_{ii} and μ_{iii} Insert (A10) (after eliminating the part that becomes zero in optimum) into (A2):

$$\mu_{ii} = \frac{v'(c_N)\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)\,d\theta - \left[-\frac{\mu_{ii}D'(\theta_{NT})}{(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)}\right]\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)\,d\theta}{v'(c_N)}$$

$$\Rightarrow \mu_{ii} - \frac{\mu_{ii}D'(\theta_{NT})\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)\,d\theta}{v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)} = \int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)\,d\theta$$

$$\Rightarrow \mu_{ii}\left(\frac{\begin{array}{c} v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)+ \\ -D'(\theta_{NT})\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)\,d\theta \end{array}}{v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)}\right) = \int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)\,d\theta$$

$$\Rightarrow \mu_{ii} = \frac{\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)\,d\theta v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)}{v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)-D'(\theta_{NT})\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)\,d\theta} \tag{A13}$$

Insert (A10) (after eliminating the part that becomes zero in optimum) into (A5) and substitute the μ_{ii} with (A13):

$$\mu_{iii} = \frac{v'\left(c_N^T\right)\int\limits_{\theta_T}^{1} p\left(\theta\right) f\left(\theta\right) d\theta - \left[-\frac{\mu_{ii} D'(\theta_{NT})}{(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)}\right]\int\limits_{\theta_T}^{1} p\left(\theta\right) f\left(\theta\right) d\theta}{v'\left(c_N^T\right)} =$$

$$= \int\limits_{\theta_T}^{1} p\left(\theta\right) f\left(\theta\right) d\theta +$$

$$+ \frac{\left[\frac{\int\limits_{\theta_{NT}}^{1} (1-p(\theta))f(\theta)d\theta v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)}{v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)-D'(\theta_{NT})\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta}\right] D'\left(\theta_{NT}\right)\int\limits_{\theta_T}^{1} p\left(\theta\right) f\left(\theta\right) d\theta}{v'\left(c_N^T\right)\left(1-p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right)\left(c_W-1-c_N\right)} \Rightarrow$$

$$\mu_{iii} = \int\limits_{\theta_T}^{1} p\left(\theta\right) f\left(\theta\right) d\theta +$$

$$+ \frac{v'\left(c_N\right) D'\left(\theta_{NT}\right)\int\limits_{\theta_{NT}}^{1}\left(1-p\left(\theta\right)\right) f\left(\theta\right) d\theta \int\limits_{\theta_T}^{1} p\left(\theta\right) f\left(\theta\right) d\theta}{\begin{array}{c} v'\left(c_N^T\right) v'\left(c_N\right)\left(1-p\left(\theta_{NT}\right)\right) f\left(\theta_{NT}\right)\left(c_W-1-c_N\right) + \\ -v'\left(c_N^T\right) D'\left(\theta_{NT}\right)\int\limits_{\theta_{NT}}^{1}\left(1-p\left(\theta\right)\right) f\left(\theta\right) d\theta \end{array}} \tag{A14}$$

Insert (A14) and (A13) into (A12):

$$D'(\theta_{NT})\frac{\left[\dfrac{\int_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)}{v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)-D'(\theta_{NT})\int_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta}\right]}{(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)}=$$

$$D'(\theta_T)\frac{\left[\int_{\theta_T}^{1}p(\theta)f(\theta)d\theta+\dfrac{v'(c_N)D'(\theta_{NT})\int_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta\int_{\theta_T}^{1}p(\theta)f(\theta)d\theta}{v'\left(c_N^T\right)v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)-v'\left(c_N^T\right)D'(\theta_{NT})\int_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta}\right]}{p(\theta_T)f(\theta_T)(c_W^T-1-c_N^T)}$$

$\Longleftrightarrow$

$$\frac{v'(c_N)D'(\theta_{NT})\int_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta}{\left(v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)-D'(\theta_{NT})\int_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta\right)}=$$

$$=\frac{D'(\theta_T)\int_{\theta_T}^{1}p(\theta)f(\theta)d\theta\begin{pmatrix}v'\left(c_N^T\right)v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)+\\-v'\left(c_N^T\right)D'(\theta_{NT})\int_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta\end{pmatrix}}{v'\left(c_N^T\right)p(\theta_T)f(\theta_T)\left(c_W^T-1-c_N^T\right)\begin{pmatrix}v'(c_N)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)+\\-D'(\theta_{NT})\int_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta\end{pmatrix}}+$$

$$+\frac{v'(c_N)D'(\theta_T)D'(\theta_{NT})\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta\int\limits_{\theta_T}^{1}p(\theta)f(\theta)d\theta}{v'\left(c_N^T\right)p(\theta_T)f(\theta_T)\left(c_W^T-1-c_N^T\right)\left(\begin{array}{c} v'\left(c_N\right)\left(1-p\left(\theta_{NT}\right)\right)f\left(\theta_{NT}\right)\left(c_W-1-c_N\right)+ \\ -D'\left(\theta_{NT}\right)\int\limits_{\theta_{NT}}^{1}\left(1-p\left(\theta\right)\right)f\left(\theta\right)d\theta \end{array}\right)}$$

$$\Longleftrightarrow$$

$$v'\left(c_N\right)D'\left(\theta_{NT}\right)\int\limits_{\theta_{NT}}^{1}\left(1-p\left(\theta\right)\right)f\left(\theta\right)d\theta=$$

$$=\frac{D'(\theta_T)v'(c_N)v'\left(c_N^T\right)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)\int\limits_{\theta_T}^{1}p(\theta)f(\theta)d\theta}{v'\left(c_N^T\right)p(\theta_T)f(\theta_T)\left(c_W^T-1-c_N^T\right)}+$$

$$-\frac{D'(\theta_T)D'(\theta_{NT})v'\left(c_N^T\right)\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta\int\limits_{\theta_T}^{1}p(\theta)f(\theta)d\theta}{v'\left(c_N^T\right)p(\theta_T)f(\theta_T)\left(c_W^T-1-c_N^T\right)}+$$

$$+\frac{D'(\theta_T)D'(\theta_{NT})v'(c_N)\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta\int\limits_{\theta_T}^{1}p(\theta)f(\theta)d\theta}{v'\left(c_N^T\right)p(\theta_T)f(\theta_T)\left(c_W^T-1-c_N^T\right)}$$

$$\Longleftrightarrow$$

$$D'\left(\theta_{NT}\right)\frac{\begin{array}{c}v'\left(c_N\right)v'\left(c_N^T\right)p\left(\theta_T\right)f\left(\theta_T\right)\left(c_W^T-1-c_N^T\right)+\\ D'\left(\theta_T\right)v'\left(c_N^T\right)\int\limits_{\theta_T}^{1}p\left(\theta\right)f\left(\theta\right)d\theta+\\ -D'\left(\theta_T\right)v'\left(c_N\right)\int\limits_{\theta_T}^{1}p\left(\theta\right)f\left(\theta\right)d\theta\end{array}}{v'(c_N)v'\left(c_N^T\right)p(\theta_T)f(\theta_T)\left(c_W^T-1-c_N^T\right)}=$$

$$=D'\left(\theta_T\right)\frac{(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)\int\limits_{\theta_T}^{1}p(\theta)f(\theta)d\theta}{p(\theta_T)f(\theta_T)\left(c_W^T-1-c_N^T\right)\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta}$$

$$\Longleftrightarrow$$

$$D'\left(\theta_{NT}\right)=$$

$$=D'\left(\theta_T\right)\frac{\int\limits_{\theta_T}^{1}p(\theta)f(\theta)d\theta}{\int\limits_{\theta_{NT}}^{1}(1-p(\theta))f(\theta)d\theta}\frac{v'(c_N)v'\left(c_N^T\right)(1-p(\theta_{NT}))f(\theta_{NT})(c_W-1-c_N)}{\begin{array}{c}v'\left(c_N\right)v'\left(c_N^T\right)p\left(\theta_T\right)f\left(\theta_T\right)\left(c_W^T-1-c_N^T\right)+\\ +D'\left(\theta_T\right)\left(\begin{array}{c}v'\left(c_N^T\right)\int\limits_{\theta_T}^{1}p\left(\theta\right)f\left(\theta\right)d\theta+\\ -v'\left(c_N\right)\int\limits_{\theta_T}^{1}p\left(\theta\right)f\left(\theta\right)d\theta\end{array}\right)\end{array}}$$

Which may be rewritten as:

$$D'(\theta_{NT}) = D'(\theta_T) \frac{\int_{\theta_T}^{1} p(\theta) f(\theta) \, d\theta \, (1 - p(\theta_{NT}))}{\int_{\theta_{NT}}^{1} (1 - p(\theta)) f(\theta) \, d\theta p(\theta_T)} \frac{f(\theta_{NT})}{f(\theta_T)} \frac{(c_W - 1 - c_N)}{(c_W^T - 1 - c_N^T) + \Phi} \tag{A15}$$

where

$$\Phi = \frac{D'(\theta_T) \int_{\theta_T}^{1} p(\theta) f(\theta) \, d\theta \left(v'\left(c_N^T\right) - v'(c_N)\right)}{v'(c_N) v'(c_N^T) p(\theta_T) f(\theta_T)}$$

Notably $\Phi < 0$ when $c_N^T > c_N$

Consumption Allocations and Relation (A15)

$\Omega > \Omega_T$ $(\theta_{NT} > \theta_T)$ 1) $c_W^T > c_W \lesseqqgtr c_N^T > c_N$, 2) $c_W > c_W^T > c_N^T > c_N$ and 3) $c_W = c_W^T > c_N^T > c_N$

Check that the rankings 1-3 do not contradict (A15) under the assumption $\theta_{NT} > \theta_T$. First, note that $\Phi < 0$ in (A15) since $c_N^T > c_N$, and it is obvious that the first term on the right-hand-side (RHS) of (A15) is greater than one i.e.

$$\frac{\int_{\theta_T}^{1} p(\theta) f(\theta) \, d\theta \, (1 - p(\theta_{NT}))}{\int_{\theta_{NT}}^{1} (1 - p(\theta)) f(\theta) \, d\theta p(\theta_T)} > 1$$

If the second and third terms of RHS, $\frac{f(\theta_{NT})}{f(\theta_T)}$ and $\frac{(c_W - 1 - c_N)}{\left(c_W^T - 1 - c_N^T\right) + \Phi}$, are also greater than one then there is no contradiction since this implies $D'(\theta_{NT}) > D'(\theta_T)$ (implying $\theta_{NT} > \theta_T$) with certainty. Now note that $(c_W - 1 - c_N) > \left(c_W^T - 1 - c_N^T\right)$ for rankings 2) and 3) (possibly for 1) as well). Both these sums are negative and thus $(c_W - 1 - c_N) > \left(c_W^T - 1 - c_N^T\right) + \Phi$ when $\Phi < 0$, implying that:$|(c_W - 1 - c_N)| < \left|\left(c_W^T - 1 - c_N^T\right) + \Phi\right| \Rightarrow \frac{(c_W-1-c_N)}{\left(c_W^T-1-c_N^T\right)+\Phi} < 1$

Furthermore, $\frac{f(\theta_{NT})}{f(\theta_T)}$ is greater than (or equal to) one unless $f(\theta)$ is decreasing in θ. Even $f(\theta)$ is decreasing and $\frac{(c_W-1-c_N)}{\left(c_W^T-1-c_N^T\right)+\Phi} < 1$ it cannot be excluded that RHS is greater than one, making it difficult to draw any conclusions without more information.

Concerning ranking 1), it is also possible that $(c_W - 1 - c_N) \leq \left(c_W^T - 1 - c_N^T\right)$ and this would ensure that $\frac{(c_W-1-c_N)}{\left(c_W^T-1-c_N^T\right)+\Phi} > 1$ and that the whole RHS is greater than one unless $f(\theta)$ is decreasing at a sufficient rate.

$\Omega > \Omega_T$ $(\theta_{NT} = \theta_T)$ $\theta_T = \theta_{NT}$ implies that $D'(\theta_{NT}) = D'(\theta_T)$ and that $f(\theta_{NT}) = f(\theta_T)$ and hence that (A15) becomes:

$$1 = \frac{\int_{\theta_T}^{1} p(\theta) f(\theta) d\theta \left(1 - p(\theta_T)\right)}{\int_{\theta_T}^{1} (1 - p(\theta)) f(\theta) d\theta p(\theta_T)} \frac{(c_W - 1 - c_N)}{(c_W^T - 1 - c_N^T) + \Phi}$$

$\Omega > \Omega_T$ $(\theta_{NT} < \theta_T)$

$$D'(\theta_{NT}) = D'(\theta_T) \frac{\int_{\theta_T}^{1} p(\theta) f(\theta) d\theta \left(1 - p(\theta_{NT})\right)}{\int_{\theta_{NT}}^{1} (1 - p(\theta)) f(\theta) d\theta p(\theta_T)} \frac{f(\theta_{NT})}{f(\theta_T)} \frac{(c_W - 1 - c_N)}{(c_W^T - 1 - c_N^T) + \Phi}$$

for $D'(\theta_{NT}) < D'(\theta_T)$ it is needed that:

$$\frac{\int_{\theta_T}^{1} p(\theta) f(\theta) d\theta \left(1 - p(\theta_{NT})\right)}{\int_{\theta_{NT}}^{1} (1 - p(\theta)) f(\theta) d\theta p(\theta_T)} \frac{f(\theta_{NT})}{f(\theta_T)} \frac{(c_W - 1 - c_N)}{(c_W^T - 1 - c_N^T) + \Phi} < 1$$

and to ensure this it is necessary that e.g. $\frac{f(\theta_{NT})}{f(\theta_T)} < 1$ i.e. $f(\theta_T) > f(\theta_{NT})$, $p(\theta)$ is constant over θ and $(c_W - 1 - c_N) < \left(c_W^T - 1 - c_N^T\right) + \Phi$ where $\Phi < 0$ when $c_N^T > c_N$.

Bibliography

Akerlof G A, 1978, "The Economics of Tagging as Applied to Optimal Income Tax, Welfare Programs, and Manpower Planning, American Economic Review 68, 8-19

Benitez-Silva H, Buchinsky M & Rust J, 2004, “How large are the classification errors in the social security disability award process”, NBER Working paper 10219, http://www.nber.org/papers/w10219

Diamond P A & Mirrlees J A, 1978, “A Model of Social Insurance with Variable Retirement”, Journal of Public Economics 10, 295-336

Diamond P A & Sheshinski E, 1995, “Economic Aspects of Optimal Disability Benefits”, Journal of Public Economics 57, 1-23

Nagi S Z, 1969, “Disability and rehabilitation: Legal, clinical and self-concepts and measurement”, Ohio State University Press

Parsons D O, 1996, “Imperfect ‘tagging’ in social insurance programs”, Journal of Public Economics 62, 183-207

Salanié B, 2002, "Optimal Demogrants with Imperfect Tagging", Economic Letters 75, 319-324

Chapter 3

Trilateral Trade and Asset Allocation

Abstract: Chapter 3 uses the basics from the property rights approach to organization (Grossman-Hart-Moore model) to develop a model for trilateral trade transactions. In this transaction a downstream producer produces the final good using inputs from two upstream suppliers. Moreover one of the upstream suppliers needs an input from the other for its production. The optimal way to organize this transaction depends on the characteristics of assets, human capital and investments. The general finding is that it is more demanding to find a unique Pareto optimal organization in the trilateral model than in the bilateral version of the model. In addition, it is found that the starting point for the analysis affects the results, suggesting that choosing suitable starting points is important for applications of the model.

Keywords: Trilateral Trade, Property Rights Approach to Organizations, Partial Integration

JEL Classifications: D23, L23

3.1 Introduction

The property rights approach to organization, also known as the Grossman-Hart-Moore model, was developed in the seminal articles by Grossman &

Hart (1986) and Hart & Moore (1990). In a later book Hart (1995) presents the basic ideas of the property rights approach in a tractable model focusing on bilateral transactions, e.g. buyer-seller relationships (similar to Grossman & Hart (1986)). This model forms the basis for the exposition in this paper and is hereinafter referred to as the *bilateral model*. In particular, this paper extends the bilateral-model to the case of trilateral trade.

In general, the property rights approach to organization (PRA) focuses on the importance of asset ownership for investments in trade relationships (transactions). This is done in a world of incomplete contracting, i.e. a contract cannot specify all possible contingencies and some outcomes may not be verifiable for a third party (e.g. courts). The aim is to find an optimal organizational structure for the analyzed transaction. Each party in the transaction is assumed to own a physical asset and have some human capital. In this setup integration is essentially the acquisition of a trading party's assets, and with this acquisition follows residual control rights over contingencies, concerning the assets, not specified in the contract. Greater control makes the investing party less vulnerable to hold-ups and thus provides incentives for greater investments in the trade relationship. These investments are beneficial for the trading parties and may either be in human capital or in physical assets. The investments may be thought of as modifications that ensure a smoother (more efficient) trade between the trading parties. Hence, the organizational structure that best supports relationship-specific investments is optimal.

Hart & Moore (1990) use these basic features of PRA to analyze the optimal control structure (i.e. ownership structure) of coalitions of agents in a multi-agent and multi-asset setting. The focus is on trade as cooperative projects and the actual structure of the trade is generally not specified.[1] In contrast, this chapter presents a deeper analysis of a specified structure. Moreover, Hart & Moore (1990) analyze worker incentives, while the bilateral model and the model presented here focus on management incentives.

[1] Except in the application to many assets where there is one downstream firm and many upstream firms; see also Bolton & Whinston (1993) for this kind of multi-agent analysis.

This paper focuses on trade relationships characterized by trilateral trade; trilateral trade is a transaction where one downstream party produces the final good and two upstream parties supply inputs for this production. One of the upstream suppliers also supplies the other with an input. It might simplify intuition to think about the following situation: a downstream producer of cars, M, needs both circuit boards and products containing circuit boards, e.g. a travel computer, in their cars. Let S be the producer of circuit boards and A be the producer of travel computers. Now trilateral trade, as depicted here, is the situation where S supplies circuit boards, significantly different types of circuit boards, to M and A, while A supplies travel computers to M. Herein the optimal organization for this kind of transaction is analyzed by extending the bilateral model (as presented by Hart (1995)).

The main conclusion is that finding a unique Pareto optimal organization, for a given set of assumptions, is more demanding in the trilateral model than in the bilateral model. Moreover, the relative productivity of investments, in terms of contribution to total surplus, may be of greater importance in the trilateral model - not least as a tie-breaker between two or more Pareto optimal organizational forms - but may not always be relevant (see section 3.3). In addition, the tendency is that some form of partial integration, if any integration, is optimal in most cases. Full integration is more of an exception.

This paper also addresses a number of modelling issues. It discusses the starting point for the analysis, i.e. the benchmark to compare hypothetical organizational changes with e.g. integration, and the effects of changing the starting point. It is found, quite unsurprisingly, that the starting point matters for the analysis. Besides, the bargaining setup is similar but distinct from the symmetric Nash bargaining solution applied in the bilateral model (see Grossman & Hart (1986) and Hart (1995)). Distinct because it is assumed that the division of surplus is determined by simultaneous bilateral negotiations (as opposed to one common bargaining with all three parties), and similar because the symmetric Nash bargaining solution is applied to these negotiations. In this context it is shown that none of the parties have incentives to deviate from this bargaining outcome. Notably, the bargaining outcome would be different if the division of surplus was decided by the

Shapley value as in Hart & Moore (1990). Hart & Moore (1990), hence, take a cooperative approach to their bargaining problem while a non-cooperative approach is taken in this paper (more about this in section 3.2). Given the bargaining setup used in this chapter, it is found that the incentives for investments in the trilateral model are entirely given by the (best) disagreement points. This contrasts with the bilateral model where the investment incentives are given by a combination of the incentives within the relationship (when the parties trade with each other) and the incentive in the disagreement point (when the parties do not trade with each other).

This chapter is organized as follows: section 3.2 provides the basics of model and section 3.3 presents the analysis and its extensions. Section 3.4 concludes the chapter.

3.2 Basics

A firm M produces a final good that is sold on the market for final goods. For this production M needs inputs from two other firms A and S. Moreover, for A to be able to produce this input, they also need an input from S. Thus S produces inputs for both M and A. Now assume that the production of these two inputs makes use of different parts of S's human capital, such that the cost of producing one is independent of producing the other (separability assumption). Note also that the input produced for A cannot be used by M and *vice versa*.[2] The parties initially own only one asset each, denoted: p_M, p_S and p_A. Figure 3.1 gives a schematic presentation of the trade relationships.

[2] One may think of S as e.g. a computer support company that provides hardware programming to A's machines and IT-applications to M's sales division. The separability assumption creates a situation that is similar to M and A having one supplier each, with the important difference that the same asset p_S is used in both production processes; which would be unnatural if the model was dealing with two suppliers.

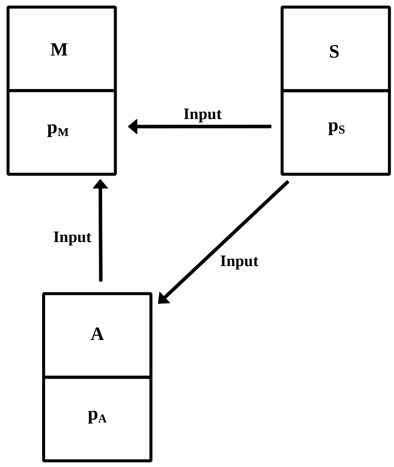

Figure 3.1: The trilateral trade

3.2.1 The Basic Model

The model presented here uses, with appropriate modifications (which will be noted), the basic assumptions made by Hart (1995) and most of the PRA literature.[3] It consists of two periods, period 0 and period 1. The timing of the model is the following (see figure 1.1 in chapter 1): in period 0 each party makes investments in their human capital[4] that are relationship-specific to each relation; M makes the investments μ_A and μ_S, A makes α_M and α_S and S makes σ_A and σ_M. These investments are complementary and observable for all three parties, but not verifiable to outsiders (not enforceable) (cf. Grossman & Hart, 1986, Hart & Moore, 1990, Hart, 1995). The investment cost per unit of investment is assumed to be one, thus μ_A represents both the level and the cost of this investment.

The trade takes place in period 1, and after bargaining S and A are reimbursed for their inputs. One may ask why there is bargaining over the reimbursement; why is the price of the input not contracted on in advance?

[3] Notably, there is no uncertainty about costs and benefits, and no asymmetric information in this model. Moreover, the parties can make correct calculations about the expected return of any action and have unlimited wealth (i.e. assets can be transferred between parties as if they were without cost).

[4] This paper abstracts from investments in physical assets discussed in e.g. Hart (1995:chapter 3).

The reason is uncertainty, due to *unforeseen contingencies*, i.e. contingencies that cannot be contracted on, and to *ex ante* uncertainty about the relevant characteristics of a suitable input. That is, the input cannot be described in a contract and thus not priced in a relevant manner. This uncertainty makes effective long-term contracts infeasible. Nevertheless, the uncertainty is resolved in period 1 and the parties then bargain over the price of the input. The specifics of the bargaining will be returned to below. For future reference note that the reimbursement from M to S is called v, the reimbursement from M to A is called m and the reimbursement from A to S is called y.

Trade Patterns and Payoffs from Trade In the trilateral model eight possible patterns of trade may be realized in period 1, depending on whether agreements are reached or not. In case of disagreement, a party has two options available at all times (not foreclosing a future agreement): either trade with neither of the other two parties (no-trade) and buy/sell on a spot-market for inputs, or reach an agreement with either one of them (partial trade) and buy/sell the other input on the spot-market. These considerations yield the different patterns of trade. The two extreme cases (full trade and no-trade) and one intermediate case (partial trade) are presented below (all trade patterns and resulting payoff structures are found in the appendix). Notably, all the intermediate cases follow the same basic outline.

The benefit from full/trilateral trade for M is denoted $T(\mu_A, \mu_S)$ while the production cost under full trade for A is denoted $K(\alpha_M, \alpha_S)$; both M's benefit and A's cost depend on their relationship-specific investments. S's production cost is separable into the two inputs S produces, the cost for the input to M's production is denoted $C(\sigma_M)$ and the cost of producing A's input is $G(\sigma_A)$. Thus, under full trade all three parties' assets and human capital are available (to all three) rendering the following payoffs:

$$U_M = T(\mu_A, \mu_S) - v - m \quad (3.1)$$

$$U_S = v - C(\sigma_M) + y - G(\sigma_A) \quad (3.2)$$

$$U_A = m - K(\alpha_M, \alpha_S) - y \quad (3.3)$$

As noted all the assets are available to all the parties under full trade, i.e. assets can be combined with all parties' human capital. This is not so in the no-trade case. The no-trade benefits and costs are denoted by $t(\mu_A, \mu_S; P_M)$, $k(\alpha_M, \alpha_S; P_A)$, $c(\sigma_M; P_S)$ and $g(\sigma_A; P_S)$, respectively, where P_i, $i = M, A, S$, denote the assets owned by each party under no-trade. Consequently the allocation of assets matters for the benefit/costs in the no-trade case.[5] Under no-trade the benefit function, $t(\mu_A, \mu_S; P_M)$, reflects that neither S's nor A's human capital is available to M. The cost functions $c(\sigma_M; P_S)$ and $g(\sigma_A; P_S)$ are S's cost of producing generic inputs and reflects that the other parties human capital is not available to S. Similarly, $k(\alpha_M, \alpha_S; P_A)$ is A's cost of producing a generic "A-type" input. Moreover, let $\bar{v}$ be the spot-market price for a generic input (of the type that S could have provided for M), $\bar{m}$ be the spot-market price for a generic "A-type" input and $\bar{y}$ be the spot-market price for the input that A needs for its production. This yields the following payoffs:

$$\begin{aligned}
u_M &= t(\mu_A, \mu_S; P_M) - \bar{v} - \bar{m} \\
u_S &= \bar{v} - c(\sigma_M; P_S) + \bar{y} - g(\sigma_A; P_S) \\
u_A &= \bar{m} - k(\alpha_M, \alpha_S; P_A) - \bar{y}
\end{aligned}$$

In the intermediate cases trade is realized between some but not all the parties, e.g. it could be that A trades with both S and M, but M does not trade with S - M's benefit from trade is then called $T_A(\mu_A, \mu_S; P_M)$ and depends not only on the relationship-specific investment but also on the assets owned by M. Here T_A indicates that A's human capital is available

[5] To limit the number of feasible ownership structures cases where a party does not own its own asset but owns other assets are abstracted from. Notably, it is found in Hart & Moore (1990) that an agent who does not own his "own" asset should not own any other assets either. Thus, the excluded configurations are found suboptimal by Hart & Moore (1990), and also suggested by Hart (1995).

The ownership configurations dealt with in this paper are the following:

$P_M = \begin{matrix} \{p_M, p_S, p_A\}, \{p_M, p_S\} \\ \{p_M, p_A\}, \{p_M\}, \varnothing \end{matrix}$, $P_S = \begin{matrix} \{p_S, p_M, p_A\}, \{p_S, p_M\} \\ \{p_S, p_A\}, \{p_S\}, \varnothing \end{matrix}$,

$P_A = \begin{matrix} \{p_A, p_S, p_M\}, \{p_A, p_S\} \\ \{p_A, p_M\}, \{p_A\}, \varnothing \end{matrix}$

to M, but S's human capital is not. As mentioned above, the payoffs from all the partial trades are similarly configured, e.g. $K_S(\alpha_M, \alpha_S; P_S)$ is A's production cost when only trading with S. An example of partial trade is that M & A and A & S trade, but not M & S. This situation yields the following payoffs:

$$\begin{aligned} U_M &= T_A(\mu_A, \mu_S; P_M) - \bar{v} - m \\ U_S &= \bar{v} - c(\sigma_M; P_S) + y - G(\sigma_A) \\ U_A &= m - K(\alpha_M, \alpha_S) - y \end{aligned}$$

It is important to note that the investments are beneficial in all the trading patterns but to different degrees. The positive effect of investments depends on the assets owned and number of parties involved in trade.

Concerning the functional form of the benefit functions and cost functions, it is assumed, similar to Hart (1995), that $T(\cdot)$ is strictly concave in both its arguments and that $K(\cdot)$,$C(\cdot)$ and $G(\cdot)$ are strictly convex in all their arguments (i.e. in the investments). Furthermore, $T_i(\cdot)$, $i = A, S$ and $t(\cdot)$ are concave in both their arguments, while $K_i(\cdot)$, $i = M, S$, $k(\cdot)$, $c(\cdot)$ and $g(\cdot)$ are convex in all their arguments. For formal statements of these assumptions see appendix ((A1)-(A10)).

Trade Surplus and Marginal Benefits Assume, as in the basic PRA setup, that trade is beneficial for all the parties reflecting that the investments are relationship-specific. In line with this it is natural to assume, in the extension to the trilateral model, that full trade is the most beneficial form of trade i.e. produces the greatest total surplus. Furthermore, partial trade is also beneficial but to a lesser extent and assume that the total surplus, under partial trade, is growing in the number of trades.[6] These assumptions give a partial ranking of the total surplus from trade (see appendix (A11)); partial since the relation between two surpluses with the same num-

[6]Example of number of trades: the partial trade with M and S trading and no other trades entails one trade, while the partial trade in the example above entails two trades, and full trade entails, obviously, three trades.

ber of trades cannot be determined. However, it is also assumed that the difference between two surpluses with the same number of trades is smaller than the difference between two surpluses with a different number of trades (see appendix ($A11.a$)). This assumption strengthens the notion that more trade is better, because it implies that it is always better to move e.g. from one trade to three trades (i.e. full trade), than from one trade to two trades independently of the initial trade pattern.

The assumptions about marginal benefits from investments, here called *marginal conditions*[7] (see appendix (A12)-(A17)), reflect that investments are more valuable in trade than in no-trade. Moreover, they show that the relationship-specific investments are at least partly specific to the other parties' assets. The interpretation of these assumptions is here exemplified by M's investment in the relationship with S.

The marginal benefit of the investment, μ_S, is at least as high or higher (under full trade) in all types of trade than in no-trade irrespective of ownership. In all forms of trade the marginal benefit of the investment is growing in the number of assets that M owns. Furthermore, the marginal benefit of an investment in the relationship with S when trading with S is at least as high as the marginal benefit of this investment when trading with A, and this holds for all ownership structures. Let this be the *trade effect.* Moreover, the investment is at least partly specific to S's asset. For an equal number of assets owned, the marginal benefit of μ_S, when trading with S and owning p_S, is greater or equal to the marginal benefit when not owning p_S. Let this be the *asset effect.* The asset effect is weaker than the trade effect; i.e. the marginal benefit of μ_S when trading with S is greater or equal to the marginal benefit when trading with A, even if M owns p_S in the latter case. The marginal conditions for the other investments may be interpreted in a similar way.[8] [9]

[7]Notably, the term *marginal conditions* is here used in a non-standard way i.e. it is not used for the FOCs from a maximization problem but for assumptions about the marginal benefit from investments in different settings.

[8]These assumptions are, in spirit, the same as assumption 6 in Hart & Moore (1990) and assumptions 2.2 and 2.3 in Hart (1995). The distinctions of a trade effect and an asset effect are new and seem natural in a trilateral setting.

[9]Notably, the trade effect expresses itself differently in S's marginal conditions than

If the parties decide to trade with each other, which they do in equilibrium, there is, by assumption, a surplus from trade to be divided among them. The division of this surplus is decided by bargaining.

3.2.2 The Bargaining

All parties negotiate bilaterally at the same time, with rational expectations about the outcome in the other negotiations. The negotiations are bilateral since it is assumed that an agreement between two parties cannot be conditioned on the participation of the third party; i.e. M negotiates with S and A at the same time but not at the same "table". Notably, this negotiation differs from a setup where all three parties are involved in a "common" and cooperative negotiation; i.e. all three parties sit at the same table at the same time. Such a negotiation is characterized by *pure bargaining* i.e. the only possible outcomes are complete cooperation (trilateral trade) or complete breakdown of cooperation (no-trade) (cf. e.g. Hart & Mas-Colell, 1996). Applying the symmetric Nash bargaining solution to this cooperative bargaining, each party receives one third of the surplus from trade.[10]

The setup of the negotiations, with three parties involved in simultaneous bilateral bargaining, entails a specification of available options as well as payoffs for each party. This implies a non-cooperative approach to bargaining contrary to Hart & Moore (1990), who take a cooperative approach by applying the Shapley value to their bargaining problem. In an attempt to keep it simple the symmetric Nash Bargaining solution (good textbook references are Osborne & Rubinstein (1990) and Muthoo (1999)) is applied to the bilateral

in M's and A's since S's payoff function is separable in the arguments connected to M and A respectively. It becomes the basic effect that a relationship-specific investment is, on the margin, more beneficial under bilateral trade than under no-trade. Moreover, the trade effect does not vary with asset allocation and, therefore, does not affect the optimal organization decision.

[10] Applying the Nash bargaining solution to the common negotiation gives one third of the surplus from trade to each party i.e.:

$$\frac{1}{3}\begin{bmatrix} (T(\cdot)-t(\cdot))-(G(\cdot)-g(\cdot))+ \\ -(K(\cdot)-k(\cdot))-(C(\cdot)-c(\cdot)) \end{bmatrix}$$

This is also the division given by the Shapley value in a situation when either no-trade or trilateral trade is the only feasible outcome i.e. when only forming the grand coalition implies a marginal contribution of value.

negotiations. This approach yields a different division compared to applying Shapley's principle to the problem.[11] Although Nash bargaining is chosen to make it simple, it is not an unreasonable choice for a non-cooperative setup because it is the equilibria (at the limit) of some non-cooperative setups - most notably in Rubinstein's alternating offers model when the time between offers approaches zero (Rubinstein, 1982, Binmore, 1987). To further strengthen the appeal of using the Nash bargaining solution, it can be shown that, given the bargaining outcomes, none of the players want to deviate from this outcome, i.e. deviate from trilateral trade to partial trade or no-trade (see the Bargaining/Trade game in the appendix).

Nash bargaining implies *ex post* efficiency i.e. that the parties will renegotiate until an efficient agreement is reached (cf. e.g. Osborne & Rubinstein, 1990). The assumption of *ex post* efficiency is found in most contributions to the literature dealing with the PRA (most notably in Grossman & Hart (1986), Hart & Moore (1990) and Hart (1995)). This assumption has been regarded (cf. Hart & Moore, 2008, Matouschek, 2004) as being unsuitable for negotiations between e.g. managers since they often have private information (for example about their effort). This is, of course, a relevant objection, but (using the standard argument) the focus here is on *ex ante* inefficiencies and hence *ex post* inefficiencies are abstracted from.[12] [13]

Inside Options vs. Outside Options In the symmetric Nash bargaining each party has a disagreement position in case they fail to reach an agreement (at that time). The disagreement positions are characterized throughout this paper as *inside options* (threat points or *status quo* options) i.e. it is the pay-

[11] For future reference note that e.g. M receives:
$\frac{1}{3}(T(\cdot) - K(\cdot) + K_S(\cdot)) - \frac{1}{2}(C(\cdot) - c(\cdot)) +$
$-\frac{2}{3}t(\cdot) + \frac{1}{6}(T_A(\cdot) + T_S(\cdot) - K_M(\cdot) + k(\cdot))$.
when forming subcoalitions (i.e. bilateral trade) implies a marginal contribution.

[12] Hart & Moore (1990) admit that it is a strong assumption to assume *ex post* efficiency, but argue that some effects of *ex post* inefficiencies might be captured by focusing on *ex ante* inefficiencies.

[13] Admittedly a full-fledged model of organizations would probably include both *ex ante* and *ex post* inefficiencies, but keeping as close as possible to previous literature using PRA, this is left for future research.

off that the parties receive when they temporarily disagree as opposed to the payoff they receive when they stop bargaining permanently (*outside option*) (e.g. Muthoo, 1999, De Meza & Lockwood, 1998, Chiu 1998). Does this distinction matter for the analysis? The short answer is that it does. Two contemporaneous articles Chiu (1998) & De Meza & Lockwood (1998), using the *outside option principle*[14], show that the results of the PRA are vulnerable to changes in bargaining assumptions. In particular they show that interpreting the disagreement position as an outside option instead of an inside option changes the results of the PRA.[15] However, both articles contend that the results of the PRA continue to hold if the disagreement positions are inside options. When it comes to the restrictiveness of assuming inside options, Chiu (1998) thinks it is restrictive because people often seek outside options to strengthen their bargaining position. De Meza & Lockwood (1998) state that the restrictiveness depends on the context analyzed; if the analysis deals with a situation where there is a natural inside option, e.g. when the agents are already trading with other parties and will continue to do so throughout the negotiation, then it is obviously not restrictive.

As already stated, the disagreement positions are seen as inside options in this paper. Thus, the model is applicable to situations where this is natural e.g. in a hospital setting (see chapter 4 below) or wage negotiations where the employee already has an alternative employment(s). The model, however, introduces an unusual feature; i.e. that each party initially has two inside options, that is if they fail to reach an agreement they have two alternative routes to take - trade with one of the other parties (third party trade or partial trade) or trade with none of them (no-trade). The next subsection presents the Nash bargaining solution when third party trade is the parties' inside option. Subsequently, it is shown that third party trade is the only credible alternative as inside option.

[14] See e.g. Muthoo (1999).

[15] E.g. they show that parties giving up assets may invest more if their outside options are binding. This is contrary to the basic notion in the GHM-model that investments follow asset ownership.

Nash Bargaining The Nash Bargaining product for the negotiation between M and A is:

$$[(T(\cdot) - v - m) - (T_S(\cdot) - v - \bar{m})] \times [(m - y - K(\cdot)) - (\bar{m} - y - K_S(\cdot))] \tag{3.4}$$

The Nash bargaining product is used to determine the optimal reimbursement between M and A, i.e. the optimal m, by maximizing (3.4) w.r.t. m. In this case m is found to be:

$$m = \bar{m} + (T(\cdot) - T_S(\cdot) + K(\cdot) - K_S(\cdot))/2 \tag{3.5}$$

In the same manner v and y can be found, and it is straightforward that:

$$v = \bar{v} + (T(\cdot) - T_A(\cdot) + C(\cdot) - c(\cdot))/2 \tag{3.6}$$

is the outcome of the bilateral negotiation between M and S when both parties have trade with A as inside option, and that:

$$y = \bar{y} + (K_M(\cdot) - K(\cdot) + G(\cdot) - g(\cdot))/2 \tag{3.7}$$

is the outcome from the bilateral negotiation between A and S when both parties have trade with M as inside option.

Credible Inside Options The possible inside options for each party in the negotiations are *no-trade* and *third party trade*; the latter is trade with only the party not involved in the particular bilateral negotiation. Hence, if the parties have not reached an agreement, then each party may choose to deviate from the negotiations in either one or two dimensions; leave one of the bilateral negotiations or both. The question that arises is: which inside option is most credible; i.e. rational to expect? A credible inside option must fulfil the following requirements: 1) it must give a higher payoff than all other inside options[16], 2) it must be achievable/feasible; if other parties are involved, as in third party trade, then they must be willing to trade. An example may clarify these requirements: imagine that M and A negotiate; then M's best inside option would be *trade with S only* because it gives a higher payoff than no trade (requirement 1). If trade with S is a credible inside option then S must prefer trade with M to no-trade and prefer trade with M and A to trade with only A; S must prefer, be willing, to trade with M in all circumstances. Given this, it turns out that third party trade is the best inside option for all three parties (see appendix).

It is obvious from the discussion above and the proof in the appendix that the inside option is defined by the idea that deviating, i.e. not reaching an agreement, from one trading-relationship is more attractive than deviating from both. The intuition behind this is the following: if the parties end up in a situation where they must choose between third party trade (i.e. reaching a bilateral agreement) and no-trade, in the sense that they fail by some odd chance to reach an agreement implying trilateral trade, then they will choose the most attractive of these options. All three parties realize that third party trade is more beneficial than no-trade and thus have incentives to reach bilateral agreements (furthermore they are free to make this choice as there is nothing in the model that prevents them from choosing partial trade).[17]

3.2.3 Individual Payoffs from Trade

The payoffs from trade are calculated by inserting the reimbursements into the payoff functions (equations (3.1)-(3.3)). The resulting payoffs from trade are denoted by U_i $i = M, A, S$, and are given by:

[16] In general it should suffice that the payoff is greater than or equal as long as requirement 2) is fulfilled. However, in this setting the payoff will be strictly greater, hence equality is not an issue.

[17] Another way to express this is : in the three simultaneous negotiations, which are only connected through payoffs, any deviation from the equilibrium (trilateral trade) will lead to a new equilibrium with third party trade (for the deviating party) and there are no incentives to deviate to no-trade in this equilibrium.

For M

$$T(\cdot) - v - m \text{ where}$$
$$m = \bar{m} + \frac{T(\cdot) - T_S(\cdot) + K(\cdot) - K_S(\cdot)}{2} \text{and } v = \bar{v} + \frac{T(\cdot) - T_A(\cdot) + C(\cdot) - c(\cdot)}{2}$$
$$\Rightarrow U_M = \frac{T_S(\cdot) + T_A(\cdot) - K(\cdot) + K_S(\cdot) - C(\cdot) + c(\cdot)}{2} - \bar{m} - \bar{v} \qquad (3.8)$$

For S

$$v - C(\cdot) + y - G(\cdot) \text{ where}$$
$$v = \bar{v} + \frac{T(\cdot) - T_A(\cdot) + C(\cdot) - c(\cdot)}{2} \text{ and } y = \bar{y} + \frac{K_M(\cdot) - K(\cdot) + G(\cdot) - g(\cdot)}{2}$$
$$\Rightarrow U_S = \frac{-C(\cdot) - c(\cdot) + T(\cdot) - T_A(\cdot)}{2} + \frac{-G(\cdot) - g(\cdot) + K_M(\cdot) - K(\cdot)}{2} + \bar{v} + \bar{y} \qquad (3.9)$$

For A

$$m - y - K(\cdot) \text{ where}$$
$$m = \bar{m} + \frac{T(\cdot) - T_S(\cdot) + K(\cdot) - K_S(\cdot)}{2} \text{and } y = \bar{y} + \frac{K_M(\cdot) - K(\cdot) + G(\cdot) - g(\cdot)}{2}$$
$$\Rightarrow U_A = \frac{-K_S(\cdot) - K_M(\cdot) + T(\cdot) - T_S(\cdot) - G(\cdot) + g(\cdot)}{2} + \bar{m} - \bar{y} \qquad (3.10)$$

The individual *ex post* benefits from trade thus become[18]:

$$U_M - \mu_A - \mu_S \text{ for } M \qquad (3.11)$$
$$U_S - \sigma_M - \sigma_A \text{ for } S \qquad (3.12)$$
$$U_A - \alpha_M - \alpha_S \text{ for } A \qquad (3.13)$$

The next section uses the basic setup presented in this section to discuss and analyze the optimal organization of the trilateral transaction.

[18] Here the analysis differs from the basic model in Hart (1995) in a non-substantial way. Hart calculates the *ex post* benefit by first subtracting the payoff from no-trade from the payoff from trade to create the surplus from trade, and then adds this to the payoff from no-trade minus the investments costs - getting the same expression for the *ex post* benefit.

3.3 Analysis, Results and Extensions

The analysis in this section primarily provides examples of Pareto optimal ways to organize the trilateral transaction. It also discusses what is here called *the undbounded analysis* of organizational choice (to be explained below) and the effects of changing the starting point of the analysis. Since the parties' choice of investment levels are fundamental, as will be seen, to the analysis, this section starts by discussing the optimal investment decision.

3.3.1 Optimal Investment Decision

First-Best Choice of Investments

The *ex post* negotiations are always efficient under any organizational structure, but the investments in date 0 might not be efficient (cf. Hart, 1995). In a first-best situation the parties can coordinate their investments to maximize the net present value of their trading relationship at date 0. Consequently they choose their investments to maximize:

$$T(\mu_A, \mu_S) - \mu_A - \mu_S - K(\alpha_M, \alpha_S) - \alpha_M - \alpha_S - C(\sigma_M) - \sigma_M - G(\sigma_A) - \sigma_A \tag{3.14}$$

Coordination of investments increases the benefit that is divided between the actors *ex post*; any choice of investments that does not maximize (3.14) can be improved on by choosing to maximize (3.14) (cf. ibid). The first order conditions for this maximization are:

$$\frac{\partial T(\mu_A, \mu_S)}{\partial \mu_A} - 1 = 0 \tag{3.15}$$

$$\frac{\partial T(\mu_A, \mu_S)}{\partial \mu_S} - 1 = 0 \tag{3.16}$$

$$-\frac{\partial C(\sigma_M)}{\partial \sigma_M} - 1 = 0 \tag{3.17}$$

$$-\frac{\partial G(\sigma_A)}{\partial \sigma_A} - 1 = 0 \tag{3.18}$$

$$-\frac{\partial K(\alpha_M,\alpha_S)}{\partial \alpha_M}-1=0 \tag{3.19}$$

$$-\frac{\partial K(\alpha_M,\alpha_S)}{\partial \alpha_S}-1=0 \tag{3.20}$$

For M, S and A respectively. For future reference let $\mu_A^*, \mu_S^*, \sigma_M^*, \sigma_A^*, \alpha_M^*, \alpha_S^*$ denote the first-best investments. However, as the incomplete contracting in the model renders the first-best impossible, the model depicts a second-best world.

Second-Best Choice of Investments

In the second-best each of the trading parties will choose period 0 investments to maximize their *ex post* benefit, which produces the following first order conditions:

For M

$$\frac{1}{2}\frac{\partial T_S(\mu_A,\mu_S;P_M)}{\partial \mu_A}+\frac{1}{2}\frac{\partial T_A(\mu_A,\mu_S;P_M)}{\partial \mu_A}-1 = 0 \tag{3.21}$$

$$\frac{1}{2}\frac{\partial T_S(\mu_A,\mu_S;P_M)}{\partial \mu_S}+\frac{1}{2}\frac{\partial T_A(\mu_A,\mu_S;P_M)}{\partial \mu_S}-1 = 0 \tag{3.22}$$

For S[19]

$$-\frac{1}{2}\frac{\partial C(\sigma_M)}{\partial \sigma_M}-\frac{1}{2}\frac{\partial c(\sigma_M;P_S)}{\partial \sigma_M}-1 = 0 \tag{3.23}$$

$$-\frac{1}{2}\frac{\partial G(\sigma_A)}{\partial \sigma_A}-\frac{1}{2}\frac{\partial g(\sigma_A;P_S)}{\partial \sigma_A}-1 = 0 \tag{3.24}$$

For A

$$-\frac{1}{2}\frac{\partial K_S(\alpha_M,\alpha_S;P_A)}{\partial \alpha_M}-\frac{1}{2}\frac{\partial K_M(\alpha_M,\alpha_S;P_A)}{\partial \alpha_M}-1 = 0 \tag{3.25}$$

$$-\frac{1}{2}\frac{\partial K_S(\alpha_M,\alpha_S;P_A)}{\partial \alpha_S}-\frac{1}{2}\frac{\partial K_M(\alpha_M,\alpha_S;P_A)}{\partial \alpha_S}-1 = 0 \tag{3.26}$$

[19]Notably, if the production cost for S was not separable in the two inputs and was instead given by a function $C(\sigma_M,\sigma_A)$ (just as for A), the first order conditions for S would have the same form as the first order conditions for A and M i.e. :
$-\frac{1}{2}\frac{\partial C_A(\sigma_M,\sigma_A;P_S)}{\sigma_j}-\frac{1}{2}\frac{\partial C_M(\sigma_M,\sigma_A;P_S)}{\sigma_j}-1=0$ for $j=A,M$.

The fact that the first-order conditions for M and A depend only on the inside options can be attributed to the Nash bargaining solution. Given the non-separable benefit/cost functions and the fact that there are three parties in the transaction - each party is involved in two simultaneous bilateral negotiations - Nash bargaining implies that each party gives up half its trilateral trade benefit/cost in each negotiation. In general it may be noted that bilateral and symmetric Nash bargaining always imply that the surplus from trade with the other party is split in half. Hence, the division of surplus implies that the equilibrium outcome (trilateral trade) does not affect the incentives for investments (for M and A). This contrasts with the bilateral model where the incentives are affected by both the equilibrium outcome and the inside option (bilateral trade and no-trade in the bilateral model). Therefore, the strongest positive effect on investments (see the marginal conditions in appendix) is absent from the first order conditions, implying that the incentives would have been stronger if the bargaining had not entailed a transfer of the entire trilateral trade benefit. Notably, if the transaction had involved four parties, the impact of the full trade benefit on the first order conditions would have been negative.[20] Thus, Nash bargaining in this variant of multilateral transactions affects incentives by either muting the effect of the equilibrium outcome (three parties) or by making this effect negative (more than three parties; e.g. four parties where each party gives up three halves of the benefit from full trade in the bargaining).

Inherent Underinvestments in the Second-Best As in the model presented by Hart (1995) the trilateral variant of the model exhibits underinvestments in second-best.

Proposition 3.1: *All parties make underinvestments in the second-best compared to the first-best.*

Proof. See appendix ■

[20]For example, if adding a party H to the transaction and making suitable adjustments of the model, then M's F.O.C w.r.t μ_A would be:

$$-\frac{1}{2}\frac{\partial T(\cdot)}{\partial \mu_A}+\frac{1}{2}\frac{\partial T_{AH}(\cdot)}{\partial \mu_A}+\frac{1}{2}\frac{\partial T_{SH}(\cdot)}{\partial \mu_A}+\frac{1}{2}\frac{\partial T_{AS}(\cdot)}{\partial \mu_A}=1$$

Where the subscripts AH, SH and AS denote the trades going on in the inside options.

3.3.2 Optimal Organization

There are several ways to organize the transaction analyzed in this paper. Organization in this model is the allocation of physical assets; i.e. the ownership structure. The optimal ownership structure is the ownership structure that supports the greatest relationship-specific investments, and thus the greatest surplus from trade. Any change in ownership structure that entails higher investments from one or more of the parties and equal investments from the others, is an improvement, because it implies a move towards the first-best. This due to the inherent underinvestments in the model.

Ten Ways to Organize the Transaction

The analysis abstracts, as already noted, from the cases where party J, $J = M, A, S$, does not own its own asset but some, or all, of the others' assets.[21] In spite of this limitation there are ten feasible ownership structures:

1 M-integration where M owns all the assets; first order conditions (cf. equations (3.21) - (3.26)) in this case are the following (let the superscript 1 denote the investments under M-integration):

$$\frac{1}{2}\frac{\partial T_S\left(\mu_A^1,\mu_S^1;p_M,p_S,p_A\right)}{\partial\mu_A}+\frac{1}{2}\frac{\partial T_A\left(\mu_A^1,\mu_S^1;p_M,p_S,p_A\right)}{\partial\mu_A} = 1 \tag{3.27}$$

$$\frac{1}{2}\frac{\partial T_S\left(\mu_A^1,\mu_S^1;p_M,p_S,p_A\right)}{\partial\mu_S}+\frac{1}{2}\frac{\partial T_A\left(\mu_A^1,\mu_S^1;p_M,p_S,p_A\right)}{\partial\mu_S} = 1 \tag{3.28}$$

$$-\frac{1}{2}\frac{\partial C\left(\sigma_M^1\right)}{\partial\sigma_M}-\frac{1}{2}\frac{\partial c\left(\sigma_M^1;\varnothing\right)}{\partial\sigma_M} = 1 \tag{3.29}$$

$$-\frac{1}{2}\frac{\partial G\left(\sigma_A^1\right)}{\partial\sigma_A}-\frac{1}{2}\frac{\partial g\left(\sigma_A^1;\varnothing\right)}{\partial\sigma_A} = 1 \tag{3.30}$$

[21] Compare also with Hart & Moore's (1990) finding, in the application to three assets, that an agent who does not own her own asset (i.e. the asset that is essential for her production) should not own any other assets.

$$-\frac{1}{2}\frac{\partial K_S\left(\alpha_M^1,\alpha_S^1;\varnothing\right)}{\partial\alpha_M}-\frac{1}{2}\frac{\partial K_M\left(\alpha_M^1,\alpha_S^1;\varnothing\right)}{\partial\alpha_M} = 1 \qquad (3.31)$$

$$-\frac{1}{2}\frac{\partial K_S\left(\alpha_M^1,\alpha_S^1;\varnothing\right)}{\partial\alpha_S}-\frac{1}{2}\frac{\partial K_M\left(\alpha_M^1,\alpha_S^1;\varnothing\right)}{\partial\alpha_S} = 1 \qquad (3.32)$$

2 S-integration where S owns all the assets; the first order conditions for S-integration follow the same pattern as above but with S owning all the assets and the investments in this case are denoted by the superscript 2 i.e. $\mu_A^2,\mu_S^2,\sigma_M^2,\sigma_A^2,\alpha_M^2,\alpha_S^2$

3 A-integration where A owns all the assets; once again the pattern from cases 1 and 2 repeats itself, this time with A owning all the assets and the investments are: $\mu_A^3,\mu_S^3,\sigma_M^3,\sigma_A^3,\alpha_M^3,\alpha_S^3$

4 Non-integration where M, S and A own their respective assets p_M, p_S and p_A, in this case the first order conditions are somewhat different from the previous cases:

$$\frac{1}{2}\frac{\partial T_S\left(\mu_A^4,\mu_S^4;p_M\right)}{\partial\mu_A}+\frac{1}{2}\frac{\partial T_A\left(\mu_A^4,\mu_S^4;p_M\right)}{\partial\mu_A} = 1 \qquad (3.33)$$

$$\frac{1}{2}\frac{\partial T_S\left(\mu_A^4,\mu_S^4;p_M\right)}{\partial\mu_S}+\frac{1}{2}\frac{\partial T_A\left(\mu_A^4,\mu_S^4;p_M\right)}{\partial\mu_S} = 1 \qquad (3.34)$$

$$-\frac{1}{2}\frac{\partial C\left(\sigma_M^4\right)}{\partial\sigma_M}-\frac{1}{2}\frac{\partial c\left(\sigma_M^4;p_S\right)}{\partial\sigma_M} = 1 \qquad (3.35)$$

$$-\frac{1}{2}\frac{\partial G\left(\sigma_A^4\right)}{\partial\sigma_A}-\frac{1}{2}\frac{\partial g\left(\sigma_A^4;p_S\right)}{\partial\sigma_A} = 1 \qquad (3.36)$$

$$-\frac{1}{2}\frac{\partial K_S\left(\alpha_M^4,\alpha_S^4;p_A\right)}{\partial\alpha_M}-\frac{1}{2}\frac{\partial K_M\left(\alpha_M^4,\alpha_S^4;p_A\right)}{\partial\alpha_M} = 1 \qquad (3.37)$$

$$-\frac{1}{2}\frac{\partial K_S\left(\alpha_M^4,\alpha_S^4;p_A\right)}{\partial\alpha_S}-\frac{1}{2}\frac{\partial K_M\left(\alpha_M^4,\alpha_S^4;p_A\right)}{\partial\alpha_S} = 1 \qquad (3.38)$$

There is, as can be seen, an obvious structure in the first order conditions for the different cases; therefore the first order conditions for partial integration cases will be represented by only one example even though there are six different cases. Note that the investments in each case are denoted by the case-number in the superscript, and to make this clear the investments are presented with each case.

5 Partial M-integration type one: M owns p_M and p_S, A owns p_A.First order conditions:

$$\frac{1}{2}\frac{\partial T_S\left(\mu_A^5,\mu_S^5;p_M,p_S\right)}{\partial \mu_A}+\frac{1}{2}\frac{\partial T_A\left(\mu_A^5,\mu_S^5;p_M,p_S\right)}{\partial \mu_A} = 1 \quad (3.39)$$

$$\frac{1}{2}\frac{\partial T_S\left(\mu_A^5,\mu_S^5;p_M,p_S\right)}{\partial \mu_S}+\frac{1}{2}\frac{\partial T_A\left(\mu_A^5,\mu_S^5;p_M,p_S\right)}{\partial \mu_S} = 1 \quad (3.40)$$

$$-\frac{1}{2}\frac{\partial C\left(\sigma_M^5\right)}{\partial \sigma_M}-\frac{1}{2}\frac{\partial c\left(\sigma_M^5;\varnothing\right)}{\partial \sigma_M} = 1 \quad (3.41)$$

$$-\frac{1}{2}\frac{\partial G\left(\sigma_A^5\right)}{\partial \sigma_A}-\frac{1}{2}\frac{\partial g\left(\sigma_A^5;\varnothing\right)}{\partial \sigma_A} = 1 \quad (3.42)$$

$$-\frac{1}{2}\frac{\partial K_S\left(\alpha_M^5,\alpha_S^5;p_A\right)}{\partial \alpha_M}-\frac{1}{2}\frac{\partial K_M\left(\alpha_M^5,\alpha_S^5;p_A\right)}{\partial \alpha_M} = 1 \quad (3.43)$$

$$-\frac{1}{2}\frac{\partial K_S\left(\alpha_M^5,\alpha_S^5;p_A\right)}{\partial \alpha_S}-\frac{1}{2}\frac{\partial K_M\left(\alpha_M^5,\alpha_S^5;p_A\right)}{\partial \alpha_S} = 1 \quad (3.44)$$

6 Partial M-integration type two: M owns p_M and p_A, S owns p_S. Investments: $\mu_A^6,\mu_S^6,\sigma_M^6,\sigma_A^6,\alpha_M^6,\alpha_S^6$

7 Partial S-integration type one: S owns p_S and p_M, A owns p_A. Investments: $\mu_A^7,\mu_S^7,\sigma_M^7,\sigma_A^7,\alpha_M^7,\alpha_S^7$

8 Partial S-integration type two: S owns p_S and p_A, M owns p_M. Investments: $\mu_A^8,\mu_S^8,\sigma_M^8,\sigma_A^8,\alpha_M^8,\alpha_S^8$

9 Partial A-integration type one: A owns p_A and p_M, S owns p_S. Investments: $\mu_A^9,\mu_S^9,\sigma_M^9,\sigma_A^9,\alpha_M^9,\alpha_S^9$

10 Partial A-integration type two: A owns p_A and p_S, M owns p_M. Investments: $\mu_A^{10},\mu_S^{10},\sigma_M^{10},\sigma_A^{10},\alpha_M^{10},\alpha_S^{10}$

The Basics of the Analysis

As already mentioned, the optimal organization of the trilateral trade depends on how well the organization supports investments. Any of the ten structures above may be optimal, depending on the nature of the assets, human capital and investments involved in the transaction. Some possible characteristics of assets, human capital and investments are described in definitions 1-5 (see appendix). The effects of these definitions on the optimal organizational structure are exemplified in coming subsections.

The analysis is divided into two analytical frameworks, a Pareto analysis and an *unbounded analysis* where the relative productivity of investments[22] and related issues are included. The Pareto analysis produces results that hold no matter what the pattern of relative productivity looks like, but only gives a partial ranking of organizational structures because it deals with Pareto improvements from a given starting point. The unbounded analysis opens up for asset allocations in accordance with the relative productivity of investments, and this may settle the choice between organizational structures that are found to be optimal in the Pareto analysis i.e. complement the Pareto analysis. Moreover, the starting point in the unbounded analysis does not matter and generically a full ranking of organizational structures is obtained. The unbounded analysis may either be used as an extension to the Pareto analysis or as an analytical tool on its own. However, relative productivity of investments is an elusive concept in the model and therefore the results generally become less clear-cut. The demands on the parties' information and calculative abilities are also greater in the unbounded analysis than in the Pareto analysis.

This amounts to an important insight; namely that the characteristics of the asset market may affect the applicability of the unbounded analysis. The Pareto analysis may be used on both rigid and more flexible asset markets, while the relative productivity of investments becomes more relevant the greater the flexibility. If assets can be transferred without friction be-

[22] For example, M's investment being ten times more productive than S's investment in monetary terms i.e. in its contribution to total surplus.

tween the parties, they will be able, and willing, to do so to maximize the investments' contribution to the trilateral trade surplus.

Examples of Pareto Analysis

The focus in this section is on changes in organizational structure that are better, yielding a higher level of investments, than a given starting point irrespective of the relative productivity of investments. To analyze these changes one has to find a natural starting point for the analysis. Here, that starting point is non-integration (following Hart, 1995), and the effects of changing the starting point are discussed in section 3.3.3. The analysis investigates whether some other organizational structure, given the assumptions imposed, increases the level of investments in the trilateral trade relationship. The results are strong in the sense that the relative productivity of investments does not matter.

Full Strict Complementarity of Assets Assume that all three assets are strictly complementary and that this strict complementarity is characterized by A being indifferent between all ownership structures that do not contain all three assets (see the second condition of definition 1 in the appendix). The formal statement of this assumption is:

$$\frac{\partial K_i\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_j} \equiv \frac{\partial K_i\left(\alpha_M, \alpha_S; \varnothing\right)}{\partial \alpha_j} \text{ and } \frac{\partial k\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_j} \equiv \frac{\partial k\left(\alpha_M, \alpha_S; \varnothing\right)}{\partial \alpha_j} \tag{3.45}$$

$$\text{where } i = M, S,\ j = M, S \text{ and } P_A = \{p_A\},\ \{p_A, p_S\} \text{ or } \{p_A, p_M\}$$

To investigate how this assumption affects the optimal organization of the trilateral trade scrutinize the first order conditions for the different ownership structures (1-10 above). First note that the assumption implies that all of A's first order conditions, except for case 3, are equal. This implies that $\alpha_M^i = \alpha_M^i$ and $\alpha_S^i = \alpha_S^i$ for $i = 1, 2, 4, ..., 10$ and furthermore (A16) and (A17) give that $\alpha_M^3 \geq \alpha_M^i$ and $\alpha_S^3 \geq \alpha_S^i$ for $i = 1, 2, 4, ..., 10$. The marginal conditions (A12)-(A15) provide the rationale for how M and S adapt their investment levels to changes in ownership structure. It is easily seen that M will invest more under M-integration (1) and both cases of partial M-integration (5 & 6) than under non-integration. M's investment level is the same for non-integration (4), partial S-integration of type two (8) and partial A-integration of type two (10). For the other cases (2,3,7, & 9) M will invest less than under non-integration. S, on the other hand, invests more than under non-integration in cases 2, 7, and 8, less in cases 1, 3, 5, and 10. S makes the same level of investments in cases 6 and 9 as under non-integration.

This means that the total level of investments is increased in two of the cases above, implying a move towards the first-best, namely partial M-integration of type two and partial S integration of type two. In the first case M will invest more than under non-integration while S's and A's investments remain the same; in the second case S will invest more and A and M will keep their investments constant.

To determine whether case 6 or case 8 is optimal, more information about the characteristics of the trilateral trade is needed. It could e.g. be the case that M's human capital is more valuable for the transaction than S's human capital - the next example takes this to the extreme and assumes that M's human capital is *essential* for the production of the final good.

Essential Human Capital In this example it is assumed that M's human capital is essential for the production of the final good. The formal implications of this assumption are found in the first part of definition 5 (see appendix):

$$\frac{\partial K_S(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j} \equiv \frac{\partial K_S(\alpha_M, \alpha_S; \varnothing)}{\partial \alpha_j} \text{ and } \frac{\partial k(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j} \equiv \frac{\partial k(\alpha_M, \alpha_S; \varnothing)}{\partial \alpha_j} \tag{3.46}$$

where $j = M, S$ and $P_A = \{p_A, p_S, p_M\}$, $\{p_A, p_S\}$, $\{p_A, p_M\}$ or $\{p_A\}$, and:

$$\frac{\partial c(\sigma_M; P_S)}{\partial \sigma_M} \equiv \frac{\partial c(\sigma_M; \varnothing)}{\partial \sigma_M} \tag{3.47}$$

where $P_S = \{p_A, p_S, p_M\}$, $\{p_S, p_M\}$,$\{p_S, p_A\}$ or $\{p_S\}$.

As in the example above this assumption has implications for the parties' first order conditions. An obvious implication is that A will have weaker incentives for investments, compared to the basic setup, in all cases where A owns assets. This stems from the basic assumption that $\frac{\partial K_S(\alpha_M,\alpha_S;\varnothing)}{\partial \alpha_j} \geq \frac{\partial K_S(\alpha_M,\alpha_S;P_A)}{\partial \alpha_j}$[23] for $j = M, S$ and P_A being a non-empty set, meaning that investments are at least as valuable under asset ownership as when not owning any assets, i.e. that incentives for investments are weakly higher when owning assets. Notably, this changes when condition (3.46) is imposed on the problem, weakly lowering and equalizing the incentives for investments, which can be attributed to the trade with S, for all ownership structures. The incentives for investments which can be attributed to the trade with M are unchanged and follow the general idea that more assets imply higher investments. Thus the incentives for investments are qualitatively the same, as if the assumption had not been imposed, but somewhat weaker. Compared to non-integration, A will invest more in cases 3, 9, and 10; less in cases 1,2, 6, and 8; and make the same investments in cases 5 and 7.

Condition (3.47) states that S will make the same investment in the relation with M, σ_M, irrespective of ownership. The incentive for investments in the relation with A is unaffected by the assumption that M's human capital is essential. Hence, definition 5 implicitly adds to the separability assumption about S's production cost, this by stating that if M's human capital is essential it is only so for the production of the input to M. S investment in the relationship with A, on the other hand, follows ownership. S will invest more than under non-integration in cases 2, 7 and 8; less in cases 1, 3, 5, and 10; and at the same level in cases 6 and 9. When it comes to M's investment it is obvious that M will invest more in cases 1, 5, and 6; less in cases 2, 3, 7, and 9; and equal amounts in cases 8 and 10.

Given that the investments are made in human capital and that M's human capital is essential for the production of the final good, it seems reasonable to find an organizational structure that increases M's investment compared to non-integration. In the bilateral model this follows immediately from the definition of essential human capital, that is if M's human capital

[23]Remember that K_S is convex i.e. $\frac{\partial K_S(\alpha_M,\alpha_S;P_A)}{\partial \alpha_j}$ $j = M, S$ is non-positive and the second derivative is non-negative.

is essential then M-integration is optimal (see Hart, 1995). In the trilateral trade case it is not possible to reach this conclusion without imposing further assumptions on the model. Specifically, all changes in ownership structure away from non-integration entail a lower level of investments by one or two of the parties (and a higher level for the third). Focusing on the cases that produce higher investments by M (cases 1, 5, and 6) it is easily seen that: 1) M-integration gives the following relationships between investments: $\sigma_M^1 = \sigma_M^4$, $\sigma_A^1 \leq \sigma_A^4$, $\alpha_M^1 \leq \alpha_M^4$, $\alpha_S^1 \leq \alpha_S^4$ (and of course $\mu_A^1 \geq \mu_A^4$, $\mu_S^1 \geq \mu_S^4$), 2) partial M-integration of type 1 gives: $\sigma_M^5 = \sigma_M^4$, $\sigma_A^5 \leq \sigma_A^4$, $\alpha_M^5 = \alpha_M^4$, $\alpha_S^5 = \alpha_S^4$, and finally that partial M-integration of type 2 gives: $\sigma_M^6 = \sigma_M^4$, $\sigma_A^6 = \sigma_A^4$, $\alpha_M^6 \leq \alpha_M^4$, $\alpha_S^6 \leq \alpha_S^4$.

It is apparent that no organizational structure entails an improvement over non-integration in this example. However, if the assumption made here is coupled with an assumption that p_M and p_S are strictly complementary such that S is indifferent between all ownership structures where S does not own p_M, as in the second condition of definition 2, the optimal organizational structure would be partial M-integration of type one.

Two Parties with Essential Human Capital Assume that both M's and A's human capital is essential for the production of the final good, which leaves S indifferent over ownership i.e. S will make the same investments, σ_M and σ_A, irrespective of ownership structure. M and A will have somewhat weaker incentives for investments (see the discussion for A in the previous example), but they will invest more the more assets they own. Two ownership structures improve the level of investments compared to non-integration, namely partial A-integration of type two and partial M-integration of type one (case 10 and case 5). Once again it is difficult to get a clear-cut result, and adding further characteristics, e.g. that p_A and p_S are strictly complementary, could actually make additional ownership structures Pareto improvements (in the example case 1 and case 6 is added). This resembles the result in the bilateral model that all organizational forms are equally good when both parties have essential human capital, with the important difference that M is still not indifferent over all ownership structures.

Independent assets Independence of assets implies that two or all three assets are strictly non-complementary in the transaction. Now, it is obvious that full independence (definition 3 in the appendix) implies that non-integration is optimal. This conclusion is straightforward since all three parties are indifferent between owning one, two or all three assets, while owning assets implies weakly greater incentives for investments than not owning any assets (see the marginal conditions). Partial independence is the situation where two of the assets are independent, e.g. p_A and p_S, and definition 4 gives the formal representation of this situation. Imagine that p_A and p_S are independent implying that S and A are indifferent between owning both assets and only owning their own asset, then non-integration is optimal. To see this consider the first-order conditions for the different organizational structures and note that:

- S will invest more in cases 2, 7; less in cases 1, 3, 5, 10; and make the same investments in cases 6, 8, 9.
- A will invest more in cases 3, 9; less in cases 1, 2, 6, 8; and make the same investments in cases 5, 7, 10
- M will invest more in cases 1, 5, 6; less in cases 2, 3, 7, 9; and make the same investments in cases 8, 10

From this it is apparent that there is no change in ownership. implying an obvious improvement over non-integration; i.e. increasing the investments by one or more parties without lowering the investments by one or more of the other parties.

Full Integration There is an apparent bias towards different types of partial integration in the examples above, begging the question what it takes to make "full" integration optimal. The answer is, while easy to find, not straightforward, it requires either quite special combinations of partial complementarity or a combination of two parties with essential human capital and partial complementarity.

M-integration, for example, is a Pareto Improvement over non-integration in two types of circumstances: first, consider a situation where *a)* assets p_M and p_S are complementary and that this complementarity implies that $\frac{\partial c(\sigma_M;P_S)}{\partial \sigma_M} \equiv \frac{\partial c(\sigma_M;\varnothing)}{\partial \sigma_M}$ and $\frac{\partial g(\sigma_A;P_A)}{\partial \sigma_A} \equiv \frac{\partial g(\sigma_A;\varnothing)}{\partial \sigma_A}$ where $P_A = p_S$ or $\{p_S, p_A\}$; *b)* assets p_M and p_A are complementary such that $\frac{\partial K_i(\alpha_M,\alpha_S;P_A)}{\partial \alpha_j} \equiv \frac{\partial K_i(\alpha_M,\alpha_S;\varnothing)}{\partial \alpha_j}$ where $i = M, S$, $j = M, S$ and $P_A = p_A$ or $\{p_A, p_S\}$. That is, both S and A are indifferent over ownership unless they own M's asset and a move from non-integration to M-integration, therefore, increases M's investment level without lowering A and S's investment levels.

Second, consider a scenario[24] where both M's and A's human capital are essential and assets p_M, and/or p_S, and p_A are complementary such that A becomes indifferent over ownership structures not containing p_i, $i = M, S$. This would also make M-integration a possible Pareto improvement over non-integration. Examples of A- and S-integration may be constructed in a similar manner, i.e. by letting the other two parties be indifferent between all ownership structures where they do not own the acquiring party's asset.

Compared to the bilateral model, this naturally adds an extra requirement to the complementarity i.e. that both the other parties must be indifferent to make full integration optimal. A more interesting observation is that assumptions about essential human capital cannot ensure that full integration is optimal as it does in the more clear cut bilateral model; even in the presence of two parties with essential human capital assumptions about partial complementarity of assets are needed to ensure that full integration is optimal.

Applying the Unbounded Analysis

As is seen above, e.g. in the example with full integration, it is somewhat cumbersome to reach clear-cut conclusions about the optimal organizational structure in the Pareto analysis, but the conclusions are strong if reached; i.e. the they are valid for the whole range of possible extensions. In this section two possible extensions are discussed, namely the asset effect and,

[24]Which actually can be broken down into three similar scenarios.

most importantly, the relative productivity of investments. It is assumed that the asset market is sufficiently flexible for these kinds of extensions to be considered.

Relative Productivity of Investments Hart defines an investment as relatively unproductive if its contribution to the net surplus from trade goes to zero (see Hart, 1995:44-46). This type of assumption may be made in the trilateral trade case as well, although making the formal statements of the kind made in Hart (1995) seems superfluous. Instead, it is natural to assume that investments are sometimes misdirected and not as beneficial for the relationship as first believed, and that this may be discerned by a low marginal contribution to the net benefit of trilateral trade. In the extreme this contribution goes to zero, implying that all marginal contributions, in partial and no-trade as well, of this investment go to zero. This makes the other investments relatively more important for the transaction i.e. for the net benefit.

Now, return for a moment to the example where M's human capital is essential and assume that S's investment in the relationship with A, σ_A, is relatively unproductive, which implies that the net benefit from trilateral trade is virtually unaffected by this investment. M's investments, on the other hand, contribute to this benefit and are thus relatively more important than σ_A. Taking this into consideration, the fact that M's human capital is essential leads to M-integration of type 1, instead of non-integration, being optimal. Under M-integration of type 1 M invests more than under non-integration and A makes the same investments, while S invests less in the relationship with A. The fact that S invests less is more than outweighed by M's increased investments since S's investment is relatively unproductive.[25]

Relative productivity could also affect the conclusions in other examples.

[25] Here is one case where the separability assumption potentially, matters for the result. In most cases both specifications yield the same results when it comes to the optimal organizational structure, but this is an exception. With separability, M-integration of type one is optimal if σ_A is relatively unproductive. If the production cost is non-separable then non-integration cannot be improved on (M-integration of type one is optimal if both of S's investments are unproductive).

In the example with full strict complementarity the choice between case 6 and 8 could be made on the basis of relative productivity. Moreover it could in fact, in this example, be optimal with A-integration if A's investments are relatively more productive than S's and M's investments. Notably, this conclusion can be reached irrespective of starting point and is totally dependent on the relative productivity. However, a move from non-integration to case 6 or 8 would still constitute a Pareto improvement, even if A's investments are relatively more productive, since it would increase M's or S's investments without decreasing A's investment. Thus if there is friction in the asset market, which prohibits the transfer of M's or S's asset (or both) to A then 6 or 8 would still be feasible improvements over non-integration.

For independent assets relative productivity of investments could foster other results than non-integration. It could be the case that e.g. M's investment is relatively more productive than S's investment, implying that case 5 could be an improvement over non-integration.

Clearly, the introduction of relative productivity opens up the analysis and provides new possibilities, but it also makes the analysis less straightforward and creates a need for more assumptions, e.g. that an investment is somewhat more important, to limit the alternatives. This might not always be intuitive and rewarding.

Asset Effect Finally, something needs to be said about the *asset effect* that appears in the marginal conditions. The asset effect reflects the fact that an investment is partially specific to another party's asset e.g. that M's investment μ_A is partially specific to A's asset p_A. This is reflected by the assumption that $\frac{\partial T_j(\mu_A,\mu_S;P_M)}{\partial \mu_A} \geq \frac{\partial T_j\left(\mu_A,\mu_S;\bar{P}_M\right)}{\partial \mu_A}, |P_M| = \left|\bar{P}_M\right|, p_A \in P_M$, $p_A \notin \bar{P}_M$ irrespective of the asset replacing p_A, i.e. the marginal benefit of the investment in the inside options is weakly greater if M owns p_A than if M does not own this asset, given that the total number of assets owned is the same in both cases. With appropriate adoptions this assumption may be imposed on all parties and their investments. This assumption becomes important for the optimal organizational structure if: I) the choice is between one party's different forms of partial integration (e.g. partial M-integration of type 1 and type 2), II) one of the party's investments is relatively unproductive or preferred over the other. The discussion about starting points below provides an example of this choice.

3.3.3 Changing the Starting Point

The examples in the Pareto analysis use non-integration as the starting point for the analysis, following Hart (1995), in a rather unquestioning manner. In this section the choice of starting point will be discussed and some short examples of how this might affect the conclusions are provided.

Stigler (1951) suggests that non-integration is rarely manifested in infant industries, characterized by small scale production, because the value-added from specialization is small on such immature markets. Instead vertical integration of tasks is prevalent. However, when the market matures it becomes profitable for firms to specialize in certain tasks and sell their services to other firms, i.e. a disintegration of tasks is carried out. On the other hand a declining market, where some tasks are not carried out at a sufficient rate to support an independent firm, is characterized by more vertical integration (cf. Stigler, 1951).[26] Thus, it is reasonable to suggest that the starting point of the analysis performed here should vary with the type of market analyzed, and if, as is done in many instances, one deals with a mature market then non-integration is the natural starting point.

There may also be other institutional reasons for choosing a starting point other than non-integration. It could be that M is a public agency and that the analysis focuses on what parts of its operations that should/could be outsourced. In addition, inertia and/or tradition could provide a rationale for a different starting point to non-integration. Irrespective of the underlying rationale there is reason to consider starting points other than non-integration and to investigate how this affects the analysis and conclusions.

Now, consider a situation where all the assets are strictly complementary

[26] Hart (1995, p. 52) suggest that his analysis is consistent with the ideas of Stigler i.e. that integration is likely in immature markets (due to complementarities) and so on. Hart does not, however, discuss the possibility of different starting points depending on the maturity of the market.

and that A as a consequence is indifferent between all the ownership structures where A does not own all the assets (as in the first example above). Furthermore, imagine that the starting point is M-integration i.e. that M owns all the assets needed for the production of the final good.[27] In this situation there is no change in ownership structure that does not entail a lower investment from M; S will, on the other hand, invest more or equal amounts if the ownership structure is changed. In fact; S will invest more in most cases, only in cases 3, 5 and 10 will S make the same investment as under M-integration. Changing to a structure where M invests less, while A and S invest the same, is clearly suboptimal - implying a move away from the first-best. Thus cases 5 and 10 can be ruled out. Do any of the remaining ownership structures imply an improvement over M-integration? This depends on the relative productivity of the investments. If all investments are equally productive, or if the asset market is rigid such that the possible gains of relative productivity cannot be reaped, then M-integration cannot be improved on. Infant industries may experience that kind of friction in the asset markets, but as the industry evolves the asset markets become more flexible and an eventual difference in relative productivity may be taken into consideration when choosing organizational structure.

For example, if M's investments are relatively unproductive compared to either S's or A's investments, then improvements can be made. If S's investments are relatively more productive than M's investments, but M's investments are still somewhat productive, then either partial S-integration (of either type) or partial M-integration of type two could be improvements, depending on how productive M's investments are.[28]

Finally, one more example. Again assets are strictly complementary in the

[27] It might be useful to think of M as a public agency controlling all assets needed for the production of a public service. But M does not control the human capital needed (besides its own human capital) i.e. A and S are free to use their human capital any way they want to. Allowing ownership by A or S (outsourcing) might be beneficial for the transaction since it increases the incentives for relationship-specific investments.

[28] Obviously, all structures where S invests more, and all structures where S owns one or more assets, may constitute an improvement, but it seems realistic that M's investments are somewhat beneficial for the transaction since the whole transaction initially is integrated.

manner described above, but now the starting point is partial M-integration of type one (case 5). In this setup S will invest more in a number of cases among them case 6. Case 6 is interesting since it entails M owning the same number of assets as in case 5, implying that M's incentives for investments are basically the same in the both cases. This means that the incentives in both cases are stronger than in the cases where M owns less assets and weaker than under M-integration, but there might be some difference in incentive strength for each of M's investments between the two cases. In fact the asset effect ensures that M invests weakly less in μ_S and weakly more in μ_A in case 6 compared to case 5. Thus, case 6 is an improvement compared to the starting point, if the increase in S's investments outweighs the loss of investments in μ_S, i.e. if S's investments are relatively more productive than the investment μ_S.

It is apparent that the starting point matters for the analysis, and an extensive list of examples may be constructed in a similar manner to the above; it is also obvious that relative productivity of investments becomes more important with other starting points than non-integration. Taking the starting point into account might enable a more thorough analysis of firm organization by heeding the institutional surroundings and how it affects the organization decision.

3.4 Concluding Discussion

A natural, but noteworthy, difference compared to the bilateral model is that it is more difficult to find a unique Pareto optimal organizational structure in the trilateral trade model; more difficult in the sense that more information about asset and investment characteristics is needed to pin down only one structure. In terms of the formal model this implies that, in most cases, at least two assumptions about e.g. assets need to be imposed for the analysis to produce a clear-cut result. In the bilateral model, on the other hand, one assumption, e.g. essential human capital, is sufficient in most cases. However, in the cases when it is possible to get clear-cut results, these results are strong in the sense that the relative productivity of investments does not matter.

If it is reasonable to assume that asset markets work without friction, then the relative productivity of investments may be used to choose between different organizational structures each of which implies a Pareto improvement. In cases where there are no obvious Pareto improvements, relative productivity could also be used to find the optimal organizational structure, given low or no friction in the asset market.

By changing the starting point of the Pareto analysis, the trilateral trade model may be able to mimic actual situations in different industries. Taking more of the institutional surroundings into account, when using the model, may potentially give new insights into and produce better predictions about the optimal organizational structure for a certain transaction. Changing the starting point of the analysis is a small step in that direction, but still provides some useful results. Obviously more work needs to be done on these issues.

There is also a tendency towards partial integration in the trilateral model, while full integration is more of an exception.[29] This result suggests a more general finding i.e. that a downstream firm, in many instances, should not treat all of its suppliers the same way (integrate all or integrate none). Instead it is beneficial to integrate one party and let the other party remain an independent contractor, implying that the latter party's investment incentive generally is greater than it would be under full integration. Which party to integrate depends on the characteristics of assets, human capital and investments, generally the party that is least sensitive, indicated by unchanged incentives for investments, to losing control over its asset may be integrated.

In the bilateral model the incentives for investments are created partly by the marginal trade benefit and partly by the marginal no-trade benefit. Thus, the marginal benefit in equilibrium and marginal benefit in the inside option create the incentives for investments. In the trilateral model the incentives are given by the marginal benefits in the credible inside options

[29] Adding an extra assumption/characteristic to the examples where non-integration cannot be improved onleads to partial integration in most cases . Full integration requires, as already mentioned, quite specific assumptions.

i.e. third party trade (when the benefit/cost functions are non-separable in investments) - trades that are not realized in equilibrium. This is an effect of the Nash bargaining assumption.

Hart & Moore's (1990) application to three assets (i.e. three firms) is the most adjacent model to the trilateral model presented in this paper. Their application in essence produces two results. First, they find that parties not owning assets that are essential to them (i.e. their own asset) should not own any other assets. This serves as a justification for the limitation of possible ownership patterns in the trilateral model. Second, they show that partial integration affects the incentives of the party that is not integrated e.g. that S's incentives are changed if M acquires p_A, i.e. M and A become integrated. This effect is not present in the trilateral model where the incentives are unchanged for the non-integrated party. This difference stems from the different bargaining setups in the two models. In particular the Shapley value, mechanically, takes into account all the possible subcoalitions for a given ownership structure, while the non-cooperative approach taken in this paper results in the incentives being given by third party trade (i.e. coalitions involving two firms) irrespective of the ownership structure.

Many real world transactions involve multiple dependencies, e.g. the manufacturing industry and hospital care, and the trilateral trade model presented here may shed some light on the complexities that arise in these situations. Obviously, it does not provide a complete picture. A number of possible extensions could improve its intuitive appeal; for example it would be interesting to introduce the possibility for powerful parties in the transaction to change their dependence on the other parties, i.e. change the technology. This would enable an analysis of corporate structures with one leading party that dominates over others, i.e. the other parties are dependent on the leading party but not the other way around, possibly answering the question of whether it is beneficial to break the bilateral dependence This, and many other things, are left for future research.

3.A Appendix

Eight patterns of trade

i. All the parties trade with each other (full trade or trade):

$$\begin{aligned} U_M &= T(\mu_A, \mu_S) - v - m \\ U_S &= v - C(\sigma_M) + y - G(\sigma_A) \\ U_A &= m - K(\alpha_M, \alpha_S) - y \end{aligned}$$

ii. None of the parties trade with each other (no-trade):

$$\begin{aligned} u_M &= t(\mu_A, \mu_S; P_M) - \bar{v} - \bar{m} \\ u_S &= \bar{v} - c(\sigma_M; P_S) + \bar{y} - g(\sigma_A; P_S) \\ u_A &= \bar{m} - k(\alpha_M, \alpha_S; P_A) - \bar{y} \end{aligned}$$

iii. M & A, A & S trades, but not M & S (no-trade M & S):

$$\begin{aligned} \dot{U}_M &= T_A(\mu_A, \mu_S; P_M) - \bar{v} - m \\ \ddot{U}_S &= \bar{v} - c(\sigma_M; P_S) + y - G(\sigma_A) \\ U_A &= m - K(\alpha_M, \alpha_S) - y \end{aligned}$$

where T_A indicates that A's human capital is available to M, but not S's human capital.

iv. M & A, M & S trade, but not A & S (no-trade A & S):

$$\begin{aligned} U_M &= T(\mu_A, \mu_S) - v - m \\ \dot{U}_S &= v - C(\sigma_M) + \bar{y} - g(\sigma_A; P_S) \\ \ddot{U}_A &= m - K_M(\alpha_M, \alpha_S; P_A) - \bar{y} \end{aligned}$$

v. M & S, A & S trade, but not A & M (no-trade A & M):

$$\begin{aligned}\ddot{U}_M &= T_S(\mu_A, \mu_S; P_M) - v - \bar{m} \\ U_S &= v - C(\sigma_M) + y - G(\sigma_A) \\ \dot{U}_A &= \bar{m} - K_S(\alpha_M, \alpha_S; P_A) - y\end{aligned}$$

vi. A & S trade, but not A & M and not M & S (trade A & S):

$$\begin{aligned}u_M &= t(\mu_A, \mu_S; P_M) - \bar{v} - \bar{m} \\ \ddot{U}_S &= \bar{v} - c(\sigma_M; P_S) + y - G(\sigma_A) \\ \dot{U}_A &= \bar{m} - K_S(\alpha_M, \alpha_S; P_A) - y\end{aligned}$$

vii. M & S trades, but not A & M and not A & S (trade M & S):

$$\begin{aligned}\ddot{U}_M &= T_S(\mu_A, \mu_S; P_M) - v - \bar{m} \\ \dot{U}_S &= v - C(\sigma_M) + \bar{y} - g(\sigma_A; P_S) \\ u_A &= \bar{m} - k(\alpha_M, \alpha_S; P_A) - \bar{y}\end{aligned}$$

viii. M & A trades, but not M & S and not A & S (trade M & A):

$$\begin{aligned}\dot{U}_M &= T_A(\mu_A, \mu_S; P_M) - \bar{v} - m \\ u_S &= \bar{v} - c(\sigma_M; P_S) + \bar{y} - g(\sigma_A; P_S) \\ \ddot{U}_A &= m - K_M(\alpha_M, \alpha_S; P_A) - \bar{y}\end{aligned}$$

Assumptions about the Benefit Functions

$$\begin{gathered}T(\mu_A, \mu_S) \textit{ is strictly concave in both } \mu_A \textit{ and } \mu_S \textit{ i.e..} \\ \frac{\partial T(\mu_A, \mu_S)}{\partial \mu_j} > 0 \textit{ and } \frac{\partial^2 T(\mu_A, \mu_S)}{\partial \mu_j \partial \mu_j} < 0 \textit{ for } j = A, S \\ \textit{and the cross derivative is bounded in the standard way.}\end{gathered} \tag{A1}$$

$$C(\sigma_M) \ \textit{is strictly convex in } \sigma_M \textit{ i.e.}$$
$$\frac{\partial C(\sigma_M)}{\partial \sigma_M} < 0 \textit{ and } \frac{\partial^2 C(\sigma_M)}{\partial \sigma_M \partial \sigma_M} > 0 \tag{A2}$$

$$G(\sigma_A) \ \textit{is strictly convex in } \sigma_A \text{ i.e.}$$
$$\frac{\partial G(\sigma_A)}{\partial \sigma_A} < 0 \textit{ and } \frac{\partial^2 G(\sigma_A)}{\partial \sigma_A \partial \sigma_A} > 0 \tag{A3}$$

$$K(\alpha_M, \alpha_S) \ \textit{is strictly convex in both } \alpha_M \textit{ and } \alpha_S \textit{ i.e.}$$
$$\frac{\partial K(\alpha_M, \alpha_S)}{\partial \alpha_j} < 0 \textit{ and } \frac{\partial^2 K(\alpha_M, \alpha_S)}{\partial \alpha_j \partial \alpha_j} > 0 \textit{ for } j = M, S \tag{A4}$$
$$\textit{and the cross derivative is bounded in the standard way.}$$

$$T_i(\mu_A, \mu_S; P_M), \ i = A, S, \textit{ is concave in both } \mu_A \textit{ and } \mu_S \textit{ i.e..}$$
$$\frac{\partial T_i(\mu_A, \mu_S; P_M)}{\partial \mu_j} \geq 0 \textit{ and } \frac{\partial^2 T_i(\mu_A, \mu_S; P_M)}{\partial \mu_j \partial \mu_j} \leq 0 \textit{ for } j = A, S \tag{A5}$$
$$\textit{and the cross derivative in bounded in the standard way.}$$

$$K_i(\alpha_M, \alpha_S; P_A) \ , i = M, S, \textit{ is convex in both } \alpha_M \textit{ and } \alpha_S \textit{ i.e.}$$
$$\frac{\partial K_i(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j} \leq 0 \textit{ and } \frac{\partial^2 K_i(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j \partial \alpha_j} \geq 0 \textit{ for } j = M, S \tag{A6}$$
$$\textit{and the cross derivative is bounded in the standard way.}$$

$$t(\mu_A, \mu_S; P_M) \ \textit{is concave in both } \mu_A \textit{ and } \mu_S \textit{ i.e..}$$
$$\frac{\partial t(\mu_A, \mu_S; P_M)}{\partial \mu_j} \geq 0 \textit{ and } \frac{\partial^2 t(\mu_A, \mu_S; P_M)}{\partial \mu_j \partial \mu_j} \leq 0 \textit{ for } j = A, S \tag{A7}$$
$$\textit{and the cross derivative is bounded in the standard way.}$$

$$
\begin{gathered}
c\left(\sigma_M; P_S\right) \text{ is convex in } \sigma_M \text{ i.e.} \\
\frac{\partial c\left(\sigma_M; P_S\right)}{\partial \sigma_M} \leq 0 \text{ and } \frac{\partial^2 c\left(\sigma_M; P_S\right)}{\partial \sigma_M \partial \sigma_M} \geq 0
\end{gathered} \tag{A8}
$$

$$
\begin{gathered}
g\left(\sigma_A; P_S\right) \text{ is convex in } \sigma_A \text{ i.e.} \\
\frac{\partial g\left(\sigma_A; P_S\right)}{\partial \sigma_A} \leq 0 \text{ and } \frac{\partial^2 g\left(\sigma_A; P_S\right)}{\partial \sigma_A \partial \sigma_A} \geq 0
\end{gathered} \tag{A9}
$$

$$
\begin{gathered}
k\left(\alpha_M, \alpha_S; P_A\right) \text{ is convex in both } \alpha_M \text{ and } \alpha_S \text{ i.e.} \\
\frac{\partial k\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_j} \leq 0 \text{ and } \frac{\partial^2 k\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_j \partial \alpha_j} \geq 0 \text{ for } j = M, S \\
\text{and the cross derivative is bounded in the standard way.}
\end{gathered} \tag{A10}
$$

Ranking of Total Surplus from Trade

$$
\begin{aligned}
T\left(\mu_A, \mu_S\right) - C\left(\sigma_M\right) - G\left(\sigma_A\right) - K\left(\alpha_M, \alpha_S\right) &\equiv a > \\
T_A\left(\mu_A, \mu_S; P_M\right) - c\left(\sigma_M; P_S\right) - G\left(\sigma_A\right) - K\left(\alpha_M, \alpha_S\right) &\equiv b \lesseqgtr \\
T\left(\mu_A, \mu_S\right) - C\left(\sigma_M\right) - g\left(\sigma_A; P_S\right) - K_M\left(\alpha_M, \alpha_S; P_A\right) &\equiv c \lesseqgtr \\
T_S\left(\mu_A, \mu_S; P_M\right) - C\left(\sigma_M\right) - G\left(\sigma_A\right) - K_S\left(\alpha_M, \alpha_S; P_A\right) &\equiv d > \\
t\left(\mu_A, \mu_S; P_M\right) - c\left(\sigma_M; P_S\right) - G\left(\sigma_A\right) - K_S\left(\alpha_M, \alpha_S; P_A\right) &\equiv e \lesseqgtr \\
T_S\left(\mu_A, \mu_S; P_M\right) - C\left(\sigma_M\right) - g\left(\sigma_A; P_S\right) - k\left(\alpha_M, \alpha_S; P_A\right) &\equiv f \lesseqgtr \\
T_A\left(\mu_A, \mu_S; P_M\right) - c\left(\sigma_M; P_S\right) - g\left(\sigma_A\right) - K_M\left(\alpha_M, \alpha_S; P_A\right) &\equiv g > \\
t\left(\mu_A, \mu_S; P_M\right) - c\left(\sigma_M; P_S\right) - g\left(\sigma_A; P_S\right) - k\left(\alpha_M, \alpha_S; P_A\right) &\equiv h
\end{aligned} \tag{A11}
$$

Now, let α and $\bar{\alpha}$ denote total surpluses with the same number of trades i.e. two or one (since three trades/full trade and no-trade yield a single total surplus each). Furthermore let β and $\bar{\beta}$ be any two surpluses with a different number of trades. Then the following is assumed to hold:

$$
\left|\alpha - \bar{\alpha}\right| < \left|\beta - \bar{\beta}\right| \tag{A11.a}
$$

Assumption $A11.a$ says that the difference between two surpluses with the same number of trades is smaller than the difference between two surpluses with a different number of trades. Furthermore it implies that if $\beta > \bar{\beta} > \alpha > \bar{\alpha}$ then $\beta - \alpha > \bar{\beta} - \bar{\alpha}$. The intuition is e.g. that the difference in surplus between three trades (a) and one trade (say e) is greater than the difference in surplus between two trades (say b) and one trade (say f) even if $e > f$.

Marginal Conditions

Assumption A12: For $j = A, S$

$$\frac{\partial T(\mu_A, \mu_S)}{\partial \mu_S} > \frac{\partial T_j(\mu_A, \mu_S; P_M)}{\partial \mu_S} \geq \frac{\partial t(\mu_A, \mu_S; P_M)}{\partial \mu_S}$$

$$P_M = \begin{array}{c} \{p_M, p_S, p_A\}, \{p_M, p_S\} \\ \{p_M, p_A\}, \{p_M\}, \{\varnothing\} \end{array}$$

$$\frac{\partial T_j(\mu_A, \mu_S; P_M)}{\partial \mu_S} \geq \frac{\partial T_j\left(\mu_A, \mu_S; \bar{P}_M\right)}{\partial \mu_S}, \bar{P}_M \subset P_M$$

$$\frac{\partial T_j(\mu_A, \mu_S; P_M)}{\partial \mu_S} \geq \frac{\partial T_j\left(\mu_A, \mu_S; \bar{P}_M\right)}{\partial \mu_S}, |P_M| = \left|\bar{P}_M\right|, p_S \in P_M,\ p_S \notin \bar{P}_M$$

and this holds irrespective of the asset replacing p_S

(Asset effect)

$$\frac{\partial T_S(\mu_A, \mu_S; P_M)}{\partial \mu_S} \geq \frac{\partial T_A(\mu_A, \mu_S; P_M)}{\partial \mu_S} \qquad \text{(Trade effect)}$$

$$\frac{\partial t(\mu_A, \mu_S; P_M)}{\partial \mu_S} \geq \frac{\partial t\left(\mu_A, \mu_S; \bar{P}_M\right)}{\partial \mu_S}, \bar{P}_M \subset P_M \qquad \text{(A12)}$$

Assumption A13: For $j = A, S$

$$\frac{\partial T(\mu_A, \mu_S)}{\partial \mu_A} > \frac{\partial T_j(\mu_A, \mu_S; P_M)}{\partial \mu_A} \geq \frac{\partial t(\mu_A, \mu_S; P_M)}{\partial \mu_A}$$

$$P_M = \begin{array}{c} \{p_M, p_S, p_A\}, \{p_M, p_S\} \\ \{p_M, p_A\}, \{p_M\}, \{\varnothing\} \end{array}$$

$$\frac{\partial T_j\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_A} \geq \frac{\partial T_j\left(\mu_A,\mu_S;\bar{P}_M\right)}{\partial \mu_A}, \bar{P}_M \subset P_M$$

$$\frac{\partial T_j\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_A} \geq \frac{\partial T_j\left(\mu_A,\mu_S;\bar{P}_M\right)}{\partial \mu_A}, \left|P_M\right| = \left|\bar{P}_M\right|, p_A \ \in P_M,\ p_A \ \notin \bar{P}_M$$

$$\text{and this holds irrespective of the asset replacing } p_A \qquad \text{(Asset effect)}$$

$$\frac{\partial T_A\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_A} \geq \frac{\partial T_S\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_A} \qquad \text{(Trade effect)}$$

$$\frac{\partial t\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_S} \geq \frac{\partial t\left(\mu_A,\mu_S;\bar{P}_M\right)}{\partial \mu_S}, \bar{P}_M \subset P_M \qquad \text{(A13)}$$

Assumption A14:

$$\frac{\partial C\left(\sigma_M\right)}{\partial \sigma_M} < \frac{\partial c\left(\sigma_M;P_S\right)}{\partial \sigma_M} \leq \frac{\partial c\left(\sigma_M;\bar{P}_S\right)}{\partial \sigma_M},$$

$$\bar{P}_S \subset P_S,\ P_S,\bar{P}_S = \begin{array}{c}\left\{p_S,p_M,p_A\right\},\left\{p_S,p_M\right\}\\ \left\{p_S,p_A\right\},\left\{p_S\right\},\left\{\varnothing\right\}\end{array}$$

$$\frac{\partial c\left(\sigma_M;P_S\right)}{\partial \sigma_M} \leq \frac{\partial c\left(\sigma_M;\bar{P}_S\right)}{\partial \sigma_M}, \left|P_S\right| = \left|\bar{P}_S\right|,\ p_M \in P_S,\ p_M \notin \bar{P}_S$$

$$\text{and this holds irrespective of the asset replacing } p_M \qquad \text{(A14)}$$

Assumption A15:

$$\frac{\partial G\left(\sigma_A\right)}{\partial \sigma_A} < \frac{\partial g\left(\sigma_A;P_S\right)}{\partial \sigma_A} \leq \frac{\partial g\left(\sigma_A;\bar{P}_S\right)}{\partial \sigma_A},$$

$$\bar{P}_S \subset P_S,\ P_S,\bar{P}_S = \begin{array}{c}\left\{p_S,p_M,p_A\right\},\left\{p_S,p_M\right\}\\ \left\{p_S,p_A\right\},\left\{p_S\right\},\left\{\varnothing\right\}\end{array}$$

$$\frac{\partial g\left(\sigma_A;P_S\right)}{\partial \sigma_A} \leq \frac{\partial g\left(\sigma_A;\bar{P}_S\right)}{\partial \sigma_A}, \left|P_S\right| = \left|\bar{P}_S\right|,\ p_A \in P_S,\ p_A \notin \bar{P}_S$$

$$\text{and this holds irrespective of the asset replacing } p_A \qquad \text{(A15)}$$

Assumption A16: For $i = M, S$

$$\frac{\partial K\left(\alpha_M, \alpha_S\right)}{\partial \alpha_M} < \frac{\partial K_i\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_M} \leq \frac{\partial k\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_M},$$

$$P_A = \begin{array}{c} \left\{p_A, p_M, p_S\right\}, \left\{p_A, p_S\right\} \\ \left\{p_A, p_M\right\}, \left\{p_A\right\}, \left\{\varnothing\right\} \end{array}$$

$$\frac{\partial K_i\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_M} \leq \frac{\partial K_i\left(\alpha_M, \alpha_S; \bar{P}_A\right)}{\partial \alpha_M}, \bar{P}_A \subset P_A$$

$$\frac{\partial K_i\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_M} \leq \frac{\partial K_i\left(\alpha_M, \alpha_S; \bar{P}_A\right)}{\partial \alpha_M}, \left|P_A\right| = \left|\bar{P}_A\right|,\ p_M \in P_A,\ p_M \notin \bar{P}_A$$

and this holds irrespective of the asset replacing p_M

(Asset effect)

$$\frac{\partial K_M\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_M} \leq \frac{\partial K_S\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_M} \qquad \text{(Trade effect)}$$

$$\frac{\partial k\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_M} \leq \frac{\partial k\left(\alpha_M, \alpha_S; \bar{P}_A\right)}{\partial \alpha_M}, \bar{P}_A \subset P_A \qquad \text{(A16)}$$

Assumption A17: For $i = M, S$

$$\frac{\partial K\left(\alpha_M, \alpha_S\right)}{\partial \alpha_S} < \frac{\partial K_i\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_S} \leq \frac{\partial k\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_S},$$

$$P_A = \begin{array}{c} \left\{p_A, p_M, p_S\right\}, \left\{p_A, p_S\right\} \\ \left\{p_A, p_M\right\}, \left\{p_A\right\}, \left\{\varnothing\right\} \end{array}$$

$$\frac{\partial K_i\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_S} \leq \frac{\partial K_i\left(\alpha_M, \alpha_S; \bar{P}_A\right)}{\partial \alpha_S}, \bar{P}_A \subset P_A$$

$$\frac{\partial K_i\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_S} \leq \frac{\partial K_i\left(\alpha_M, \alpha_S; \bar{P}_A\right)}{\partial \alpha_S}, \left|P_A\right| = \left|\bar{P}\right|,\ p_S \in P_A,\ p_S \notin \bar{P}_A$$

and this holds irrespective of the asset replacing p_S

(Asset effect)

$$\frac{\partial K_S(\alpha_M, \alpha_S; P_A)}{\partial \alpha_S} \leq \frac{\partial K_M(\alpha_M, \alpha_S; P_A)}{\partial \alpha_S} \qquad \text{(Trade effect)}$$

$$\frac{\partial k(\alpha_M, \alpha_S; P_A)}{\partial \alpha_S} \leq \frac{\partial k\left(\alpha_M, \alpha_S; \bar{P}_A\right)}{\partial \alpha_S}, \bar{P}_A \subset P_A \qquad \text{(A17)}$$

Credible Inside Option:

The focus here is on the negotiation between M and A, and similar reasoning would show that trade with A is a credible inside option for both parties in the negotiation between M and S, and that M is a credible inside option for both parties in the negotiation between A and S. The derivation of the reimbursements in the exposition below is left out but easy to check by finding the Nash bargaining solution for each negotiation.

Proposition 3.2: S is a credible inside option for M

Proof. For this to be true then the following must hold: i) M must prefer trade with S to no-trade, ii) S must prefer trade with M to no-trade, and iii) S must prefer trade with M and A to trade with A only. These statements are reformulated below and shown to hold:

i) It is beneficial for M to trade with S (compared to no-trade) if M and A fail to reach an agreement

This so because trade with S gives:

$T_S - \bar{m} - v$ where $v = \bar{v} + (T_S - t + C - c)/2$.

and the benefit from no-trade is:

$t(\cdot) - \bar{m} - \bar{v}$. Hence the gain from trading with S compared to no-trade is:

$T_S(\cdot) - \bar{m} - (\bar{v} + (T_S - t + C - c)/2) - (t(\cdot) - \bar{m} - \bar{v}) =$

$(T_S(\cdot) - t(\cdot) - C(\cdot) + c(\cdot))/2 > 0$ since

$T_S(\cdot) - C(\cdot) - g(\cdot) - k(\cdot) > t(\cdot) - c(\cdot) - g(\cdot) - k(\cdot)$ in assumption $A11$.

ii) It is better for S to trade with M than not trade at all when trading with A is not an option (i.e. the negotiation with A has not reached an agreement).

This is so because to trade with M gives:

$-C - g + v + \bar{y}$ where $v = \bar{v} + (T_S - t + C - c)/2$

and no-trade gives:$-c-g+\bar{v}+\bar{y}$. Hence the surplus from trade is:

$-C-g+\bar{y}+\bar{v}+(T_S-t+C-c)/2-(-c-g+\bar{y}+\bar{v})=$

$(T_S-t-C+c)/2>0$ which is greater than zero since assumption $A11$ states that:

$T_S-C-G-K_S>t-c-G-K_s$

iii) It is better for S to trade with M and A than only with A (implying that when considering trade with A, S's best option is to trade with M as well).

S.1b) in the proof of proposition 4 , below, shows that this is true. ■

Proposition 3.3: *S is a credible inside option for A*

Proof. For this to be true then the following must hold: i) A must prefer trade with S to no-trade, ii) S must prefer trade with A to no-trade, and iii) S must prefer trade with M and A to trade with M only. These statements are reformulated below and shown to hold:

i) It is beneficial for A to trade with S (compared to no-trade) if M and A fail to reach an agreement

To see that this is true compare the case where A only trades with S:

$\bar{m}-y-K_S$ where $y=\bar{y}+(k-K_S+G-g)/2$

with no-trade:

$\bar{m}-\bar{y}-k$.

The difference between the two benefits is:

$\bar{m}-\bar{y}-(k-K_S+G-g)/2-K_S-[\bar{m}-\bar{y}-k]=$

$=(k-K_S+g-G)/2>0$ where the strict inequality stems from the assumption

that $t-c-G-K_S>t-c-g-k$ see $A11$.

ii) It is better for S to trade with A than not trade at all when trading with M is not an option (i.e. the negotiation with M has not reached an agreement)

To see this compare the case where S only trades with A:

$\bar{v}-c+y-G$ where $y=\bar{y}+(k-K_S+G-g)/2$

with no-trade:

$\bar{v}-c+\bar{y}-g$.

It is obvious from the preceding comparison that the assumptions $A11$ above ensures that trade is beneficial. In this case, as above, the difference between trade with A and no-trade at all is:

$(k - K_S + g - G)/2 > 0$

iii) It is better for S to trade with M and A than only with M (implying that when considering trade with M, S's best option is to trade with A as well)

S.1a) in the proof of proposition 4 **,** below, shows that this is true. ■

The Bargaining/Trade Game

Given the bargaining outcomes from Nash Bargaining or, equivalently, from Rubinstein's alternating-offers model (when the time between offers approaches zero) (cf. Rubinstein, 1982, Binmore, 1987), for all feasible bilateral negotiations, a bargaining/trade game can be constructed - call this game G. G involves three players A, M, and S. Each player i has four strategies: trade with both j and k (Z_i^{jk}), trade with only $j(Z_i^j)$, trade only with k (Z_i^k), or trade with neither i.e. no-trade (Z_i^0) and corresponding payoffs u_i^{jk}, u_i^j, u_i^k and u_i^0 where i, j , k = A, M, S and $i \neq j \neq k$. As above, the derivation of the reimbursements is left out but easy to check by finding the Nash bargaining solution for each negotiation.

Proposition 3.4: Given the bargaining outcomes, full trade i.e. $\left(Z_A^{MS}, Z_M^{AS}, Z_S^{AM}\right)$ is a strict Nash equilibrium in G

Proof. First note that given the bargaining outcome in simultaneous bilateral negotiations with third party trade as inside option, the payoffs under full trade are the following:

$$u_M^{AS} = U_M = \frac{T_S + T_A - K + K_S - C + c}{2} - \bar{m} - \bar{v}$$

$$u_S^{AM} = U_S = \frac{-C - c + T - T_A}{2} + \frac{-G - g + K_M - K}{2} + \bar{v} + \bar{y}$$

$$u_A^{MS} = U_A = \frac{-K_S - K_M + T - T_S - G + g}{2} + \bar{m} - \bar{y}$$

Thus, this is the payoff for each party in the candidate for Nash equilibrium $\left(Z_A^{MS}, Z_M^{AS}, Z_S^{AM}\right)$. Do any of the parties have incentives to deviate from this outcome?

Notably, there are two levels of deviation: **1)** deviate from full trade and trade with only one of the parties, (Z_i^j) or (Z_i^k) and **2)** deviate from full trade to no-trade, (Z_i^0).

First consider M; does M want to deviate?

M.1a) Deviate and only trade with S give M:
$u_M^S = T_S - \bar{m} - v$ where $v = \bar{v} + (T_S - t + C - c)/2 \Rightarrow$
$u_M^S = (T_S + t - C + c)/2 - \bar{m} - \bar{v}$. Hence:
$u_M^{AS} - u_M^S = \begin{pmatrix} T_S + T_A - K + \\ +K_S - C + c \end{pmatrix} /2 - \bar{m} - \bar{v} - (T_S + t - C + c)/2 + \bar{m} + \bar{v} \Rightarrow$
$2\left(u_M^{AS} - u_M^S\right) = T_A - K - (t - Ks) > 0$ since assumption $A11$ states that:
$T_A - c - G - K > t - c - G - K_S \Rightarrow T_A - K - (t - Ks) > 0$

M has no incentive to deviate and only trade with S since this clearly gives a lower payoff

M.1b) Deviate and only trade with A give M:
$u_M^A = T_A - m - \bar{v}$ where $m = \bar{m} + (T_A - t + K_M - k)/2 \Rightarrow$
$u_M^A = (T_A + t - K_M + k)/2 - \bar{m} - \bar{v}$. Hence:
$u_M^{AS} - u_M^A = \begin{pmatrix} T_S + T_A - K + \\ +K_S - C + c \end{pmatrix} /2 - \bar{m} - \bar{v} - (T_A + t - K_M + k)/2 + \bar{m} + \bar{v} \Longrightarrow$

$\nabla \equiv 2\left(u_M^{AS} - u_M^A\right) = T_S - C - k - (T_A - c - K_M) + \{T_A - K - (t - K_S)\}$

From assumption $A11$ it is obvious that the second part of ∇ (within curly brackets) is greater than zero i.e.:

$T_A - K - (t - K_S) > 0$ since $T_A - c - G - K > t - c - G - K_S$

$A11$ also states that:

$T_S - C - g - k \lesseqgtr T_A - c - g - K_M \Rightarrow T_S - C - k - (T_A - c - K_M) \lesseqgtr 0$ implying that ∇ is greater than zero when this difference is greater than or equal to zero, but also when it is smaller than zero and $|T_A - K - (t - K_S)| > |T_S - C - k - (T_A - c - K_M)|$ which is ensured by assumption $A11.a$. Hence M has no incentive to deviate and only trade with A.

M.2) Deviate and not trade at all give M:
$u_M^0 = t - \bar{m} - \bar{v}$: Hence
$u_M^{AS} - u_M^0 = \begin{pmatrix} T_S + T_A - K + \\ +K_S - C + c \end{pmatrix} /2 - \bar{m} - \bar{v} - t + \bar{m} + \bar{v} \Rightarrow$
$2\left(u_M^{AS} - u_M^0\right) = T_S + T_A - K + K_S - C + c - 2t =$
$= T_S - C - (t - c) + T_A - K - (t - K_S) > 0$ since $A11$ states that:
$T_S - C - G - K_S > t - c - G - K_S \Leftrightarrow T_S - C - (t - c) > 0$, and
$T_A - c - G - K > t - c - G - K_S \Leftrightarrow T_A - K - (t - K_S) > 0$
Once again M has no incentive to deviate and not trade at all.

Second, does S want to deviate?

S.1a) Deviate and only trade with M give S:
$u_S^M = -C - g + v + \bar{y}$ where $v = \bar{v} - (T_S - t + C - c)/2 \Rightarrow$
$u_S^M = (-C - c + T_S - t)/2 - g + \bar{v} + \bar{y}$ Hence:
$u_S^{MA} - u_S^M = \begin{pmatrix} -C - c + T - T_A - G - g + \\ +K_M - K + C + c - T_S + t \end{pmatrix} /2 + g + \bar{v} + \bar{y} - \bar{v} - \bar{y} \Rightarrow$
$\Theta_1 \equiv 2\left(u_S^{MA} - u_S^M\right) = T - C - G - K - (T_A - c - g - K_M) + \{t - c - (T_S - C)\}$
The second part of θ (within curly brackets) is smaller than zero since:
$T_S - C - G - K_S > t - c - G - K_S$ (see $A11$)
while the first part of θ is positive since:
$T - C - G - K > T_A - c - g - K_M$ ($A11$)
Obviously S will not deviate as long as:

$\Theta_1 > 0$ and this is always the case since assumption $A11.a$ ensures that the positive difference in the first part of Θ_1 is greater (in absolute terms) than the negative difference in the second part of Θ_1. Thus S does not want to deviate and only trade with M.

S.1b) Deviate and only trade with A give S:
$u_S^A = -c - G + \bar{v} + y$ where $y = \bar{y} + (k - K_S + G - g)/2 \Rightarrow$
$u_S^A = -c + (-G - g + k - K_S)/2 + \bar{v} + \bar{y}$ Hence:
$u_S^{MA} - u_S^A = \begin{pmatrix} -C - c + T - T_A + \\ -G - g + K_M - K \end{pmatrix} /2 + \begin{pmatrix} c + G + \\ +g - k + K_S \end{pmatrix} /2 + \bar{v} + \bar{y} - \bar{v} - \bar{y} \Rightarrow$

$\Theta_2 \equiv 2\left(u_S^{MA} - u_S^A\right) = T - C - G - K - (T_A - c - g - K_M) + \{G - k - (g - K_S)\}$

Once again the first part of θ_2 is positive (see above) and the second part is negative since:

$T_S - C - G - K_S > T_S - C - g - k$ $(A11)$ implying that $-G - K_S > -g - k \Leftrightarrow g - K_S > G - k$

However, $A11.a$ once again ensures that $\Theta_2 > 0 \Rightarrow S$ does not want to deviate and only trade with A

S.2) Deviate and not trade at all give S:

$u_S^0 = -c - g + \bar{v} + \bar{y}$ Hence:

$u_S^{MA} - u_S^0 = \begin{pmatrix} -C - c + T - T_A + \\ -G - g + K_M - K \end{pmatrix} /2 + c + g + \bar{v} + \bar{y} - \bar{v} - \bar{y} \Rightarrow$

$2\left(u_S^{MA} - u_S^0\right) = T - C - (T_A - c) + K_M - G - (K - g) > 0$

This difference is positive since (see $A11$):

$T - C - G - K > T_A - c - G - K \Rightarrow T - C - (T_A - c) > 0$

and

$T - C - G - K > T - C - g - K_M \Rightarrow -G - K > -g - K_M \Rightarrow$

$K_M - G - (g - K) > 0$

Hence S does not want to deviate and not trade at all.

Finally, does A want to deviate?

A.1a) Deviate and only trade with S give A:

$u_A^S = -K_S - y + \bar{m}$ where $y = \bar{y} + (k - K_S + G - g)/2 \Rightarrow$

$u_A^S = (-K_S - k - G + g)/2 + \bar{m} - \bar{y}$ Hence:

$u_A^{MS} - u_A^S = \begin{pmatrix} -K_S - K_M + T + \\ -T_S - G + g \end{pmatrix} /2 + (K_S + k + G - g)/2 + \bar{m} - \bar{y} - \bar{m} + \bar{y} \Rightarrow$

$2\left(u_A^{MS} - u_A^S\right) = T - K_M - (T_S - k) > 0$

Since $A11$ states that:

$T - C - g - K_M > T_S - C - g - k \Leftrightarrow T - K_M - (T_S - k) > 0$

Hence A does not want to deviate and only trade with S

A.1b) Deviate and only trade with M give A:

$u_A^M = -K_M - \bar{y} + m$ where $m = \bar{m} + (T_A - t + K_M - k)/2 \Rightarrow$

$u_A^M = (-K_M + T_A - t - k)/2 - \bar{y} + m$ Hence:

$u_A^{MS} - u_A^M = \begin{pmatrix} -K_S - K_M + T+ \\ -T_S - G + g \end{pmatrix} /2 + (K_M - T_A + t + k)/2 - \bar{y} + m$
$+\bar{y} - m \Rightarrow$

$\Omega \equiv 2\left(u_A^{MS} - u_A^M\right) = \{T - K_M - (T_S - k)\} + t - G - K_S - (T_A - g - K_M)$

The first part of Ω (within curly brackets) is greater than zero (again according to $A11$) i.e.

$T - C - g - K_M > T_S - C - g - k \Rightarrow T - K_M - (T_S - k) > 0$

The second part of Ω may either be zero, negative or positive since:

$t - c - G - K_S \lesseqgtr T_A - c - g - K_M \Leftrightarrow t - G - K_S - (T_A - g - K_M) \lesseqgtr 0$

Obviously $\Omega > 0$ when $t - G - K_S - (T_A - g - K_M) \geq 0$ and $A11.a$ ensures than this is the case even when $t - G - K_S - (T_A - g - K_M) < 0$

Hence A does not want to deviate and only trade with M

A.2) Deviate and not trade at all give A:

$u_A^0 = -k - \bar{y} + \bar{m}$ Hence:

$u_A^{MS} - u_A^0 = \begin{pmatrix} -K_S - K_M + T+ \\ -T_S - G + g \end{pmatrix} /2 + k - \bar{y} + m + \bar{y} - m \Rightarrow$

$2\left(u_A^{MS} - u_A^0\right) = T - K_M - (T_S - k) - K_S - G - (-k - g) > 0$

This difference is greater than zero since (see $A11$ again):

$T - C - g - K_M > T_S - C - g - k \Rightarrow T - K_M - (T_S - k) > 0$

and

$T_S - C - G - K_S > T_S - C - g - k \Leftrightarrow -G - K_S > -g - k \Leftrightarrow$

$-K_S - G - (-k - g) > 0$

Hence A does not want to- deviate and not trade at all.

Conclusion: None of the parties want to deviate from the equilibrium candidate where all the parties trade and the surplus is divided by simultaneous bilateral negotiations (Nash Bargaining). ■

Proof, Inherent Underinvestments in the Second-Best

Proof of Proposition 3.1. The proof follows Hart (1995:41) and is here only presented for μ_A and σ_M since the reasoning is identical for the other investments.

First μ_A: it is assumed that

$$\frac{\partial T(\mu_A, \mu_S)}{\partial \mu_A} > 0, \frac{\partial^2 T(\mu_A, \mu_S)}{\partial \mu_A \partial \mu_A} < 0$$

and that

$$\frac{\partial T(\mu_A, \mu_S)}{\partial \mu_A} > \frac{\partial T_i(\mu_A, \mu_S; P_M)}{\partial \mu_A} i = S, A$$

.

Now suppose that $\hat{\mu}_A$ solves the second-best problem then:

$$\frac{\partial T(\hat{\mu}_A, \mu_S)}{\partial \mu_A} > \frac{1}{2}\frac{\partial T_S(\hat{\mu}_A, \mu_S; P_M)}{\partial \mu_A} + \frac{1}{2}\frac{\partial T_A(\hat{\mu}_A, \mu_S; P_M)}{\partial \mu_A} = 1$$

and it follows that:

$$\frac{\partial T(\hat{\mu}_A, \mu_S)}{\partial \mu_A} > \frac{\partial T(\mu_A^*, \mu_S)}{\partial \mu_A} = 1 \Rightarrow \mu_A^* > \hat{\mu}_A$$

since $\frac{\partial^2 T(\mu_A, \mu_S)}{\partial \mu_A \partial \mu_A} < 0$
Second σ_M: it is assumed that

$$\frac{\partial C(\sigma_M)}{\partial \sigma_M} < 0, \frac{\partial^2 C(\sigma_M)}{\partial \sigma_M \partial \sigma_M} > 0$$

and that

$$\frac{\partial C(\sigma_M)}{\partial \sigma_M} < \frac{\partial c(\sigma_M; P_S)}{\partial \sigma_M}$$

Now suppose that $\hat{\sigma}_M$ solves the second-best problem then:

$$\frac{\partial C(\hat{\sigma}_M)}{\partial \sigma_M} < \frac{1}{2}\frac{\partial C(\hat{\sigma}_M)}{\partial \sigma_M} + \frac{1}{2}\frac{\partial c(\hat{\sigma}_M; P_S)}{\partial \sigma_M} = -1$$

and from this it follows that:

$$\frac{\partial C(\hat{\sigma}_M)}{\partial \sigma_M} < \frac{\partial C(\sigma_M^*)}{\partial \sigma_M} = -1 \Rightarrow \sigma_M^* > \hat{\sigma}_M$$

since $\frac{\partial^2 C(\sigma_M)}{\partial \sigma_M \partial \sigma_M} > 0$, showing that second-best leads to underinvestments.

■

Notably, this holds for all investments by all parties.

Definitions - Optimal Ownership Structure

Definition 1: Full Strict Complementarity of assets: the assets p_M, p_S and p_A are strictly complementary if:

$$\frac{\partial T_i\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_j}\equiv\frac{\partial T_i\left(\mu_A,\mu_S;\varnothing\right)}{\partial \mu_j}\text{ and }\frac{\partial t\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_j}\equiv\frac{\partial t\left(\mu_A,\mu_S;\varnothing\right)}{\partial \mu_j}$$
$$\text{where } i=S,A,\ j=S,A \text{ and } P_M=\{p_M\}\,,\ \{p_M,p_S\}\ \text{ or }\ \{p_M,p_A\}$$

or if:

$$\frac{\partial K_i\left(\alpha_M,\alpha_S;P_A\right)}{\partial \alpha_j}\equiv\frac{\partial K_i\left(\alpha_M,\alpha_S;\varnothing\right)}{\partial \alpha_j}\text{ and }\frac{\partial k\left(\alpha_M,\alpha_S;P_A\right)}{\partial \alpha_j}\equiv\frac{\partial k\left(\alpha_M,\alpha_S;\varnothing\right)}{\partial \alpha_j}$$
$$\text{where } i=M,S,\ j=M,S \text{ and } P_A=\{p_A\}\,,\ \{p_A,p_S\}\ \text{ or }\ \{p_A,p_M\}$$

or if:

$$\frac{\partial c\left(\sigma_M;P_S\right)}{\partial \sigma_M}\equiv\frac{\partial c\left(\sigma_M;\varnothing\right)}{\partial \sigma_M}\text{ and }\frac{\partial g\left(\sigma_A;P_A\right)}{\partial \sigma_A}\equiv\frac{\partial g\left(\sigma_A;\varnothing\right)}{\partial \sigma_A}$$
$$\text{where } P_A=\{p_S\}\,,\ \{p_S,p_M\}\ \text{ or }\ \{p_S,p_A\}$$

Definition 2: Partial Strict Complementarity of assets: the assets p_M and p_S are strictly complementary if:

$$\frac{\partial T_i\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_j}\equiv\frac{\partial T_i\left(\mu_A,\mu_S;\varnothing\right)}{\partial \mu_j}\text{ and }\frac{\partial t\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_j}\equiv\frac{\partial t\left(\mu_A,\mu_S;\varnothing\right)}{\partial \mu_j}$$
$$\text{where } i=S,A,\ j=S,A \text{ and } P_M=\{p_M\}\ \text{ or }\ \{p_M,p_A\}$$

or if:

$$\frac{\partial c\left(\sigma_M;P_S\right)}{\partial \sigma_M}\equiv\frac{\partial c\left(\sigma_M;\varnothing\right)}{\partial \sigma_M}\text{ and }\frac{\partial g\left(\sigma_A;P_A\right)}{\partial \sigma_A}\equiv\frac{\partial g\left(\sigma_A;\varnothing\right)}{\partial \sigma_A}$$
$$\text{where } P_A=\{p_S\}\ \text{ or }\ \{p_S,p_A\}$$

The assets p_M and $p_A$are strictly complementary if:

$$\frac{\partial T_i(\mu_A, \mu_S; P_M)}{\partial \mu_j} \equiv \frac{\partial T_i(\mu_A, \mu_S; \varnothing)}{\partial \mu_j} \text{ and } \frac{\partial t(\mu_A, \mu_S; P_M)}{\partial \mu_j} \equiv \frac{\partial t(\mu_A, \mu_S; \varnothing)}{\partial \mu_j}$$
$$\text{where } i = S, A,\ j = S, A \text{ and } P_M = \{p_M\} \text{ or } \{p_M, p_S\}$$

or if:

$$\frac{\partial K_i(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j} \equiv \frac{\partial K_i(\alpha_M, \alpha_S; \varnothing)}{\partial \alpha_j} \text{ and } \frac{\partial k(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j} \equiv \frac{\partial k(\alpha_M, \alpha_S; \varnothing)}{\partial \alpha_j}$$
$$\text{where } i = M, S,\ j = M, S \text{ and } P_A = p_A \text{ or } \{p_A, p_S\}$$

The assets p_S and $p_A$are strictly complementary if:

$$\frac{\partial c(\sigma_M; P_S)}{\partial \sigma_M} \equiv \frac{\partial c(\sigma_M; \varnothing)}{\partial \sigma_M} \text{ and } \frac{\partial g(\sigma_A; P_A)}{\partial \sigma_A} \equiv \frac{\partial g(\sigma_A; \varnothing)}{\partial \sigma_A}$$
$$\text{where } P_A = \{p_S\} \text{ or } \{p_S, p_M\}$$

or if:

$$\frac{\partial K_i(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j} \equiv \frac{\partial K_i(\alpha_M, \alpha_S; \varnothing)}{\partial \alpha_j} \text{ and } \frac{\partial k(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j} \equiv \frac{\partial k(\alpha_M, \alpha_S; \varnothing)}{\partial \alpha_j}$$
$$\text{where } i = M, S,\ j = M, S \text{ and } P_A = \{p_A\} \text{ or } \{p_A, p_M\}$$

Definition 3: Full Independence of assets: the assets p_M, p_S and p_A are independent if:

$$\frac{\partial T_i(\mu_A, \mu_S; P_M)}{\partial \mu_j} \equiv \frac{\partial T_i(\mu_A, \mu_S; p_M)}{\partial \mu_j} \text{ and } \frac{\partial t(\mu_A, \mu_S; P_M)}{\partial \mu_j} \equiv \frac{\partial t(\mu_A, \mu_S; p_M)}{\partial \mu_j}$$
$$\text{where } i = S, A,\ j = S, A \text{ and } P_M = \{p_M, p_S, p_A\},\ \{p_M, p_S\} \text{ or } \{p_M, p_A\}$$

and

$$\frac{\partial K_i(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j} \equiv \frac{\partial K_i(\alpha_M, \alpha_S; p_A)}{\partial \alpha_j} \text{ and } \frac{\partial k(\alpha_M, \alpha_S; P_A)}{\partial \alpha_j} \equiv \frac{\partial k(\alpha_M, \alpha_S; p_A)}{\partial \alpha_j}$$
$$\text{where } i = M, S,\ j = M, S \text{ and } P_A = \{p_M, p_S, p_A\},\ \{p_A, p_M\} \text{ or } \{p_A, p_S\}$$

and

$$\frac{\partial c\left(\sigma_M; P_S\right)}{\partial \sigma_M} \equiv \frac{\partial c\left(\sigma_M; p_S\right)}{\partial \sigma_M} \text{ and } \frac{\partial g\left(\sigma_A; P_A\right)}{\partial \sigma_A} \equiv \frac{\partial g\left(\sigma_A; p_S\right)}{\partial \sigma_A}$$
$$\text{where } P_A = \{p_S, p_M, p_A\} \ \ \{p_S, p_M\} \text{ or } \{p_S, p_A\}$$

Definition 4: Partial Independence of assets: the assets p_M and p_S are independent if:

$$\frac{\partial T_i\left(\mu_A, \mu_S; P_M\right)}{\partial \mu_j} \equiv \frac{\partial T_i\left(\mu_A, \mu_S; p_M\right)}{\partial \mu_j} \text{ and } \frac{\partial t\left(\mu_A, \mu_S; P_M\right)}{\partial \mu_j} \equiv \frac{\partial t\left(\mu_A, \mu_S; p_M\right)}{\partial \mu_j}$$
$$\text{where } i = S, A,\ j = S, A \text{ and } P_M = \{p_M, p_S\}$$

and

$$\frac{\partial c\left(\sigma_M; P_S\right)}{\partial \sigma_M} \equiv \frac{\partial c\left(\sigma_M; p_S\right)}{\partial \sigma_M} \text{ and } \frac{\partial g\left(\sigma_A; P_A\right)}{\partial \sigma_A} \equiv \frac{\partial g\left(\sigma_A; p_S\right)}{\partial \sigma_A}$$
$$\text{where } P_A = \{p_M, p_S\}$$

The assets p_M and p_A are independent if:

$$\frac{\partial T_i\left(\mu_A, \mu_S; P_M\right)}{\partial \mu_j} \equiv \frac{\partial T_i\left(\mu_A, \mu_S; p_M\right)}{\partial \mu_j} \text{ and } \frac{\partial t\left(\mu_A, \mu_S; P_M\right)}{\partial \mu_j} \equiv \frac{\partial t\left(\mu_A, \mu_S; p_M\right)}{\partial \mu_j}$$
$$\text{where } i = S, A,\ j = S, A \text{ and } P_M = \{p_M, p_A\}$$

and

$$\frac{\partial K_i\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_j} \equiv \frac{\partial K_i\left(\alpha_M, \alpha_S; p_A\right)}{\partial \alpha_j} \text{ and } \frac{\partial k\left(\alpha_M, \alpha_S; P_A\right)}{\partial \alpha_j} \equiv \frac{\partial k\left(\alpha_M, \alpha_S; p_A\right)}{\partial \alpha_j}$$
$$\text{where } i = M, S,\ j = M, S \text{ and } P_A = \{p_A, p_M\}$$

The assets p_S and p_A are independent if:

$$\frac{\partial c\left(\sigma_M; P_S\right)}{\partial \sigma_M} \equiv \frac{\partial c\left(\sigma_M; p_S\right)}{\partial \sigma_M} \text{ and } \frac{\partial g\left(\sigma_A; P_A\right)}{\partial \sigma_A} \equiv \frac{\partial g\left(\sigma_A; p_S\right)}{\partial \sigma_A}$$
$$\text{where } P_A = \{p_S, p_A\}$$

and

$$\frac{\partial K_i\left(\alpha_M,\alpha_S;P_A\right)}{\partial \alpha_j}\equiv\frac{\partial K_i\left(\alpha_M,\alpha_S;p_A\right)}{\partial \alpha_j}\text{ and }\frac{\partial k\left(\alpha_M,\alpha_S;P_A\right)}{\partial \alpha_j}\equiv\frac{\partial k\left(\alpha_M,\alpha_S;p_A\right)}{\partial \alpha_j}$$
$$\text{where } i=M,S,\ j=M,S \text{ and } P_A=\{p_A,p_S\}$$

Definition 5: Essential Human Capital: M's human capital is essential if:

$$\frac{\partial K_S\left(\alpha_M,\alpha_S;P_A\right)}{\partial \alpha_j}\equiv\frac{\partial K_S\left(\alpha_M,\alpha_S;\varnothing\right)}{\partial \alpha_j}\text{ and }\frac{\partial k\left(\alpha_M,\alpha_S;P_A\right)}{\partial \alpha_j}\equiv\frac{\partial k\left(\alpha_M,\alpha_S;\varnothing\right)}{\partial \alpha_j}$$
$$\text{where } j=M,S \text{ and } P_A=\{p_A,p_S,p_M\},\ \{p_A,p_S\},\ \{p_A,p_M\}\ \text{ or }\ \{p_A\}$$

and:

$$\frac{\partial c\left(\sigma_M;P_S\right)}{\partial \sigma_M}\equiv\frac{\partial c\left(\sigma_M;\varnothing\right)}{\partial \sigma_M}$$
$$\text{where } P_S=\{p_A,p_S,p_M\},\ \{p_S,p_M\},\{p_S,p_A\}\ \text{ or }\ \{p_S\}$$

S's human capital is essential if:

$$\frac{\partial T_A\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_j}\equiv\frac{\partial T_A\left(\mu_A,\mu_S;\varnothing\right)}{\partial \mu_j}\text{ and }\frac{\partial t\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_j}\equiv\frac{\partial t\left(\mu_A,\mu_S;\varnothing\right)}{\partial \mu_j}$$
$$\text{where } j=S,A \text{ and } P_M=\{p_A,p_S,p_M\},\ \{p_M,p_S\},\ \{p_M,p_A\}\ \text{ or }\ \{p_M\}$$

and

$$\frac{\partial K_M\left(\alpha_M,\alpha_S;P_A\right)}{\partial \alpha_j}\equiv\frac{\partial K_M\left(\alpha_M,\alpha_S;\varnothing\right)}{\partial \alpha_j}\text{ and }\frac{\partial k\left(\alpha_M,\alpha_S;P_A\right)}{\partial \alpha_j}\equiv\frac{\partial k\left(\alpha_M,\alpha_S;\varnothing\right)}{\partial \alpha_j}$$
$$\text{where } j=M,S \text{ and } P_A=\{p_A,p_S,p_M\},\ \{p_A,p_S\},\ \{p_A,p_M\}\ \text{ or }\ \{p_A\}$$

A's human capital is essential if:

$$\frac{\partial T_S\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_j}\equiv\frac{\partial T_S\left(\mu_A,\mu_S;\varnothing\right)}{\partial \mu_j}\text{ and }\frac{\partial t\left(\mu_A,\mu_S;P_M\right)}{\partial \mu_j}\equiv\frac{\partial t\left(\mu_A,\mu_S;\varnothing\right)}{\partial \mu_j}$$
$$\text{where } j=S,A \text{ and } P_M=\{p_A,p_S,p_M\},\ \{p_M,p_S\},\ \{p_M,p_A\}\ \text{ or}\,\{p_M\}$$

and

$$\frac{\partial g\left(\sigma_A; P_S\right)}{\partial \sigma_A} \equiv \frac{\partial g\left(\sigma_A; \varnothing\right)}{\partial \sigma_A}$$

where $P_A = \{p_A, p_S, p_M\}\,,\ \{p_S, p_M\}\,, \{p_S, p_A\}$ or $\{p_S\}$

Bibliography

Binmore, K, 1987, "Perfect Equilibria in Bargaining Models", in Binmore & Dasgupta eds., The Economics of Bargaining, Basil Blackwell, Oxford

Bolton, P & Whinston, M D, 1993, "Incomplete Contracts, Vertical Integration and Supply Assurance", Review of Economic Studies 60:121-148

Chiu, S, 1998, "Noncooperative Bargaining, Hostages, and Optimal Asset Ownership", American Economic Review 88:882-901

De Meza, D & Lockwood, B, 1998, "Does Asset Ownership Always Motivate Managers? Outside Options and the Property Right Theory of the Firm, Quarterly Journal of Economics 113:361-386

Grossman, S J & Hart, O, 1986, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", Journal of Political Economy 94:691-719

Hart, O & Moore, J, 1990, "Property Rights and the Nature of the Firm", Journal of Political Economy 98:1119-1158

Hart, O, 1995, *Firms, Contracts and Financial Structure*, Claredon Lectures in Economics, Claredon Press, Oxford

Hart, O & Moore, J, 2008, "Contracts as Reference Points", Quarterly Journal of Economics CXXIII:1-48

Hart, S & Mas-Colell, A, 1996, "Bargaining and Value", Econometrica 64:357-380

Matouschek, N, 2004, "Ex post Inefficiencies in a Property Rights Theory of the Firm", Journal of Law, Economics and Organization 20:125-147

Muthoo, A, 1999, *Bargaining Theory with Applications*, Cambridge University Press, Cambridge

Osborne, M J & Rubinstein, A, 1990, *Bargaining and Markets*, Academic Press, San Diego

Rubinstein, A, 1982, "Perfect Equilibrium in a Bargaining Model", Econometrica 50:97-110

Stigler, G., 1951, "The Division of Labor Is Limited by the Extent of the Market", Journal of Political Economy 59:185-193

Chapter 4

Public Hospitals - Incentives and Organization

Abstract: Chapter 4 presents a novel way to analyze the organization of hospitals, with special focus on public hospitals. The novelty is that the property rights approach to organization (PRA) is used to analyze the problem. It is proposed that PRA is suitable for the analysis of all hospitals and especially so for public hospitals. The analysis explores issues concerning privatization and integration of public hospital services. The findings are generally supportive of integration as long as the public principal's human capital is essential for the production of hospital care.

Keywords: Public Hospitals, The Property Rights Approach to Organizations, Joint Production, Integration, Privatization

JEL Classifications: D23, I18

4.1 Introduction

Public hospitals are a central feature of public health care systems. Providing specialized care, they are at the centre of attention of citizens as well as politicians, and they represent the bulk of health care expenditure.

Hospital care, both in private and public settings, is the joint production of care by several different specialities (e.g. Harris, 1977). Obviously joint production requires cooperation and better cooperation renders higher quality of hospital care. Good cooperation entails that the cooperating parties make investments in their relationship e.g. learn about the other parties' needs and modify human capital and assets to suit these needs. Undoubtedly, a hospital organization that supports good cooperation is a prerequisite, but not a guarantee, for efficient and high quality hospital care.

The economic literature dealing with hospitals typically focuses on three issues: ownership (public, not-for-profit or for-profit see overview in Sloan (2000)), economies of scale (optimal scale of hospitals, see overview in Posnett (2002)), or reimbursement issues (see e.g. Ma, 1994). The actual modelling of hospitals has received less interest, with Newhouse's (1970) model of not-for-profit hospitals and Harris' (1977) conceptual model of the internal organization of US hospitals being the obvious exceptions. The internal organization of public hospitals has aroused even less interest among economists.

This paper proposes, and argues, that the property rights approach to organization (PRA)[1], developed in Grossman & Hart (1986), Hart & Moore (1990) and Hart (1995), is conducive to the analysis of public hospitals. Notably, this approach is new to the hospital literature. Moreover, PRA (based on the exposition in Hart (1995)) is used to analyze the organization of public hospitals. Using this approach, a comparative study is performed, yielding insights into privatization and integration of hospital services. This is done in two scenarios: *privatization*, where an integrated structure serves as the starting point for the analysis and *construction*, where non-integration serves as the starting point. In the latter scenario it is envisioned that the hospital principal wants to construct a new treatment unit within a hospital or a new hospital, while the principal in the first scenario considers reorganizing an already integrated structure.

There are three broad types of departments/services in most hospitals, *medical, support* and *hotel* services. The focus of the paper is on the bilateral relationships between departments within public hospitals, e.g. between a surgery department and a radiology department, while relationships involv-

[1] Also know as the Grossman-Hart-Moore model.

ing more than two parties are only briefly discussed. The focus of the analysis is on *between-department-cooperation* while *within-department-cooperation* is abstracted from.

This chapter is organized as follows: next, the fundamentals of hospital care and health care markets are introduced followed by a short introduction of the property rights approach and an illustrative example. Section 4.2 discusses the conduciveness of PRA to analyzing public hospital organization. Section 4.3 analyzes two different bilateral relationships within public hospitals, and section 4.4 takes a brief look at the extension to trilateral relationships. Section 4.5 concludes the chapter.

Health Care Markets, Hospital Production and Organization When studying the organization of hospital care one must take into account the specific nature of health care markets. Typically, they are markets characterized by decentralized decision-making, asymmetric information, unforeseen contingencies, and irreducible uncertainty.[2] [3] This implies that neither health care outcomes nor provider behavior can be contracted on in a contingent manner. These features of the health care market also apply to the internal organization of hospitals. It is not possible for hospital managers to write complete contracts with the different parts of the hospital specifying treatment outcomes, all input characteristics, and individual efforts.

In general a hospital consists of a wide variety of services and assets, all to some extent needed in the production of hospital care. The most evident services are of course the medical services such as surgery, cardiology,

[2] Irreducible uncertainty is the absence of information about the consequences of health care treatment that is shared equally by the providers, patients and payers (cf. McGuire, 2000). It essentially implies that recovery from disease is a random event. While health care influences this randomness, it will never be a sure thing that a treatment always gives the same outcome or requires the same inputs.

[3] The characteristics of the health care markets have inspired ample research yielding three strands of literature: insurance literature dealing with the relation payer-patient (e.g. Arrow 1963, Pauly 1968, Zeckhauser 1970 and Nyman 1999), physician agency literature dealing with the relationship between patient and provider (e.g. Evans 1974, Dranove 1988, and a nice overview in McGuire 2000), and reimbursement literature dealing with the payer-provider relation (e.g. Ellis & McGuire 1986, Ma 1994, Chalkley & Malcomson 1998 and Eggelston 2004).

and oncology. Without these medical services there would be no production of hospital care. However the production of the final good, hospital care, requires a number of inputs from other services that are usually found in hospitals, foremost support services (e.g. radiology, pharmacy, blood-bank and laboratories) but also hotel services (cooking, laundry, cleaning, transports and so on).

The production of hospital care is the *joint production* of care carried out by a number of interdependent medical departments and support services (and hotel services) (Harris, 1977). In the joint production of hospital care every patient "receives customized attention"[4], due to unforeseen contingencies (e.g. complications) and patient characteristics, consisting of inputs from different parts of the hospital. The customized attention that each patient receives requires decision making to be decentralized. That is, that patient care decisions are made by the treating hospital departments because they have information about the patient that the hospital management does not have - there is asymmetric information about the treatment needs of the patient.[5] Moreover, the irreducible uncertainty about treatment outcomes makes it difficult for the hospital management to assess the departments' actions *ex post*. This implies that the hospital management cannot contract, in an enforceable manner, the hospital departments to take certain actions e.g. to ensure qualitative cooperation in the joint production of hospital care. Instead the organization of the joint production in itself must be such that it ensures a high level of cooperation and coordination. The most common organizational structure for hospitals is vertical integration (see e.g. Coles & Hesterly, 1998, Söderström & Lundbäck, 2002) possibly in response to this challenge, but there might also be other explanations such as inertia and/or political control (public hospitals). This paper evaluates how different organizational forms (typically vertical integration and non-integrated contractual networks) cope with the challenges of joint production.

[4]See Harris (1977).

[5]Physicians (medical departments) will, in most instances, have an informational advantage over both hospital managements and patients (see e.g. Arrow, 1963).

The Property Rights Approach PRA focuses on the importance of asset ownership for the *relationship-specific investments* made in a transaction (trade relationship). In the model, investments follow asset ownership and the organizational form is defined by asset distribution, e.g. under integration one party owns all the assets used in the transaction. Relationship-specific investments are investments that are more valuable in the transaction than outside the transaction.[6] That is, the investments ensure that the trade relationship becomes more efficient and beneficial (e.g. through greater coordination and better cooperation) for the trading parties. The transaction is carried out in a world of incomplete contracting, which creates a potential for hold-ups, i.e. by making the investment, the investing party becomes vulnerable to withdrawal from trade by a party that does not invest or invests less.

In contrast to transaction cost theory,[7] PRA suggests that integration does not automatically solve/reduce the hold-up problem. Instead PRA contends that opportunistic behavior may prevail within firms (cf. Hart 1995). Heeding this possibility, PRA provides a framework for understanding the boundaries of a firm. The optimal organizational structure is the structure that yields the greatest incentives for relationship-specific investments.

The relationship-specific investments may be interpreted in terms of effort for coordination and cooperation; thus PRA provides a good framework for analyzing the joint production of hospital care and its organization. The complete rationale for using PRA in the case of hospital care is discussed in section 4.2.

Cardiovascular Intensive Care - an Illustrative Example A hospital manager may face two types of organizational decisions: reorganization of old treatment units, or whole hospitals, and organization of new treatment units (construction). For public hospitals the former usually includes priva-

[6] Relationship-specific investments are investments by some party A in the relationship with another party B, and vice versa, that increase the mutual dependence between the parties - ensuring a more rewarding cooperation between A and B; i.e. increasing the value of the relationship.

[7] E.g. Klein *et al* 1978, Williamson 1985.

tization or, if you like, disintegration because public hospitals in most cases are vertically integrated structures. Thus, two scenarios are discussed in this example: privatization and construction. With privatization the initial organization is vertical integration and with construction it is non-integration.

To keep things simple consider the hospital treatment of *acute myocardial infarction* (heart attack, AMI) by a cardiovascular intensive care unit, and assume that the public principal organizing the treatment also functions as the cardiology department. The cardiology department produces the final good, hospital care for AMI, using inputs from radiology and thoracic surgery. [8] The production of the final good is a combination of treatment at the cardiology department and decisions on what inputs to use. The cardiology department receives information from ultrasonography, supplied by the radiology department, to determine the size and severity of the thrombosis. Then they administer thrombolytic therapy but find that the treatment is ineffective and opt for bypass surgery. The bypass surgery is performed by a thoracic surgeon, using information from an *angiography (diagnostic cardiac catherization)* to locate the infarct.[9] Figure 4.1 gives a schematic depiction of the hospital treatment of AMI i.e. the involved departments, the inputs and the assets that each department uses for the production of their input.

For this treatment sequence to function smoothly the three departments need to invest in their relationship, e.g. the cardiology department needs to invest in knowledge about thoracic surgery and interpretation of sonograms, the thoracic surgeon needs to be able to specify requirements regarding diagnostic technologies and know how thrombolytic therapy affects the bypass surgery, and the radiology department needs to invest in knowledge about ultrasonography and CT-scanners, MRI *etc.* and AMI. In this paper it is hypothesized that the more the parties invest in knowledge the better is the joint production of hospital care. The question that arises is: how should the cardiovascular intensive care unit be organized to foster the greatest in-

[8] Abstracting from e.g. first-responder care, anesthetics, post-operative care and rehabilitation.

[9] The cardiology department might also opt for angioplasty, in this case bypass surgery would serve as an emergency backup. Moreover the angiography results are supplied to the cardiology department in the case of PCI and possibly in this case as well.

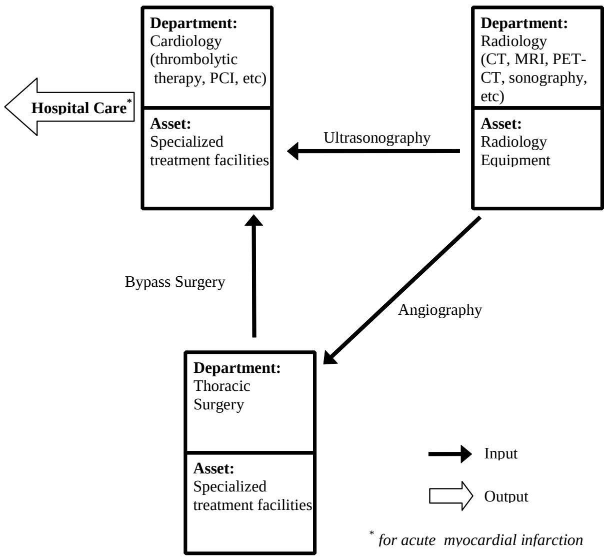

Figure 4.1: A possible scenario for the treatment of AMI

vestments?

The answer (as will be seen in the subsequent sections) will typically depend on the starting point (i.e. the initial organization), the characteristics of assets (independent or complementary) and the characteristics of human capital (essential or non-essential). Starting with the characteristics of assets, it is obvious that the assets are complementary, i.e. it is hard to imagine that the treatment could be performed without all three parties' assets e.g. without a heart-lung machine (for the bypass surgery), a sonograph and facilities for thrombolytic therapy. Moreover, one might argue that the human capital of the cardiology department as well as the thoracic surgeon is essential for the treatment, while the radiology department's human capital may be important but not essential. How do these suggestions affect the organizational choice?

In the construction case, vertical integration (i.e. the cardiology department owning all assets) should be opted for if the complementarity between assets implies that the thoracic surgeon is indifferent between owning and not owning its facilities. In all other circumstances either integration between thoracic surgery and radiology or integration between cardiology and radiology, with the third party as an outside contractor, would be a better option.

Moreover, privatization of parts of an already existing, vertically integrated, treatment unit is only an improvement if both the radiology department's and the thoracic surgeon's human capital is essential, while the cardiology department's human capital is relatively unimportant. Given the setup in this example, this is an unlikely situation.

There are ample examples of treatments performed at public hospitals, besides the treatment of AMI, that could fit this picture. Any treatment involving two treating departments and one support service fits directly e.g. the treatment of breast cancer involving a radiology department (e.g. mammography), a surgeon (removal of the tumor) and an oncology department (e.g. chemotherapy). Furthermore, with proper adaptations, examples of treatments using one treating department and one or two support services may be constructed e.g. hip arthroplasty involving an orthopedic surgeon, a radiology department (in evaluation stage) and an anesthesiologist (for anesthesia during the operation).

4.2 The Property Rights Approach and Hospital Organization

PRA - Assumptions and the Basic Mechanism The simplest version of the PRA-model is a two-period model with two parties, one producing the final product and the other an input to this production (Grossman & Hart, 1986, Hart, 1995). These two parties plan to trade with each other; but if they fail to reach an agreement at that time, then each party has a *disagreement option*. Typically this option is to sell and buy the input on a spot-market. In this paper the disagreement option is modeled, explicitly, as a threat point (or if you like *status quo* option or inside option) i.e. it is an option that is always available to the parties until they reach an agreement (start trading) and its existence does not foreclose a future agreement.[10]

There is no uncertainty about costs and benefits, and no asymmetric information in this model. It is also assumed that the parties have unlimited wealth to ensure that any transfer of assets is possible; this is equivalent to assuming that asset transfers are costless. Moreover, the parties can make correct calculations about the expected return of any action. However, there is *ex ante* uncertainty about the quality of the input - its characteristics cannot be contracted on in period 0.

Figure 1.1 in chapter 1 describes the timing of the model. Notably assets are already allocated, i.e. the organizational structure for the transaction is decided when the investments are made in period 0. In period 1 the parties trade with each other, the uncertainty about input quality is resolved, and the parties bargain over the division of surplus given the threat points.[11]

In PRA the interaction between asset ownership and relationship-specific investments determines the organizational structure. All investments are beneficial for the investing party irrespective if she trades with the other party or not. However, the benefit is greater if the investing parties trade with each other.[12] Furthermore, these investments are typically assumed to be complementary and to be made in the parties' respective human capital.[13] Notably, human capital cannot be transferred between the parties (cannot, in fact, be owned by any other party) and investments in human capital are typically connected to the assets e.g. acquiring knowledge about how to use an asset for a certain purpose.

Asset ownership gives the asset owner control over contingencies concerning the assets that are not specified in the contract (residual control rights).

[10] Notably, De Meza & Lockwood (1998) and Chiu (1998) show that the predictions of the PRA-model are vulnerable to changes in bargaining assumptions i.e. that the results of the model change if the disagreement options are viewed as outside options instead of threat points. Outside options are options that are available only after negotiations have permanently broken down.

[11] The bargaining solution used is symmetric Nash bargaining.

[12] Investments make the use of assets more productive, especially in conjunction with the other party's asset if this party makes similar investments, but also if the investing parties do not trade with each other - albeit to a lesser extent.

[13] Investments in physical capital are discussed briefly by Hart (1995) and are addressed in section 4.3.2 of this chapter.

Residual control rights are important given the incomplete contracting in the model stemming from the presence of uncertainty and unforeseen contingencies. In general, asset ownership creates greater incentives for investments because the increased control makes the disagreement option more attractive and more sensitive to investments, making the asset owner less vulnerable to hold-ups (withdrawal from trade by the other party).

As noted, a *period* 0*-contract*, e.g. establishing the basic conditions for the transaction, cannot specify the relevant characteristics (the quality) of the input in a verifiable manner.[14] If this was possible the parties would internalize the effect of investments (on surplus) and hence choose their investments to maximize the *ex post* surplus - giving the *first-best* investments. The model thus depicts a *second-best* situation where the parties cannot fully internalize the effect of investments on surplus due to contractual incompleteness.[15]

In the basic setup the incentives for investments crucially depend on asset allocation, which enables a comparative analysis of the organizational structure. In the comparative analysis the best organizational structure is the structure that supports the greatest relationship-specific investments, and thus creates the greatest surplus.

Finally, three additional features of PRA deserve mentioning before moving on to specific issues about PRA and hospitals. First, any contract may be renegotiated in the model at any time (until the transaction is carried out) at zero cost, implying that there is a lack of commitment in the model i.e. no period 0-contract is binding. This is due to the unforeseen contingencies. Second, the bargaining is assumed to be *ex post* efficient, in line with the renegotiation argument above, and thus the focus is on *ex ante* inefficiencies in the investment decisions. Finally, the model focuses on issues concerning organization and abstracts from demand (consumer side) issues that might affect the benefit of a certain organization. This is of course a shortcoming of the model, but an arguably inconsequential shortcoming in the case of public

[14] E.g. it is difficult to *ex ante* describe all relevant characteristics of a specific X-ray image, e.g. in terms of precision, workload and interpretability - these might vary considerably from patient to patient.

[15] Note that the investments are observable to both parties, but not verfiable to outsiders (not enforceable).

hospitals.[16]

PRA and Hospitals There are two fundamental circumstances that decide the optimal organization of hospital care, both in public and private settings: hospital care being joint production and complete contracting being impossible. The property rights approach to organization captures both these aspects of hospital care. First, PRA is an incomplete contracting model where the quality of inputs cannot be contracted on and any contract may be renegotiated i.e. the parties cannot commit themselves to a contract - an effect of the unforeseen contingencies and the uncertainty in the transaction. Both unforeseen contingencies and uncertainty seem to be prevalent features of hospital/health care. Second, and most profoundly, the emphasis on relationship-specific investments in PRA captures the essence of joint production; the coordination and mutual dependence of the different parties contributing to hospital care.

It is obvious that if the joint production does not work properly the hospital care will be of poor quality. Joint production requires coordination of efforts. The individual efforts are generally not verifiable, implying that the quality of each party's input to the production of hospital care cannot be contracted on. One way to coordinate the efforts of the parties is to make them mutually dependent on each other. Relationship-specific investments have the potential to ensure a mutual dependence of different parts of the hospital and thus ensure coordination of efforts, even if the investments are made non-cooperatively, i.e. even if they are chosen to maximize individual benefit. The relationship-specific investments in this setting are modifications of human capital and/or assets to meet the special demands of a specific transaction. These investments ensure that the joint production of hospital care runs smoothly and are therefore of great importance.

Example: *A radiology department, for example, has to supply the surgical department with specific diagnostic information and, to be able to do*

[16] There are of course a number of other shortcomings in the model that apply to the entire strand of literature and need not be repeated here. See e.g. Hart & Moore (2008), Holmström (1999) and Whinston (2003).

this, they have to make investments in human capital - learn the special requirements of the surgeon - or in the asset - buy equipment that produces the type of image that the surgery requires. The surgeon, on the other hand, must invest in the interpretation of the diagnostic information (sometimes the interpretation of the actual images).

How does this relate to hospital organization? The link between relationship-specific investments and organization goes through the adaptation to changing circumstances, and realtionship-specific investments make the parties, vulnerable to withdrawal from trade (unforeseen contingencies), since contracts are incomplete.

Example: *If the radiology equipment needs to be modified, e.g. through the purchase of new appliances, to fit a special surgical procedure in a way that is not specified in the contract (between the radiology department and the surgical department), then the radiology department will have the ability to do so if it owns the equipment. The surgical department, however, is dependent on the radiology department's willingness to make this alteration. Since the surgical department has made relationship-specific investments in the relation with this specific radiology department it now may be "held up" for some of the cost of the modification. The threat of being held up by the radiology department reduces the surgical department's incentives to make investments in this specific relation.*

The ability to change and the cost of changing follow ownership, and it is apparent that ownership also affects the relationship-specific investments. It is likely that the surgical department would have greater incentives to invest if it owned the radiology equipment, since it would then have control over the adaptation needed in the new circumstances (residual control rights).

Relationship-specific investments are, or at least seem to be, beneficial for the joint production of hospital care, and hospitals should therefore be organized in a way that promotes these investments. This fits well with the conclusion of PRA that the organizational structure that supports the greatest relationship-specific investments is the optimal way to organize a certain transaction.

Moreover PRA provides a framework for capturing salient features of hospital care such as the medical departments' human capital, in most cases, being essential for the treatment outcome and that the different departments' assets are likely to be complementary in the treatment. An example of the latter could be the radiology equipment used for mammography and cytologist's equipment used to examine the cells; these two assets are complementary in the diagnosis and treatment of breast cancer.

PRA and Public Hospitals This paper contends that PRA is especially suitable for analyzing public hospitals, which like many other public agencies, are characterized by politically set goals, public funding and ownership, and lack of competition.[17] The public funding of hospital care is in most cases, e.g. in Sweden, based on a fixed budget principle. When it comes to the internal funding of the departments' services the pattern is similar, i.e. the funding is given in terms of a budget. This total cost may be based on historical as well as predicted treatment costs given the internal prices for the different services needed for a treatment. There is room in this system for disagreements and negotiations over the actual cost, i.e. resources used, for a treatment and over the quality of a service given the price. PRA, in a sense, captures these negotiations since it entails bargaining over the reimbursement for a service - this bargaining may be interpreted in terms of negotiations over both price for a given quality and quality for a given price.[18]

Lack of competition between hospitals is a significant feature of public health care systems, e.g. the Swedish health care system where the public hospital serves a regulated part of the population. Thus demand does not discipline hospital behavior and affect hospital organization as it potentially does in a private health care system. PRA's focus on organizational issues and the lack of a demand side fits well with analyzing public hospitals. Just as it would be natural to include demand issues in an analysis of private hospitals, it is natural to abstract from demand issues in the analysis of public hospitals with regional monopolies.

[17] See Dixit (2002) for a general discussion of characteristics of public agencies.

[18] The fact that contracts may be renegotiated without cost, in the model, captures another public feature namely that public agencies often have the political power to breach and renew agreements at any time.

Since the threat points are important for the results in the PRA it is important to establish the existence and nature of threat points in public hospitals. There are two prolific ways to think of threat points in a public hospitals setting: 1) as an internal spot-market for hospital services or 2) as an external spot-market for hospital services.[19] 1) implies that each department involved in a transaction may trade with other departments outside the transaction, but within the realms of the hospital, until an agreement is reached.

Example: *A medical department (M) needs diagnostic information from a radiology department (imagine that there are two radiology departments in the hospital: radiology I and II). They plan to acquire this information from radiology I and both departments make investments in this relationship to ensure that their cooperation runs smoothly. When negotiating the reimbursement for this diagnostic information, the medical department's threat point is to acquire diagnostic information (without the specific features that I's investment entails) from radiology II - in fact they acquire information from radiology II until an agreement is reached with radiology I. Radiology I, on the other hand, may prioritize its transactions with other medical departments, and not benefit from the investments made in the relationship with M, over the transaction with M until an agreement is reached. This threat point may be interpreted as radiology I is putting M's request at the bottom of the "to-do list".*

The second interpretation of threat points implies that there is a demand for special hospital services outside the actual hospital - typically from other hospitals.

Example: *The medical department (M) and radiology I may both sell their services to (nearby) hospitals that do not have M's specialist abilities or a radiology department, until they reach an agreement.*[20]

[19]Spot-market in the sense that there a near immediate delivery of the services, but not in the sense that they are paid for in cash.

[20]Both interpretations abstract from the need to reach an agreement with the "threat-point" trading parties. This is, however, reasonable since these trades entail a transfer of a generic input and there is typically an agreed upon price-list for these inputs. This is in line with thinking about the negotiations as dealing with quality for given price, where quality stems from the relationship-specific investments.

Either of these interpretations of threat points seem valid in most public settings. The specifics of the threat points, in terms of payoffs *etc.*, irrespective of interpretation, are discussed in the next section.

The choice of PRA is also justified by its ability to address relevant issues such as privatization and restrictions on ownership. A major issue in the political debate concerning public hospitals is whether certain activities could, and should, be privatized or, put in other words, disintegrated - this in a political environment where the need for political control spurs a preference for hierarchic public organizations, mainly because this limits the autonomy of the organization (cf. Williamson 1999). The PRA may be used to evaluate the benefits from disintegration by using integration as the starting point for the analysis. A related issue is politically assigned restrictions on ownership. A common characteristic of public ownership, from a historical point of view, is that public assets are seldom for sale. Obviously, some public assets are never or seldom sold to private interests, e.g. military defense facilities, and, at least in the Swedish case, hospital facilities.[21] This characteristic is easily captured in PRA by assuming some restrictions on ownership. This kind of assumption will be used in the analysis below.

Finally, the PRA assumes that the parties have unlimited wealth, which is an unjust assumption in many settings. In the case of public hospitals it seems less restrictive and in some sense even reasonable. In particular, hospital departments typically have a soft budget constraint; publicly funded hospital departments may run a deficit years on end without going bankrupt. Moreover, the implication of the assumption is that assets may always be transferred so that they are put to their best use (cf. Hart, 1995) and it is reasonable to believe that public hospital principals have the authority and the means to transfer assets if needed.

Limitations of the PRA There are some obvious limitations to applying PRA to public hospitals; this subsection is devoted to a discussion of these

[21] To my knowledge there are only three private hospitals in Sweden - St Göran, Lundby and Simrishamn - the latter two are minor hospitals. They are private in the sense that the provision of care is by a private entrepreneur, but they are publicly funded and the infrastructure/buildings, as I understand it, are still publicly owned.

limitations. There are two types of limitations: one type that arises because assumptions or mechanisms in PRA are unsuitable for the analysis of public hospitals, and a second type that stems from public hospital characteristics not being readily captured by the PRA.

There are two limitations of the first type:

- There is an assumption of no asymmetric information between trading parties in the PRA, which is at odds with the contention that there is asymmetric information between, at least, hospital managers and treating departments. It might be argued that the information asymmetry is smaller between departments that interact with each other on a daily basis than between the management and the departments. In spite of the latter it is an obvious simplification of the informational structure between departments to assume no asymmetric information.

- In PRA the incentives for investments are provided by asset ownership. This does not give a full-fledged picture of the incentives within hospitals. Incentives for investments in human capital are to a great extent provided by career opportunities, salary, peer reviews and so on. However, to make investments it is important for the investing party to know that this investment pays off, and asset ownership might be a way to ensure this since it gives control over the use of the asset. Hence, the residual-control right, to say it in PRA-terms, which stems from asset ownership, plays a part in providing the incentives for investments, but it is a simplification to assume that it is the sole source of investment incentives.

There are three limitations of the second type:

- As hospital care is typically characterized by team-work not only between departments but also within departments, the workers' incentives within the departments are important for the outcome. The PRA generally focuses on management incentives i.e. treats the department as a black box implicitly assuming incentive alignment between the department management and the department staff.[22] This is an obvious drawback of using PRA to analyze public hospitals, but also a simplification that highlights the focus on between-department cooperation.

- The PRA-model presented here, and elsewhere, describes a quite simple structure while the actual structure of public hospitals is complex. Still, this kind of simplification is inevitable when dealing with a theoretical model. The discussion of trilateral transactions in section 4.4 might be considered as a step in the right direction.

- Investment decisions are generally centralized in public hospitals. The public hospital principal decides on investments and prioritizes among investments, but the initiatives for investments are likely to stem from the individual departments. The departments appeal for resources, from the principal, for investments and are responsible for their implementation (in particular investments in human capital). These appeals are made non-cooperatively (i.e. in the private interest of the departments) and it is unlikely that the principal can force cooperative investment upon the departments (at least given the decentralized implementation of investment). In the PRA-model the investment decisions are decentralized and non-cooperative, not capturing the centralized decision-making. Nonetheless, it may well capture the potential internal conflicts arising from appeals for investments and that investments, typically, are implemented by departments (one might consider the implementation as an investment - once again think about investments in human capital). However, the fact that PRA does not capture centralized investment decisions is an obvious limitation.

In spite of these limitations it is suggested in this chapter that applying PRA to the organization of public hospitals might be rewarding, in particular since it nicely captures the joint production of hospital care. Moreover, this

[22] Hart & Moore (2008) state that the PRA-model is unsuitable for analyzing the internal organization of firms in terms of workers' incentives, authority and so on.

approach is novel to literature and might serve as a stepping stone towards a full-fledged model of public hospitals. In the next section PRA is applied to the most basic transaction within a public hospital namely a *bilateral transaction* where two departments cooperate to produce hospital care.

4.3 Bilateral Transactions in Public Hospitals

As discussed and exemplified in previous sections, hospital care is the joint production of medical services, support services and hotel services. In its most basic form the joint production involves two parties, i.e. it is a *bilateral transaction*, e.g. a medical department and a support service or a medical department and a hotel service. There are many examples of bilateral transactions of this kind within public hospitals e.g. the treatment of a simple leg or arm fracture (orthopedic department and radiology department) or a treatment requiring a special diet (medical department and hospital kitchen) and so on.[23] The plenitude of potential bilateral transactions within a hospital makes the organization of these transaction both important for hospital performance and interesting to analyze.

In the analysis below one may think of the bilateral transactions either as the simplest incarnation of a hospital or as a treatment unit within a larger hospital; both interpretations are valid. Below two, somewhat different, bilateral transactions are considered. First, a transaction with one medical department and one support service, where both parties make investments in human capital. Second, a transaction involving a medical department and a hotel service is considered. In both cases the medical department produces the final good and also functions as the public principal i.e. this party formulates the overall goals, the priorities etc. for hospital care (this is not modelled). In the latter transaction one party invests in its physical asset while the other party invests in its human capital. The basic features of PRA, discussed in section 4.2, apply in both cases and in both cases the

[23] The example in section 1, of course, deals with a more complex situation.

comparative analysis is performed *vis-a-vis* two starting points: integration (to analyze the privatization of hospital services) and non-integration (to analyze the construction of treatment units). Concerning the analysis, the setup for the first transaction is similar to the setup used by Hart (1995), while the analysis of the second transaction contains some novel features.

4.3.1 Medical Department plus Support Service

Consider a setting with one support service (e.g. radiology department) and one medical department (e.g. surgery department). Denote the support service S. Hence S supplies an input to M's production of the final good (hospital care). The support service uses one asset, p_S, to produce the input and the medical department uses one asset, p_M, and the input to finalize the hospital care. Figure 4.2 gives a visualization of the production process, the assets used for the production and introduces the relationship-specific investments (M's investment in the relation with μ_S and S's investment in the relation with σ_M).

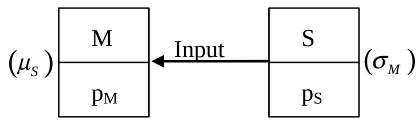

Figure 4.2: "The M & S Transaction"

S's investment (σ_M) enables it to produce an input suitable for M's production of hospital care. Moreover, M's investment (μ_S) allows it to make efficient use of the input. These investments are in human capital. The assumption that they are made in the respective parties' human capital is important and implies that they cannot be transferred from one party to the other. That is S cannot make M's investment and vice versa - irrespective of ownership structure. Assume that these investments reflect both the level and the cost of the investment (cf. Hart, 1995).

The Model

If both parties decide to enter this particular transaction (i.e. reach an agreement over the terms of trade), where S supplies the input and M uses the input to produce hospital care, then their payoffs are the following:

$$U_M = T(\mu_S) - v \tag{4.1}$$
$$U_S = v - C(\sigma_M) \tag{4.2}$$

where $T(\mu_S)$ is the treatment outcome and v is the reimbursement paid to the support department for the input. Assume that the treatment outcome can be interpreted in monetary units. $T(\mu_S)$ is the treatment outcome when S's human capital is available to M, i.e. when the parties trade with each other. $C(\sigma_M)$ is the support department's production cost for the input. The treatment outcome is improved when M invests more (i.e. μ_S increases) and S's production cost falls with greater σ_M. That is, the more the parties invest in their relationship the greater is the surplus from trade (i.e. treatment outcome minus cost). This reflects the benefits of increased coordination in the joint production of hospital care. Notably, both assets are accessible for both parties, irrespective of ownership structure, if they reach an agreement. However, the party owning the asset has the final say over its usage (i.e. residual control right).

Until they reach an agreement both parties may trade with other departments within or outside the hospital (see section 2) - these transactions (hereafter labeled *spot-trade*) are their *status quo* options or threat points. In the threat point M acquires (buys) a generic input and S supplies (sells) a generic input, which is supposed to be priced according to an agreed-upon price-list for hospital services - let this price be denoted $\bar{v}$. Here generic means that the input is not adjusted to fit the transaction between M and S. The production cost for a generic input is $c(\sigma_M; P_S)$ and P_S are the assets owned by S in spot-trade where the parties may only access the assets that they own. Notably, S makes a relationship-specific investment, in period 0, in this case as well.[24] Given this investment, S has to incur a cost to make the input generic, i.e. although the investment is always beneficial it creates an extra cost in spot-trade. It is assumed that $c(\sigma_M; P_S) > C(\sigma_M)$ since it is costly to make the input generic. In spot-trade M still produces the final good using a generic input and in the absence of S's human capital. The treatment outcome in this cases is $t(\mu_S; P_M)$, where P_M are the assets owned by M. Given the need for coordination in the joint production of health care it is assumed that $t(\mu_S; P_M) < T(\mu_S)$.

Notably, asset ownership affects the treatment outcome and the cost in spot-trade. Asset ownership matters because the parties do not have access to the other party's asset in their threat points unless they own it. Here M's asset ownership (P_M) may equal $\{p_M\}$ or $\{p_M, p_S\}$ while S's ownership (P_S) may equal $\varnothing$ or $\{p_S\}$. That is, it is assumed that S may never own M's assets, since public assets are infrequently for sale, reflecting inertia or a political preference for public ownership.[25] This assumption is called *restricted ownership.* The payoffs under spot-trade are given by u_M and u_S.

$$
\begin{aligned}
u_M &= t(\mu_S; P_M) - \bar{v} \qquad (4.3)\\
u_S &= \bar{v} - c(\sigma_M; P_S) \qquad (4.4)
\end{aligned}
$$

As already been hinted it is assumed that $T(\mu_S)$ is strictly concave in μ_S and that $C(\sigma_M)$ is strictly convex in σ_M. Moreover, assume that $t(\mu_S; P_M)$ is concave in μ_S and that $c(\sigma_M; P_S)$ is convex in σ_M.[26] [27]

[24] As already noted, investments, for both parties, have a higher value within the transaction than in the disagreement point, but the investments are valuable in this case as well. That is, making the investment is beneficial in both cases but to different degrees.

[25] The analysis below deals with the non-integration case when M owns p_M and S owns p_S and the full integration case when M owns both $p_M$and p_S. It abstracts from the case where S owns both p_M and p_S, which is called *type 2 integration* by Hart (1995)

[26] That is, $T'(\mu_S) > 0, T''(\mu_S) < 0$, $C'(\sigma_M) < 0, C''(\sigma_M) > 0$ and $t'(\mu_S) \geq 0, t''(\mu_S) \leq 0$, $c'(\sigma_M) \leq 0, c''(\sigma_M) \geq 0$.

[27] Concerning the payoffs it could be argued that S's payoff(s) should include treatment

Relationship-specific investments are beneficial in any industry and may be especially so for the hospital industry, given the joint production and the complementaries among different services in hospital care. The assumptions made above imply that the surplus from trade is greater than the surplus from spot-trade:

$$T(\mu_S) - C(\sigma_M) > t(\mu_S; P_M) - c(\sigma_M; P_S) \quad (4.5)$$
$$\text{for } \forall\ P_M, P_S \text{ where } P_M \cap P_S = \varnothing, P_M \cup P_S = \{p_M, p_S\}$$

Moreover it is assumed (as in Hart (1995)) that the marginal benefit from an increased investment is greater, or at least as great, the more assets the party making the investment has access to. The ranking of the first derivatives with respect to investments, hereafter called the *marginal conditions*, are the following:

$$\begin{array}{ccc} \dfrac{\partial T(\mu_S)}{\partial \mu_S} & > & \dfrac{\partial t(\mu_S; p_M, p_S)}{\partial \mu_S} \geq \dfrac{\partial t(\mu_S; p_M)}{\partial \mu_S} \geq \dfrac{\partial t(\mu_S; \varnothing)}{\partial \mu_S} \\ \dfrac{\partial C(\sigma_M)}{\partial \sigma_M} & < & \dfrac{\partial c(\sigma_M; p_M, p_S)}{\partial \sigma_M} \leq \dfrac{\partial c(\sigma_M; p_S)}{\partial \sigma_M} \leq \dfrac{\partial c(\sigma_M; \varnothing)}{\partial \sigma_M} \end{array} \quad (4.6)$$

The strict inequalities in (4.6) mean that M's investment is at least partly specific to S's human capital and that S's investment is at least partly specific to M's human capital. For example, if the medical department is a specialist in neuro surgery and the radiology department invests in increased knowledge about brain tomography, then the latter's investment is at least specific to the medical department's human capital. The weak inequalities mean that the value of the assets, in conjunction with investments, in the threat points is undetermined.

In equilibrium the parties will trade with each other. As already stated,

outcome or a fraction of treatment outcome to reflect that support services are also interested in the treatment outcome. This will not change the optimal investments as long as S's investment only has an indirect effect on treatment outcome i.e. that the presence of S implies $t(\mu_S; P_M) < T(\mu_S)$. Hence adding an explicit treatment outcome component to S's payoff would not alter the conclusions unless S's investment also had a direct effect on the treatment outcome. Moreover, S implicitly cares about treatment outcome since the monetary rewards from a treatment outcome is divided through bargaining.

trade creates a surplus compared to spot-trade, and the division of this surplus is decided by negotiation and, following Hart (1995), symmetric Nash bargaining is applied. This negotiation decides the reimbursement from M to S for the input that S supplies, in which case the reimbursement is:[28]

$$v = \bar{v} - \frac{T(\mu_S) - t(\mu_S; P_M) + C(\sigma_M) - c(\sigma_M; P_S)}{2} \tag{4.7}$$

That is, the parties each get half the surplus from trade, relative to the threat points. The individual *ex post* benefits (U_M and U_S respectively) from trade are equal to the payoff from trade, after inserting the reimbursement v, minus the investment cost i.e.:

$$\begin{aligned} U_M &= T(\mu_S) - \bar{v} - \tfrac{T(\mu_S) - t(\mu_S; P_M) + C(\sigma_M) - c(\sigma_M; P_S)}{2} - \mu_S = \\ &= \frac{T(\mu_S) + t(\mu_S; P_M) - C(\sigma_M) + c(\sigma_M; P_S)}{2} - \bar{v} - \mu_S \end{aligned} \tag{4.8}$$

$$\begin{aligned} U_S &= \bar{v} - \tfrac{T(\mu_S) - t(\mu_S; P_M) + C(\sigma_M) - c(\sigma_M; P_S)}{2} - C(\sigma_M) - \sigma_M = \\ &= \frac{T(\mu_S) - t(\mu_S; P_M) - C(\sigma_M) - c(\sigma_M; P_S)}{2} + \bar{v} - \sigma_M \end{aligned} \tag{4.9}$$

In the second-best world of incomplete contracting M and S choose investments at date 0 to maximize (4.8)and (4.9) respectively. Hart (1995) shows that this choice of investments leads to under-investments, for any ownership structure, compared to the first-best. Notably, Hart's proof also applies for the assumptions made here. The second-best first order conditions are:

$$\frac{1}{2}\frac{\partial T(\mu_S)}{\partial \mu_S} + \frac{1}{2}\frac{\partial t(\mu_S; P_M)}{\partial \mu_S} - 1 = 0 \tag{4.10}$$

$$-\frac{1}{2}\frac{\partial C(\sigma_M)}{\partial \sigma_M} - \frac{1}{2}\frac{\partial c(\sigma_M; P_S)}{\partial \sigma_M} - 1 = 0 \tag{4.11}$$

In this model ownership (organization) matters because it affects the marginal benefit of spot-trade. The marginal benefit of an investment is greater, or at least as great, depending on the nature of the assets and investments,

[28] Calculated by maximizing the Nash Bargaining product:
$NBP = [(T(\cdot) - v) - (t(\cdot) - \bar{v})] \times [(v - C(\cdot)) - (\bar{v} - c(\cdot))]$

if the investing party owns more assets, see (4.6). The intuition is that the investing party will invest more if the investment improves the threat point, and which generally is the case when the investing party owns more of the assets.

Organizational Choice

The aim of the analysis is to decide the optimal organization for the bilateral transaction under certain circumstances. Typically these circumstances are characteristics of assets and human capital, but the characteristics of investments may also be a factor. Definitions 1 and 2 define the effects of two important characteristics for the production of hospital care.[29]

Definition 1: *Assets* p_M *and* p_S *are strictly complementary if either* $t'(\mu_S; p_M) \equiv t'(\mu_S; \varnothing)$ *or* $c'(\sigma_M; p_S) \equiv c'(\sigma_M; \varnothing)$

Definition 2: *M's human capital (S's human capital) is essential if* $c'(\sigma_M; p_M, p_S) \equiv c'(\sigma_M; \varnothing)$ $(t'(\mu_S; p_M, p_S) \equiv t'(\mu_S; \varnothing))$

Strict complementarity means that the incentive for investment, for one of the parties, in the threat point (i.e. where ownership matters) is unaffected by ownership unless the party owns both assets. That is, owning only one of the assets does not increase the marginal return of investments. Essential human capital, on the other hand, means that the investment incentive for the non-essential party is equalized over ownership structures.

In the analysis below two organizational forms are considered: *non-integration* and *integration.* The incentives for investments are given by the first order conditions in the different organizational structures:

Under non-integration (N) the first-order conditions become:

$$\frac{1}{2}\frac{\partial T\left(\mu_S^N\right)}{\partial \mu_S^N} + \frac{1}{2}\frac{\partial t\left(\mu_S^N; p_M\right)}{\partial \mu_S^N} = 1 \tag{4.12}$$

$$-\frac{1}{2}\frac{\partial C\left(\sigma_M^N\right)}{\partial \sigma_M^N} - \frac{1}{2}\frac{\partial c\left(\sigma_M^N; p_S\right)}{\partial \sigma_M^N} = 1 \tag{4.13}$$

[29] Definitions 4 and 5 in Hart (1995) with appropriate adaptations to the current model.

Under full integration (F) the first-order conditions become:

$$\frac{1}{2}\frac{\partial T\left(\mu_S^F\right)}{\partial \mu_S^F}+\frac{1}{2}\frac{\partial t\left(\mu_S^F;p_M,p_S\right)}{\partial \mu_S^F} = 1 \quad (4.14)$$

$$-\frac{1}{2}\frac{\partial C\left(\sigma_M^F\right)}{\partial \sigma_M^F}-\frac{1}{2}\frac{\partial c\left(\sigma_M^F;\varnothing\right)}{\partial \sigma_M^F} = 1 \quad (4.15)$$

The marginal conditions (4.6) imply that M will invest at least as much under integration as under non-integration, while it is the other way around for S. That is: $\mu_S^F \geq \mu_S^N$ and $\sigma_M^F \leq \sigma_M^N$.

The PRA enables, as noted, a comparative analysis of organizational structures; in this analysis the organizational structures can be ranked in terms of "surplus-outcomes" (where the organizational structure yielding the greatest total surplus is the best structure), as well as qualitatively based on the parties' investments. In the latter case any organizational change that increases investments by one party, and does not decrease the other party's investment is beneficial for the transaction since it approaches the first-best (remember the inherent underinvestments in the model) - giving a set of organizational changes that imply an (Pareto) improvement over the original organization.

An internal ranking of these Pareto improvements, if there are multiple alternatives, cannot be achieved without knowing the relative contribution of the different investments to total surplus (i.e. the ranking in terms of surplus-outcomes) (see chapter 3 for a discussion about the relative contribution (productivity) of investments to total surplus). In the analysis of bilateral transactions below the set of Pareto improvements is typically limited to one (or no) alternative depending on the original organization (unless both parties human capital is essential). Hence, the focus here is on analyzing whether there are possible Pareto improvements to be made given the original organization or, as it is called hereinafter, starting point. Issue concerning the relative productivity of investments are hence abstracted from in this paper.

Two Starting Points for the Analysis The starting point for the analysis of Pareto improvements will obviously affect the results, and an organizational change may entail a Pareto improvement given one starting point but not given some other starting point. When thinking about public hospitals it is natural to think of them as large integrated entities, possibly inefficient and in need of disintegration i.e. privatization of certain activities. Thus one natural starting point is (full) integration. One could also think of a public principal wanting to construct a new *hospital division* or *treatment unit* consisting of support services and medical services. In this case non-integration is the natural starting point. The two starting points yield two different strands of analysis labeled *privatization and construction.* In this section this conceptual distinction may be somewhat superfluous because non-integration and M-integration (M owning all the assets) are the only available alternatives (with one being the starting point), but the concepts are deemed important enough to warrant an explicit exposition of the alternatives, not least because they clarify the reasoning in the analysis - moreover, the distinction comes back and becomes more interesting in the next two sections. In this section both strands of analysis are exemplified using definitions 1 and 2.

Strict Complementarity In the production of hospital care it is reasonable to assume that the assets are complementary and thus that either $t'(\mu_S; p_M) = t'(\mu_S; \varnothing)$ or $c'(\sigma_M; p_S) = c'(\sigma_M; \varnothing)$. That is, the complementarity of assets either implies that M is indifferent between owning p_M and not owning p_M, or that S is indifferent between owning p_S and not owning p_S. The reason for this is that p_M is useless without p_S, and *vice versa*, when the assets are strictly complementary. Complementarity implies that some form of integration is optimal (Hart, 1995). The assumption of restricted ownership makes the case when $t'(\mu_S; p_M) = t'(\mu_S; \varnothing)$ redundant. Thus, if the assets are complementary, then the complementarity implies that $c'(\sigma_M; p_S) = c'(\sigma_M; \varnothing)$, i.e. that S is indifferent. The intuition is that S as a support service is dependent on access to the other party's (the party S is supporting) asset to be able to realize a higher marginal benefit of the investment when the assets are complementary. M, on the other hand, is able to realize a greater marginal benefit, compared to the case when M does not own any assets, by only owning p_M in spot-trade, especially given the preference for public ownership (i.e. the restricted ownership assumption).

Construction Here the starting point for the analysis is non-integration. Obviously strict complementarity, which makes S indifferent in the sense described above, implies that (4.13) and (4.15) are the same, which in turn implies that $\sigma_M^F = \sigma_M^N$. That is, S invests as much under integration as under non-integration. M, on the other hand, invests at least weakly more under integration ($\mu_S^F \geq \mu_S^N$) and thus integration entails a (at least weak) Pareto improvement over non-integration.

Privatization Does disintegration increase the total relationship-specific investments, i.e. the joint production/ coordination, compared to full integration when assets are strictly complementary? The answer is no. As in the construction case strict complementarity implies that S makes the same investment under both non-integration and full integration, while M will invest weakly less under non-integration than under full integration. Thus privatization does not constitute a Pareto improvement in this setting.

Essential Human Capital One or both parties' human capital may be essential for the production of hospital care, e.g. the doctors and nurses at the medical department. The question is how this affects the choice of organization. Definition 2 formalizes the idea of essential human capital; it says that if one party's human capital is essential, then ownership does not matter for the other party in its threat point. That is, if e.g. M's human capital is essential for the production of hospital care, then S's marginal (spot-trade) benefit of investments is independent of ownership i.e. $c'(\sigma_M; p_M, p_S) = c'(\sigma_M; p_S) = c'(\sigma_M; \varnothing)$. In short the absence of M's human capital makes asset ownership irrelevant.

Construction Starting from non-integration and assuming that M's human capital is essential gives that integration is a Pareto improvement. To see this, note that the solutions to (4.13) and (4.15) are the same (asset ownership has no effect on S's investment incentive), while $\mu_S^F \geq \mu_S^N$ still holds. Notably, if M's human capital is essential, this effect overrides any effect from the relation between assets; that is, integration is optimal irrespective of assets being complements or independent when M's human capital is pivotal for the production of hospital care. This begs the question whether M's human capital is essential. Given that M is a medical department and medical care is the primary output from a hospital, it is natural to assume that M's human capital is essential. What about S's human capital? It is of course important in the joint production of hospital care but in most cases not essential.[30] However, even if S's human capital is essential as well, i.e. if both parties' human capital is essential, this does not alter the conclusion that integration is optimal, in the sense that it is still in the set of optimal organizational forms, because organizational form does not matter in this case - "... neither party's investment will pay off in the absence of agreement with the other" (Hart, 1995:48).

Privatization Clearly, disintegration or privatization does not constitute a Pareto improvement over full integration when M's human capital is essential. That is, S will make the same investment but M will invest weakly less if the asset p_S is reallocated from M to S. If both M's and S's human capital is essential, then privatization is an option. However, all other organizational forms are equally good because the incentives for investments are the same in all ownership structures for both parties. Thus if privatization is opted for in this case, it can be made without cost, in terms of the model, but it is not an improvement as such.

[30] One might argue (concerning essential human capital) that a surgeon should be able to produce usable diagnoses and/or X-ray images if she has the equipment needed (assets), while a radiologist would most likely fail at performing surgery even if she had the equipment.

Conclusion M & S

The analysis provides a strong case for integration given that assets are strictly complementary in the way described and that the medical department's human capital is essential for the production of hospital care. Both assumptions are likely to hold for many treatments performed in hospitals. Interestingly privatization of the support service is weakly Pareto dominated by keeping the vertical structure in all instances. Furthermore, integration should also be opted for when constructing new treatment units.[31]

4.3.2 Medical Department plus Hotel Service

Now consider a different setup in which the transaction involves a hotel service, H, and a medical department/public principal, M. M owns one asset p_B and H owns one asset p_H. H's asset is for example a kitchen (kitchen equipment) where the patients' food is produced. M uses this input in the production of hospital care, e.g. during *pre-* and *post*-operative care. The medical department pays H a reimbursement h for the input. Furthermore, M makes an investment in its physical asset while H invests in human capital.

First, the case when M makes a generic investment is discussed, followed by the case when M makes a specific investment. In both cases M's physical asset may be thought of as a building or some asset not directly used in the production of hospital care. Thus the public principal might be less restricted in disposing of this asset than other assets used directly in the production of health care. One might actually think of M having two assets p_M and p_B. Hence the assumption of restricted ownership is relaxed for p_B in the two scenarios below.

[31] This might be interpreted in terms of *lean production* where complementary assets and competencies should be close to each other to enable a quick and comprehensive treatment of patients. (see e.g. Kollberg et al (2006) for a discussion of lean thinking and health care)

A Generic Investment by M

Here, H makes a relationship-specific investment δ in period 0, which is an investment in human capital e.g. educating the kitchen personnel about suitable food for different diseases and/or the hospitals special requirements about nutrition values and cooking procedures. The hospital management (the medicine department) makes a generic investment β in the physical asset p_B, e.g. a building where the kitchen may be placed, in period 0. The investment is generic in the sense that it is not specific to H's asset, i.e. any kitchen equipment may be placed in this building. This investment increases the value of the asset in the transaction, but also in all other uses - the increase in value is independent of H's and M's participation in the transaction once the investment is made (cf. Hart, 1995). The payoffs from trade are the following:

$$U_M = Z(\beta) - h \tag{4.16}$$

$$U_H = h - L(\delta) \tag{4.17}$$

where $Z(\beta)$ is the treatment outcome when H's human capital is available and $L(\delta)$ is the cost of producing hotel services when the investment β (in practice a part of p_B) and M's human capital are available to H. If the two parties do not trade with each other they have to buy and sell the hotel service on a spot-market; here, it is suitable to think about an external spot-market e.g. a market for catering. $\bar{h}$ is the market price for a generic hotel service. The payoffs in spot-trade (the threat point) are the following:

$$u_M = z(\beta; P_B) - \bar{h} \tag{4.18}$$

$$u_H = \bar{h} - l(\delta; P_H) \tag{4.19}$$

Here $z(\beta; P_B)$ is the treatment outcome in the absence of H's human capital and $l(\delta; P_H)$ is the production cost in the absence of M's human capital. P_B denotes the assets available to M in the threat point, and $P_B = \varnothing$, $P_B = \{p_B\}$ or $P_B = \{p_B, p_H\}$. Similarly, P_H is the assets available to the supplier of hotel services if the parties do not trade with each other, and

$P_H = \{p_B, p_H\}$, $P_H = \{p_H\}$ or $P_H = \varnothing$. The marginal conditions in this case are:

$$\frac{\partial Z(\beta)}{\partial \beta} = \frac{\partial z(\beta; p_B, p_H)}{\partial \beta} = \frac{\partial z(\beta; p_B)}{\partial \beta} > \frac{\partial z(\beta; \varnothing)}{\partial \beta} = 0 \tag{4.20}$$

$$\frac{\partial L(\delta)}{\partial \delta} < \frac{\partial l(\delta; p_B, p_H)}{\partial \delta} \leq \frac{\partial l(\delta; p_H)}{\partial \delta} \leq \frac{\partial l(\delta; \varnothing)}{\partial \delta} \tag{4.21}$$

The derivative $\frac{\partial z(\beta;\varnothing)}{\partial \beta}$ equals zero because an investment in p_B is of no value to M, in spot-trade, when M does not own p_B . Furthermore, the equalities in (4.20) are explained by the assumption that β is a generic investment in the physical asset (p_B) - thus the presence of H's human capital and physical asset has no effect on the marginal benefit of this investment. However, the presence of H's human capital has a positive level effect on the treatment outcome and there is a surplus from trade i.e.: $Z(\beta) - L(\delta) > z(\beta; P_B) - l(\delta; P_H)$ for all ownership structures. This surplus is divided through negotiations and once again the symmetric Nash bargaining solution is applied to find the reimbursement h, which in this case is:

$$h = \bar{h} + \frac{Z(\beta) - z(\beta; P_B) + L(\delta) - l(\delta; P_H)}{2} \tag{4.22}$$

Inserting this h in the payoffs from trade and subtracting the investment cost produces the *ex post* benefits from trade. The first order conditions are given by maximizing these benefits with respect to the investments. In this setting three organizational forms are considered; M-integration, H-integration and non-integration. The first order conditions under non-integration (N), M-integration (M) and H-integration (H) are:

$$\frac{1}{2}\frac{\partial Z(\beta^N)}{\partial \beta^N} + \frac{1}{2}\frac{\partial z(\beta^N; p_B)}{\partial \beta^N} = 1 \tag{4.23}$$

$$-\frac{1}{2}\frac{\partial L(\delta^N)}{\partial \delta^N} - \frac{1}{2}\frac{\partial l(\delta^N; p_H)}{\partial \delta^N} = 1 \tag{4.24}$$

$$\frac{1}{2}\frac{\partial Z\left(\beta^{M}\right)}{\partial\beta^{M}}+\frac{1}{2}\frac{\partial z\left(\beta^{M};p_{B},p_{H}\right)}{\partial\beta^{M}}=1 \tag{4.25}$$

$$-\frac{1}{2}\frac{\partial L\left(\delta^{M}\right)}{\partial\delta^{M}}-\frac{1}{2}\frac{\partial l\left(\delta^{M};\varnothing\right)}{\partial\delta^{M}}=1 \tag{4.26}$$

$$\frac{1}{2}\frac{\partial Z\left(\beta^{H}\right)}{\partial\beta^{H}}+\frac{1}{2}\frac{\partial z\left(\beta^{H};\varnothing\right)}{\partial\beta^{H}}=\frac{1}{2}\frac{\partial Z\left(\beta^{H}\right)}{\partial\beta^{H}}=1 \tag{4.27}$$

$$-\frac{1}{2}\frac{\partial L\left(\delta^{H}\right)}{\partial\delta^{H}}-\frac{1}{2}\frac{\partial l\left(\delta^{H};p_{B},p_{H}\right)}{\partial\delta^{H}}=1 \tag{4.28}$$

Organizational Choice Since M owns p_B under both non-integration and M-integration and the investment β increases the value of this asset irrespective of the ownership structure they will make the same investments in both cases i.e. $\beta^N = \beta^M$. The same reasoning gives that M will invest less under H-integration. When M owns p_B H does not get a part of the increase in value from the investment, i.e. the asset owner receives the full increase in value (cf. Hart 1995). H, on the other hand, will invest weakly more the more assets it owns; hence $\delta^H \geq \delta^N \geq \delta^M$ (see (4.21)). Under H-integration M will invest strictly less and H will invest weakly more than under both M-integration and non-integration, thus H-integration is not a Pareto improvement irrespective of the starting point. Non-integration, however, implies that H will invest weakly more than under M-integration while M will make the same investment; hence non-integration is a Pareto improvement over M-integration. This reasoning also gives that non-integration cannot be improved on if this is the starting point.

It is obvious that the best organizational choice for this transaction is non-integration. This is in line with casual observation of modern public hospitals suggesting that hotel services like laundry, cleaning and cooking are often outsourced i.e. not integrated in the hospital organization.

A Specific Investment by M

Once again H makes the relationship-specific investment δ in period 0, but in this scenario M's investment β in the asset p_B is specific to H's asset p_H.

Imagine for example that M invests in adapting a building such that it fits the specific type of kitchen inventory that H works with better than most competitors' equipment - i.e. there is some significant difference between H's equipment and generic equipment. The more M invests in making p_B suitable for p_H the costlier it will be not to trade with H (as long as H owns p_H). That is, the cost of making p_B generic again increases with the investment. Call this cost b (the *readjustment cost*). Let b be an increasing function in the investment β such that $\frac{db(\beta)}{d\beta} \geq 0$ and $\frac{d^2b(\beta)}{d^2\beta} \geq 0$. This cost affects M's payoff from spot-trade (the payoff from trade is unaffected):

$$z(\beta, b(\beta); P_B) - \bar{h} \tag{4.29}$$

Obviously the cost lowers the benefit from spot-trade compared to the case when investments do not have a readjustment cost. Moreover, the cost b has a negative effect on M's payoff in spot-trade:

$$\frac{\partial z(\beta, b; P_B)}{\partial b} \leq 0 \text{ for a given } \beta$$

This implies that the investment β is more valuable (less costly) when M and H trade with each other. However, since β is an investment in p_B that is specific to p_H (H's physical capital), the increase in value is independent of H and M's participation in the transaction, but will depend on the access to both assets. Specifically, assume that the negative effect of the readjustment cost on M's spot-trade benefit is zero when M owns both assets, i.e.:

$$0 = \frac{\partial z(\beta, b; p_B, p_H)}{\partial b} \geq \frac{\partial z(\beta, b; p_B)}{\partial b} \text{ for a given } \beta \tag{4.30}$$

Now assume that the inequality in (4.30) is strict, implying that the effect of the readjustment cost is strictly negative when M only owns p_B.

The total effect on M's benefit from the investment under spot-trade is a composite of a benefit and a cost effect:[32]

$$\frac{dz\left(\beta,b\left(\beta\right);P_B\right)}{d\beta}=\frac{\partial z\left(\beta,b\left(\beta\right);P_B\right)}{\partial\beta}+\frac{\partial z\left(\beta,b\left(\beta\right);P_B\right)}{\partial b}\frac{db(\beta)}{d\beta} \tag{4.31}$$

where $\frac{\partial z(\beta,b(\beta);P_B)}{\partial\beta}\geq 0$ [33] is the benefit effect and it is already assumed that $\frac{db(\beta)}{d\beta}\geq 0$, i.e. the negative effect is multiplied by the impact of increased investments on the readjustment cost. The total effect of the investment also depends on the assets owned by M. The total effect on benefit (given the assumptions above) is strictly greater when M owns both p_B and p_H, unless the readjustment cost does not rise with investment i.e. unless $\frac{db(\beta)}{d\beta}=0$. To make things interesting, assume that the readjustment cost is strictly increasing in the investment β. In this case the marginal conditions are (note that marginal cost structure for H is unchanged and still given by relation (4.21)):

$$\frac{\partial Z\left(\beta\right)}{\partial\beta}=\frac{dz\left(\beta,b\left(\beta\right);p_B,p_H\right)}{d\beta}>\frac{dz\left(\beta,b\left(\beta\right);p_B\right)}{d\beta}\geq\frac{dz\left(\beta,b\left(\beta\right);\varnothing\right)}{d\beta}=0$$

$$\frac{\partial L\left(\delta\right)}{\partial\delta}<\frac{\partial l\left(\delta;p_B,p_H\right)}{\partial\delta}\leq\frac{\partial l\left(\delta;p_H\right)}{\partial\delta}\leq\frac{\partial l\left(\delta;\varnothing\right)}{\partial\delta} \tag{4.32}$$

The equality in (4.32) stems from H's human capital being unimportant, in a marginal sense, as long as M has access to the asset p_H. The inequality in (4.32) is strict since it is assumed that $\frac{db(\beta)}{d\beta}>0$, and once again the incentive for M to make the investment in p_B, in spot-trade, is zero if it does not own this asset. However, trade is beneficial for both parties, as in the previous setup. That is, the presence of H's human capital is beneficial (given H's investments in human capital) for M and the other way around. This creates a surplus from trade. To divide the surplus Nash bargaining is also applied

[32] Assume that $\frac{dz(\beta,b(\beta);P_B)}{d\beta}\geq 0$, i.e. the total effect of making investments is positive, so that it is worthwhile making the investment. Furthermore note that $\frac{d^2z(\beta,b(\beta);P_B)}{d\beta^2}\leq 0$ if $\frac{\partial z(\beta,b(\beta);P_B)}{\partial\beta\partial b}\leq 0$ and $\frac{\partial^2 z(\beta,b(\beta);P_B)}{\partial b^2}\leq 0$ - assume that this is the case, implying that the marginal benefit of the investment β is decreasing.

[33] Assume $\frac{\partial^2 z(\beta,b(\beta);P_M)}{\partial^2\beta}\leq 0$, i.e. that benefit is decreasing in β.

in this case, yielding similar *ex post* benefits to the generic setup, and thus similar first order conditions (identical for H).The first order conditions for non-integration, M-integration and H-integration respectively become:

$$\frac{1}{2}\frac{\partial Z\left(\beta^N\right)}{\partial \beta^N}+\frac{1}{2}\frac{dz\left(\beta^N,b\left(\beta^N\right);p_B\right)}{d\beta^N}=1 \tag{4.33}$$

$$-\frac{1}{2}\frac{\partial L\left(\delta^N\right)}{\partial \delta^N}-\frac{1}{2}\frac{\partial l\left(\delta^N;p_H\right)}{\partial \delta^N}=1 \tag{4.34}$$

$$\frac{1}{2}\frac{\partial Z\left(\beta^M\right)}{\partial \beta^M}+\frac{1}{2}\frac{dz\left(\beta^M,b\left(\beta^M\right);p_B,p_H\right)}{d\beta^M}=1 \tag{4.35}$$

$$-\frac{1}{2}\frac{\partial L\left(\delta^M\right)}{\partial \delta^M}-\frac{1}{2}\frac{\partial l\left(\delta^M;\varnothing\right)}{\partial \delta^M}=1 \tag{4.36}$$

$$\frac{1}{2}\frac{\partial Z\left(\beta^H\right)}{\partial \beta^H}+\frac{1}{2}\frac{dz\left(\beta^H,b\left(\beta^H\right);\varnothing\right)}{d\beta^H}=\frac{1}{2}\frac{\partial Z\left(\beta^H\right)}{\partial \beta^H}=1 \tag{4.37}$$

$$-\frac{1}{2}\frac{\partial L\left(\delta^H\right)}{\partial \delta^H}-\frac{1}{2}\frac{\partial l\left(\delta;p_B,p_H\right)}{\partial \delta^H}=1 \tag{4.38}$$

Organizational Choice It is obvious from (4.32) that $\beta^M > \beta^N \geq \beta^H$ - since the benefit of investments is assumed to be concave in β. Moreover, the investment β is an adjustment of p_B to make it complementary (specific) to p_H. Now, assume that the investment β ensures strict complementarity between the assets, i.e. if this investment is made, then p_B and p_H are strictly complementary. Definition 1 says that if the assets are strictly complementary then either 1) $\frac{\partial l(\delta;p_H)}{\partial \delta} \equiv \frac{\partial l(\delta;\varnothing)}{\partial \delta}$ or 2) $\frac{dz(\beta,b(\beta);p_B)}{d\beta} \equiv \frac{dz(\beta,b(\beta);\varnothing)}{d\beta}$.

If 1) holds then H is indifferent between owning and not owning p_H and M will obviously have greater incentives for investments if it owns all the assets. Thus, a move from non-integration between M and H to M-integration would entail a Pareto improvement. By the same token a move from M-integration to non-integration with H owning its asset (privatization) is not a Pareto improvement - H will make the same investment and M will invest less.

Obviously H-integration does not constitute a Pareto improvement either, since M then would invest less (strictly or weakly) than under both M-integration and non-integration.

If 2) applies then M, in anticipation of this, will not make the investment unless owning both assets, since the marginal value of an investment in a physical asset only accrues to the owner of the asset and therefore the party not owning the asset will make a zero investment (cf. Hart, 1995). In this case the marginal value of the investment when owning the asset, p_B, is equal to the marginal value when not owning the asset, hence no investment is made unless M owns both assets. It is a radical conclusion that M will only invest in its own asset if it also owns p_H, and it implies that M-integration gives M incentive to make the investment, but not non-integration. But M-integration is not a Pareto improvement over non-integration, since H will invest weakly more if it owns p_H than if it does not, i.e. $\delta^M \leq \delta^N$. However, H-integration would be an improvement, albeit with no investment from M, over non-integration since H invests more. Obviously, a move from M-integration to a disintegrated structure (or H-integration) does not constitute a Pareto improvement since M would not make the investment in this case.

Thus if 2) applies, then H-integration is a Pareto improvement in the construction scenario (i.e. with non-integration as the starting point) if zero investments by M are acceptable. Furthermore, if the starting point is M-integration, then privatization (non-integration or H-integration) is not a Pareto improvement. Thus, depending on the starting point, the optimal organizational structure is one of the extremes i.e. either H-integration or M-integration.

Interacting Investments It seems somewhat strange that the optimal organizational structure involves zero investments from M as it does in the construction scenario in 2), especially since M's investment is specific to H's asset that should be valuable to both parties if made. One way to make the investment valuable to both parties is to assume that the parties investments interact. For example, an investment in educating the kitchen personnel in hygiene would be more valuable to H if M at the same time made an investment in the special equipment needed to put this knowledge into use e.g. disinfection equipment.[34]

If the benefit of H's investment is closely linked to the investment made by M, then a zero investment by M is a rather unsatisfactory solution. Unsatisfactory in the sense that benefit from trade for H would be higher if M made an investment. This kind of interaction is implicitly assumed in the statement that $L(\delta)$ is the production cost when M's human capital and β are available to H (see the previous subsection). Thus, it is in line with the rest of the model to assume that:

$$L(\delta : \beta) < L(\delta : 0) \tag{4.39}$$

where $L(\delta : \beta)$ is the production cost when some investment β is made, and $L(\delta : 0)$ is the production cost when no investment is made.[35] However, this realization by itself does not give M any incentives to make investments, but if it is coupled with some ownership of p_H it might.

Joint ownership of the asset p_H could be a way to ensure positive investments from M and at the same time maintain H's level of investments. If the parties decide on joint ownership of p_H and both have *veto* power or the use of the asset, any increase in surplus from trade, from the investment, will be shared $50:50$ under Nash bargaining (see Hart, 1995). Here the increase in surplus from trade equals the surplus from trade when the investment is made (and p_H is jointly owned) minus the surplus from trade when the investment is not made (and p_H is owned by H) i.e.:

$$\begin{gathered} Z(\beta) - h - z(\beta, b(\beta); \{p_B, p_H\}) + \bar{h} + h - L(\delta : \beta) - \bar{h} + l(\delta; p_H) + \\ - \left[Z(0) - h - z(0; p_B) + \bar{h} + h - L(\delta : 0) - \bar{h} + l(\delta; p_H)\right] = \\ Z(\beta) - z(\beta, b(\beta); \{p_B, p_H\}) - L(\delta : \beta) + l(\delta; p_H) + \\ - \left[Z(0) - z(0; p_B) - L(\delta : 0) + l(\delta; p_H)\right] = \end{gathered}$$

[34] Admittedly not a very specific investment, but for the sake of argument assume that it is specific to the kitchen equipment used by H.

[35] The intuition is straightforward; if M does not make the necessary adjustments of the building H either has to make them itself or experience a greater production cost because the building is ill-fitting for its production.

$$= Z(\beta) - Z(0) - z(\beta, b(\beta); \{p_B, p_H\}) + z(0; p_B) + L(\delta : 0) - L(\delta : \beta) \quad (4.40)$$

where 0 indicates that M does not make any investment when it only owns p_B. Thus M will get

$$\frac{1}{2}\left[Z(\beta) - Z(0) - z(\beta, b(\beta); \{p_B, p_H\}) + z(0; p_B) + L(\delta : 0) - L(\delta : \beta)\right]$$

if it makes the investment, and it will make the investment if and only if this covers the cost of investment, i.e.:

$$\beta > 0 \text{ iff } \frac{1}{2}\left[\begin{array}{c} Z(\beta) - Z(0) - z(\beta, b(\beta); \{p_B, p_H\}) + \\ +z(0; p_B) + L(\delta : 0) - L(\delta : \beta) \end{array}\right] \geq \beta \qquad (4.41)$$

When (4.41) applies then joint ownership of H's asset is an improvement over non-integration, i.e. it increases M's investment without lowering H's investment.[36] This may be confirmed by looking at the first-order and marginal conditions for M and H.

Is joint ownership an improvement over M-integration? Given that M has *veto* power over the use of assets its incentive for investments is unaltered by the move from M-integration to joint ownership, while H's investment increases. Thus, joint ownership as a Pareto improvement, irrespective of the starting point, if the investments δ and β interact in the sense described here.

Conclusion M & H

The optimal way to organize a transaction between a medical department/public principal and a hotel service depends on the characteristics of investments. In this section it is assumed that the hotel service makes an investment in its human capital and that the medical department invests in its physical asset. If the medical department's investment is generic, non-integration is optimal. If the investment is specific to the hotel service's asset, then the transaction may be organized in three ways depending on the starting point,

[36] H's incentive for investments is unaltered since they have *veto* power over the use of the asset.

the complementarity of assets and the interaction between investments.

When the assets are strictly complementary, such that H is indifferent between owning and not owning p_H, M-integration is the best organizational structure irrespective of the starting point. If the strict complementarity means that M is indifferent over owning or not owning p_B, then H-integration is an improvement over non-integration and M-integration cannot be improved on if it is the starting point; this if zero investments by M are acceptable. However, if M's investment matters to H, i.e. the two parties investments are interacting, then joint ownership of p_H is the best option irrespective of the starting point.

4.4 Trilateral Transactions in Public Hospitals

This chapter has so far analyzed two different bilateral transactions, but as already mentioned, the joint production of hospital care in many instances involves more than two parties. The effects of an extension to a trilateral transaction is briefly discussed in this section.

In general a hospital consists of many different medical departments, support departments and hotel services. These services and departments are all involved, to different degrees, in the joint production of hospital care. Many of the factions of the hospital are not only involved in bilateral transactions but also in multilateral transactions. Medical departments, for example, supply inputs to each other using inputs from support services and hotel services in a complex pattern of internal demand and internal supply. Thus, a natural extension to the bilateral model is to analyze the organization of a more complex transaction e.g. a trilateral transaction. Chapter 3 of this thesis studies a PRA-model that encompasses trilateral transactions of the type described in figure 4.3.

Now, assume that A (in figure **??**) is a medical department (e.g. a cardiology department). M is, as above, a medical department (e.g. thoracic

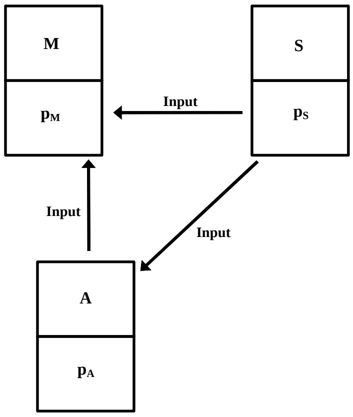

Figure 4.3: The trilateral transaction

surgery) and the public principal and S is a support service (e.g. radiology department). Issues concerning hotel services are abstracted from in this section for two reasons: first, trilateral transactions involving two medical departments and one support service seem commonplace and important in hospitals, and second, hotel services are further from the core activities of hospitals than medical departments, and thus less interesting to include in the trilateral transaction.

As above, M produces the final good but now it uses inputs from both S and A. Moreover, A uses an input produced by S in its production. Thus S produces inputs to both M and A and the two inputs are not the same product. The three parties make relationship-specific investments in their human capital and use one asset each for their production (see figure 4.3). In these aspects and in all other aspects, but one, the model replicates the model in chapter 3. The exception here is that restricted ownership is assumed, i.e. that M will always own the asset p_M. This limits the number of possible organizational forms to six (instead of ten), but besides this the results presented here stem directly from chapter 3.

In a trilateral model it is possible to create large number of examples of organizational choices by varying starting points, characteristics of assets, human capital and investments, but here the focus is on two questions: *When is privatization optimal?* and *When should full integration be opted for?*

When is Privatization Optimal? Here, the starting point is M-integration, i.e. that the public principal owns all the assets (p_M, p_A and p_S). Thus, privatization/disintegration entails that one or two (not p_M) of M's assets are reallocated to the other parties and because investments increase with asset ownership, as in the bilateral model, something special is required to make privatization optimal. In fact privatization is only a Pareto improvement if both A's and S's human capital is essential. If this is the case, any form of organization other than M-integration is a Pareto improvement i.e. increases the level of investments by at least one party without changing the other parties' investments. In all other cases M-integration cannot be improved on.

When Should Full Integration be Opted For? Full integration in this setting is equivalent to M-integration since ownership of p_M is restricted. Full integration is interesting because medical services and support services are often vertically integrated in public hospitals. This provides a rationale for using full integration as the starting point of the analysis, as above, but also for investigating the circumstances under which full integration should be used when creating a new hospital division or treatment unit involving a trilateral transaction. There are two cases where M-integration constitutes a Pareto improvement over non-integration: 1) if both M's and A's human capital is essential and p_M and p_A are strictly complementary such that A is indifferent over ownership that does not include p_M. 2) if p_M and p_S are strictly complementary and p_M and p_A are strictly complementary in a way that makes A and S indifferent. Apparently M-integration is only an improvement under special combinations of characteristics, implying that choosing full integration without information about all the characteristics might be suboptimal. Thus the public principal constructing the new division needs information about the characteristics of both human capital and assets before deciding to go for full integration of these activities.

For most other combinations of characteristics the optimal organization is either partial integration of some form (e.g. M owning S's asset and A owning its own asset) or non-integration (see chapter 3).

4.5 Concluding Remarks

This chapter suggests and argues that the PRA is conducive to the analysis of hospital organization, and especially so for public hospitals. It is argued that while some assumptions of the PRA are ill-fitting for this analysis, the main mechanisms are quite suitable. The need for coordination and cooperation in the joint production of hospital care and the inherent uncertainty in health care (and other "minor" features) are captured by the PRA.

The analysis develops a straightforward model, based on Hart (1995), of the joint production of hospital care. Most interestingly, it is found, when studying a transaction involving a medical department (public principal) and a support service, that integration should be opted for when constructing new treatment units as well as when considering privatization of the support service. Both results are foremost driven by the reasonable assumption that the medical department's human capital is essential for the production of hospital care. This is intuitively appealing and somewhat trivial; without the cooperation with medical departments most support services would experience difficulties in giving patients suitable treatment. Furthermore, this proposes that public ownership of all the assets in the transaction is the best option as long as the public principal's human capital is essential.[37]

Hospitals also consist of hotel services that contribute to the joint production, e.g. by producing food and doing laundry. The analysis of the transaction between the medical department and the hotel service reveals that the organization of the transaction depends on the characteristics of investments. In certain circumstances integration should be opted for, in other circumstances the best option is non-integration and in yet other circumstances joint ownership could be the solution. The fundamental lesson concerning this transaction is that being dependent on access to the other party's asset and/or investment lowers the incentive for investments unless some residual control rights over assets can be granted - either through in-

[37] Obviously, if the public principal were not also a medical department, as is assumed here, it would be more questionable that its human capital is essential for hospital care and this would change the conclusion made here.

tegration or joint ownership. Specific investments in physical capital and complementary investments create such dependence.

If the analysis is extended to trilateral transactions, it is found that privatization, i.e. disintegrating an integrated structure, is a Pareto improvement only in very special circumstances. However, it is also found that. when constructing new treatment units, full integration, in most instances, is not an improvement over non-integration or partial integration. Thus trilateral transactions that are already integrated should remain integrated and new treatment units should either be non-integrated or partially integrated.

Generally, the predictions of the model are supportive of medical departments (with essential human capital) owning support services' assets but not other medical departments' assets (unless the other medical departments are indifferent over ownership, which seems unlikely).

This chapter and most of the PRA-literature assume that investments in human capital are complementary and typically aimed at improving the existing stock of human capital, e.g. by education. An interesting topic for future research is the effect of allowing for investments in human capital that are substitutes, e.g. what happens if M hires the specialist nurse that S also wants to hire? Furthermore, investments in public hospitals may also be substitutes in the sense that hospital management may decide to approve only one department's appeal for investments. Hence modelling investments as substitutes is a reasonable and interesting extension to analysis in this paper.

The results in this chapter depend on assumptions about the characteristics of investments, human capital and assets. These assumptions need to be scrutinized and compared to the actual characteristics of investments, human capital and assets in public hospitals to enable a conclusion applicable to a specific hospital. This research lies in the future. This paper, however, provides a basic framework for thinking about public hospital organization from a new perspective by applying the property rights approach to organization.

Bibliography

Arrow, K J., 1963, "Uncertainty and the Welfare Economics of Medical Care", American Economic Review, 53:941-973

Chalkley, M & Malcomson, J M., 1998, "Contracting for Health Services when Patient Demand does not Reflect Quality", Journal of Health Economics, 17:1-19

Chiu, S, 1998, "Noncooperative Bargaining, Hostages, and Optimal Asset Ownership", American Economic Review 88:882-901

De Meza, D & Lockwood, B, 1998, "Does Asset Ownership Always Motivate Managers? Outside Options and the Property Right Theory of the Firm, Quarterly Journal of Economics, 113:361-386

Coles, J W & Hesterly, W S, 1998, "Transaction costs, quality, and economies of scale: examining contracting choices in the hospital industry", Journal of Corporate Finance, 4:321-345

Dixit, A, 2002, "Incentives and Organizations in the Public Sector: An Interpretative Review" Journal of Human Resources, 37:696-727

Dranove, D, 1988, "Demand Inducement and the Physician/Patient Relationship", Economic Inquiry, 26:281-298

Eggelston, K, 2005, "Multitasking and Mixed Systems for Provider Payment", Journal of Health Economics. 24:211-223

Ellis, R P & McGuire, T G, 1986, "Provider Behavior under Prospective Reimbursement - Cost Sharing and Supply", Journal of Health Economics, 5:129-151

Evans, R G, 1974. "Supplier-induced demand: some empirical evidence and implications" in *The Economics of Health and Medical Care,* Perlman, M. ed, Macmillan, Edinburgh, 162–173.

Grossman, Sanford J. & Hart, Oliver D., 1986, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration", Journal of Political Economy, 94:691-719

Harris, J E, 1977, "The Internal Organization of Hospitals: Some Economic Implications", Bell Journal of Economics, 8:467-482

Hart, O & Moore, J, 1990, "Property Rights and the Nature of the Firm", Journal of Political Economy, 98:1119-1158

Hart, O 1995, *Firms, Contracts and Financial Structure*, Claredon Lectures in Economics, Claredon Press, Oxford

Hart, O & Moore, J, 2008, "Contracts as Reference Points", Quarterly Journal of Economics CXXIII:1-48

Holmström, B, 1999, "The Firm as a Subeconomy", Journal of Law, Economics, & Organization, 15:74-102

Klein, B, Crawford, R G & Alchian, A, 1978, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", Journal of Law and Economics, 21:297-326

Kollberg, B, Dahlgaard, J J & Brehmer, P-O, 2007, "Measuring lean initiatives in health care services: issues and findings", International Journal of Productivity and Performance Management, 56:7-24

Ma, A, 1994,"Health Care Payment Systems: Cost and Quality Incentives", Journal of Economics & Management Strategy, 3:93-112

McGuire, T G, 2000, "Physician Agency", in *Handbook of Health Economics*, vol:1A, Culyer A. J. & Newhouse J. P. eds, 462-536

Newhouse, Joseph P., 1970, "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital", The American Economic Review, 60:64-74

Nyman, J A, 1999, "The economics of moral hazard revisited", Journal of Health Economics, 18:811-824

Pauly, M V, 1968, "The Economics of Moral Hazard: Comment", American Economic Review, 58:531-537

Posnett, J, 2002, "Are Bigger Hospitals Better?", in *Hospitals in a Changing Europe*, McKee M & Healy J eds., Open University Press, Buckingham, 100-118

Sloan, F, 2000, "Not-for-profit ownership and hospital behavior" in Handbook of Health Economics, vol.1, Culyer A J & Newhouse J P eds, Elsevier, Amsterdam, 1141-1174

Söderström, L & Lundbäck, M, 2002, *Stor och liten - Om vårdeffektiva sjukhus*, Reformera Nummer 15, Reforminstitutet, Bromma

Whinston, M D, 2003, "On the Transaction Cost Determinants of Vertical Integration", Journal of Law, Economics, & Organization, 19: 1-23

Williamson, O E, 1985, *The Economic Institutions of Capitalism*, Free Press, New York

Williamson, O E, 1999, "Public and Private Bureaucracies: A Transaction Cost Economics Perspective", Journal of Law, Economics and Organization, 15:306-342

Zeckhauser, R, 1970, "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives", Journal of Economic Theory, 2:10-26

Lund Economic Studies

1. Guy Arvidsson	Bidrag till teorin för verkningarna av räntevariationer, 1962
2. Björn Thalberg	A Trade Cycle Analysis. Extensions of the Goodwin Model, 1966
3. Bengt Höglund	Modell och observationer. En studie av empirisk anknytning och aggregation för en linjär produktionsmodell, 1968
4. Alf Carling	Industrins struktur och konkurrensförhållanden, 1968
5. Tony Hagström	Kreditmarknadens struktur och funktionssätt, 1968
6. Göran Skogh	Straffrätt och samhällsekonomi, 1973
7. Ulf Jakobsson och Göran Norman	Inkomstbeskattningen i den ekonomiska politiken. En kvantitativ analys av systemet för personlig inkomstbeskattning 1952-71, 1974
8. Eskil Wadensjö	Immigration och samhällsekonomi. Immigrationens ekonomiska orsaker och effekter, 1973
9. Rögnvaldur Hannesson	Economics of Fisheries. Some Problems of Efficiency, 1974
10. Charles Stuart	Search and the Organization of Marketplaces, 1975
11. S Enone Metuge	An Input-Output Study of the Structure and Resource Use in the Cameroon Economy, 1976
12. Bengt Jönsson	Cost-Benefit Analysis in Public Health and Medical Care, 1976
13. Agneta Kruse och Ann-Charlotte Ståhlberg	Effekter av ATP - en samhällsekonomisk studie, 1977
14. Krister Hjalte	Sjörestaureringens ekonomi, 1977
15. Lars-Gunnar Svensson	Social Justice and Fair Distributions, 1977
16. Curt Wells	Optimal Fiscal and Monetary Policy - Experiments with an Econometric Model of Sweden, 1978
17. Karl Lidgren	Dryckesförpackningar och miljöpolitik - En studie av styrmedel, 1978
18. Mats Lundahl	Peasants and Poverty. A Study of Haiti, London, 1979
19. Inga Persson-Tanimura	Studier kring arbetsmarknad och information, 1980
20. Bengt Turner	Hyressättning på bostadsmarknaden - Från hyresreglering till bruksvärdesprövning, Stockholm 1979
21. Ingemar Hansson	Market Adjustment and Investment Determination. A Theoretical Analysis of the Firm and the Industry, Stockholm 1981
22. Daniel Boda Ndlela	Dualism in the Rhodesian Colonial Economy, 1981
23. Tom Alberts	Agrarian Reform and Rural Poverty: A Case Study of Peru,

	1981
24. Björn Lindgren	Costs of Illness in Sweden 1964-75, 1981
25. Göte Hansson	Social Clauses and International Trade. An Economic Analysis of Labour Standards in Trade Policy, 1981
26. Noman Kanafani	Oil and Development. A Case Study of Iraq, 1982
27. Jan Ekberg	Inkomsteffekter av invandring, 1983
28. Stefan Hedlund	Crisis in Soviet Agriculture?, 1983
29. Ann-Marie Pålsson	Hushållen och kreditpolitiken. En studie av kreditrestrik-tioners effekt på hushållens konsumtion, sparande och konsumtionsmönster, 1983
30. Lennart Petersson	Svensk utrikeshandel, 1871-1980. En studie i den intraindustriella handelns framväxt, 1984
31. Bengt Assarsson	Inflation and Relative Prices in an Open Economy, 1984
32. Claudio Vedovato	Politics, Foreign Trade and Economic Development in the Dominican Republic, 1985
33. Knut Ödegaard	Cash Crop versus Food Crop Production in Tanzania: An Assessment of the Major Post-Colonial Trends, 1985
34. Vassilios Vlachos	Temporära lönesubventioner. En studie av ett arbetsmarknadspolitiskt medel, 1985
35. Stig Tegle	Part-Time Employment. An Economic Analysis of Weekly Working Hours in Sweden 1963-1982, 1985
36. Peter Stenkula	Tre studier över resursanvändningen i högskolan, 1985
37. Carl Hampus Lyttkens	Swedish Work Environment Policy. An Economic Analysis, 1985
38. Per-Olof Bjuggren	A Transaction Cost Approach to Vertical Integration: The Case of Swedish Pulp and Paper Industry, 1985
39. Jan Petersson	Erik Lindahl och Stockholmsskolans dynamiska metod, 1987
40. Yves Bourdet	International Integration, Market Structure and Prices. A Case Study of the West-European Passenger Car Industry, 1987
41. Krister Andersson and Erik Norrman	Capital Taxation and Neutrality. A study of tax wedges with special reference to Sweden, 1987
42. Tohmas Karlsson	A Macroeconomic Disequilibrium Model. An Econometric Study of the Swedish Business Sector 1970-84, 1987
43. Rosemary Vargas-Lundius	Peasants in Distress. Poverty and Unemployment in the Dominican Republic, 1989
44. Lena Ekelund Axelson	Structural Changes in the Swedish Marketing of Vegetables,

1991

45. Elias Kazarian — Finance and Economic Development: Islamic Banking in Egypt, 1991

46. Anders Danielson — Public Sector Expansion and Economic Development. The Sources and Consequences of Development Finance in Jamaica 1962-84, 1991

47. Johan Torstensson — Factor Endowments, Product Differentiation, and International Trade, 1992

48. Tarmo Haavisto — Money and Economic Activity in Finland, 1866-1985, 1992

49. Ulf Grönkvist — Economic Methodology. Patterns of Reasoning and the Structure of Theories, 1992

50. Evelyne Hangali Maje — Monetization, Financial Development and the Demand for Money, 1992

51. Michael Bergman — Essays on Economic Fluctuations, 1992

52. Flora Mndeme Musonda — Development Strategy and Manufactured Exports in Tanzania, 1992

53. Håkan J. Holm — Complexity in Economic Theory. An Automata Theoretical Approach, 1993

54. Klas Fregert — Wage Contracts, Policy Regimes and Business Cycles. A Contractual History of Sweden 1908-90, 1994

55. Per Frennberg — Essays on Stock Price Behaviour in Sweden, 1994

56. Lisbeth Hellvin — Trade and Specialization in Asia, 1994

57. Sören Höjgård — Long-term Unemployment in a Full Employment Economy, 1994

58. Karolina Ekholm — Multinational Production and Trade in Technological Knowledge, 1995

59. Fredrik Andersson — Essays in the Economics of Asymmetric Information, 1995

60. Rikard Althin — Essays on the Measurement of Producer Performance, 1995

61. Lars Nordén — Empirical Studies of the Market Microstructure on the Swedish Stock Exchange, 1996

62. Kristian Bolin — An Economic Analysis of Marriage and Divorce, 1996

63. Fredrik Sjöholm — R&D, International Spillovers and Productivity Growth, 1997

64. Hossein Asgharian — Essays on Capital Structure, 1997

65. Hans Falck — Aid and Economic Performance - The Case of Tanzania, 1997

66. Bengt Liljas — The Demand for Health and the Contingent Valuation

		Method, 1997
67.	Lars Pålsson Syll	Utility Theory and Structural Analysis, 1997
68.	Richard Henricsson	Time Varying Parameters in Exchange Rate Models, 1997
69.	Peter Hördahl	Financial Volatility and Time-Varying Risk Premia, 1997
70.	Lars Nilsson	Essays on North-South Trade, 1997
71.	Fredrik Berggren	Essays on the Demand for Alcohol in Sweden - Review and Applied Demand Studies, 1998
72.	Henrik Braconier	Essays on R&D, Technology and Growth, 1998
73.	Jerker Lundbäck	Essays on Trade, Growth and Exchange Rates, 1998
74.	Dan Anderberg	Essays on Pensions and Information, 1998
75.	P. Göran T. Hägg	An Institutional Analysis of Insurance Regulation – The Case of Sweden, 1998
76.	Hans-Peter Bermin	Essays on Lookback and Barrier Options - A Malliavin Calculus Approach, 1998
77.	Kristian Nilsson	Essays on Exchange Rates, Exports and Growth in Developing Countries, 1998
78.	Peter Jochumzen	Essays on Econometric Theory, 1998
79.	Lars Behrenz	Essays on the Employment Service and Employers' Recruitment Behaviour, 1998
80.	Paul Nystedt	Economic Aspects of Ageing, 1998
81.	Rasha M. Torstensson	Empirical Studies in Trade, Integration and Growth, 1999
82.	Mattias Ganslandt	Games and Markets – Essays on Communication, Coordination and Multi-Market Competition, 1999
83.	Carl-Johan Belfrage	Essays on Interest Groups and Trade Policy, 1999
84.	Dan-Olof Rooth	Refugee Immigrants in Sweden - Educational Investments and Labour Market Integration, 1999
85.	Karin Olofsdotter	Market Structure and Integration: Essays on Trade, Specialisation and Foreign Direct Investment, 1999
86.	Katarina Steen Carlsson	Equality of Access in Health Care, 1999
87.	Peter Martinsson	Stated preference methods and empirical analyses of equity in health, 2000
88.	Klas Bergenheim	Essays on Pharmaceutical R&D, 2000
89.	Hanna Norberg	Empirical Essays on Regional Specialization and Trade in Sweden, 2000
90.	Åsa Hansson	Limits of Tax Policy, 2000
91.	Hans Byström	Essays on Financial Markets, 2000

92. Henrik Amilon	Essays on Financial Models, 2000
93. Mattias Lundbäck	Asymmetric Information and The Production of Health, 2000
94. Jesper Hansson	Macroeconometric Studies of Private Consumption, Government Debt and Real Exchange Rates, 2001
95. Jonas Månsson	Essays on: Application of Cross Sectional Efficiency Analysis, 2001
96. Mattias Persson	Portfolio Selection and the Analysis of Risk and Time Diversification, 2001
97. Pontus Hansson	Economic Growth and Fiscal Policy, 2002
98. Joakim Gullstrand	Splitting and Measuring Intra-Industry Trade, 2002
99. Birger Nilsson	International Asset Pricing, Diversification and Links between National Stock Markets, 2002
100. Andreas Graflund	Financial Applications of Markov Chain Monte Carlo Methods, 2002
101. Therése Hindman Persson	Economic Analyses of Drinking Water and Sanitation in Developing Countries, 2002
102. Göran Hjelm	Macroeconomic Studies on Fiscal Policy and Real Exchange Rates, 2002
103. Klas Rikner	Sickness Insurance: Design and Behavior, 2002
104. Thomas Ericson	Essays on the Acquisition of Skills in Teams, 2002
105. Thomas Elger	Empirical Studies on the Demand for Monetary Services in the UK, 2002
106. Helena Johansson	International Competition, Productivity and Regional Spillovers, 2003
107. Fredrik Gallo	Explorations in the New Economic Geography, 2003
108. Susanna Thede	Essays on Endogenous Trade Policies, 2003
109. Fredrik CA Andersson	Interest Groups and Government Policy, A Political Economy Analysis, 2003
110. Petter Lundborg	Risky Health Behaviour among Adolescents, 2003
111. Martin W Johansson	Essays on Empirical Macroeconomics, 2003
112. Joakim Ekstrand	Currency Markets - Equilibrium and Expectations, 2003
113. Ingemar Bengtsson	Central bank power: a matter of coordination rather than money supply, 2003
114. Lars Pira	Staples, Institutions and Growth: Competitiveness of Guatemalan Exports 1524-1945, 2003
115. Andreas Bergh	Distributive Justice and the Welfare State, 2003

116. Staffan Waldo	Efficiency in Education - A Multilevel Analysis, 2003
117. Mikael Stenkula	Essays on Network Effects and Money, 2004
118. Catharina Hjortsberg	Health care utilisation in a developing country -The case of Zambia, 2004
119. Henrik Degrér	Empirical Essays on Financial Economics, 2004
120. Mårten Wallette	Temporary Jobs in Sweden: Incidence, Exit, and On-the-Job Training, 2004
121. Tommy Andersson	Essays on Nonlinear Pricing and Welfare, 2004
122. Kristian Sundström	Moral Hazard and Insurance: Optimality, Risk and Preferences, 2004
123. Pär Torstensson	Essays on Bargaining and Social Choice, 2004
124. Frederik Lundtofte	Essays on Incomplete Information in Financial Markets, 2005
125. Kristian Jönsson	Essays on Fiscal Policy, Private Consumption and Non-Stationary Panel Data, 2005
126. Henrik Andersson	Willingness to Pay for a Reduction in Road Mortality Risk: Evidence from Sweden, 2005
127. Björn Ekman	Essays on International Health Economics: The Role of Health Insurance in Health Care Financing in Low- and Middle-Income Countries, 2005
128. Ulf G Erlandsson	Markov Regime Switching in Economic Time Series, 2005
129. Joakim Westerlund	Essays on Panel Cointegration, 2005
130. Lena Hiselius	External costs of transports imposed on neighbours and fellow road users, 2005
131. Ludvig Söderling	Essays on African Growth, Productivity, and Trade, 2005
132. Åsa Eriksson	Testing and Applying Cointegration Analysis in Macroeconomics, 2005
133. Fredrik Hansen	Explorations in Behavioral Economics: Realism, Ontology and Experiments, 2006
134. Fadi Zaher	Evaluating Asset-Pricing Models in International Financial Markets, 2006
135. Christoffer Bengtsson	Applications of Bayesian Econometrics to Financial Economics, 2006
136. Alfredo Schclarek Curutchet	Essays on Fiscal Policy, Public Debt and Financial Development, 2006
137. Fredrik Wilhelmsson	Trade, Competition and Productivity, 2006
138. Ola Jönsson	Option Pricing and Bayesian Learning, 2007

139. Ola Larsson	Essays on Risk in International Financial Markets, 2007
140. Anna Meyer	Studies on the Swedish Parental Insurance, 2007
141. Martin Nordin	Studies in Human Capital, Ability and Migration, 2007
142. Bolor Naranhuu	Studies on Poverty in Mongolia, 2007
143. Margareta Ekbladh	Essays on Sickness Insurance, Absence Certification and Social Norms, 2007
144. Erik Wengström	Communication in Games and Decision Making under Risk, 2007
145. Robin Rander	Essays on Auctions, 2008
146. Ola Andersson	Bargaining and Communication in Games, 2008
147. Marcus Larson	Essays on Realized Volatility and Jumps, 2008
148. Per Hjertstrand	Testing for Rationality, Separability and Efficiency, 2008
149. Fredrik NG Andersson	Wavelet Analysis of Economic Time Series, 2008
150. Sonnie Karlsson	Empirical studies of financial asset returns, 2009
151. Maria Persson	From Trade Preferences to Trade Facilitation, 2009
152. Eric Rehn	Social Insurance, Organization and Hospital Care, 2009
153. Peter Karpestam	Economics of Migration, 2009