Stockholms universitet

Bureaucracy, Informality and Taxation: Essays in Development Economics and Public Finance

Anders Fredriksson

ISBN 978-91-7155-874-9
ISSN 0346-6892

Printed in Sweden by Universitetsservice, US-AB, Stockholm 2009
Distributor: Institute for International Economic Studies

Doctoral Dissertation
Department of Economics
Stockholm University

Abstract

This thesis consists of three self-contained essays.

Essay 1, "Dispatchers", is a study of a specialized service sector that has arisen in many developing countries, assisting individuals and firms in their contacts with the government bureaucracy.

It is a well-established fact that the government bureaucracy in many developing countries is large, difficult to understand, non-transparent and time-consuming. However, "de jure" bureaucratic procedures sometimes have little to do with how firms or individuals actually go about when dealing with the government bureaucracy. One institution that has emerged in many countries is a specialized intermediary, henceforth called dispatcher, that assists individuals and firms in their contacts with the public sector. It is often the workings of this "de facto" institution, rather than the de jure procedure, that determines outcomes. A model where firms demand a license from the government bureaucracy is developed in order to address two sets of questions related to the use of dispatchers. First, what is the impact of dispatchers on time and resources that firms spend in obtaining licenses and what is the impact on the degree of informality, i.e. on the fraction of firms that choose to not get the license? How do these results depend on the organization of bureaucrats and dispatchers, the regulatory framework and the extent of corruption in the bureaucracy? Second, what are the incentives of corrupt bureaucrats and dispatchers to try to make regulation more/less complicated? When are the incentives of bureaucrats and dispatchers to create "red tape" aligned? Ultimately and ideally, the answers to these questions can help explain why reforms of the public sector have been so difficult.

Essay 2, "Informal firms, investment incentives and formalization", studies investment, growth and possible formalization of small informal firms in developing countries.

In a typical developing country, the majority of small firms are informal and entry costs into formality are high. This paper is motivated by these two observations.

It asks the question of what can be expected in terms of firm investment, growth and formalization in such a setting. It also studies the effects of policies towards the informal sector on formalization decisions. I show that the investment paths and growth trajectories differ substantially between firms that choose to formalize and those (ex-ante almost identical firms) that do not. Second, the formalization decision depends non-trivially on the productivity of the informal firm, due to the balancing of an accumulation effect and a threshold effect. This, in turn, has an effect on how policies towards the informal sector should be designed. Third, when aggregating over firms, the long-run firm size distribution exhibits a range of small firms and a range of larger firms, but also a "missing middle", much in line with actual firm size distributions observed in developing countries. Fourth, the long-run firm-size distribution turns out to depend on the initial firm-level stock of capital, a result that can be interpreted as a poverty/informality trap.

Essay 3, "Compositional and dynamic Laffer effects in a model with constant returns to scale", studies dynamic effects of tax cuts.

There is a renewed interest in the dynamic effects of tax cuts on government revenue. The possibility of tax cuts paying for themselves over time definitely seems like an attractive option for policy makers. This paper looks at what conditions are required for reductions in capital taxes to be fully self-financing. This is done in a model with constant returns to scale in broad capital. Such a framework exhibits growth; the scope for self-financing tax cuts is therefore different than in the neoclassical growth model, most recently studied by Mankiw and Weinzierl (2006) and Leeper and Yang (2008). Compared to previous literature, I make a methodological contribution in the definition of Laffer effects and clarify the role of compositional and dynamic effects in making tax cuts self-financing. I also provide simple analytical expressions for what tax rates are required for tax cuts to be fully self-financing. The results show that large distortions are required to get Laffer effects. Introducing a labor/leisure choice into the model opens up a new avenue for such effects, however.

Preface

One of the essays in this thesis, chapter 4, was written at the Department of Economics at Stockholm University. It constituted my licentiate thesis, defended during the autumn of 2005. It was written under the guidance of Jonas Agell, who passed away in 2007.

During the short time I knew Jonas, I experienced not only a very skilled and analytically minded researcher and advisor. Jonas also cared much for the people in his surroundings; colleagues, students and others alike. Whenever I had meetings with Jonas he was very positive and encouraging, from which I benefited much. I often think of Jonas.

I think it was the curiosity to learn and understand more, combined with experiences from different situations while traveling and living in a few countries in Latin America, that made me take the decision to quit my job at Ericsson Brazil and apply to the PhD program in Economics. At some stage into the program, when I first started talking to Jakob Svensson about the informal sector in Latin America, I knew that Development Economics was going to be my main field and that I wanted to have Jakob as an advisor.

If I were to single out one aspect - out of many - for which to thank Jakob, it would be for your encouragement, support and for believing in me and my projects. It has made all the difference over the past years. For this I am forever indebted.

I am sincerely grateful, not only for discussions, feedback, careful reading and thoughtful advice on how to move forward. I would also like to thank you for supporting and rendering possible the different ways in which this thesis has come about; including much useful and highly appreciated work periods in the U.S. and Brazil. Your advice has been the decisive factor in my process of learning how to formulate a research idea and take it to a paper.

I also want to thank you for showing support in a broader sense than the purely work-related, for listening and for being accessible. I am sincerely grateful for your constant encouragement and advice.

I would like to thank all Professors, graduate students and administrative staff at the IIES. Thanks to its people, "The Institute" is not only an exceptional place to do research but is also characterized by a friendly and encouraging atmosphere. I have discussed my work with Professors and graduate students, which I acknowledge

in the relevant chapters. I am indebted to David Strömberg and Fabrizio Zilibotti for guiding and supporting me at an early stage of the program and to John Hassler, Ethan Kaplan and Masayuki Kudamatsu for always being supportive and available to discuss my work. I would like to thank Christina Lönnblad and Annika Andreasson for highly professional administration and for assistance in finalizing this thesis. I would also like to thank Christina for excellent editorial assistance.

I want to express my gratitude to the Department of Economics, where I started my PhD-studies. I am indebted to Ingela Arvidsson and Karl Eriksson for support. Financial support from Handelsbankens Forskningsstiftelser, Stockholms Universitets Donationsstipendier and Knut och Alice Wallenbergs Stiftelse is gratefully acknowledged.

A few years back, I had the opportunity to spend one semester at the University of California at Berkeley. I want to thank its Economics faculty and PhD-students. I also spent six weeks at Fundação Getulio Vargas (FGV) in Rio de Janeiro. Maria Cristina Terra and Samuel Pessôa made my stay there possible, for which I am very grateful. I also thank FGV's PhD-students.

During my PhD-studies, I have received different kinds of support, at various stages, from Anita Karlsson, Christina Lönnblad, Hans Wijkander and Jonas Häckner. I want you to know that I am very grateful.

I am forever indebted to Leena and Seija. Without your support, this thesis would never have been written.

When I started the program, I remember having the expectation of getting to know many interesting people, thinking that an international PhD-program in Economics would bring together people with a lot of different backgrounds and interests. This turned out to be very much true. It has been a true privilege to be in the environment that the PhD-program and the "Economics environment" in Stockholm constitutes!

I am indebted to Ulrika Stavlöt and Andreas Madestam, who guided me prior to starting the program. Many thanks also to fellow PhD-students at Stockholm University from the first years: Daria Finocchiaro, Helena Holmlund, Irina Slinko, Jaewon Kim, Magnus Wiberg, Martina Björkman, Mirco Tonin, Virginia Queijo von Heideken and, in particular, Anna Larsson, for much appreciated encouragement and friendship.

Another colleague, José Mauricio Prado, became a good friend from the start.

Whether it was fruitful work discussions - from which I benefited much, walking around Brunnsviken or having a beer in his native Santos while watching the "Receita Federal" speedboat, I enjoyed it. Thanks a lot! I also would like to thank Sara Åhlén and Martin Bech Holte, that I shared an office with, at different stages. To Sara, Anders Åkerman, Gülay Özcan, Karl Brenklert and Mema: All the dinners and activities were fun!

I would like to thank Lars Johansson. Your friendship truly makes a difference and it means a lot to me. I would also like to thank Fredrik Brunes and Fredrik Wilander for much valued friendship.

The past few years has meant a strictly increasing number of friends indeed. To Agatha Murgoci, Annika Nilsson, Bo Larsson, Carl Wilkens, Carlos Razo, Caterina Mendicino, Elena Paltseva, Erika Färnstrand Damsgaard, Gisela Waisman, Jan Pettersson, Mia Hinnerich, Raquel Gaspar, Shon Ferguson and everyone else: it is a true pleasure to know you all!

One of the salient features of the PhD-program in Stockholm, to which all the different countries of origin of my colleagues above bear testimony, is how international it is. The same holds for my friends outside academia and former colleagues - both in country of origin and in current location. I would like to thank all; you are a source of much inspiration and I dearly value your friendship. A special thanks goes to Pär Egnell, not only for the different running experiences, but also for always listening to, and showing interest in, my different projects.

Finally, I want to thank my family - my parents Gulli and Leif, my sisters Annika and Åsa and my brother Peter - for your interest in my work and its progress. Your support, patience, understanding and love means a lot to me.

Stockholm, May 2009

Anders Fredriksson

Table of Contents

Chapter 1:	Introduction	1
Chapter 2:	Dispatchers	9
Chapter 3:	Informal firms, investment incentives and formalization	55
Chapter 4:	Compositional and dynamic Laffer effects in a model with constant returns to scale	95

Chapter 1

Introduction

The first essay in this thesis, "Dispatchers", is a study of a specialized service sector that has arisen in many developing countries, assisting individuals and firms in their contacts with the government bureaucracy. The second essay, "Informal firms, investment incentives and formalization", studies investment, growth and possible formalization of small informal firms in developing countries. The third essay, "Compositional and dynamic Laffer effects in a model with constant returns to scale", studies dynamic effects of tax cuts.

The motivation for the first two essays, "Dispatchers" and "Informal firms, investment incentives and formalization", comes from the observation that the government bureaucracy in many developing countries is large, difficult to understand, non-transparent and time-consuming. Going through procedures at the government bureaucracy can be very costly, both in terms of financial resources and time costs.

For individuals, procedures such as getting an ID, getting a passport, getting birth certificates, opening a bank account, paying utility bills, receiving one's pension, getting the license plates to a new car and regulating the purchase of a used car, frequently involve many interactions with a multitude of government offices. For firms, procedures such as starting the firm itself, maintaining the legal status of the firm, obtaining licenses and permits, employing workers, clearing goods through customs, paying taxes and even closing the firm, are all examples of procedures that can be very costly. As for the time it takes individuals and firms to go through bureaucratic procedures, the difference between a typical developing country and, say, the average OECD-country, can be large indeed.

Take the procedure to start a firm as an example. Information on such start-up procedures, i.e. the procedures a firm must go through, at the government bureaucracy, in order to register as a legal entity, are reported by a World Bank project ("Doing Business"). The Doing Business project has documented different aspects of the business environment, across the globe, over the last 5-10 years (see Djankov et al., 2002; World Bank, 2009a). Whereas the financial cost to start a firm is USD 370 in the United States, the average cost in Latin America is around USD 1240, The average monthly income per capita was USD 3840 in the United States in 2007, meaning that three days of work generate an income equal to the firm start-up cost. In Latin America, the average monthly income was one tenth as much, or USD 380. Thus, it takes more than three months of work to generate an income equal to the firm start-up cost.

One shortcoming in the discussion of bureaucracy above is that often the "de facto" way of firms' and individuals' dealings with the bureaucracy has little to do with formal "de jure" procedures. One institution that has emerged in many countries is a specialized intermediary, henceforth called "dispatcher", which assists individuals and firms in bureaucratic contacts. In many cases firms and individuals simply do not go to the bureaucracy, but use a dispatcher instead. Consider, for instance, the procedure to start up a business in Brazil: According to the official procedure, the potential entrant has to go through up to 17 different steps. Although the total financial burden is not extremely high, the time cost – both in completing all the steps and in waiting to get the actual license (152 days) - is high (World Bank, 2009a).

However, registered businesses in Brazil report little problems in going through the procedure (Stone et al, 1996). An intermediary agent such as a "contador" or a "despachante" acts – at least in some instances - much as a "one stop shop" for entrant firms and virtually all firms go through such an intermediary, instead of through the bureaucracy itself (Stone et al, 1996; Zylbersztajn and Graça, 2003; Zylbersztajn et al., 2007). More generally, in Brazil and elsewhere there are intermediaries to assist with most, if not all, interactions with the bureaucracy that firms, or individuals, may need to go through. What this means is thus that in many circumstances, the "de jure" procedure has been replaced by the "de facto" institution of "dispatchers".

In the first essay, "Dispatchers", I build a model where firms demand a license from the government bureaucracy and where the licenses can be acquired either by going through the "de jure" bureaucratic procedure, by bribing corrupt bureaucrats or through dispatchers. Firms can also remain informal, i.e. choose to not get the license at all. To the best of my knowledge, this is one of the first papers to set up a model of bureaucracy intermediation, discuss and formally model the interaction between bureaucrats and dispatchers, discuss and model the organization of bureaucrats and of dispatchers, and study the effects of dispatchers on license allocations and informality.

In the paper, firms want to use dispatchers because these save time for firms. Bureaucrats, on the other hand, use dispatchers as a means of increasing corruption revenue. The paper is centered around two sets of questions: First, what is the impact of dispatchers on time and resources that firms spend in obtaining licenses and what is the impact on the degree of informality, i.e. on the fraction of firms that choose to not get the license? How do these results depend on the organization of bureaucrats and dispatchers, the regulatory framework and the extent of corruption in the bureaucracy?

The second set of questions relates to the incentives of bureaucrats and dispatchers. What are the incentives of corrupt bureaucrats and dispatchers to try to make regulation more/less complicated? When are the incentives of bureaucrats and dispatchers to create "red tape" aligned? To what extent do the answers to these questions depend on the organization of bureaucrats and dispatchers? Ultimately and ideally, the answers to these questions can help explain why reforms of the public sector have been so difficult.[1]

The analysis shows that dispatchers can contribute substantially to improve welfare, something which contrasts with the few previous papers on the subject. I also show that dispatchers can reduce informality in bureaucratic procedures, whereas this extensive margin is often ignored in papers concerned with license allocations. The conditions under which bureaucrats and dispatchers have incentives to create red tape, possibly reversing welfare gains, are also analyzed. I argue that the incentives to create red tape can be stronger when there are dispatchers, as compared to the situation when such middlemen do not exist. As a consequence, although dispatchers ceteris paribus increase welfare in the model, their existence can also make reform of the public sector more difficult.

[1] Although the model in chapter 2 concerns *firms*, it can be broadly interpreted. That is, it can hopefully serve as a model for also understanding *individuals*' use of dispatchers.

In an extension of the main analysis, I show that the welfare implications of the dispatcher sector change dramatically when the role of dispatchers is to facilitate "rule breaking", rather than to save time for firms.

Having analyzed the "de facto" institution of dispatchers, in the next essay, “Informal firms, investment incentives and formalization”, I instead take the "de jure" costs of the bureaucracy literally. This essay is concerned with one specific bureaucratic procedure: firm formalization. It studies firms’ decision of whether to become formal, i.e. to go through the procedure at the government bureaucracy to register as a legal entity.

An important motivation for chapter 3 is that, in a typical developing country, the majority of small firms are informal. That is, firms have not gone through the procedure to register as a legal entity. This observation is well-established. A recent enterprise survey in Brazil shows that 90% of the smallest firms, i.e. of firms with 1-5 employees, have not gone through the procedure to register (SEBRAE, 2005). An enterprise survey in Mexico, the other large Latin American economy, shows similar values (INEGI, 2003). Studies and accounts from other developing countries indicate similar degrees of informality among the smallest firms in the economy.

I take literally the observation that going through the registration procedure at the government bureaucracy is very costly in terms of financial resources, as compared to typical capital stock replacement values or profit levels of the smallest firms. Using a dynamic model, I address the question of what is the effect of large formalization costs on investment, growth and the formalization decision for small informal firms. I also study the effects of policies towards the informal sector, on the same investment and formalization decisions.[2]

How do formalization costs, to be paid at some future date, affect investment today? At what firm size and when do firms choose, if at all, to become formal? What are the crucial parameters affecting firm formalization? What is the effect

[2] The informal sector has interested academics for more than 35 years. For the questions I study, the most important reference is probably de Soto (1989) who focused on the widespread informality observed in Peru and how such informality can result from a complex bureaucratic and legal system that does not correspond to the needs of ordinary citizens. The work by Djankov et al. (2002) and World Bank (2009a) follows in the vein of de Soto in that it focuses on the amount of regulation in different countries and the effects of (over-) regulating on e.g. business start-up and informality.

of credit constraints on the formalization decision? Can formalization costs lead to poverty traps? How can the government affect the formalization decision?

Similarly to the other two essays in this thesis, the analysis in chapter 3 is mainly positive rather than normative, although the model also includes a policy parameter by which the government can affect the informal sector's incentives to invest and formalize.

The contribution of chapter 3 is that, to the best of my knowledge, it is the first paper to use a dynamic framework to explicitly focus on investment incentives and growth of informal firms, in anticipation of formalization. The main results emerging from the model are that the investment paths and growth trajectories differ substantially between firms that choose to formalize and those (ex-ante almost identical firms) that do not. Second, the formalization decision depends non-trivially on the productivity of the informal firm, due to the balancing of an accumulation effect and a threshold effect. This, in turn, has an effect on how policies towards the informal sector should be designed. Third, when aggregating over firms, the long-run firm size distribution exhibits a range of small firms and a range of larger firms but also a "missing middle", much in line with actual firm size distributions observed in developing countries (Bigsten et al. 2004, Tybout, 2000). Fourth, the long-run firm-size distribution turns out to depend on the initial firm-level stock of capital, a result that can be interpreted as a poverty/informality trap.

Chapter 4, "Compositional and dynamic Laffer effects in a model with constant returns to scale", is a public finance paper and not connected to the two previous essays. It studies dynamic effects of tax cuts.

In an AK-model, Agell and Persson (2001) both defined dynamic Laffer effects of taxes on physical capital and identified what conditions are required to get self-financing tax cuts. I modify the analysis of these authors by introducing human capital and a labor/leisure choice in the AK-model to make three main points. First, I further define "Laffer effects" in the constant returns models by dividing effects of tax cuts into dynamic and compositional effects. This is crucial when there is more than one factor of production. Second, simple analytical expressions for when tax cuts in AK-style models will fully finance themselves are provided. Third, the introduction of the labor/leisure choice, which follows both the endogenous growth literature and Mankiw and Weinzierl (2006), gives rise to a new margin for self-financing tax cuts; these effects are worked out.

Having added leisure to the model, we have a framework with three incentive margins that, as a result of tax cuts, can create Laffer effects on their own or in combination. The three incentive margins are: 1) dynamic effects of taxes on interest and growth rates, 2) compositional effects of taxes on production (an "uneven playing field") and 3) the labor/leisure choice. In a world with the first – dynamic – effect only, there is a direct revenue effect of a tax cut and an indirect effect of different interest and growth rates. The second – compositional – effect comes in when physical and human capital are taxed differently; the current tax base is then also affected by tax cuts, adding to the direct revenue effect and the growth effect. Adding the third margin – leisure – there is an additional effect on the tax base through a different labor/leisure choice after a tax cut and there is also an additional growth effect.

In this setup, I derive what combinations of tax rates on physical and human capital are required for a tax cut to be self-financing. The results suggest that dynamic and compositional distortions will need to be large if there are to be Laffer effects; less so, however, if the model contains a labor/leisure choice. I show that the margin opened up by the endogenous labor/leisure choice may be quantitatively important.

Bibliography

[1] Agell, J., Persson, M., 2001. On the analytics of the dynamic Laffer curve. Journal of Monetary Economics 48, 397-414.

[2] Bigsten, A., Kimuyu, P., Lundvall, K., 2004. What to Do with the Informal Sector? Development Policy Review 22, 701-715.

[3] de Soto, H., 1989. The Other Path. New York: Harper and Row.

[4] Djankov, S., La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2002. The regulation of entry. The Quarterly Journal of Economics 117, 1-37.

[5] INEGI - Instituto Nacional de Estadística, Geografía, e Informática, 2003. Encuesta nacional de micronegocios 2003.

[6] Mankiw, G., Weinzierl, M., 2006. Dynamic scoring: A back-of-the-envelope guide. Journal of Public Economics 90, 1415-1433.

[7] SEBRAE, 2005. Economia Informal Urbana 2003.

[8] Stone, A., Levy, B., Paredes, R., 1996. Public Institutions and Private Transactions: A Comparative Analysis of the Legal and Regulatory Environment for Business Transactions in Brazil and Chile, in "Empirical studies in institutional change" (Alston, Eggertsson & North eds.)

[9] Tybout, J., 2000. Manufacturing Firms in Developing Countries: How Well Do They Do, and Why? Journal of Economic Literature 38, 11-44.

[10] World Bank, 2009a. http://www.doingbusiness.org

[11] Zylbersztajn, D., Faccioli, F., Silveira, R., 2007. Measuring the start up costs in Brazilian small firms. Revista de Administração da Universidade de São Paulo 42, 293-301.

[12] Zylbersztajn, D., Graça, C., 2003. Los costos de la formalización empresarial: medición de los costos de transacción en Brasil. Revista de Economía Institucional 5, 146-165.

Chapter 2

Dispatchers*

1 Introduction

It is a well-established fact that the government bureaucracy in many developing countries is large, difficult to understand, non-transparent and time-consuming. In the countries with the longest delays, it takes approximately a factor of one hundred as long time to start up a firm as it does in the countries with the shortest delays. In the most expensive countries, it costs approximately a factor of one thousand as much as in the least expensive countries (measured in relation to each country's GNI). For this same procedure (starting a firm), the fastest and least expensive procedures are found in the developed world, while the longest delays and most expensive procedures are found in the developing world. Similar differences hold for other bureaucratic procedures (Djankov et al., 2002; World Bank, 2009a).

However, such reported "de jure" procedures sometimes have little to do with how firms or individuals actually go about when dealing with the government bureaucracy. One institution that has emerged in many countries is a specialized intermediary, henceforth called dispatcher, that assists individuals and firms in their

* This project has benefited from discussions, feedback and comments from several people. In particular, thanks to Jeanet Bentzen, Nabanita Datta Gupta, Shon Ferguson, John Hassler, Seema Jayachandran, Ethan Kaplan, Jan Klingelhöfer, Per Krusell, Masayuki Kudamatsu, Victor Lapuente, Marcela Lilia Camargo, Rocco Macchiavello, Dilip Mookherjee, Annika Nilsson, Maria Perrotta, José Mauricio Prado, Sylvia Saes, Bruno Salama, Claes Sandgren, Paulo Eduardo da Silva, David Strömberg, Jakob Svensson, Eric Verhoogen and Decio Zylbersztajn. I also thank seminar participants at the IIES and the Department of Economics at Stockholm University and participants at the Nordic Conference of Development Economics 2008, ISNIE 2008, the CEPR/BREAD development summer school in Verona 2008 and at LACEA 2008. Thanks also to Christina Lönnblad for editorial assistance. Any remaining errors are my own.

contacts with the public sector. It is often the workings of this "de facto" institution, rather than the de jure procedure, that determines outcomes such as how many firms or individuals go through a certain bureaucratic procedure, processing times, waiting times and financial costs.

In the strict meaning of the word, a dispatcher is someone that expedites tasks, that gets tasks out of his hands, that gets things done. Here, it also takes the meaning of facilitator as well as proxy and someone with power of attorney. That is, the dispatcher is an intermediary that can represent an individual or a firm at the government bureaucracy, in order to expedite and facilitate tasks that the individual or firm needs to get done.

This paper is centered around two sets of questions: First, what is the impact of dispatchers on time and resources that firms spend in obtaining licenses and what is the impact on the degree of informality, i.e. on the fraction of firms that choose to not go through a particular bureaucratic procedure? How do these results depend on the organization of bureaucrats and dispatchers, the regulatory framework and the extent of corruption in the bureaucracy?

The second set of questions relates to the incentives of bureaucrats and dispatchers. What are the incentives of corrupt bureaucrats and dispatchers to try to make regulation more/less complicated? When are the incentives of bureaucrats and dispatchers to create "red tape" aligned? To what extent do the answers to these questions depend on the organization of bureaucrats and dispatchers? Ultimately and ideally, the answers to these questions can help explain why reforms of the public sector have been so difficult.

To address these questions, I develop a model in which there are firms that demand a license from the government bureaucracy. The license brings a production benefit to the firms and is acquired by going through a procedure at the government bureaucracy. The procedure consists of a number of steps to be completed, where each step involves an interaction with one bureaucrat that is a monopolist in this specific step of the procedure. Each step has a financial cost, a time cost at the bureaucrat ("standing in line") and a time cost of going to the bureaucrat ("transport"). The number of steps in the procedure is an important parameter in the model.

The license can also be acquired through a dispatcher. Firms have an incentive to use dispatchers to get the license because these eliminate the time cost of going through the bureaucratic procedure. Dispatchers function as a "one stop shop" to get the license. In section 3, I motivate this way of modeling dispatchers by providing evidence from different parts of the world on the time saving nature of the dispatcher activity. I also discuss reasons for why dispatchers have a "superior technology" in handling bureaucratic procedures.

A closer look at the dispatching activity reveals that demand-side reasons related to time saving are not the full story. The dispatching activity is common in countries which also rank high on bureaucratic corruption and, as already indicated above, the modeling of corruption is a key element of the paper.

Depending on the type of corruption, the reasons for the existence of dispatchers differ. I model both the case where the bureaucracy is characterized by "according to rules corruption" and the case where there is instead "bending the rules corruption". In addition, the number of steps of the procedure where there are corrupt bureaucrats is another parameter in the analysis.

When corruption is "according to rules", which is the main model in the paper (sections 3-7), corrupt bureaucrats accept bribes directly from firms and let these avoid the time cost at the bureaucracy (i.e. avoid "standing in lines"). Bureaucrats still perform their job, however. Bribing corrupt bureaucrats to avoid standing in line is thus also a means through which to acquire the license. The supply-side reason for the existence of dispatchers in this scenario is that by accepting bribes not only directly but also through dispatchers, corrupt bureaucrats can earn a higher bribe revenue.

The license can thus be acquired either through the de jure bureaucracy, through bribing or through dispatchers. The firm can also choose to not get the license at all, i.e. to remain informal.

To the best of my knowledge, this paper is the first to make explicit the demand-side role of dispatchers as time-savers and the supply-side role as a channel for "according to rules corruption", and use this framework to analyze the effects on time/resources spent on getting licenses, on informality and on incentives to create red tape. Together with Bose and Gangopadhyay (2008), it is also one of the first papers that explicitly models profit maximizing dispatchers ("intermediary" in their terminology). In addition, the effects of different organization structures of corrupt bureaucrats as well as of different degrees of competition in the dispatcher sector are analyzed.[1]

In an extension in section 8.2, the model is modified and reinterpreted, and corruption is instead "bending the rules". That is, firms bribe to get a reduction in regulation. As an example of clearly detrimental "bending the rules" corruption involving intermediaries, Bertrand et al. (2007) provide evidence that the way to obtain a driver´s license in Delhi, India, without actually learning how to drive, is through an intermediary ("agent" in their terminology).

With bending the rules corruption, there are additional reasons for the existence of dispatchers. From the demand side, if firms trying to bribe honest bureaucrats can get fined, it will be safer to work through dispatchers because these possess a superior knowledge about which bureaucrats are corrupt. The expected penalty for firms will then be lower. Another argument is that indirect bribing, i.e. going through dispatchers, can provide the reduction in regulation for sure, whereas direct bribing cannot (Hasker and Okten, 2007; Bose and Gangopadhyay, 2008). In both cases, the expected benefit from bribing is higher when going through dispatchers than when bribing directly.

From the supply side, bureaucrats that bend the rules can get caught from not following the law.[2] By working through intermediaries, with whom bureaucrats have a repeated relationship, the risk of getting caught is reduced. This is because bureaucrats and dispatchers are "partners in crime", both earning corruption profits. They are therefore unlikely to reveal each other.

In section 8.2, these demand-side and supply-side reasons for the existence of dispatchers are included in the model.

The main contributions of the paper are fourfold. First, it provides one of the first

---

[1] In its analysis of red tape, the paper is also somewhat related to the work by Lui (1985). Whereas in Lui's model, corrupt bureaucrats may or may not have an incentive to slow down/speed up a bureaucratic procedure, in this paper, red tape works through corrupt bureaucrats' incentives to increase or decrease the number of steps in the procedure. Besides the channels being different to those in Lui (1985), the value added here is also to add intermediaries into the red tape analysis.

[2] Bending the rules corruption is thus considered to be a more severe type of corruption. I abstract from a government and penalties altogether in the "according to rules" case.

formal analyses of the dispatcher function and of the organization of bureaucrats and dispatchers. Second, the paper explicitly models an extensive margin – informality – and shows that dispatchers can reduce informality in bureaucratic procedures. Third, the analysis shows that dispatchers can contribute substantially to improve welfare, something which contrasts with earlier literature on the subject. Fourth, the conditions under which bureaucrats and dispatchers have incentives to create red tape, possibly reversing welfare gains, are analyzed.

The contribution of the "bending the rules" case is the contrast it provides with the main model. Instead of the possibility that dispatchers may improve welfare, the results resemble those of Hasker and Okten (2007). With a different view on regulation altogether, these authors provide a model with "bending the rules" corruption and show that intermediaries can have a detrimental impact on the amount of regulation effectively faced by firms and that anticorruption policies become ineffective in the presence of intermediaries.

The paper proceeds as follows. Some stylized facts about dispatchers are presented in section 2. The main model, with "according to rules" corruption, is presented in section 3. In section 4, the model is solved for the case without dispatchers and in section 5 for the case with dispatchers. The solution is discussed in section 6. In section 7, the incentives for bureaucrats/dispatchers to create "red tape" are discussed. The extension/reinterpretation of the model, with "bending the rules" corruption, in addition to some further IO aspects of the model, are presented in section 8. Section 9 discusses the results and concludes. Finally, some of the derivations are found in the appendix.

2 Stylized facts about dispatchers

Corruption and red tape in the government bureaucracy are phenomena that are fairly well-understood, at least from a theoretical viewpoint (see for instance Bardhan, 1997; Rose-Ackerman, 1999 and Svensson, 2003). Much less is known about dispatchers in general, the relation to corruption and the interplay between bureaucrats and dispatchers. This section presents stylized facts about dispatchers in different parts of the world and provides a rationale for the model to be presented.

Different types of intermediaries assisting with bureaucratic contacts, "dispatchers", are common throughout the developing world. Myrdal (1968) documents their existence in India and Oldenburg (1987) goes further with a more formal account of the role of intermediaries in a land consolidation program in Uttar Pradesh. Oldenburg identifies different roles of intermediaries within and outside the bureaucracy and details the functions of "brokers", "touts", "scribes", "consolidators", "helpers" and "barkers" within the land consolidation program. Levine (1975) documents the existence of intermediaries in the interface between the Ghanaian bureaucracy and firms and individuals.

The prevalence of "despachantes", used in bureaucratic contacts in Brazil, is documented by Rosenn (1971) and, from a sociological and anthropological viewpoint, by DaMatta (1979, 1984). When studying the formalization of firms, Stone et al. (1996), Zylbersztajn and Graça (2003) and Zylbersztajn et al. (2007) provide evidence that using "despachantes" is the most common way to formalize a firm in Brazil. Husted (1994) describes how "coyotes" help individuals obtain drivers' licenses in Mexico. Such "coyotes" are an example of "tramitadores", a more general and widely used term for (mostly) informal intermediaries present in most of (Spanish-speaking) Latin America, assisting individuals and firms with bureaucratic procedures ("tramites"). Proética (2006) documents, for Peru, the degree of individuals' usage of "tramitadores" in different bureaucratic contacts. Lambsdorff (2002) refers to "tramitadores" helping out with the bureaucracy in El Salvador. Examples of reports documenting the use of such intermediaries by firms, at formalization are CIET (1998a, b) and IFC (2007b) for Bolivia, CIEN (2001) for Guatemala, IFC (2008) for Honduras and IFC (2007a) for Peru.[3] Gancheva (1999) describes the use of similar intermediaries by firms in Bulgaria.

Although none of the papers above, with the possible exception of Oldenburg (1987), is a specific study of the dispatcher function, they point at different functions performed by such intermediaries. In some settings, the main reason why individuals use dispatchers seems to be the dispatchers' knowledge of how bureaucratic procedures actually work. In many countries with large and non-transparent bu-

[3] Another generic name, much in use in some parts of (Spanish-speaking) Latin America, for the type of intermediary in mind, is "gestor".

reaucracies, actually finding out what is required in order to get, say, a passport, is a challenge in itself. Rosenn (1971) writes: "The despachante functions effectively because he knows how to fill out the bewildering variety of forms, to whom the copies should be delivered, and what documentation will be required" (p. 537). Honduran firms claim that they use "tramitadores", when becoming formal, because of lack of unified information from the authorities regarding the formalization procedure (IFC, 2008). The same holds in a small sample of microenterprises in Guatemala (CIEN, 2001). For Bulgarian firms obtaining an operations permit, "the procedures are not clear, nor are they easily accessible to potential licenses applicants" (Gancheva, 1999, p. 22).

Time-saving in bureaucratic procedures is a related reason why individuals and firms use dispatchers. Data from the World Bank Enterprise surveys on senior management time spent in dealing with requirements of government regulation confirm that the time spent with regulation varies a great deal between different parts of the world. Whereas the OECD average is 1.2% of a work week, the world average is 7.5% and the Latin American/Caribbean average is 11.4% (World Bank, 2009b). A 1996 report studying only a few countries showed similar values for the Latin American countries (World Bank, 1996). The numbers confirm earlier work by de Soto (1989).

By frequent interactions, dispatchers learn how to handle the procedures at the government offices and can solve the bureaucratic matters faster than a particular individual or firm. Dispatchers' processing of many applications at the same time and having personal relations with bureaucrats are additional reasons why these intermediaries possess a "superior technology". As a result, the dispatchers' cost for acquiring licenses are lower.

Furthermore, Stone et al. (1996) and Zylbersztajn and Graça (2003) indicate that firms use despachantes to become formal because these act much like "one stop shops". The time-saving achieved by using dispatchers thus consists of two parts: for dispatchers at the bureaucracy itself and for firms by eliminating the need to visit multiple offices. These two time-saving components will be made explicit in the model that will be presented here.[4]

[4] From the supply side, one possible argument for the existence of dispatchers is that the gov-

3 The model

The model is static and has three actors. First, there are firms that demand a license that brings a production benefit to the firm. The license can be obtained by going through a bureaucratic procedure consisting of n steps. The second component of the model is thus the bureaucracy, consisting of n bureaucrats. The third component is the dispatcher sector, offering the completion of the bureaucratic procedure as a "one stop shop".

3.1 Firms

Firms, indexed by superscript i, differ in their production parameter A^i, which is uniformly distributed on the unit interval, $0 < A^i \leq 1$. There is a total measure of 1 of firms. Firms can get a production increase, from A^i to gA^i, where $g > 1$, if they go through the bureaucratic procedure. Firms maximize profits, which in this model is the same as maximizing production, by choosing if and through which means to acquire the license.

3.2 Bureaucracy

The "de jure" bureaucratic procedure is modeled as n equal steps that the firm must go through to obtain the license. Each bureaucratic step consists of one visit to the government bureaucracy in which the firm interacts with one bureaucrat who is a monopolist in this specific step and responsible for this step only. The bureaucrat completes this one step and then the firm has to undertake the next step of the procedure, facing a new bureaucrat.

Each step is associated with a financial cost p that is the actual cost faced by the bureaucrat. In addition, going through the bureaucratic procedure also means

ernment allows dispatchers to exist as a means of helping individuals and firms going through bureaucratic procedures. Bureaucracy intermediaries then become a "second best" option in societies where the government can do little to reform its bureaucracy. Another supply-side argument explaining the existence of dispatchers may be that dispatchers are easier to work with for bureaucrats because they "always have their papers in order". That is, the cost for handling applications from dispatchers is lower. Bureaucrats would then be able to serve more customers of the bureaucracy in less time, which would be socially beneficial.

that the firm spends time in licensing rather than in production. The time cost has two components, a waiting time k per bureaucratic step ("standing in line") and a transport time t per bureaucratic step. The per step time costs are modeled as a cost per unit of production, i.e. $A^i(k+t)$. The assumption of the time cost being proportional to firm productivity can be interpreted as firm management (rather than "office boys") being involved in the licensing procedure. Firms always have the option of getting the license through the de jure procedure.[5]

Corruption in the bureaucracy is now introduced. Let m $(\leq n)$ of the bureaucrats be corrupt and, consequently, $n-m$ honest. Corruption is "according to rules", meaning that bureaucrats always complete the actual step and face the cost p.[6]

In the direct interaction with firms, each corrupt bureaucrat completes his step but lets the firm avoid the time cost k, i.e. avoid "standing in line". Direct corruption is thus an alternative to get the license through which firms avoid standing in k lines. Each corrupt bureaucrat incurs the cost p, and charges a profit maximizing price b_c to the firm. The total financial cost of a license, from bureaucrats to firms, becomes $B_c \equiv mb_c + (n-m)p$ and the total time cost is $A^i(n-m)k + A^i nt$.

In the indirect interaction with firms, i.e. through dispatchers, each corrupt bureaucrat completes his step according to the rules, incurs the cost p, and charges a profit-maximizing price b_d to dispatchers. The total financial cost of a license, from bureaucrats to dispatchers, is $B_d \equiv mb_d + (n-m)p$.

By assumption, corrupt bureaucrats cannot price discriminate and honest bureaucrats always charge the price p, irrespective of whether the interaction is with a firm (direct) or with a dispatcher (indirect).[7]

[5] I abstract from sequential bargaining problems in the model. In studying roadblocks holding up truck transports in Banda Aceh, Indonesia, Olken (2007) discusses a model in which the bargaining problem at each roadblock changes as a truck travels along the road. In analyzing central planning vs. transition n-step production chains in Russia, Blanchard and Kremer (1997) also use a sequential bargaining model.

[6] Throughout the paper (except section 8.2), the cost p faced by a bureaucrat can be considered as being delivered by the bureaucrat to his superior, without any stealing by the bureaucrat. The actual step is always completed by the bureaucrat.

[7] The starting point is corrupt bureaucrats that set prices independently of each other ("decentralized bureaucrats"). I also allow for price setting in a coordinated ("centralized") fashion. In this latter case, bureaucrats internalize the horizontal externality that arises when prices are set independently. Centralized/decentralized is the terminology of Shleifer and Vishny (1993, 1998). Bureaucrats always incurring the cost p constitutes the case "without theft". The terminology used here is "according to rules"-corruption.

3.3 Dispatchers

Dispatchers are the third alternative through which to get the license and it is a "one stop shop" alternative.[8] Dispatchers incur the cost B_d for the license and charge firms a profit maximizing price d for their service. This is the only cost that firms face when going to a dispatcher.

There are x dispatchers. Through Cournot competition, a mark-up is added by the dispatcher sector. This specification allows us to study, in a convenient way, how the market structure of dispatchers affects license allocations. By assumption, dispatchers cannot price discriminate.[9]

Regarding notation and terminology in what follows, the de jure *bureaucracy* is subindexed b, the direct *corruption*/bureaucracy bribe case is subindexed c and the *dispatcher* case is subindexed d. Going through the procedure, using any of the three means, is referred to as "getting the license". Not getting the license is referred to as "remaining informal". If all firms were to get the license, the number of licenses awarded would be 1 and informality would be 0. Informality is thus one minus licenses awarded. Figure 1 summarizes the problem set-up.

[8] As discussed in section 2, dispatchers have a superior technology for handling the bureaucratic procedure.

[9] Three qualifications for the specification of the dispatcher activity are given here. First, the zero time cost of dispatchers at the bureaucracy is obviously a limit case but a small time cost ($< k$) at honest bureaucrats would not change the qualitative results of the analysis. Similarly, adding a small time cost of firms' interaction with dispatchers would not change the qualitative results and is therefore omitted. Third, although x, the number of dispatchers, is exogenous in the model, one can think of x as inversely related to entry costs into the sector. Such entry costs can consist of getting to know the bureaucratic procedure, how to handle it in an expeditious way, knowing the corrupt bureaucrats, etc.

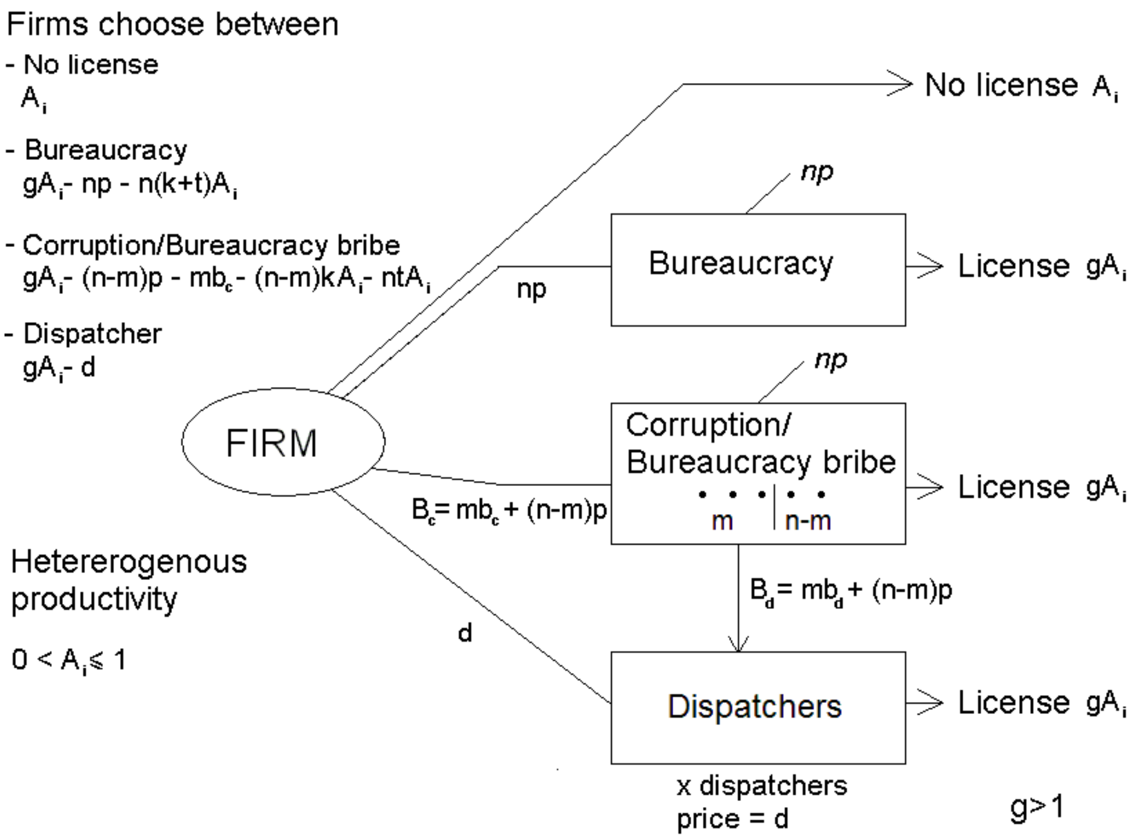

Figure 1. The problem set-up.

3.4 Timing, equilibrium concept and solution method

The timing is as follows. First, each bureaucrat sets two bribe levels, b_c and b_d, taking the decisions of the other $m-1$ corrupt bureaucrats as given. Due to symmetry, all bureaucrats set the same bribe levels. Firms and dispatchers, respectively, take the total bribe levels B_c and B_d as given. Second, given B_d, each of the x dispatchers simultaneously sets a dispatcher price d (which is equal for all dispatchers due to symmetry). Finally, firms take B_c and d as given and maximize the profits. That is, firms decide if and through which means to acquire the license. A subgame perfect equilibrium is derived by backward induction: direct and indirect firm demand for licenses is derived, then the profit maximization problem of dispatchers is solved and, finally, the corrupt bureaucrats solve their profit maximization problem.[10]

[10] If corruption is centralized, the m corrupt bureaucrats take one joint decision on total direct and indirect bribe levels.

4 Solving the model – without dispatchers

In this section, the model is solved for the case without dispatchers. It serves as a point of comparison for the full analysis in sections 5, 6 and 7 and most of the discussion is postponed until the full solution, including dispatchers, has been presented.

4.1 Firm demand

Firms choose between getting a license the de jure way, through bribing or to remain informal. This section derives the demand function for acquiring the license through bribing, which is the part of demand that is relevant for profit maximizing corrupt bureaucrats.

There are two productivity thresholds which are relevant in constructing this demand curve and which of these that applies depends on the optimal bribe set by corrupt bureaucrats. The relevant threshold for the choice between the de jure bureaucracy and corruption is A^i_{bc}, which is the solution to

$$gA^i - np - A^i\left(nk + nt\right) = gA^i - B_c - A^i\left(\left(n-m\right)k + nt\right) \Rightarrow A^i_{bc} = \frac{B_c - np}{mk}. \tag{1}$$

The LHS in the first equality in (1) is net production when getting the license the de jure way and the RHS is net production when getting the license through bribing to avoid the time cost k at the m corrupt bureaucrats. As the firm obtains the license in both cases, A^i_{bc} does not depend on the production gain, $g-1$.

The relevant threshold for the choice between corruption and informality is A^i_c, which is the solution to

$$gA^i - B_c - A^i\left(\left(n-m\right)k + nt\right) = A^i \Rightarrow \qquad A^i_c = \frac{B_c}{g - 1 - \left(n-m\right)k - nt}. \tag{2}$$

The LHS in (2) is net production when bribing, while the RHS is net production when remaining informal. It is instructive to also derive the threshold for the choice between the de jure bureaucracy and informality, A^i_b:

$$gA^i - np - A^i n\left(k+t\right) = A^i \Rightarrow \qquad A^i_b = \frac{np}{g - 1 - nk - nt}. \tag{3}$$

Intuitively, if B_c is low, such that $A^i_c < A^i_b$, all firms that get the license will get it by bribing and the relevant threshold for the demand function is A^i_c. When B_c is instead high, such that $A_c^i > A_b^i$, some firms that get the license find it profitable to go through the de jure bureaucracy instead. The threshold relevant for corrupt bureaucrats' demand function is instead A_{bc}^i.

At the point where A_c^i equals A_b^i, by construction A_{bc}^i takes the same value. By noting that for small B_c we have $A_{bc}^i < A_c^i < A_b^i$ and for large B_c we have $A_{bc}^i > A_c^i > A_b^i$, and by using the fact that the firm distribution is uniform, the demand function for corrupt bureaucrats can be written as follows:

$$Q_c(B_c) = 1 - \text{Max}\{A_c^i, A_{bc}^i\} = 1 - \text{Max}\left\{\frac{B_c}{g-1-(n-m)\,k-nt}, \frac{B_c - np}{mk}\right\}. \tag{4}$$

Finally, the cutoff bribe level for which we have $A_b^i = A_c^i = A_{bc}^i$, is given by

$$\bar{B} = np\frac{g-1-(n-m)\,k-nt}{g-1-nk-nt}. \tag{5}$$

4.2 Bureaucrats' profit maximization

With decentralized corruption, each corrupt bureaucrat sets his price b_c without taking into account the individual prices $\tilde{b}_c$ set by the other $m-1$ corrupt bureaucrats (Bardhan, 1997; Shleifer and Vishny, 1993).[11]

The profit function of the individual corrupt bureaucrat is bribe revenue minus cost, i.e. $b_c - p$, times demand. Using the demand function from (4), the profit maximization problem becomes:

Choose b_c to maximize

$$(b_c - p)\times\left(1 - \text{Max}\left\{\frac{B_c}{g-1-(n-m)\,k-nt}, \frac{B_c - np}{mk}\right\}\right). \tag{6}$$

Note that this formulation incorporates a constraint on each optimal candidate in relation to $\bar{B}$. That is, solving the problem with the assumption of $A_{bc}^i > A_c^i$ gives a candidate for optimum that must be larger than $\bar{B}$, and vice versa if we instead start by assuming that $A_{bc}^i < A_c^i$. If neither of the candidates fulfills its corresponding condition, the optimal solution is instead $\bar{B}$ (which, by construction, gives $A_c^i = A_{bc}^i = \text{Max}\{A_c^i, A_{bc}^i\}$).

The solution to (6) is obtained by taking the first-order conditions with respect to b_c, applying symmetry between bureaucrats ($\tilde{b}_c = b_c$) and solving for b_c, aggregating

[11] The total bureaucracy price B_c is then written as $B_c = b_c + (m-1)\,\tilde{b}_c + (n-m)\,p$.

to get B_c and then comparing with $\bar{B}$ to check when the solution applies. As the paper focuses on how the size of the bureaucracy affects license allocations, I equate the expression for each candidate for optimal B_c with the threshold bribe level $\bar{B}$ and solve for the values of n for which the solution applies. The solution is presented in figure 2 below and in the discussion that follows.[12] The left-hand panel shows the optimal bureaucracy price as a function of the size of the bureaucracy. The right panel shows, for each value of n, the allocation of licenses.[13]

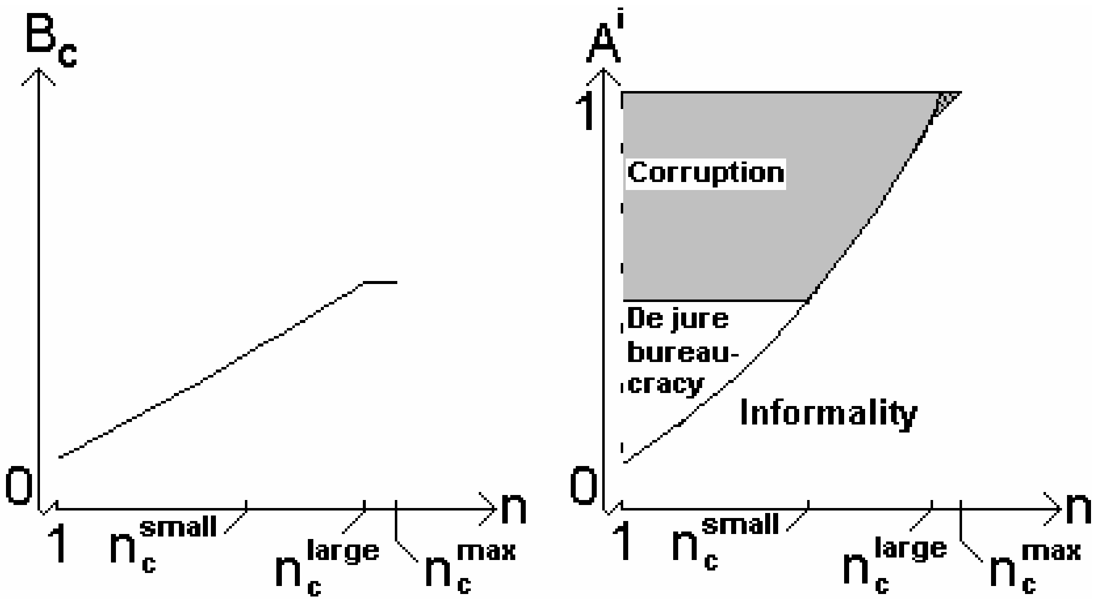

Figure 2. Optimal bribe level (left-hand panel) and the resulting license allocations (right-hand panel) in the model without dispatchers.

The solution displays a small-, an intermediate- and a large bureaucracy region. For small bureaucracies, going through the de jure procedure is not very costly. The profit maximizing total bribe level, $B_c^{\text{small}} = np + \dfrac{m}{1+m}(mk)$, will be such that a fraction $\dfrac{m}{1+m}$ of all firms get the license through bribing, but some firms also get the license through the de jure procedure ($m = 1$ was used in figure 2, thus half of

[12] See appendix 1 for a statement of the first-order conditions and for the complete solution to the problem.

[13] The vertical axis in the right-hand panel is firm productivity A^i. The graph thus shows the productivity spectrum of firms that get the license (above the curve) and remain informal (below the curve) for each level of n. Although n represents the number of steps, I treat it here as a continuous parameter.

Figures 2-4 have parameters as follows: $g = 2$, $p = 0.1$, $k = 0.04$, $t = 0.05$ and $m = 1$. In addition, $x = 3$ in figures 3-4.

the firms bribe). Because the marginal firm chooses between two ways of getting the license, the optimal bribe level contains a component reflecting the time saving from bribing (mk), but not the gain of the license itself. The small bureaucracy region, which holds over an interval $1 \leq n \leq n_c^{\text{small}}$, corresponds to the second term in brackets in the demand function.

The threshold between the de jure bureaucracy and direct bribes, $\dfrac{m}{1+m}$, is independent of np. This is due to the fact that np is both the cost of the license for bureaucrats and, for small bureaucracies, the opportunity cost for firms. As long as the marginal firm continues to choose between bribing and the de jure bureaucracy, increases in np can be transferred to firms without losing any demand (B_c^{small} is additive in np).

Above n_c^{small}, it is profit maximizing for corrupt bureaucrats to set prices such that all licenses are awarded through corruption. In an intermediate range, $n_c^{\text{small}} < n \leq n_c^{\text{large}}$, prices are set such that it is just as costly to bribe as it is to go through the de jure procedure: $B_c^{\text{intermediate}} = \bar{B}$. For even larger bureaucracy sizes, $n > n_c^{\text{large}}$, the marginal firm will choose between informality and getting the license through corruption, corresponding to the first term in brackets in the demand function. The optimal price in this region is $B_c^{\text{large}} = np + \dfrac{m}{1+m}\left(g - 1 - np - (n-m)\,k - nt\right)$.

Above n_c^{large}, the time cost mk is avoided by all firms that get the license. Corrupt bureaucrats can make positive profits up to the bureaucracy size $n_c^{\max} = \dfrac{g-1+mk}{p+k+t}$ at which the most productive firm's gain from the license, $g-1$, equals the costs of the license that cannot be avoided, $np + (n-m)\,k + nt$. The left-hand panel shows that the corrupt bureaucrats can no longer raise bribe levels as much (B_c flattens out), and, as $n \to n_c^{\max}$, the total bribe B_c^{large} approaches np.

If corruption had not existed, the maximum size of bureaucracy for which any licenses had been awarded would be smaller. The effect of an expansion in the largest possible bureaucracy size for which any licenses will be awarded, due to corruption, is shown by the striped area in the right-hand panel. We will return to this issue in more detail when discussing the case with dispatchers.

An increase in m, the amount of corrupt bureaucrats, will have two effects. First, more time can be saved by bribing and $n_c^{\max}$ will increase. Because corruption is decentralized however, the total mark-up over mk increases. This results in more firms instead acquiring the license through the de jure bureaucracy: the threshold between bureaucracy and corruption shifts up and n_c^{small} shifts to the right. Fewer firms will then benefit from the time saving offered by bribing.[14]

5 Solving the model - with dispatchers

When we add dispatchers to the framework in the previous section, firms choose between getting a license the de jure way, through bribing, through dispatchers or to remain informal. Naturally, the demand function relevant for dispatchers is only the licenses awarded through dispatchers. For corrupt bureaucrats, there is bribe revenue both from direct corruption and from licenses awarded via dispatchers.

5.1 Firm demand for the dispatcher service

There are three new productivity thresholds to consider, derived in the same fashion and with the same intuition as above. For the choice between direct corruption and dispatcher, we get

$$gA^i - B_c - A^i\left((n-m)k + nt\right) = gA^i - d \Rightarrow \qquad A^i_{cd} = \frac{d - B_c}{(n-m)k + nt}, \tag{7}$$

for the choice between de jure bureaucracy and dispatcher,

$$gA^i - np - A^i\left(nk + nt\right) = gA^i - d \Rightarrow \qquad A^i_{bd} = \frac{d - np}{nk + nt}, \tag{8}$$

and for the choice between informality and dispatcher,

$$gA^i - d = A^i \Rightarrow \qquad A^i_d = \frac{d}{g-1}. \tag{9}$$

For later reference, and similar to expression (5) above, note that $A^i_b = A^i_d = A^i_{bd}$ when d equals $\bar{d}$, given by

[14] This effect is due to the decentralization of corruption. If corruption had instead been centralized, the solution is obtained by replacing "m" in the solution by "1", except in the terms (mk) or $(n-m)\,k$. As an example, we would get $B_c^{\text{small}} = np + \frac{1}{2}(mk)$. The fraction of firms that gets the license through bribing, for small bureaucracies, would then be $\frac{1}{2}$ instead of $\frac{1}{1+m}$.

$$\bar{d} \equiv np\frac{g-1}{g-1-nk-nt}. \tag{10}$$

At this dispatcher price, the firm which is indifferent between going through bureaucracy the de jure way and remaining informal is also indifferent between going to the dispatcher and remaining informal. Consequently, the same firm is also indifferent between going through the procedure the de jure way and using dispatchers. In addition, if $B_c = \bar{B}$, it is also equally costly to bribe: A^i_{cd} equals $A^i_b = A^i_d = A^i_{bd}$.

In constructing a demand curve, note that the highest productivity firms will always use dispatchers. The larger time-saving ($nk + nt$ instead of mk) makes firms willing to pay a higher price, from which bureaucrats and dispatchers can earn additional profits.

Given that the highest productivity firms always use dispatchers, the demand curve for the dispatcher service can be derived in a fashion similar to the derivation of (4). Depending on the dispatcher price d, the relevant demand curve is either a choice between bribing and dispatchers, between the de jure bureaucracy and dispatchers or between informality and dispatchers. From the three thresholds (7)-(9) above, we get the dispatcher demand function:

$$Q_d(B_c,d) =$$
$$1-\mathrm{Max}\{A^i_{cd},A^i_{bd},A^i_d\} = 1-\mathrm{Max}\left\{\frac{d-B_c}{(n-m)\,k+nt},\frac{d-np}{nk+nt},\frac{d}{g-1}\right\}. \tag{11}$$

5.2 Dispatcher profit maximization

The dispatcher sector is modeled as characterized by Cournot competition. This makes it possible to study the effects of a mark-up in the dispatcher sector in a simple way. Each of the x identical dispatchers takes the cost of permits that they face at the bureaucracy, B_d, as given. They maximize profits by choosing their profit maximizing quantity, q_d, taking the total quantity choice of the other $x-1$ dispatchers, $(x-1)\,\tilde{q}_d$, as given. By writing Q_d from (11) as $Q_d = q_d + (x-1)\,\tilde{q}_d$, the individual dispatcher's inverse demand function $d\,(q_d)$ can be derived[15] .

[15] $d\,(q_d) = \mathrm{Min}\left\{\begin{array}{c} B_c + ((n-m)\,k+nt)\,(1-(q_d+(x-1)\,\tilde{q}_d)) \\ np + (nk+nt)\,(1-(q_d+(x-1)\,\tilde{q}_d)) \\ (g-1)\,(1-(q_d+(x-1)\,\tilde{q}_d)) \end{array}\right\}.$

Maximizing profits, each dispatcher solves:
Choose q_d to maximize $q_d\left(d\left(q_d\right)-B_d\right)$

After taking the first-order condition with respect to q_d, applying symmetry between all dispatchers ($\tilde{q}_d = q_d$) and then solving for q_d, plugging the optimal quantities back into $d\left(q_d\right)$, we get the following optimal response functions, corresponding to the three terms in the demand function:[16]

$$
\begin{aligned}
d_1\left(B_d\right) &= \frac{x}{1+x}B_d+\frac{1}{1+x}\left(B_c+(n-m)\,k+nt\right) \\
d_2\left(B_d\right) &= \frac{x}{1+x}B_d+\frac{1}{1+x}\left(np+nk+nt\right) \\
d_3\left(B_d\right) &= \frac{x}{1+x}B_d+\frac{1}{1+x}\left(g-1\right).
\end{aligned}
\tag{12}
$$

In addition, we get
$d_4\left(B_d\right)=\bar{d}$ when neither of the above applies.

These response functions capture the standard feature of Cournot competition, that is, a mark-up over cost (B_d) that gradually declines when the number of dispatchers, x, grows.

5.3 Firm demand for direct corruption and dispatchers

Plugging back the optimal responses into expression (11), we get the indirect demand function relevant for corrupt bureaucrats:

$$
Q_d(B_c,B_d)=1-\mathrm{Max}\left\{\frac{d_1\left(B_d\right)-B_c}{(n-m)\,k+nt},\frac{d_2\left(B_d\right)-np}{nk+nt},\frac{d_3\left(B_d\right)}{g-1}\right\}. \tag{13}
$$

The first term in the bracket corresponds to the coexistence of both dispatcher demand and direct corruption. In this case, total demand therefore also contains a direct demand component, the derivation of which is similar to section 4. We get it by replacing the upper bound, 1, in (4), with $\frac{d_1\left(B_d\right)-B_c}{(n-m)\,k+nt}$, i.e.

$$
Q_c(B_c,B_d)=
(b_c-p)\left(\frac{d_1\left(B_d\right)-B_c}{(n-m)\,k+nt}-\mathrm{Max}\left\{\frac{B_c}{g-1-(n-m)\,k-nt},\frac{B_c-np}{mk}\right\}\right). \tag{14}
$$

[16] See appendix 2 for a statement of the first-order conditions and for more details on solving the problem.

5.4 Bureaucracy profit maximization

Using an indicator function I which is 1 if and only if the first term in the bracket in (13) is largest[17] , the profit maximization problem of corrupt bureaucrats can now be stated as follows:

Bureaucrats choose b_d and b_c to maximize

$$(b_d - p)\left(1 - \text{Max}\left\{\frac{d_1(B_d) - B_c}{(n-m)k + nt}, \frac{d_2(B_d) - np}{nk + nt}, \frac{d_3(B_d)}{g-1}\right\}\right) +$$
$$(b_c - p)\left(\frac{d_1(B_d) - B_c}{(n-m)k + nt} - \text{Max}\left\{\frac{B_c}{g - 1 - (n-m)k - nt}, \frac{B_c - np}{mk}\right\}\right) I. \tag{15}$$

As before, the problem formulation incorporates a constraint on the candidate for optimal B_c in relation to $\bar{B}$. In addition, there is a similar constraint on the candidate for optimal d, relative to $\bar{d}$ from (10).[18]

6 Optimal license allocations

The solution to the bureaucrats' problem in (15) and the resulting optimal dispatcher prices, bureaucracy bribes and license allocations are presented in figure 3 (bottom panels).[19] The solution displays a small-, an intermediate- and a large bureaucracy range. To allow a comparison, figure 3 also presents the previously discussed solution in the no-dispatcher case (middle panels) as well as license allocations when there is neither corruption nor dispatchers (upper panels).[20]

---

[17] $I = 1$ if $\text{Max}\left\{\frac{d_1(B_d) - B_c}{(n-m)k + nt}, \frac{d_2(B_d) - np}{nk + nt}, \frac{d_3(B_d)}{g-1}\right\} = \frac{d_1(B_d) - B_c}{(n-m)k + nt}$, $I = 0$ otherwise.

[18] The direct demand component (second row) incorporates a constraint on the candidates for optimal B_c. This is similar to the formulation in (6) and we will thus either get unconstrained solutions for B_c, or $B_c = \bar{B}$. There is also a constraint on d from the indirect demand component (first row). If there is direct demand and the constraint $B_c = \bar{B}$ binds, it means that the constraint on d simplifies to comparing d with $\bar{d}$ for the solution candidate from the $\frac{d_1(B_d) - B_c}{(n-m)k + nt}$-term: the candidate for d must be larger than $\bar{d}$ for this term to apply. If there is direct demand and $B_c \neq \bar{B}$, there will always be an unconstrained d. If there is no direct demand ($I = 0$), each of the two candidates for optimal d should be compared with $\bar{d}$.

[19] The problem is solved in appendix 3.

[20] In the choice between only de jure bureaucracy and informality, the threshold is $A_b^i = \frac{np}{g - 1 - nk - nt}$ from (3).

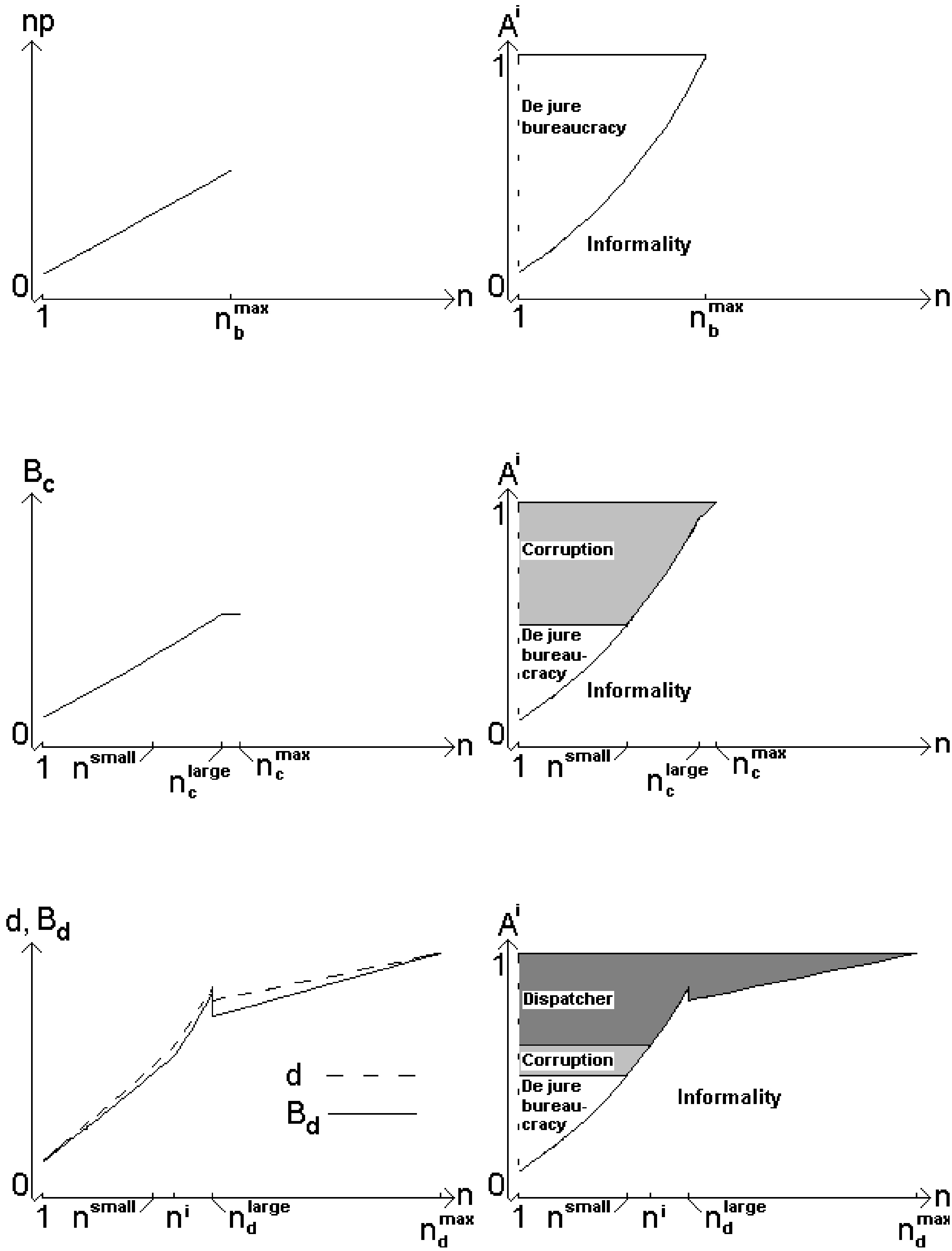

Figure 3. Optimal prices (left-hand panels) and allocations (right-hand panels) with dispatchers (bottom), direct corruption only (middle) and de jure bureaucracy only (top).

The first thing to note about the solution is that neither the direct bribe level $B_c^{\text{small}} = np + \frac{m}{1+m}(mk)$, which applies in the region $1 \leq n \leq n^{\text{small}}$, nor n^{small} itself, change when dispatchers are introduced.[21] As was the case with np in section 4.2, the direct bribe level B_c now acts as an opportunity cost, not only for firms but also for corrupt bureaucrats in their choice of indirect bribe B_d. The resulting indirect bribe level will therefore be additive in direct bribes: $B_d^{\text{small}} = B_c^{\text{small}} + \frac{m}{1+m}((n-m)k + nt)$, and indirect profits above what could have been obtained by direct corruption are independent of B_c^{small}. As a result, the introduction of dispatchers does not change the optimal values of B_c^{small}, n^{small} and the de jure/corruption threshold, $\frac{m}{1+m}$.

The threshold between direct corruption and dispatchers is instead $\frac{1+m+mx}{(1+m)(1+x)}$ over the small bureaucracy region. The larger is x, the lower will be the mark-up from the dispatcher sector and the more firms will use dispatchers.[22] The dispatcher/corruption threshold in figure 3 (lower right-hand panel) shifts down. In the limit as $x \to \infty$, it converges to $\frac{m}{1+m}$, and all corruption moves to be indirect.[23]

As was the case in the no-dispatcher case, above n^{small} no firms will use the de jure bureaucracy and the optimal direct bribe is $\bar{B}$. Over a first intermediate range $n^{\text{small}} < n \leq n^{\text{i}}$, however, the dispatcher-corruption threshold remains unchanged from the small bureaucracy region. The direct bribe level, now $\bar{B}$, still acts as an opportunity cost, both for firms and for corrupt bureaucrats in their choice of indirect bribe. The optimal indirect bribe will therefore still be additive in the direct bribe, equalling $\bar{B} + \frac{m}{1+m}((n-m)k + nt)$, without any loss of demand. Note that n^{i} converges to n^{small} as $x \to \infty$.

Above n^{i} all licenses are awarded through dispatchers. There is a second intermediate range, $n^{\text{i}} < n \leq n_d^{\text{large}}$ over which bureaucrats optimally set B_d such that dispatchers' best response is $d = \bar{d}$. As a result, over this region, it is just as

[21] I therefore drop the subindex on n^{small}.

[22] The dispatcher price over the small bureaucracy range is $d^{\text{small}} = B_d^{\text{small}} + \frac{(n-m)k + nt}{(1+m)(1+x)}$.

[23] If the vertical externality introduced by dispatchers had not been present, all corruption would always have been indirect in this model. Firms can save more time and bureaucrats can make higher profits.

costly for the marginal firm to go through the de jure bureaucracy as it is to go to a dispatcher.[24]

Because time costs are completely eliminated by going to a dispatcher, there is a large bureaucracy region $n_d^{\text{large}} < n \leq n_d^{\max} = \dfrac{g-1}{p}$ for which corrupt bureaucrats and dispatchers can make positive profits. At $n_d^{\max}$, the most productive firm's gain from the license, $g-1$, equals the costs of the license that cannot be avoided, np. The right-hand panels show that the largest bureaucracy size for which any licenses at all are awarded expands from $n_b^{\max} \equiv \dfrac{g-1}{p+k+t}$ in the upper right-hand panel, through $n_c^{\max} = \dfrac{g-1+mk}{p+k+t}$ in the direct corruption only case (middle panels) to $n_d^{\max} = \dfrac{g-1}{p}$ in the dispatcher case. In the direct corruption only case (middle panels), the number of corrupt bureaucrats, m, and the time cost of standing in line, k, determine how much larger bureaucracies can be and still award any licenses. With dispatchers, the maximum size of bureaucracy that will still award any licenses can be even larger, as time costs become irrelevant.

6.1 Informality

What is the impact of dispatchers on informality? Comparing the upper right-hand and the bottom right-hand panels in figure 3 shows that the introduction of dispatchers and direct/indirect corruption reduces informality when n is above n_d^{large}. Bureaucracies that would have been prohibitively large now award some licenses. The case with direct corruption only provides an intermediary case between the de jure bureaucracy and the dispatcher case.[25]

Below n_d^{large}, the threshold between licenses awarded and informality is determined by $A_b^i = \dfrac{np}{g-1-nk-nt}$, which is the de jure/informality threshold from

[24] Over the region $n^{\text{i}} < n \leq n_d^{\text{large}}$, the interests of bureaucrats and dispatchers are unaligned. Moving first, bureaucrats can therefore gain corruption profits at the expense of dispatchers. In figure 3, the shrinking difference between dispatcher price and the indirect bribe over this bureaucracy range is seen in the lower left-hand panel. The upper limit, n_d^{large}, is determined by bureaucrats comparing profits in this case to the unconstrained large bureaucracy case. This will generate a discontinuity in prices and also in the allocations at n_d^{large}, seen in the lower right-hand panel.

[25] In theory, the possibility exists that informality is higher with dispatchers than with direct corruption only, for some bureaucracy sizes at the lower end of the large bureaucracy region. This can be the case if there is a high-mark-up in the dispatcher sector (x is low).

(3). For these bureaucracy sizes, it is profit maximizing for bureaucrats and dispatchers to capture no more than the firms that would have acquired the licenses from the de jure bureaucracy anyway.[26]

In order to relate to actual bureaucratic procedures, consider the case of starting a firm in Brazil. The World Bank reports that it takes 17 different steps, 152 days and 8% of GNI per capita to start a firm. In line with such a time-consuming and costly de jure procedure is the fact that as much as 90% of Brazilian 1-5 person firms are informal (World Bank, 2009a; SEBRAE, 2005).

Still, however, most firms that have become formal in Brazil seem to report having done so without much problems. Stone et al. (1996), Zylbersztajn and Graça (2003) and Zylbersztajn et al. (2007) study small to medium sized firms in the garment industry. Firms have, by and large, paid one fee to a "despachante" (or an accountant) and have had all papers in order after approximately 50 days.

The use of "despachantes", as portrayed in these Brazilian studies, combined with the high degree of informal firms observed in Brazil, is in line with the large bureaucracy case of the model. All firms that get the license use dispatchers instead of going through the prohibitively costly "de jure" procedure.[27]

6.2 Time and resources spent in acquiring licenses

What is the impact of dispatchers on time/resources that firms spend in obtaining licenses? When introducing dispatchers, firms are given one more option to acquire the license, with the other means, i.e. going through the procedure the de jure way or bribing directly, unaltered. The introduction of dispatchers, while keeping the other parameters of the model as in the no-dispatcher case, can never make firms worse off.

Because of dispatchers' superior technology to acquire the license, combined with dispatchers being one stop shops, firms can make production gains when using

[26] The same threshold applies below n_c^{large} in the direct corruption only case in the middle right-hand panel.

[27] In studying the start-up procedure for firms in Bulgaria, Gancheva (1999) reports that the ratio of firms that have used an intermediary at start-up to those that have not, is positively correlated with the length of observed time of going through the de jure procedure. This observation is consistent with the present model, where the ratio of dispatcher usage increases as n grows.

dispatchers. More time is dedicated to production, rather than licensing.

Bureaucrats, on their part, will always find it in their interest to channel some firms through dispatchers. It will always be profit maximizing for corrupt bureaucrats and dispatchers to capture the firms that are most willing to avoid time costs. Therefore, it is the highest productivity firms that will gain from the introduction of dispatchers.[28]

An increase in x, the number of dispatchers, makes each firm that uses dispatchers better off and also increases the use of dispatchers. We saw this effect, for small bureaucracies, in a downward shift in the dispatcher/corruption threshold. For large bureaucracies, the threshold between dispatchers and informality, $A_d^{i,\text{large}} = \dfrac{1+m+mx}{(1+m)(1+x)} + \dfrac{npx}{(g-1)(1+m)(1+x)}$, also shifts down.

A change in m, the number of corrupt bureaucrats, has two effects. The first is the additional time saving that can be obtained from bribing one more bureaucrat to not stand in line. The second effect comes from the lack of coordination in bribe setting, which makes the horizontal externality from each bureaucrat's price setting on other bureaucrats' demand not being internalized. This increases the total prices, as seen through the $\dfrac{m}{1+m}$-term in the expressions for B_c^{small} and B_d^{small}. The sum of the two effects makes fewer firms bribe directly and fewer firms also use dispatchers.[29]

If bribes and fees to dispatchers are considered as mere transfers without any welfare impact, the value of time saving is the appropriate welfare measure in the model.

[28] The improvement for firms can be seen as an "intensive" and an "extensive" margin effect. Compare the dispatcher case (lower right-hand panel) with the de jure bureaucracy only (upper right-hand panel). We see that up to n_d^{large}, it is firms that would have got the license anyway that get it at a lower total cost. Between n_d^{large} and $n_b^{\max}$, there are also some new firms that would have been informal had there been the de jure bureaucracy only. Above $n_b^{\max}$, the whole effect comes from the extensive margin.

[29] With centralization of bureaucrats, we have $B_c^{\text{small}} = np + \frac{1}{2}(mk)$, $d^{\text{small}} = np + \frac{1}{2}(nk+nt) + \dfrac{(n-m)k+nt}{2(1+x)}$. The latter decreases in m and the former increases less than the additional time saving gained from direct bribing. There is no change in demand (the thresholds are $\frac{1}{2}$ and $\frac{2+x}{2+2x}$, respectively), and firms are thus better off.

The effect of dispatcher prices decreasing in m, whereas bureaucrats' indirect bribes are $B_d^{\text{large}} = np + \frac{1}{2}(g-1-np)$, can be seen as dispatchers having less of an advantage in comparison to direct corruption.

For bureaucracy sizes below n^{small}, the welfare improvement can be calculated as follows: It is the amount of firms that go to dispatchers, multiplied by $nk+nt$, plus the amount of firms that bribe directly, multiplied by mk. All these firms would have got the license through the de jure bureaucracy anyway (which is the comparison here). The welfare improvement is $\int_{\frac{1+m+mx}{(1+m)(1+x)}}^{1} \left((nk+nt)\, A^i\right) dA_i + \int_{\frac{m}{1+m}}^{\frac{1+m+mx}{(1+m)(1+x)}} \left(mkA^i\right) dA_i$. In this range, the welfare improvement is larger the larger is n, increasing in the amount of dispatchers x and decreasing in the amount of corrupt bureaucrats m.[30]

Between n_d^{large} and $n_b^{\max}$, both time saving and increased production, through a reduction in informality, contribute to the welfare improvement. Above $n_b^{\max}$, the welfare improvement can be calculated as $\int_{A_d^{i,\text{large}}}^{1} \left((g-1)\, A^i\right) dA_i$.

As a final remark, consider the case where $m = 0$, i.e. when there is no corruption. The direct corruption case then degenerates to the de jure bureaucracy case. The cost for dispatchers is np. The resulting allocations then depend on the mark-up in the dispatcher sector. Consider the limit case of perfect competition between dispatchers, $x \to \infty$. In this case, the cost of a license to firms is np and the threshold between informality and dispatcher will be $\dfrac{np}{g-1}$. More firms would acquire the license, as compared to the above, irrespective of the value of n.

7 Incentives to create red tape

One of the questions set out in this paper was to study the incentives of corrupt bureaucrats and dispatchers to create red tape. Can the introduction of middlemen teach us something about difficulties in reforming bureaucracies? These questions will be studied using the bureaucracy and dispatcher profit functions, starting with the individual bureaucrat and individual dispatcher profit expressions for small bureaucracies, $1 \leq n \leq n^{\text{small}}$:

[30] If corruption were centralized, demand would have been unaffected by increases in m and, as a result, the welfare gain would have been increasing in m.

Direct corruption only

$$\pi^{\text{small}}_{\text{Bureaucrat}} = \frac{m}{(1+m)^2}k$$

With dispatchers

$$\pi^{\text{small}}_{\text{Bureaucrat}} = \frac{m}{(1+m)^2}k + \frac{((n-m)k + nt)x}{(1+m)^2(1+x)}$$

$$\pi^{\text{small}}_{\text{Dispatcher}} = \frac{(n-m)k + nt}{(1+m)^2(1+x)^2}. \tag{16}$$

The first thing to note about these expressions is that with direct corruption only, there is no influence from the number of steps n on the profits of each bureaucrat. The direct corruption profit of a corrupt bureaucrat, $\frac{m}{(1+m)^2}k$, only depends on the corrupt bureaucrat saving firms the time cost of standing in one line. It is independent of n.

The additional profit for a corrupt bureaucrat, from having dispatchers, is the second term in $\pi^{\text{small}}_{\text{Bureaucrat}}$ in (16). Adding dispatchers means that firms can save more and more time as n increases, which will increase each bureaucrat's (and each dispatcher's) profits, since they capture part of the surplus.

Given these profit functions and given a small bureaucracy, it follows that each corrupt bureaucrat has an incentive to try to "add steps" to the procedure, i.e. to increase n. The increased profits are due to firms' increased willingness to pay for the dispatcher service when the time cost (nt) increases. Say that the additional step is an additional document required at a new *honest* counter. The corrupt bureaucrats, having nothing to do with this step per se, will be able to capture part of the firms' increased willingness to avoid the time cost of going through that step. This incentive to increase the number of steps is not present when there are no dispatchers.

Moreover, note that dispatchers have the same incentive to increase the number of steps. Dispatcher profits increase linearly all the way up to n^{i}. This follows from the reasoning above that dispatchers lose no demand up to this bureaucracy size (see the lower right-hand panel of figure 3).[31]

[31] Profits in the dispatcher sector are increasing with bureaucracy size for small bureaucracies. This result does not depend on the presence of corruption. Red tape incentives are thus present also when there is no corruption, although it may be harder to envision how red tape may come about when bureaucrats and dispatchers do not have a common interest.

Without dispatchers, adding steps to increase profits is different. The steps to be added must be corrupt steps (the only way of increasing firms' willingness to pay) and the m corrupt bureaucrats need to add these steps without increasing the number of corrupt bureaucrats. Otherwise, there is one more bureaucrat to share the corruption profits. More specifically, if the m corrupt bureaucrats increase the number of corrupt steps to m', there is an increase in the per bureaucrat corruption profits from $\frac{1}{m}\frac{m^2k}{(1+m)^2}$ to $\frac{1}{m}\frac{(m')^2k}{(1+m')^2}$. If the number of corrupt bureaucrats also increases, the latter expression becomes $\frac{m'k}{(1+m')^2}$, which is decreasing in m'.[32,33]

A second red tape incentive is at work as soon as there is a mark-up in the dispatcher sector: bureaucracy profits peak at a larger bureaucracy size in the dispatcher case. This result means that corrupt bureaucrats' incentive to create red tape is in place for larger sizes of bureaucracy than if there had been no dispatchers.[34]

For large bureaucracies, all firms that get a license get it through dispatchers. This means that bureaucrats and dispatchers can no longer raise prices to fully compensate themselves as np increases. Large bureaucracy profits, as stated below, will therefore decrease in n. This, in turn, implies that for large bureaucracies, there is an incentive for bureaucrats and dispatchers to reduce red tape.

$$\begin{aligned}\pi^{\text{large}}_{\text{Bureaucrat}} &= \frac{(g-1-np)^2x}{(g-1)(1+m)^2(1+x)}\\ \pi^{\text{large}}_{\text{Dispatchers}} &= \frac{(g-1-np)^2x}{(g-1)(1+m)^2(1+x)^2}\end{aligned} \tag{17}$$

[32] Assume that there is one corrupt bureaucrat controlling all steps. Certainly, there is then a red tape incentive even without dispatchers (m is fixed, choose m'). The red tape incentive is related to increasing the time cost, through the number of steps, in order to make the high productivity firms pay more, without affecting demand too much. This is different from the red tape incentive in Lui (1985). In his model, revenue is maximized by bureaucrats "optimally" working just as fast/slow so that all agents want to go through the procedure, combined with what is effectively price discrimination. Here, there is no price discrimination and the number of steps is the vehicle for corruption revenue.

[33] This argument holds in a weak sense if corruption is centralized: per bureaucrat corruption profits are constant when there is an increase in the number of corrupt steps and the number of corrupt bureaucrats.

[34] From figure 3, we know that demand for indirect bribing/dispatchers is constant all the way up to n^{i}, whereas demand for direct bribing is constant only up to n^{small} in both cases. As a consequence, bureaucracy profits will grow faster/decline less, beyond n^{small}, in the dispatcher case. I have looped through a large number of parametrizations of the problem, and consistently find that, with dispatchers, profits peak at a larger bureaucracy size.

Figure 4 displays bureaucracy and total dispatcher profits for the same parameter values as those used in figures 2-3. Bureaucracy profits peak between n^{i} and n_d^{large}. There is one bureaucrat and three dispatchers in this case ($m = 1$, $x = 3$), which explains that total dispatcher profits are much lower than the bureaucrat's profits.

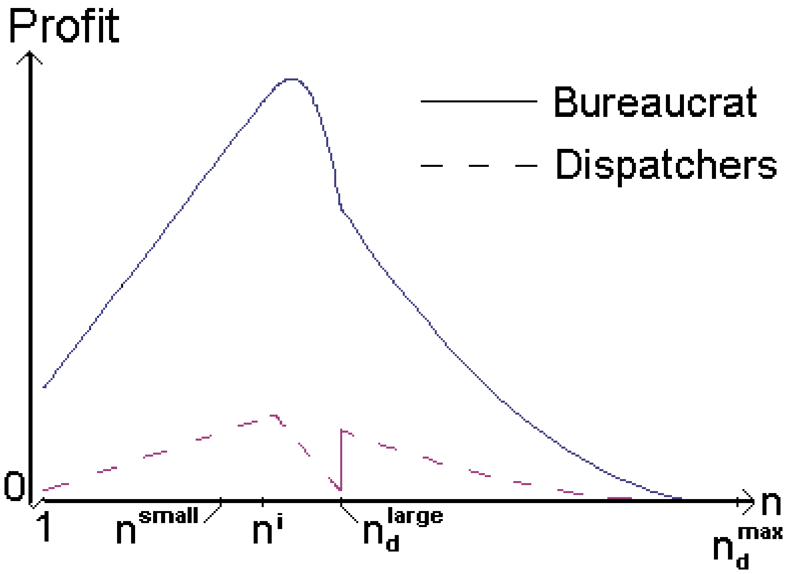

Figure 4. Bureaucracy and dispatcher profits.

Whereas the incentives of bureaucrats and dispatchers are aligned with respect to increasing n for small bureaucracies, they are obviously unaligned with respect to x, the number of dispatchers.[35] With respect to m, decentralization of corruption makes bureaucrats charge higher prices which reduces demand. In addition, an increase in m diminishes the relative advantage of going to a dispatcher and makes dispatchers worse off (over the interval where direct corruption is a relevant choice for firms).[36]

[35] More competition in the dispatcher sector expands demand, due to a lower dispatcher mark-up. This increases bureaucracy profits. As is standard in Cournot competition, both individual and total dispatcher profits decrease as competition increases.

[36] Centralized total bureaucracy profits increase in m. However, such increases in m only benefit corrupt bureaucrats under the special circumstances outlined above.

8 Extensions

8.1 Some additional IO aspects

This section lists a few IO aspects that could potentially affect outcomes in the model presented so far.

The entry of dispatchers has not been modeled. Instead, I have simply worked with an exogenous parameter x representing the number of dispatchers. This parameter could be inversely related to entry costs. Two points are made here. First, total possible profits are lower if corruption is decentralized[37] , which may then result in a lower x for given entry costs and, as a result, a higher mark-up in the dispatcher sector, thus further reducing demand. Second, if bureaucrats control entry into dispatching, they may choose to make entry either easy or difficult. The advantage of the former is that bureaucracy profits may increase due to a lower mark-up, a disadvantage is if the bureaucracy-dispatcher relation needs to be kept secret to avoid government control.

If corrupt bureaucrats and dispatchers merge, something which is more likely to happen if corruption is centralized, the vertical "double monopolization" externality is eliminated. This increases demand and total profits (over which bureaucrats and dispatchers can bargain). Another option for bureaucrats is to work with one dispatcher only and make this dispatcher the residual claimant of the profits (Wade, 1982). That is, the m corrupt bureaucrats can jointly "sell the office" to one single dispatcher, then charge the marginal cost p for each step, and then let the dispatcher maximize profits. Also in this case does centralization of bureaucrats seem a natural prerequisite.

8.2 Dispatchers and bending the rules corruption

In this section, the model is modified to account for a different type of corruption than in the paper so far. With "bending the rules"-corruption, firms bribe to avoid legislation. The question is what impact dispatchers have in such a setting. That is,

[37] For instance, $\pi^{\text{small}}_{\text{Dispatcher}} = \dfrac{(n-m)\,k + nt}{(1+m)^2\,(1+x)^2}$ would become $\dfrac{(n-m)\,k + nt}{4\,(1+x)^2}$ in the centralized case.

what is the impact of introducing dispatchers on the amount of "bending the rules"? What can we learn from adapting the model to these different circumstances?

Bertrand et al. (2007) show that the way to obtain a driver's license in Delhi, India, without actually learning how to drive, is through an intermediary. The type of interaction considered here will be similar in that we consider firms that are undeserving of some government good and use a corrupt bureaucracy to get it. This could be an environmental license for an industry, a health permit for a restaurant etc. The universe of firms will now be undeserving firms. This is less restrictive than it may seem: because undeserving firms have no other means of getting the license than bribing, the optimization problem of corrupt bureaucrats and dispatchers is therefore separate from dealing with deserving firms. If the presence of dispatchers also causes deserving firms to change their corrupt behavior, these effects come in addition to those from undeserving firms.

There is a government in the background monitoring the bureaucracy. This makes sense as the type of corruption now imagined is more severe than before (rule breaking vs. "speed money"). Dispatchers, as described in the introduction, reduce the probability for bureaucrats of getting caught. In addition, dispatchers know better than firms how the bureaucracy works and how to get the licenses.

The firm productivity distribution is as before, $0 < A^i \leq 1$, the gain is still g and the procedure consists of n steps.[38] The option of going to the bureaucracy the de jure way loses its meaning as this means following the rules. In addition, the analysis is restricted to the case when all bureaucrats are corrupt, i.e. $m = n$, otherwise the undeserving firms could never get the license. In addition, bureaucrats are now centralized. I continue to assume that there is no price discrimination. Let there be y dispatchers.

Because firms bribe to avoid regulation, bureaucrats no longer face the cost p of fulfilling the regulation.[39] Instead, however, each bureaucrat faces an expected penalty p_1 for breaking the rules, when bribed directly by a firm. The expected penalty when firms instead bribe through dispatchers is p_2 where, based on the

[38] If getting the license, say the health permit, the firm increases its production. For instance, it can market itself, post the health permit at the establishment etc. Informality, i.e. not having the permit, thus retains its meaning even when we consider that all firms are willing to break the rules.

[39] This is similar to the case "with theft" in Shleifer and Vishny (1993).

reasoning in the introduction, it is assumed that $p_2 < p_1$.

There are two time costs. One is the time cost of going through each step at the bureaucracy, "transport".[40] The second time cost will be called an investigation time cost. It is assumed that if bribing the bureaucracy directly, firms have to spend time investigating whether it will be possible to get the license and assure that they will not get caught if bribing. This time cost will be assumed to be proportional to the number of steps, n, and firm productivity, A^i. When using dispatchers, no investigation on behalf of the firms is necessary. Because dispatchers eliminate both time costs, we can work with the sum of both: define u as the sum of transport and investigation costs per step.[41]

Below, details in the derivations are skipped as they are similar to the above. I use a hat on all optimal prices and thresholds to distinguish them from the above.

8.2.1 Without dispatchers

Solving this problem is straightforward. The firm compares getting the license with informality and gets the license if its productivity is greater than $\frac{B_c}{g-1-nu}$. Bureaucrats thus choose B_c to maximize $(B_c - np_1)\left(1 - \frac{B_c}{g-1-nu}\right)$, the solution to which is

$$\hat{B}_c = np_1 + \frac{1}{2}\left(g - 1 - np_1 - nu\right). \tag{18}$$

The maximum size of bureaucracy for which any licenses will be awarded is $\hat{n}_c^{\max} = \frac{g-1}{p_1+u}$.

8.2.2 With dispatchers

Firms choose between informality, bribing and dispatchers. In the choice between dispatchers and informality, a firm gets the license if its productivity is greater than $\frac{d}{g-1}$. The firm uses dispatchers rather than bribes if its productivity is greater

[40] There is thus no longer a division between queuing and transport, alternatively one can think of firms not standing in lines for this type of interaction

[41] The "investigation cost" has a function similar to penalties on firms, or to assuming that when bribing directly, firms cannot be certain of getting a reduction in regulation.

than $\dfrac{d - B_c}{nu}$. With demand curves and dispatcher responses derived as before[42], we get the following solution:

Small bureaucracy region, $1 \leq n \leq \hat{n}^{\text{small}}$.

$\hat{B}_c^{\text{small}} = np_1 + \dfrac{1}{2}(g - 1 - np_1 - nu) \qquad \hat{B}_d^{\text{small}} = np_2 + \dfrac{1}{2}(g - 1 - np_2)$

$\hat{d}^{\text{small}} = \hat{B}_d^{\text{small}} + \dfrac{n(p_1 - p_2 + u)}{2(1+y)}$

$\hat{n}^{\text{small}} = \dfrac{(g-1)(u + (p_2 - p_1)y)}{u(p_1 + p_2 y + u)}$ solves $\dfrac{\hat{d}^{\text{small}} - \hat{B}_c^{\text{small}}}{nu} = \dfrac{\hat{B}_c^{\text{small}}}{g - 1 - nu}$

i - Intermediate bureaucracy region, $\hat{n}^{\text{small}} < n \leq \hat{n}^{\text{large}}$

$\hat{B}_c^{\text{i}} = \dfrac{(g - 1 - nu)(2nu + (g - 1 + np_2)y)}{2(nu + (g-1)y)} \qquad \hat{B}_d^{\text{i}} = np_2 + \dfrac{1}{2}(g - 1 - np_2)$

$\hat{d}^{\text{i}} = \hat{B}_d^{\text{i}} + \dfrac{nu(g - 1 - np_2)}{2(nu + (g-1)y)}$

Large bureaucracy region, $\hat{n}^{\text{large}} < n \leq \hat{n}_b^{\max}$

$$\hat{B}_d^{\text{large}} = np_2 + \frac{1}{2}(g - 1 - np_2) \qquad d^{\text{large}} = \hat{B}_d^{\text{large}} + \frac{g - 1 - np_2}{2(1+y)} \tag{19}$$

We get $\hat{n}^{\text{large}} = \dfrac{g-1}{u}$ by equating bureaucracy profits for the last two cases.

The analysis of the model is focused on how the amount of rule-breaking changes when dispatchers are introduced. The allocation of licenses to undeserving firms is presented in figure 5. The threshold between licenses awarded and informality is $A_c^{i,\text{small}} = \dfrac{1}{2}\left(1 + \dfrac{np_1}{g - 1 - nu}\right)$ up to bureaucracy sizes $\hat{n}^{\text{small}}$, the convex curve in figure 5 (which is dashed between $\hat{n}^{\text{small}}$ and $\hat{n}_c^{\max} = \dfrac{g-1}{p_1 + u}$ to indicate the informal/bribe threshold if there were no dispatchers). Above $\hat{n}^{\text{small}}$, the threshold between licenses awarded and informality is first $A_d^{i,\text{i}} = \dfrac{1}{2} + \dfrac{nu + np_2 y}{2(nu + (g-1)y)}$, then $A_d^{i,\text{large}} = \dfrac{2+y}{2+2y} + \dfrac{np_2}{2(1+y)(g-1)}$.

[42] See appendix 4.

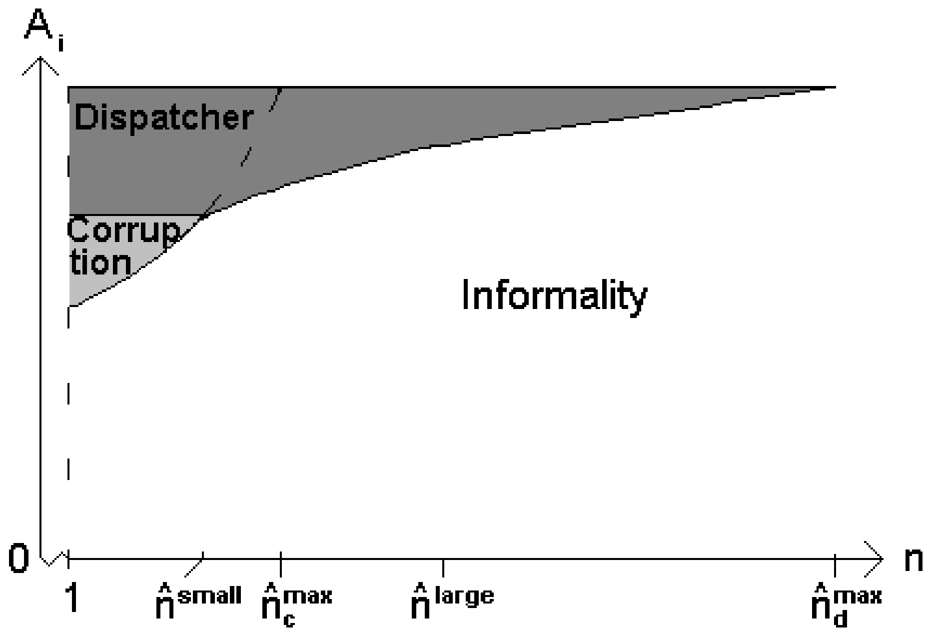

Figure 5. License allocations for undeserving firms.[43]

The main result of this section can be expressed as follows: There is a small bureaucracy range, $1 \leq n \leq \hat{n}^{\text{small}}$, in which the introduction of dispatchers does not change the amount of undeserving firms that gets licenses. Above $\hat{n}^{\text{small}}$, however, the introduction of dispatchers increases the amount of rule breaking. The threshold $\hat{n}^{\text{small}}$ is increasing in p_2 and decreasing in p_1 and y. It is increasing in u for small values of u, then decreasing. The maximum size of bureaucracy for which undeserving firms can obtain licenses by bribing increases from $\hat{n}_c^{\max} = \dfrac{g-1}{p_1+u}$ to $\hat{n}_d^{\max} = \dfrac{g-1}{p_2}$.

Dispatchers increase the extent of corruption for large bureaucracies and anti-corruption policies become largely ineffective. The latter part resembles the analysis of Hasker and Okten (2007) who point out that anticorruption policies become ineffective when intermediaries are present. As the expected penalty on bureaucrats in the direct interaction with firms, p_1, increases, corruption moves to the intermediary sector ($\hat{n}^{\text{small}}$ shifts to the left in figure 5).[44]

[43] The parameters are $g = 2$, $p_1 = 0.06$, $p_2 = 0.05$, $u = 0.1$, $y = 1$.

[44] If the small bureaucracy region existed at the onset, there is some corruption detention effect from raising penalties ($A_d^{i,\text{i}}$ replaces $A_c^{i,\text{small}}$ as the threshold to informality and $A_d^{i,\text{i}}$ must be larger than $A_c^{i,\text{small}}$ for the small bureaucracy region to exist), but as $\hat{n}^{\text{small}} \to 1$, all corruption has moved to the intermediary sector.

When interacting with dispatchers, p_2 is the bureaucrat's expected penalty. Because the relation between bureaucrats and dispatchers is secret in nature, p_2 is low. In the present set-up with bending the rules corruption, increasing p_2 would be a good anti-corruption policy and it would have two effects: less corruption for any bureaucracy size above $\hat{n}^{\text{small}}$ and a reduction in the bureaucracy size region where rule-breaking is feasible.

Note that the effect of the number of dispatchers, y, is also distinct from the previous model. Competition between dispatchers will increase the amount of rule-breaking, suggesting that a government should make it difficult for such intermediaries to operate.

9 Discussion

In two recent papers on intermediaries and corruption, Hasker and Okten (2007, in abstract) and Bose and Gangopadhyay (2008, in abstract), respectively, stress that "intermediary agents worsen the impact of corruption" and "welfare in an economy with intermediaries is lower than that in an economy without intermediaries". This paper instead stresses the possibility that bureaucracy intermediaries, here called "dispatchers", can improve welfare.

The few academic studies that exist on despachantes in Brazil, as well as reports from several countries in Latin America, stress that time saving is an important reason for using such intermediaries.

Perhaps the main difference between the present paper and earlier literature is in the view on regulation. As shown by the work following de Soto (1989) and Djankov et al. (2002), there is strong evidence of substantial red tape in many countries. In Latin America it is customary, both for individuals and firms in their interaction with the bureaucracy, to have to visit many different government offices, offices that are located at different places, have different and irregular opening hours, each requiring authentication and certification of documents, and so on. The model developed in this paper takes literally the time it takes to perform procedures in such an environment and asks what is the impact of dispatchers.

In a different strand of literature, regulation is taken to be optimal. The type of question then posed is how licenses should be allocated to the most deserving agents (Banerjee, 1997), what is the impact of intermediaries on the amount of regulation faced by firms and on the effectiveness of anticorruption policies (Hasker and Okten, 2007) and what is the impact of intermediaries on waiting times in the bureaucracy, from a larger amount of undeserving agents applying (Bose and Gangopadhyay, 2008). In section 8.2, the model was modified and reinterpreted to highlight the difference in results when the view on corruption and regulation changes.

This paper includes an extensive margin, i.e. informality. There is vast informality in many bureaucratic procedures in most developing countries, even in "compulsory" regulation. By including this margin, the analysis highlights how informality changes when the bureaucracy becomes more complicated, with and without dispatchers (with Brazilian firm start-up exemplifying, section 6.1).

The paper focuses on incentives of bureaucrats and dispatchers to create red tape, possibly reversing welfare improving effects. In São Paulo, Brazil, a recent bureaucracy reform is called "PoupaTempo" (SaveTime). The co-location of government counters at the same physical location reduces the number of visits to the bureaucracy (nt in the model). There is ample evidence, seen in newspaper reports, pending cases at the judiciary etc., that despachante organizations try to delay and block such reforms. This is consistent with the small bureaucracy case in the model where dispatchers (and corrupt bureaucrats) prefer to maintain, and increase, the amount of regulation.

Bibliography

[1] Banerjee, A. V., 1997. A theory of misgovernance. Quarterly Journal of Economics, 112, 1289-1332.

[2] Bardhan, P., 1997. Corruption and Development: A Review of Issues, Journal of Economic Literature 25, 1320–1346.

[3] Bertrand, M., Djankov, S., Hanna, R., Mullainathan, S., 2007. Obtaining a Driving License in India: An Experimental Approach to Studying Corruption, Quarterly Journal of Economics, 122, 1639-1676.

[4] Bertrand, M., Djankov, S., Hanna, R., Mullainathan, S., 2008. Corruption in Driving Licenses in Delhi, Economic and Political Weekly 43 (5), 71-76.

[5] Blanchard, O., Kremer, M., 1997. Disorganization. Quarterly Journal of Economics 112 (4), 1091-1126.

[6] Bose, G., Gangopadhyay, S., 2008. Intermediaries in corruption markets. Manuscript.

[7] CIEN (Centro de Investigaciones Económicas Nacionales), 2001. Análisis de la economía informal en Guatemala. Estudio de casos: entrevistas a micro y pequeños empresarios.

[8] CIET (Centro de Investigación de Enfermedades Tropicales), 1998a: Bolivia (REP): Perceptions of the population concerning corruption in public services.

[9] CIET, 1998b: Bolivia (REP): Perception of the private sector on corruption in public services

[10] DaMatta, R., 1979. Carnavais, malandros e heróis: para uma sociologia do dilema brasileiro. Rio de Janeiro: Rocco.

[11] DaMatta, R., 1984. O que faz o brasil, Brasil? Rio de Janeiro: Rocco.

[12] de Soto, H., 1989. The Other Path. New York: Harper and Row.

[13] Djankov, S., La Porta, R., Lopez-De-Silanes, F., Shleifer, A., 2002. The Regulation of Entry. The Quarterly Journal of Economics 117, 1-37.

[14] Fjeldstad, O., 2003. Fighting fiscal corruption: lessons from the Tanzania revenue authority. Public Administration and Development 23, 165-175.

[15] Gancheva, Y. 1999. Rules, Regulations and Transactions Costs in Transition Bulgaria. Institute for Market Economics, RSS 993/1999.

[16] Hasker, K., Okten, C., Intermediaries and corruption. Journal of Economic Behavior & Organization 67, 103–115.

[17] Husted, B. 1994. Honor among Thieves: A Transaction-Cost Interpretation of Corruption in Third World Countries. Business Ethics Quarterly 4, 17-27.

[18] IFC, 2007a. Business Simplification in Lima, Peru – An evaluation of the reform of licensing procedures.

[19] IFC, 2007b. Municipal scorecard. Midiendo las barreras burocráticas a nivel municipal - reporte Bolivia.

[20] IFC, 2008. Reforma del proceso para obtener permiso de operación de negocio en las alcaldías municipales del distrito central y San Pedro Sula.

[21] Lambsdorff, J., 2002. Making corrupt deals: contracting in the shadow of the law. Journal of Economic Behavior and Organization 48, 221-241.

[22] Levine, V., 1975. Political corruption: the Ghana case. Stanford, CA. Hoover Institution Press.

[23] Lui, F., 1985. An Equilibrium Queuing Model of Bribery. Journal of Political Economy 93, 760–781.

[24] Myrdal, G., 1968. Asian drama: An inquiry into the poverty of nations. New York: Twentieth Century Fund.

[25] Oldenburg, P., 1987. Middlemen in third-world corruption: Implications of an Indian case. World Politics 39, 508-535.

[26] Olken, B., 2007. Monitoring Corruption: Evidence from a Field Experiment in Indonesia. Journal of Political Economy 115, 200-249.

[27] Proética, 2006. Segunda encuesta nacional sobre corrupción - breve análisis de los resultados.

[28] Republica de Bolivia - Vicepresidencia de la Republica, 1998. Plan nacional de integridad - Estrategia Boliviana de desarrollo institucional y lucha contra la corrupción.

[29] Rose-Ackerman, S.,1999. Corruption and Government: Causes, Consequences and Reform. Cambridge University Press: Cambridge.

[30] Rosenn, K., 1971. The jeito, Brazil´s institutional bypass of the formal legal system and its developmental implications. The American Journal of Comparative law, 514-549.

[31] SEBRAE, 2005. Economia Informal Urbana 2003.

[32] Shleifer, A., Vishny, R., 1993. Corruption. Quarterly Journal of Economics 108, 599-617.

[33] Shleifer, A., Vishny, R., 1998. The grabbing hand: Government pathologies and their cures. Cambridge, Massachusetts. Harvard University Press.

[34] Stone, A., Levy, B., Paredes, R., 1996. Public Institutions and Private Transactions: A Comparative Analysis of the Legal and Regulatory Environment for Business Transactions in Brazil and Chile, in "Empirical studies in institutional change" (Alston, Eggertsson & North eds.)

[35] Svensson, J., 2005. Eight Questions about Corruption. Journal of Economic Perspectives, 19 (3), 19-42.

[36] Wade, R., 1982. The System of Administrative and Political Corruption: Canal Irrigation in South India. Journal of Development Studies 18 (3), 287–328.

[37] World Bank, 1996. Guatemala: building peace with rapid and equitable growth. Report No. 15352-GU.

[38] World Bank, 2009a. http://www.doingbusiness.org/

[39] World Bank, 2009b. http://www.enterprisesurveys.org/

[40] Zylbersztajn, D., Faccioli, F., Silveira, R., 2007. Measuring the start up costs in Brazilian small firms. Revista de Administração da Universidade de São Paulo 42, 293-301.

[41] Zylbersztajn, D., Graça, C., 2003. Los costos de la formalización empresarial: medición de los costos de transacción en Brasil. Revista de Economía Institucional 5, 146-165.

Appendix

A1: Bureaucrat´s problem in the no-dispatcher case

Choose b_c to maximize

$$\pi_c = (b_c - p) \times \left(1 - \text{Max}\left\{\frac{B_c}{g-1-(n-m)\,k-nt}, \frac{B_c - np}{mk}\right\}\right)$$

The first order conditions are

$$B_c > \bar{B}: \qquad \left(1 - \frac{b_c + (m-1)\,\tilde{b}_c + (n-m)\,p}{g-1-(n-m)\,k-nt}\right) - \frac{b_c - p}{g-1-(n-m)\,k-nt} = 0$$

$$B_c < \bar{B}: \qquad \left(1 - \frac{b_c + (m-1)\,\tilde{b}_c - mp}{mk}\right) - \frac{b_c - p}{mk} = 0$$

In addition, if none of these two conditions hold, we get

$$B_c = \bar{B}: \qquad \frac{b_c + (m-1)\,\tilde{b}_c + (n-m)\,p}{g-1-(n-m)\,k-nt} = \frac{b_c + (m-1)\,\tilde{b}_c - mp}{mk}$$

Applying symmetry between bureaucrats ($b_c = \tilde{b}_c$), solving for b_c and then aggregating to obtain B_c, we get

$$B_c > \bar{B}: \qquad B_c = np + \frac{m}{1+m}\,(mk)$$

$$B_c < \bar{B}: \qquad B_c = np + \frac{m}{1+m}\,(g-1-np-(n-m)\,k-nt)$$

$$B_c = \bar{B}: \qquad B_c = \bar{B} = np\frac{g-1-(n-m)\,k-nt}{g-1-nk-nt}$$

These are three different candidates for solutions. Depending on the parameters of the problem, one of these will apply. I equate the expression for each candidate for optimal B_c with the threshold bribe level $\bar{B}$ and solve for the values of n for which the solution applies. I have chosen to consider the number of corrupt bureaucrats m as fixed when n changes. The ratio of m/n will thus change when n varies. Throughout the paper, unless otherwise stated, adding one bureaucratic step will mean adding one *honest* step. The solution can be written as follows:

Small bureaucracy region, $1 \leq n \leq n_c^{\text{small}}$

$$B_c^{\text{small}} = np + \frac{m}{1+m}\,(mk)$$

Intermediate bureaucracy region, $n_c^{\text{small}} < n \leq n_c^{\text{large}}$

$$B_c^{\text{intermediate}} = \bar{B} = np\frac{g-1-(n-m)\,k-nt}{g-1-nk-nt}$$

Large bureaucracy region, $n_c^{\text{large}} < n \leq n_c^{\max}$

$$B_c^{\text{large}} = np + \frac{m}{1+m}\,(g-1-np-(n-m)\,k-nt)$$

where $n_c^{\text{small}} = \dfrac{g-1}{p+k+t+p/m}$ solves $B_c^{\text{small}} = \bar{B}$,
$n_c^{\text{large}} = \frac{g-1+k+mk}{2(k+t)} + \frac{g-1-k}{2(k+t+p)} - \frac{\sqrt{((g-1)(p+2t)+mk^2+k(2g-2+p+mp+mt))^2 - 4(g-1)(g-1+mk)(k+t)(k+t+p)}}{2(k+t)(k+t+p)}$
solves $B_c^{\text{large}} = \bar{B}$ and $n_c^{\max} = \dfrac{g-1+mk}{p+k+t}$ is the maximum bureaucracy size for which any licenses will be awarded.

A2: The dispatcher profit maximization problem

Choose q_d to maximize $q_d\left(d\left(q_d\right) - B_d\right)$

$$\text{where } d\left(q_d\right) = \text{Min}\left\{\begin{array}{c} B_c + \left(\left(n-m\right)k + nt\right)\left(1 - \left(q_d + \left(x-1\right)\tilde{q}_d\right)\right) \\ np + \left(nk + nt\right)\left(1 - \left(q_d + \left(x-1\right)\tilde{q}_d\right)\right) \\ \left(g-1\right)\left(1 - \left(q_d + \left(x-1\right)\tilde{q}_d\right)\right) \end{array}\right\}$$

The first order conditions are

Case 1: $\quad 2\left(\left(n-m\right)k + nt\right)q_d = B_c + \left(\left(n-m\right)k + nt\right)\left(1 - \left(x-1\right)\tilde{q}_d\right) - B_d$

Case 2: $\quad 2\left(nk + nt\right)q_d = np + \left(nk + nt\right)\left(1 - \left(x-1\right)\tilde{q}_d\right) - B_d$

Case 3: $\quad 2\left(g-1\right)q_d = \left(g-1\right)\left(1 - \left(x-1\right)\tilde{q}_d\right) - B_d$

If none of these three conditions apply, we get
Case 4: $d = \bar{d}$

Applying symmetry between all dispatchers ($\tilde{q}_d = q_d$) and then solving for q_d, plugging the optimal quantities back into $d\left(q_d\right)$, we get the following optimal response functions:

$$d_1\left(B_d\right) = \frac{x}{1+x}B_d + \frac{1}{1+x}\left(B_c + \left(n-m\right)k + nt\right)$$
$$d_2\left(B_d\right) = \frac{x}{1+x}B_d + \frac{1}{1+x}\left(np + nk + nt\right)$$
$$d_3\left(B_d\right) = \frac{x}{1+x}B_d + \frac{1}{1+x}\left(g-1\right)$$
$d_4\left(B_d\right) = \bar{d}$ when neither of the above applies.

Which condition applies depends on the optimal values of B_c and B_d, which are yet to be determined.

A3: Bureaucrat´s problem in the dispatcher case

As before, corruption is decentralized. We use

$B_c = b_c + (m-1)\,\tilde{b}_c + (n-m)\,p$

$B_d = b_d + (m-1)\,b_d + (n-m)\,p$

and the optimal dispatcher response functions

$d_1(B_d) = \frac{x}{1+x}B_d + \frac{1}{1+x}(B_c + (n-m)\,k + nt)$

$d_2(B_d) = \frac{x}{1+x}B_d + \frac{1}{1+x}(np + nk + nt)$

$d_3(B_d) = \frac{x}{1+x}B_d + \frac{1}{1+x}(g-1)$

Direct and indirect demand

The first problem to solve is when bureaucrats have both direct and indirect demand, this amounts to three different cases. Assume first that $\frac{B_c - np}{mk} > \frac{B_c}{g-1-(n-m)\,k-nt}$, i.e. that some firms also go to the de jure bureaucracy. Each bureaucrat chooses b_d and b_c, taking the choices of the other bureaucrats $((m-1)\,\tilde{b}_d$ and $(m-1)\,\tilde{b}_c)$ as given, to maximize:

$$\text{Max } (b_d - p)\left(1 - \frac{d_1(B_d) - B_c}{(n-m)\,k + nt}\right) + (b_c - p)\left(\frac{d_1(B_d) - B_c}{(n-m)\,k + nt} - \frac{B_c - np}{mk}\right)$$

The profit function can be rewritten as follows

$(b_d - b_c)\left(1 - \frac{d_1(B_d) - B_c}{(n-m)\,k + nt}\right) + (b_c - p)\left(1 - \frac{B_c - np}{mk}\right)$

Note that the second component of the rewritten profit function is as in the case without dispatchers. The first order condition with respect to b_d is simple. After applying symmetry between bureaucrats ($b_c = \tilde{b}_c$, $b_d = b_d$) it becomes:

$b_d = b_c + \frac{(n-m)\,k + nt}{1+m}$

In choosing b_d, the direct bribe level acts as an opportunity cost for both bureaucrats and firms. It plays the same role as p does in the choice of direct bribe level in the no-dispatcher case and b_d is therefore additive in b_c. Using this first order condition, we can again rewrite the profit function:

$\frac{((n-m)\,k + nt)\,x}{(1+m)^2\,(1+x)} + (b_c - p)\left(1 - \frac{B_c - np}{mk}\right)$

The first term, i.e. the rewritten indirect profit term, does not depend on b_c. The introduction of dispatchers has thus not changed the bureaucrat's choice of optimal direct bribe, in the case where we have coexistence of de jure bureaucracy, direct corruption and dispatchers. After aggregating over bureaucrats, we get the small bureaucracy case solution:

Small bureaucracy region, $1 \leq n \leq n^{\text{small}}$

$$B_c^{\text{small}} = np + \frac{m}{1+m}(mk) \qquad B_d^{\text{small}} = np + \frac{m}{1+m}(nk+nt)$$

$$d^{\text{small}} = B_d^{\text{small}} + \frac{(n-m)k+nt}{(1+m)(1+x)}$$

where $n^{\text{small}} = n_c^{\text{small}}$ (from the no-dispatcher case, subindex is omitted in the text) solves $B_c^{\text{small}} = \bar{B}$

The second case of indirect demand, $\dfrac{B_c - np}{mk} < \dfrac{B_c}{g-1-(n-m)k-nt}$, corresponds to the large bureaucracy case in the model without dispatchers. This case will never be optimal: bureaucrats find it more profitable, for large bureaucracies, to channel demand for licenses through dispatchers.

The third possibility of direct demand is that the $B_c = \bar{B}$-constraint binds. This will be the case in a first intermediate bureaucracy size region. The bureaucrat then chooses b_d to

$$\text{Max}\,(b_d - p)\left(1 - \frac{d_1(B_d) - B_c}{(n-m)k+nt}\right) + (b_c - p)\left(\frac{d_1(B_d) - B_c}{(n-m)k+nt} - \frac{B_c}{g-1-(n-m)k-nt}\right)$$

s.t. $B_c = \bar{B}$.

The first order condition is:

$$b_d = \frac{1}{2}\left(\tilde{b}_d - m\tilde{b}_d + p(1+m) + nt + k\left((n-m) + \frac{(1+m)np}{g-1-n(k+t)}\right)\right)$$

Applying symmetry between bureaucrats ($b_d = \tilde{b}_d$), solving for b_d and then aggregating to obtain B_d, we get the solution:

i1 - Intermediate bureaucracy region 1, $n^{\text{small}} < n \leq n^{\text{i}}$

$$B_c^{\text{i1}} = \bar{B} \qquad B_d^{\text{i1}} = \bar{B} + \frac{m}{1+m}((n-m)k+nt)$$

$$d^{\text{i1}} = B_d^{\text{i1}} + \frac{(n-m)k+nt}{(1+m)(1+x)}$$

where $n^{\text{i}} = \dfrac{g-1}{p+k+t+px/(1+m+mx)}$ solves $\dfrac{d^{\text{i1}} - B_c^{\text{i1}}}{(n-m)k+nt} = \dfrac{B_c^{\text{i1}}}{g-1-(n-m)k-nt}$

Indirect demand only

With indirect demand only, there are three cases that can apply: Either of the two demand components $\left(\frac{d_2(B_d)-np}{nk+nt}, \frac{d_3(B_d)}{g-1}\right)$ can be larger than the other, or we have $d=\bar{d}$. The $\frac{d_2(B_d)-np}{nk+nt}$-component of demand can be discarded however.[45]

With $d=\bar{d}$, we get a second intermediate bureaucracy size region[46] :

i2 - Intermediate bureaucracy region 2, $n^{\mathrm{i}} < n \leq n_d^{\mathrm{large}}$

$$B_c^{\mathrm{i2}} = \bar{B} \qquad B_d^{\mathrm{i2}} = \frac{((n-m)k+nt)(g-1-np-nk-nt)+(g-1)npx}{(g-1-nk-nt)x}$$

$$d^{\mathrm{i2}} = \bar{d}$$

Finally, there is a large bureaucracy size region in which the $\frac{d_3(B_d)}{g-1}$-term applies, i.e. the marginal firm chooses between dispatchers and informality. The bureaucrat the chooses b_d to

$$\text{Max } (b_d - p)\left(1 - \frac{d_3(B_d)}{g-1}\right)$$

The first order condition is:

$$b_d = \frac{1}{2}\left(g-1+\tilde{b}_d - m\tilde{b}_d + p(1+m) - np\right)$$

Applying symmetry between bureaucrats ($b_d = \tilde{b}_d$), solving for b_d and then aggregating to obtain B_d, we get the solution:

Large bureaucracy region, $n_d^{\mathrm{large}} < n \leq n_d^{\mathrm{max}}$

$$B_d^{\mathrm{large}} = np + \frac{m}{1+m}(g-1-np) \qquad d^{\mathrm{large}} = B_d^{\mathrm{large}} + \frac{g-1-np}{(1+m)(1+x)}$$

where n_d^{large} is obtained by equating bureaucracy profits for cases **i2** and **large.**[47]

---

[45] The case where firms choose between dispatchers and the de jure bureaucracy will only occur when there is no mark-up in the dispatcher sector, and then comes out as a solution to the intermediate 1 case when $x \to \infty$.

[46] Over the region $n^{\mathrm{i}} < n \leq n_d^{\mathrm{large}}$, the interests of bureaucrats and dispatchers are unaligned. Moving first, bureaucrats can therefore gain corruption profits at the expense of dispatchers. In this region, bureaucrats set the indirect bribe level B_d such that the best response of dispatchers is $\bar{d}$. In doing this, bureaucrats keep the direct bribe at a level $B_c = \bar{B}$, such that dispatchers would lose demand to direct corruption if they set any other price than $\bar{d}$.

The upper limit, n_d^{large}, is determined by bureaucrats comparing profits in this case to the unconstrained large bureaucracy case. This will generate a discontinuity in prices and also in the allocations at n_d^{large}, seen in the lower right-hand panel of figure 3.

[47] The analytical solution for n_d^{large} is not stated due to a large expression.

As a final remark, if corruption were centralized, the solution is obtained by replacing "m" in all four cases (small, i1, i2, large) with "1", except in the terms (mk) and $(n-m)\,k$. As an example, we would get $B_c^{\text{small}} = np + \frac{1}{2}(mk)$.

A4: Breaking the rules corruption

The dispatcher demand curve is $1-\text{Max}\left\{\dfrac{d-B_c}{nu}, \dfrac{d}{g-1}\right\}$. Solving the dispatcher problem gives the two response functions, $d_1(B_d) = \dfrac{y}{1+y}B_d + \dfrac{1}{1+y}(B_c+nu)$ and $d_2(B_d) = \dfrac{y}{1+y}B_d + \dfrac{1}{1+y}(g-1)$. The bureaucrats' profit maximization problem and solution is:

Small bureaucracy region, $1 \leq n \leq \hat{n}^{\text{small}}$. Bureaucrats set B_d and B_c to

$$\text{Max}\,(B_d - np_2)\left(1 - \frac{d_1(B_d) - B_c}{nu}\right) + (B_c - np_1)\left(\frac{d_1(B_d) - B_c}{nu} - \frac{B_c}{g-1-nu}\right) \Rightarrow$$

$$\hat{B}_c^{\text{small}} = np_1 + \frac{1}{2}(g-1-np_1-nu) \qquad \hat{B}_d^{\text{small}} = np_2 + \frac{1}{2}(g-1-np_2)$$

$$\hat{d}^{\text{small}} = \hat{B}_d^{\text{small}} + \frac{n(p_1-p_2+u)}{2(1+y)}$$

$$\hat{n}^{\text{small}} = \frac{(g-1)(u+(p_2-p_1)y)}{u(p_1+p_2y+u)} \text{ solves } \frac{\hat{d}^{\text{small}} - \hat{B}_c^{\text{small}}}{nu} = \frac{\hat{B}_c^{\text{small}}}{g-1-nu}$$

i - Intermediate bureaucracy region, $\hat{n}^{\text{small}} < n \leq \hat{n}^{\text{large}}$

$$\hat{B}_c^{\text{i}} = \frac{(g-1-nu)(2nu+(g-1+np_2)y)}{2(nu+(g-1)y)} \qquad \hat{B}_d^{\text{i}} = np_2 + \frac{1}{2}(g-1-np_2)$$

$$\hat{d}^{\text{i}} = \hat{B}_d^{\text{i}} + \frac{nu(g-1-np_2)}{2(nu+(g-1)y)}$$

Large bureaucracy region, $\hat{n}^{\text{large}} < n \leq \hat{n}_b^{\max}$. Bureaucrats choose B_d to

$$\text{Max}\,(B_d - np_2)\left(1 - \frac{d_2(B_d)}{g-1}\right) \Rightarrow$$

$$\hat{B}_d^{\text{large}} = np_2 + \frac{1}{2}(g-1-np_2) \qquad d^{\text{large}} = \hat{B}_d^{\text{large}} + \frac{g-1-np_2}{2(1+y)}$$

We get $\hat{n}^{\text{large}} = \dfrac{g-1}{u}$ by equating bureaucracy profits for the last 2 case

Chapter 3

Informal firms, investment incentives and formalization*

1 Introduction

In a typical developing country, the majority of small firms are informal and entry costs into formality are high. This paper is motivated by these two observations. It addresses the question of what can be expected in terms of firm investment, growth and formalization in such a setting. In particular, the paper focuses on firms' incentive to invest when, at some future point in time, an increase in productivity can be gained, but only after paying large entry costs into formality. The effect of a penalty policy on informal firm investment, growth and formalization is also discussed.

The observation that most small firms in developing countries are informal is well-established. A recent enterprise survey in Brazil shows that 90% of the smallest firms, i.e. firms with 1-5 employees, have not gone through the procedure to register as a legal entity (SEBRAE, 2005). An enterprise survey in Mexico, the other large Latin American economy, shows similar values (INEGI, 2003). Studies and accounts

* I thank Shon Ferguson, Harry Flam, Johan Gars, Gustav Hansson, John Hassler, Per Krusell, Ted Miguel, José Mauricio Prado, Jesper Stage, Jakob Svensson and Fabrizio Zilibotti for helpful discussions, feedback and suggestions. I also thank participants at the Stockholm University Department of Economics workshop, the Berkeley development lunch workshop, the IIES brown bag seminar, and participants at SUDSWEc and the Nordic Conference of Development Economics in Copenhagen, both held in 2007. Thanks also to Christina Lönnblad for editorial assistance. Any remaining errors are mine.

from other developing countries indicate similar degrees of informality among the smallest firms in the economy (see, for instance, Bigsten et al., 2004, for Kenya and de Soto, 1989, for Peru).

Turning to bureaucratic and legal costs facing small and medium enterprises, such costs have received considerable attention in recent development research. In particular, the work by de Soto (1989) and Djankov et al. (2002) has directed the attention to substantial government-related costs of "doing business" and entry into formality. Examples of such costs are start-up fees, financial costs incurred in order to pay taxes (except for the taxes themselves), financial costs related to hiring and laying off workers, as well as the time spent with these activities. These costs can be substantial. Whereas it costs USD 370 to start a firm in the US, the average cost in Latin America is around USD 1240, as reported by the World Bank Doing Business project. The average monthly income per capita was USD 3840 in the United States in 2007, meaning that three days of work generate an income equal to the firm start-up cost. In Latin America, the average monthly income was one tenth as much, or USD 380. It thus takes more than three months of work to generate an income equal to the firm start-up cost. Furthermore, income levels in the informal sector in Latin America are typically much lower than the official GNI figures, meaning that it takes even longer to generate an income equal to the firm start up costs.

As implied by the above, the definition of an informal firm used in this paper is a firm that has not gone through the registration procedure at the government bureaucracy.

The combination of small informal firms and large formalization costs has motivated setting up a simple dynamic model of profit-maximizing firms. Firms can invest in their capital stock, grow larger and, possibly over time, become formal. The cost of becoming formal is taken literally: at one instant in time, the firm can choose to pay the formalization fee, defined as F**;** a fee that represents all costs to register the firm at the government bureaucracy. Having paid F, the firm changes status from informal to formal and obtains a productivity benefit.

How do formalization costs, to be paid at some future date, affect investment today? At what firm size and when do firms choose to become formal, if at all? What are the crucial parameters affecting firm formalization? What is the effect of credit constraints on the formalization decision? Can formalization costs lead to poverty traps? How should policy vis-a-vis informal firms be viewed? How can the government affect the formalization decision? These questions are addressed in this paper.

Several interesting results emerge from the analysis of the tractable dynamic model. First, the investment paths and growth trajectories differ substantially between firms that choose to formalize and those (ex-ante almost identical firms) that do not. Second, the formalization decision depends non-trivially on the productivity of the informal firm, due to the balancing of an *accumulation effect* and a *threshold effect*. This, in turn, has an effect on how policy designed to incentivize informal firms to become formal should be designed. Third, when aggregating over firms, the long-run firm size distribution exhibits a range of small firms and a range of larger firms but also a "missing middle", much in line with actual firm size distributions observed in developing countries (Bigsten et al. 2004, Tybout, 2000). Fourth, the long-run firm-size distribution turns out to depend on the initial firm-level stock of capital, a result that can be interpreted as a poverty/informality trap.

The paper proceeds as follows: In section 2, the literature to which this paper relates is reviewed and the model to be presented is motivated in relation to this earlier writing. Section 3 discusses formalization costs in different countries and presents some data on income levels in the informal economy, together with typical informal firm capital stocks and profit levels from three recent studies. The dynamic model of firm investment and formalization is presented in section 4 and analyzed in section 5. Some extensions to the analysis, focusing on how the investment and formalization behavior changes when the basic assumptions are altered, are to be found in section 6. Section 7 discusses the results and concludes the paper. The appendix presents some of the details in deriving the analytical results.

2 Literature review

An important debate in the literature on the informal sector, preceding the analysis in this paper, is whether small informal entrepreneurial activities should be considered as proper "firms" at all, or merely as temporary subsistence labor while waiting for a formal job. In early writings on how the economy develops from traditional to modern, Lewis (1954), Todaro (1969) and Harris and Todaro (1971) considered the "urban traditional" sector as a source of labor supply for the "modern" sector.[1] In none of these papers is the urban traditional sector seen as an important element of economic activity or as a contributor to capital accumulation. It is rather considered as a temporary low-productivity subsistence activity.

The entrepreneurial view that informal small-scale economic activities should be considered as entrepreneurs/firms, rather than as subsistence activities, has been popularized by de Soto (1989). However, a change in terminology and focus to an informal – rather than an "urban traditional" – sector, stressing entrepreneurship and not only surplus labor, emerged with the writings of the International Labor Organization (Hart, 1973; ILO, 1972). The informal sector/informal economy started to be seen more as a permanent, increasing and diverse phenomenon – "from marginal operations to large enterprises" (Hart, p. 68).[2]

In an "occupational choice" model in the spirit of Lucas (1978), with economic agents differing in entrepreneurial ability, Rauch (1991) studies the choice between being a worker, an informal entrepreneur or a formal entrepreneur. Both types of entrepreneurs employ workers. The informal entrepreneurial sector arises as a result of a government (above-market clearing) minimum wage policy. The static general equilibrium model delivers predictions on the relative size of the informal sector, firm size distribution, and changes to these from the minimum wage level.[3]

[1] Starting with the work of Lewis (1954), the traditional urban sector plus rural-to-urban migrants were seen as a source of unlimited labor supply from which the modern sector could get labor at subsistence pay. Todaro (1969) modeled the rural–to urban migration decision, taking into account the existence of an unemployed or underemployed pool of urban traditional workers that compete for the same jobs as rural migrants. Harris and Todaro (1971) studied a minimum wage policy in a similar setting.

[2] Rauch (1991), Chen (2004) and de Mel et al. (2008) all discuss early writings on the informal economy.

[3] Rauch's paper can be seen as combining the two views above on informal activity. In recent empirical work from Sri Lanka, de Mel et al. (2008) collect data on personal characteristics from wage workers, own-account workers and owners of enterprises with 5-50 employees to address the

Minimum wages that do not clear the market constitute an example of a government intervention that may lead to an informal sector. The focus here is instead on the effects of government-imposed formalization costs. The paper takes as given the de Soto entrepreneurial view and studies investment and formalization decisions of profit maximizing informal firms.

The question of whether a firm formalizes or not in the face of large such costs involves at least two issues: the formalization costs themselves and the potential gains from formalization. In addition, a modeling choice must be made. A dynamic framework is appropriate to capture the effects of large formalization costs on small informal firms: firms must grow to a certain size to become formal. A dynamic model can also shed light on how the investment incentives and the resulting growth path today are affected by a future "non-convexity" in the production function.

To the best of my knowledge, this paper is the first to explicitly focus on the investment incentives in anticipation of a formalization cost. However, the model is similar in spirit to the literature on non-convexities and poverty traps, a literature that typically focuses on whether initial (wealth) conditions matter for long-run allocations.[4]

---

question of whether own-account workers resemble wage workers or firm-owners more. By using a "species classification" approach from biology, they classify around 70% of the own-account workers in their study as wage workers and 30% as small- and medium-size enterprise owners. On the other hand, the authors argue that given the large number of own-account workers in low-income countries, the possibilities for job creation and growth in this sector should not be ignored. In addition, by exploiting the panel structure of the data on own-account workers, for the fraction of this group that resembles enterprise owners, the growth rate implied by these own-account workers' transition into being employers is much larger than the growth rates found in a comparable study with data from the United States.

[4] Regarding terminology, the present paper discusses government imposed formalization costs, in the form of going through a firm registration procedure, as the fundamental non-convexity which is of importance for firm growth. This is different from occupational choice models, such as Banerjee and Newman (1993), Ghatak and Jiang (2002) and Buera (2008), where the non-convexity is typically a minimum scale investment. The two different types of entry costs may well operate on different levels of firm size: an individual considering starting a manufacturing "firm" may consider buying a machine ("entry"). After having grown, such an informal manufacturing firm, with an established operation and possibly with a number of employees, may consider "formalization". In a recent empirical paper on the return to capital of investment for small firms in Mexico, McKenzie and Woodruff (2006) find, in line with other papers, high returns on small investments for the smallest firms and thus, they find no evidence of "entry nonconvexities". They do find lower returns for firms with a capital stock in the USD 1000-2000 range, however, and cannot reject that there is a threshold effect, one potential explanation for which is that "fiscal and bureaucratic costs are faced only by firms above a minimum size" (McKenzie and Woodruff, 2006, page 5).

One basic insight from neoclassical theory is that non-convexities alone will not affect long-run allocations. Economic agents could simply borrow to overcome such hurdles. The analysis of models with non-convexities is therefore intimately connected with introducing some other constraint relevant for developing economies, in particular capital market imperfections which may make individuals or firms unable to converge to a common long-run steady state or a balanced growth path (Banerjee, 2001 and McKenzie and Woodruff, 2006 discuss this point). The effects of initial capital and credit constraints on the possibilities for firm formalization are discussed in this paper.

Typically, the interplay between non-convexities, credit constraints and initial wealth is studied in dynamic occupational choice models with an OLG-structure, where one generation bequests wealth to the next and where individuals have a "warm glow" utility function. In the baseline human capital investment model of Galor and Zeira (1993), this results in a direct relationship between the initial wealth of one generation of a dynasty and the long-run steady state of the same dynasty.[5] There is no intergenerational saving/investment where a current generation takes into consideration the possibility that the decision of a future generation may be affected by today's choice.

The framework in this paper is different. Firms maximize profits over the entire life span of the firm. This means that the investment decision is truly intertemporal. The firm considers whether it should build up a capital stock over time, although this may imply current losses, in order to formalize at some later point in time.[6]

A feature of the present model, as opposed to most other papers, is that it is possible to solve analytically for the shape of the investment function over time. The comparative statics of the model can thus be analyzed in a straightforward way.

Turning to the second issue, what is to be gained from formalization? This paper

[5] In an extension, as well as in the occupational choice models of Banerjee and Newman (1993) and Ghatak and Jiang (2002), the entire wealth distribution endogenously determines occupational choices and wages which, in turn, affect the bequest to the next generation and the long-run equilibrium.

[6] See Banerjee (2001 pp. 31-32) for a discussion of "joy of giving" vs. "Barro preferences". In an appendix, Galor and Zeira (1993) point out how a "utility of offspring"-approach would affect their results and show that a poverty trap would still result. The occupational choice model of Buera (2008) also uses fully intertemporal preferences.

assumes that there is a productivity gain from becoming formal and focuses on the resulting effect on investment incentives while informal, but it does not provide one specific channel through which formal productivity is higher.

A non-exhaustive list of aspects that differ between informal and formal firms, from the development literature, includes access to credit and capital, taxes, public goods provisioning, access to risk pooling mechanisms, security in business environment, property rights, marketing possibilities, access to export markets, supplier-buyer relationships and other contracting issues (see, for instance, de Soto, 1989; Tokman, 1992; Levenson and Maloney, 1998; Bigsten et al. 2004; Chen, 2004; Maloney, 2004).[7,8]

One mechanism, out of many possible, that affects the productivity of informal firms and, therefore, the incentive to become formal, is instead proposed: Penalties and enforcement vis-a-vis informal firms make these firms divert time from production, with lower total production as a result. Tokman (1992) provides ample evidence that informal firms in Latin America organize part of production so that it is "invisible". The accounts in Tokman contain numerous examples of how small informal firms organize activities to minimize the disturbance from authorities, for instance by choosing less visible and less favorable production locations, physically hiding production when authorities visit and in anticipation of such visits, meeting customers one by one due to the lack of a visible sales location and marketing possibilities, and so on. The set-up, where firms respond to penalties by diverting time from production, allows us to explicitly study the effect of changes in policy, i.e. penalties, on informal firm investments and decisions to formalize.

The main focus of the paper is to study the investment incentives of an individual firm. However, the aggregate formalization behavior of heterogenous firms – differing in an ability parameter (or in initial capital) – is also studied. The aim is not to provide an industry evolution model, as in Jovanovic (1982), Hopenhayn (1992) and

[7] The effects of taxation in the formal sector and of differences in public goods provisioning between sectors are modeled by Loayza (1995) and Garcia Penalosa and Turnovsky (2005). Differences in access to outside finance are modeled by Antunes and Cavalcanti (2007).

[8] The assumption that formality brings a productivity benefit is not uncontroversial. As an example, much of the discussion in Brazil is centered around (too high) taxation in the formal sector. This paper assumes that formality is desirable, although the framework could, in principle, allow for firms that do not desire formality.

Melitz (2003). It is rather to display the implications of the non-convexity on long-run firm sizes and formality status, when firms differ in ability and initial capital. These "aggregate" predictions of the model are outlined in section 5.

3 Formalization costs

The cost of formalizing a business consists of both monetary costs and other costs. It is well documented that these costs can be very high (Djankov et al., 2002). The most up-to date source of information on such costs is most likely the "Doing Business" project financed by the World Bank. This data set on costs to start a firm originally covered 75 countries (Djankov et al., 2002), while 181 countries are now included (World Bank, 2009a). A summary of the most recent data, from 2009, is presented in table 1, with the number of procedures to register a firm and the official time it takes. The last column measures the official cost of the different registration procedures as a percentage of official Gross National Income (GNI). The financial cost to start a business (column 3) is at least 30% of yearly GNI per capita in most of the developing world, and as much as 111% in Sub-Saharan Africa.

Region	1 Number of procedures	2 Time (days)	3 Cost to start a firm / (GNI/capita)
East Asia & Pacific	8.6	44.2	32.3
Eastern Europe & Central Asia	7.7	22.6	8.6
Latin America & Caribbean	9.7	64.5	39.1
Middle East & North Africa	8.4	23.5	41.0
South Asia	7.4	32.5	31.9
Sub-Saharan Africa	10.2	47.8	111.2
OECD	5.8	13.4	4.9
United States	6	6	0.7

Table 1. Number of procedures, duration and cost to register a business in different parts of the world. Source: World Bank, 2009a.

Table 2 presents data for the year 2007 for the Latin American countries present in the World Bank data, augmented with informal economy income figures from Schneider (2002). Columns 1-3 show that 6-17 different bureaucratic procedures with a total cost of 585-2820 USD and taking 19-152 days have to be taken to formalize a firm. The average is 12 procedures, 58 days and 1238 USD in official cost. All Latin American countries have a higher firm start-up cost than the United States and the average cost is 336% of the US cost.[9]

Column 4 shows the official 2007 GNI/capita figures from World Bank (2009b), column 5 shows the informal GNP/capita figures from Schneider (2002) and columns 6-7 show the ratio between the cost to start a firm to the monthly informal GNP/capita and the ratio between a "total cost" to the monthly informal GNP/capita, respectively.[10],[11] Columns 6 and 7 can thus be interpreted as the number of months an average informal worker would have to work to generate an income equal to the official firm start-up cost and the total cost, respectively.

If we only focus on the official cost to start a firm (column 6), then Brazil, the most favorable country, requires three months of work to generate the income required for the formalization cost. The Latin-American average is 11 times informal GNP/capita and Bolivia and Nicaragua have very high costs in terms of informal income. These costs are high and are likely to be prohibitive for many small informal firms.

[9] All averages calculated are unweighted.

[10] To get an informal economy per capita income relevant for 2007, I have multiplied Schneider's informal economy per capita GNP figures, which refer to the year 2000, with the ratio between 2007 and 2000 official income figures. The calculation thus assumes that the informal economy per capita income has changed at the same rate as the official per capita income.

[11] The total cost measure, as perceived by an informal entrepreneur, is probably a summary measure of the monetary cost + the time cost of actually fulfilling all requirements + transport costs etc. to visit the different government bodies. The calculation for total cost in column 7 is somewhat ad-hoc and, as follows: (the official cost) + (the number of procedures times half the informal average daily GNP/capita) + (an ad-hoc measure of the loss of waiting set to the duration in days divided by three times half the daily informal GNP/capita).

The daily GNP/capita is the monthly GNP/capita divided by 20. Each procedure is assumed to require one day of work. Each procedure is assumed to have a value of half an average daily informal GNP/capita. The loss due to waiting is set to be a third of the duration time times half the daily GNP/capita.

Country	1 Number of procedures	2 Time (days)	3 Cost to start a firm (USD)	4 Official monthly GNI (USD)	5 Informal monthly GNP (USD)	6 Start up cost/ (Informal monthly GNP)	7 Total cost/ (Informal monthly GNP)
Argentina	14	31	702	504	128	5,5	6,1
Bolivia	15	50	1891	105	70	26,8	27,6
Brazil	17	152	585	493	196	3,0	4,7
Chile	9	27	818	696	138	5,9	6,4
Colombia	13	44	644	271	106	6,1	6,8
Costa Rica	12	77	1279	463	121	10,5	11,5
Dom. Rep.	9	72	1072	296	95	11,3	12,1
Ecuador	14	65	979	257	88	11,1	12,0
Guatemala	13	30	1271	203	105	12,1	12,7
Honduras	13	44	970	133	66	14,7	15,4
Mexico	8	27	1184	695	209	5,7	6,1
Nicaragua	6	39	1290	82	37	34,9	35,4
Panama	7	19	1317	459	294	4,5	4,8
Peru	10	72	1121	288	172	6,5	7,4
Uruguay	10	43	2820	532	272	10,4	11,0
Venezuela	16	141	1859	610	205	9,1	10,6
Average LA	12	58	1238	380	144	11,1	11,9
United States	6	6	368	3837			

Table 2. Number of procedures, duration and cost to start a firm in Latin America (columns 1-3). Official and informal per capita income figures (columns 4-5). Ratio between the cost to start a firm and informal monthly GNP (column 6) and ratio between a total cost measure, incorporating time costs, and informal monthly GNP (column 7). The sources are Schneider (2002) and World Bank (2009a, 2009b).

To finish this section, three examples on capital stock levels and profits from small (typically informal) firms are given.

In a representative sample of 3700 firms with five employees or less in Mexico, McKenzie and Woodruff (2006) report that the median capital stock replacement value across industries is USD 963. In Mexico, typical capital stocks are thus worth less than the costs of going through the registration procedure, from table 2. In the same Mexican data set, the average reported monthly earnings for firms with less than the median capital stock are USD 172 (Woodruff, 2006). In another study, from Sri Lanka, De Mel, McKenzie and Woodruff (2008) report that the median level of invested capital for 408 firms is around USD 180.[12]

[12] In the latter study, firms with less than 1000 USD in capital stock were targeted, which caused

As an example of informal firm profits, the Brazilian study of informal 1-5 person firms cited in the introduction reports that roughly 75% of the firms say that they make profits. The average monthly profit of these profit-making firms was USD 314. The profit for firms with remunerated employees was USD 825. For own account firms/workers, that may or may not have non-remunerated employees, the profits were USD 235 (SEBRAE, 2005). These entrepreneurial activities are often the main or the sole activity of the individuals involved, indicating a small room for anything but consumption expenses.[13]

4 The model

In this section, a dynamic model of firm investment, growth and possible formalization is introduced and solved. The firm starts out as informal and the question is if, when and at what firm size the firm will become formal. The modeling is inspired by the framework in Harstad and Svensson (2009).

The production function is simple: production is linear in the capital stock (k_t). As informal, the firm produces $A^i k_t$, if it has formalized, production is instead $A^f k_t$, where $A^f > A^i$. Thus, it is assumed that formality is desirable for the firm. I first solve a dynamic profit maximization problem in sections 4.1 to 4.5. Because the focus in section 4 is on one individual firm, heterogeneity between firms is not introduced until section 4.6, after which I also discuss a possible microfoundation for A^i.

The firm can grow by investing (i_t) in its capital stock. The cost of investing is convex in the size of the investment, $\frac{z}{2} i_t^2$. This gives a profit flow (π_t) in the case of the firm being informal, as follows:

$$\pi_t = A^i k_t - \frac{z}{2} i_t^2 \qquad (1)$$

The capital stock depreciates at the rate δ. The growth of the capital stock is therefore

around 6% of the originally selected sample of entrepreneurs to be dropped. The authors argue that "we believe the resulting sample is representative of a substantial majority of the own-account workers in Sri Lanka" (page 1335, footnote 4).

[13] An average exchange rate of 2.86 Reais/USD in October 2003 was used to calculate these numbers.

$$\dot{k}_t = i_t - \delta k_t. \tag{2}$$

To get access to the higher productivity, A^f, the firm must pay a formalization fee F at some time T. After formalization, flow profits equal $A^f k_t - \frac{z}{2} i_t^2$. The firm discounts future profits at the rate ρ.

The basic dynamic problem, in an environment with no restrictions on how the firm can finance investment and formalization costs from its own lifetime revenue, is stated below. The effect of different credit restrictions on the problem set-up is discussed in section 6. This discussion is postponed because credit constraints turn out to affect the dynamic analysis in a way which can be handled within the main framework.

4.1 The firm profit maximization problem

An informal firm, starting with an initial capital stock of k_0, chooses an investment path, whether it should become formal and the time of formalization (T). The firm's profit maximization problem can be written as:

$$\begin{aligned} &\text{Choose } i_t,\, T \text{ to Max}\left[\int_0^T \left(A^i k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} dt + \int_T^\infty \left(A^f k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} dt - F e^{-\rho T}\right] \\ &\text{s.t. } \dot{k}_t = i_t - \delta k_t \text{ and } k(0) = k_0 \end{aligned} \tag{3}$$

The problem can be solved in two steps. First, we use the principle of optimality to solve backwards for the formal and then for the informal investment path (assuming that T exists). We also derive the investment path if T does not exist. By using the investment path assuming that formalization does take place, we then determine when the firm wants to formalize by solving for the optimal T. If such a T exists, we then know the optimal capital accumulation path. If it does not exist, the firm is informal forever.

4.2 Optimal investments

Assume that T exists. Solving backwards, the "formal problem" takes the capital stock at time T, defined as $\tilde{k}_T$, as an initial condition, and is solved for the investment path from T to ∞. We get a formal investment function i^{formal} and a continuation value V^{formal}, which is the optimal profit from T and onwards. V^{formal} will be a function of both T and $\tilde{k}_T$. The profit maximization problem is:[14]

Choose i_t to Max $\int_T^\infty \left(A^f k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} dt$
s.t. $\dot{k}_t = i_t - \delta k_t$ and $k(T) = \tilde{k}_T$

By defining the present-value Hamiltonian $H = \left(A^f k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} + \lambda_t (i_t - \delta k_t)$, where λ_t is the present value Lagrange multiplier on the capital accumulation constraint, and applying the first-order conditions $\frac{\partial H}{\partial i} = 0$, $\frac{\partial H}{\partial k} = -\frac{\partial \lambda}{\partial t}$ and the transversality condition $\text{Lim}_{t\to\infty} (\lambda_t k_t) = 0$, we get the optimal solution:

$$i^{\text{formal}} = \frac{A^f}{z(\delta+\rho)}$$

$$k_t^{\text{formal}} = \tilde{k}_T e^{-\delta(t-T)} + \frac{A^f}{z\delta(\delta+\rho)}\left(1 - e^{-\delta(t-T)}\right) \qquad k_\infty^{\text{formal}} = \frac{A^f}{z\delta(\delta+\rho)}$$

$$V^{\text{formal}}\left(T, \tilde{k}_T\right) = e^{-\rho T}\left(\frac{A^f \tilde{k}_T}{\delta+\rho} + \frac{\left(A^f\right)^2}{2z\rho(\delta+\rho)^2}\right) \tag{4}$$

The firm invests a constant amount each "period". The capital stock converges to its steady state value of $k_\infty^{\text{formal}} = \frac{A^f}{z\delta(\delta+\rho)}$, at which depreciation and investment offset each other.[15] The constant investment rate is due to the convexity of investment costs – the firm wants to spread investment over time. The investment rate increases in the productivity parameter A^f and decreases in the cost of investment z, the depreciation rate of capital δ and the rate of time preference ρ.

The informal investment path, for a given T, can, in turn, be determined by solving for the investment path that takes the firm from k_0 to $\tilde{k}_T$ and then maximize total profits with respect to $\tilde{k}_T$:

[14] $\tilde{k}_T$ is not a choice variable in the overall problem, it is only introduced as an auxiliary variable when we solve the formal and informal problems separately.

[15] In solving the problem, a non-explosive path of investment is profit-maximizing. Other investment paths, that fulfill the differential equations for i_t and k_t stemming from the first- order conditions on the Hamiltonian, can be ruled out for optimality reasons (and do not fulfill $\text{Lim}_{t\to\infty} (\lambda_t k_t) = 0$).

$$\text{Choose } i_t \text{ and } \tilde{k}_T \text{ to Max}\left[\int_0^T \left(A^i k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} dt + e^{-\rho T} V^{\text{formal}}\left(T, \tilde{k}_T\right)\right]$$
$$\text{s.t. } \dot{k}_t = i_t - \delta k_t,\ k(0) = k_0 \text{ and } k(T) = \tilde{k}_T \tag{5}$$

The investment path is derived as above, the only difference being the terminal constraint on capital (instead of a transversality condition). Having solved for the optimal informal i_t- and k_t-paths as functions of $\tilde{k}_T$, and having plugged these back into the profit function, we integrate to get the optimal value of informal profits as a function of $\tilde{k}_T$. The total profits are then differentiated with respect to $\tilde{k}_T$. The optimality condition with respect to $\tilde{k}_T$, stated below, is that the loss of informal profits from increasing $\tilde{k}_T$ should be exactly offset by a gain in formal profits:

$$\frac{d}{d\tilde{k}_T}\left(\int_0^T \left(A^i k_t\left(\tilde{k}_T\right) - \frac{z}{2}\left(i_t\left(\tilde{k}_T\right)\right)^2\right) e^{-\rho t} dt + e^{-\rho T} V^{\text{formal}}\left(T, \tilde{k}_T\right)\right) = 0 \tag{6}$$

This equation is solved for $\tilde{k}_T$. The optimal $\tilde{k}_T$ is then plugged back into the solution for i_t and k_t, which, after simplification, becomes

$$\begin{aligned} i_t^{\text{formalization}} &= \frac{A^i}{z(\delta+\rho)} + \frac{A^f - A^i}{z(\delta+\rho)} e^{(\delta+\rho)(t-T)} \\ k_t^{\text{formalization}} &= k_0 e^{-\delta t} + \frac{A^i\left(1 - e^{-\delta t}\right)}{z\delta(\delta+\rho)} + \frac{\left(A^f - A^i\right)\left(e^{(\delta+\rho)(t-T)} - e^{-(\delta+\rho)T - \delta t}\right)}{z(\delta+\rho)(2\delta+\rho)}. \end{aligned} \tag{7}$$

This investment path starts out close to $A^i / (z(\delta+\rho))$, and then increases up to the level of formal investments at T, i.e. $A^f / (z(\delta+\rho))$. Investment increases close to formalization because the marginal value of capital is high after formalization, which makes the firm willing to decrease its profits by accumulating more capital, while still being informal.

Now assume that T does not exist. The firm is then informal forever. Solving this problem is identical to solving the formality problem above, but productivity is A^i, time runs from 0 and the initial capital stock is k_0. The "ever-informal" problem is:

$$\text{Choose } i_t \text{ to Max} \int_0^\infty \left(A^i k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} dt$$
$$\text{s.t. } \dot{k}_t = i_t - \delta k_t \text{ and } k(0) = k_0$$

The solution, obtained as in the formal problem above, is:

$$i^{\text{informal}} = \frac{A^i}{z(\delta+\rho)}$$
$$k_t^{\text{informal}} = k_0 e^{-\delta t} + \frac{A^i}{z\delta(\delta+\rho)}\left(1-e^{-\delta t}\right) \qquad k_\infty^{\text{informal}} = \frac{A^i}{z\delta(\delta+\rho)} \tag{8}$$

As for the investment path once formal, the investment rate is constant and the capital stock converges to a steady-state value, $k_\infty^{\text{informal}} = \dfrac{A^i}{z\delta(\delta+\rho)}$. This capital stock is lower than if the firm had been formal, because productivity is lower.

4.3 Solving for the formalization time T

If T exists, the investment path before and after formalization is given above (expressions 7 and 4, respectively). The optimal T can be derived by recognizing that at the time of formalization, it must be that formalization is just as attractive as remaining informal. This determines the capital stock at which the firm wants to formalize which, in turn, with the capital accumulation prior to formalization $k_t^{\text{formalization}}$ given in (7), determines T. We get that formalization takes place when

$$\frac{d}{dT}\left(\int_0^T \left(A^i k_t - \frac{z}{2}i_t^2\right)e^{-\rho t}dt + \int_T^\infty \left(A^f k_t - \frac{z}{2}i_t^2\right)e^{-\rho t}dt - Fe^{-\rho T}\right) = 0. \tag{9}$$

As discussed above, the pre-formalization investment rate approaches the formal investment rate as $t \to T$. At T, these effects cancel out and the condition in (9) simplifies to $A^i k_T - A^f k_T + \rho F = 0$. The optimal capital stock at formalization, defined as k^F, becomes

$$k^F \equiv \frac{\rho F}{A^f - A^i}. \tag{10}$$

We get T by equating the optimal capital accumulation path at $t = T$, i.e. $k_T^{\text{formalization}}$ from (7), with k^F:

$$k_0 e^{-\delta T} + \frac{A^i\left(1-e^{-\delta T}\right)}{z\delta(\delta+\rho)} + \frac{\left(A^f - A^i\right)\left(1-e^{-(2\delta+\rho)T}\right)}{z(\delta+\rho)(2\delta+\rho)} = \frac{\rho F}{A^f - A^i} \tag{11}$$

This equation implicitly defines the optimal time of formalization, T.

Formalization means a promise of future higher profits. The firm that formalizes builds a higher capital stock while informal, in anticipation of such profits. Because $i_t^{\text{formalization}} > i^{\text{informal}}$, this period is thus associated with losses as compared to the "ever-informal" path. There is a certain amount of losses/additional investment that can be sustained in anticipation of formalization. This gets reflected in the amount of capital that is optimally accumulated prior to formalization, i.e. the LHS in (11).

The formalization decision also depends on at what capital stock it is optimal to pay F. The first-order condition in (9) implies that the marginal gain from formalization, which is $\left(A^f - A^i\right)$ times the capital stock, should equal the marginal loss of not delaying formalization, i.e. ρF.

It should be observed at this stage that although we have not restricted the time of payment of F in any sense, the firm *does not* want to pay the formalization fee at once. This is because it is only beneficial to pay F once a certain capital stock/firm size has been reached and getting to that point is costly due to the convexity of investment costs.

4.3.1 Existence of T

Determining under what conditions T exists completes the solution to the dynamic problem. Proposition 1 below states the full conditions for when a firm formalizes. The main idea in deriving this proposition is to let $T \to \infty$ in the LHS of expression (11), which gives an auxiliary maximum level of capital $k_\infty^{\text{formalization}} = \frac{A^i(\delta+\rho)+\delta A^f}{z\delta(\delta+\rho)(2\delta+\rho)}$ in anticipation of formalization and then to compare this capital level to the RHS in (11). Appendix 1 gives some further details.

Proposition 1: A firm that starts with a capital level k_0 less than $k_\infty^{\text{informal}} = \frac{A^i}{z\delta(\delta+\rho)}$ will become formal if and only if the formalization cost F is less than or equal to $\bar{F}$, where $\bar{F} \equiv \frac{\left(A^f - A^i\right)\left(A^i(\delta+\rho)+\delta A^f\right)}{z\delta\rho(\delta+\rho)(2\delta+\rho)}$. This threshold is increasing in A^f, decreasing in z, δ and ρ and increasing in A^i for small values of A^i, and then decreasing. For firms that start with k_0 larger than $k_\infty^{\text{informal}}$, formalization will take place if and only if $F \leq \bar{F} + G(k_0)$, where $G(k_0)$ is positive and a strictly increasing function of k_0.

The next subsection states the full solution. A second proposition is then presented, after which the basic comparative statics and the intuition of firm formalization are discussed. The discussion of the second part of proposition 1, the k_0-dependence, is postponed until section 5.

4.4 The full solution to the dynamic problem

The solution to the dynamic problem can be stated as follows: If the conditions in proposition 1 are satisfied, there exists a formalization time T which is the solution to equation 11. In this case, the firm follows the formalization investment path ($i_t^{\text{formalization}}$ from expression 7) and then switches to the formal investment path (i^{formal} from expression 4) at time T. Such an investment path is shown in figure 1. If instead proposition 1 is not satisfied, the firm follows an "informal-ever" investment path (i^{informal} from expression 8, the broken line in figure 1).

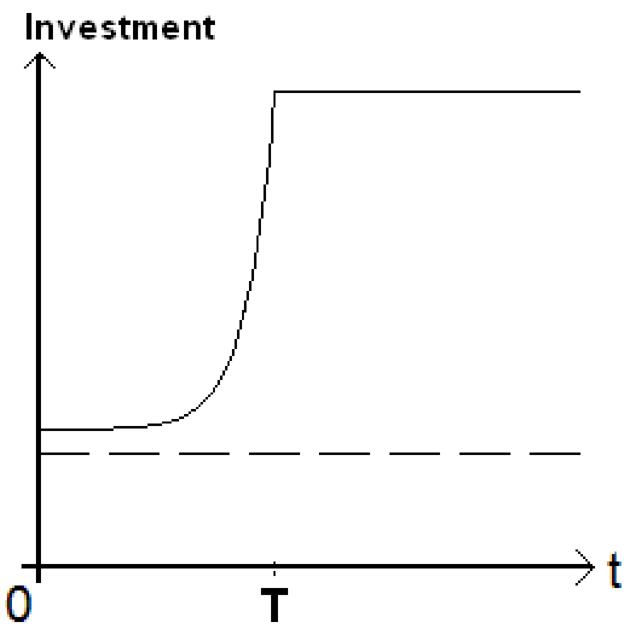

Figure 1. Investment paths.

4.5 Comparative statics of the dynamic problem

Proposition 1 was derived from expression (11). An alternative approach to the above is to use expression (11) to analyze the comparative statics of the time of formalization, T. The same parameter changes that make formalization "easier" (reflected in an increase in $\bar{F} - F$) also imply a smaller T.

Proposition 2: The formalization time T is a function of all parameters of the problem: $T\left(F,z,\delta,\rho,A^f,k_0,A^i\right)$. It is increasing in F, z, δ, and ρ and decreasing in A^f and the initial capital stock, k_0. It is decreasing in A^i for small values of A^i, then increasing.

Increases in the formalization fee F will make the necessary capital accumulation take longer time. An increased cost of investing z slows down the growth of the capital stock. Preformalization investments also decrease unambiguously in the depreciation rate δ and the discount rate ρ. In addition, an increase in ρ makes firms want to postpone formalization (the RHS in 11 increases), which makes T increase further. An increase in A^f strengthens the incentive to invest (LHS of 11). In addition, it decreases the level of capital k^F at which formalization becomes advantageous (RHS of 11). Both effects speed up formalization. The initial capital stock adds to the capital stock obtained by investing, and T is therefore smaller the higher is k_0.

With respect to the informal productivity level A^i, there are two effects: an *investment effect* and a *threshold effect*. An increase in A^i means more investment and capital accumulation (LHS in 11) but also that formalization becomes less advantageous (RHS in 11). For small values of A^i (in comparison to A^f), the investment effect dominates and formalization becomes easier ($\bar{F} - F$ increases, proposition 1) and faster (T decreases, proposition 2). For large values of A^i, the threshold effect instead dominates.

The response in T to changes in the parameter values implies that there are two effects on the investment path when a parameter changes. Consider an increase in A^f. This produces a *direct effect* by which $i_t^{\text{formalization}}$ in (7) increases, for a given T. In addition, there is an *indirect effect* through a smaller formalization time T, which further increases investment at any moment in time. In figure 1, these direct and indirect effects could be depicted as a formalization (pre-T) investment path at a higher level and with a higher slope at each point in time, a shift to the left in T, and a shift upwards in the formal (post-T) investment level.

This unambiguous multiplicative effect is also present (but goes in the other direction) for changes in z, δ, and ρ.

Before further analyzing these results, heterogeneity between firms is introduced in section 4.6 and a microfoundation for the informal productivity parameter A^i, connected to a penalty policy vis-a-vis informal firms, is provided in section 4.7.

4.6 Introducing firm heterogeneity

The discussion so far has concerned one firm. To allow a discussion in section 5 about firms that become formal versus those that do not, an assumption about firm heterogeneity is introduced. Specifically, let firms be indexed by j and assume there is a firm-specific "ability" parameter θ_j that multiplies two baseline productivity parameters, A^I and A^F. For the sake of simplicity, let θ_j be uniformly distributed on the unit interval, $0 < \theta_j \leq 1$. The baseline parameters A^I and A^F can be interpreted as the maximum productivities of the informal and formal sectors, respectively. A number of reasons why these may differ were outlined in the introduction. Through its ability parameter, each firm then has its own productivity in relation to A^I and A^F: as formal it is $A_j^f = \theta_j A^F$. The productivity of the same firm j while informal, A_j^i, contains an additional component, discussed in the next section.

4.7 The informal productivity A_j^i

As discussed in the introduction, there are potentially many different reasons for productivity differences between informality and formality. The risk of being detected and penalized by the authorities for operating "illegally" is an often-used characterization of the informal firm environment, by informal entrepreneurs themselves (Tokman, 1992) as well as in economic models (Loayza, 1995). One reason for penalizing informal firms is that these do not pay taxes; thus, the government wants firms to formalize in order to increase tax revenue. Another rationale is that formal firms put pressure on the authorities to deal with informality, claiming that competition from non-compliers is "unfair".

An explicit story for how penalties affect informal productivity, A_j^i, is through the time use of informal entrepreneurs: firms spend time "hiding" from the authorities, rather than producing. This was motivated in section 2.

Let the productivity of the informal firm be $\theta_j A^I$ from above if it can operate without hiding. Let l be the fraction of the informal entrepreneur's unitary time endowment spent trying to avoid detection, rather than in production, let $p(l)$ be the resulting probability of not getting caught and let x be the fraction of output which is taken from the informal entrepreneur if caught (the penalty/policy parameter, where $0 \leq x \leq 1$). The expected productivity when operating informally becomes $p(l)\theta_j A^I (1-l) + (1-p(l))\,\theta_j A^I (1-l)(1-x)$, which can be rewritten as[16]

$$\theta_j A^I (1-l) - x\theta_j A^I (1-l)(1-p(l)). \tag{12}$$

The first term reflects production and the second term the effect of penalties. Let the probability of not being detected be $p(l) = \sqrt{l}$. This function fulfills the natural requirements that $p(0) = 0$, $p(1) = 1$ and also $dp/dl_{l=0} = \infty$ and $d^2p/dl^2 < 0$. By solving for the optimal time allocation and detection probability from the first-order condition $-\theta_j A^I + p\'(l)x\theta_j A^I (1-l) + x\theta_j A^I (1-p(l)) = 0$, we get the informal productivity parameter.[17] It is a strictly decreasing and convex function $h(x)$ of the penalty parameter x, where $h(0) = 1$, multiplied by $\theta_j A^I$:

$$A_j^i = \theta_j A^I h(x) \tag{13}$$

The penalty parameter thus has a negative effect on productivity. Firms can shield themselves from the worst case, by allocating time to hiding rather than to production ($h(x)$ is always larger than $1-x$).[18] Figure 2 shows the resulting informal productivity, $A_j^i = \theta_j A^I h(x)$, as a function of x for an individual firm j. The formal productivity $A_j^f = \theta_j A^F$ is also shown.

The effects of penalties on formalization are discussed in section 5.

---

[16] The capital stock k is omitted because the time allocation decision to maximize expected "per period" production is static and independent of the dynamic investment decision in (3).

[17] See appendix 2.

[18] We could use a more general function for the probability; $p(l) = l^{\xi}$ with $0 < \xi < 1$. The parameter ξ would in a sense reflect the strength of enforcement of penalties x. A lower value of ξ would mean that even small amounts of time used for "hiding" are very effective in avoiding detection and could be interpreted as weak enforcement.

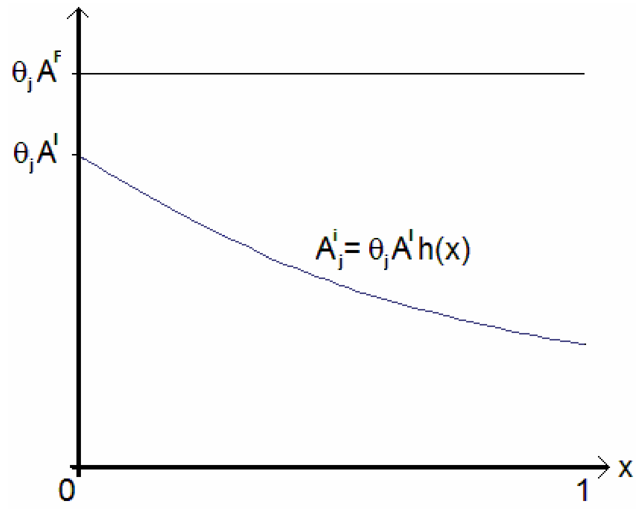

Figure 2. The informal and formal productivities of an individual firm j, as a function of penalties x.

5 Analysis of the model

What does the model imply in terms of formalization and investment? This section discusses a few predictions, starting out with a proposition about which firms that formalize, the investment paths, time of formalization and firm size at formalization.

5.1 Characteristics of firms that become formal

Expression (11) is repeated for convenience, disregarding the effect of initial capital:[19]

$$\frac{A_j^i\left(1-e^{-\delta T}\right)}{z\delta\left(\delta+\rho\right)}+\frac{\left(A_j^f-A_j^i\right)\left(1-e^{-(2\delta+\rho)T}\right)}{z\left(\delta+\rho\right)\left(2\delta+\rho\right)}=\frac{\rho F}{A_j^f-A_j^i} \tag{11'}$$

As derived in sections 4.6-4.7, the productivities are $A_j^i=\theta_j A^I h\left(x\right)$ and $A_j^f=\theta_j A^F$.

Proposition 3. Firms with an ability parameter θ_j above a threshold value $\theta^{\text{formalization}}$, i.e. firms in the range of $\theta^{\text{formalization}} \leq \theta_j < 1$ become formal. For such firms, the larger is θ_j the larger is investment, the faster is formalization and the smaller is the firm size at which formalization takes place.

[19] If not explicitly stated, I assume that the initial capital of firms is small, such that there is no k_0-dependence in whether firms formalize or not (see proposition 1).

The intuition for this proposition is straightforward: firms with high θ_j both invest more due to a higher productivity (LHS of 11 increases), and they have more to gain more from formalization (RHS of 11 decreases). In a cross section of firms, we should thus not only observe that it is high ability/productivity firms that become formal, but furthermore that their firm size at formalization is smaller and the time from firm start-up to formalization is shorter.

The threshold value $\theta^{\text{formalization}}$ is derived by plugging in the full expressions for firm productivities ($A_j^i = \theta_j A^I h(x)$ and $A_j^f = \theta_j A^F$) in the formalization criterion derived in the first part of proposition 1 and solving for θ_j, which gives

$$\theta^{\text{formalization}} \equiv \sqrt{\frac{A^F F z\delta(\delta+\rho)(2\delta+\rho)}{(A^F - A^I h(x))(A^I h(x)(\delta+\rho)+\delta A^F)}}. \tag{14}$$

5.2 Penalties

5.2.1 Policy maker

Before analyzing the effects of penalties on formalization, a highly relevant question is: Who is this policy maker? So far, the penalty policy has been connected to a somewhat diffuse "authority".

One interpretation of the penalty parameter x is that it is the government that sets (and enforces) such penalties. Then, it is assumed that the government can audit informal firms and penalize them for operating illegally. In practice, this could take place through "benevolent" tax officers, police, local authorities etc. One reason for such audits to take place may be that the government wants to increase tax revenue by making firms formal or that there is some negative externality from informal production.

An alternative view on policy, which is very different, is when there is no government in the traditional sense. Indeed, we are studying the informal sector which, by definition, consists of unregistered firms. The penalty parameter x might instead be collected by "malevolent" police, corrupt bureaucrats, local mafias etc. (De Soto, 1989; Tokman, 1992). The likely aim is then not to speed up formalization, but to maximize bribe revenue from informal firms. Although I do not provide any formal analysis of such a case, there is no reason to believe that penalties would be set as in the "benevolent" case. Instead, one can hypothesize about the effects of short time horizons of "collectors" ($\Rightarrow x \uparrow$), lack of commitment to refrain from collecting more bribes ($\Rightarrow x \uparrow$), risk of being detected if collecting too much ($\Rightarrow x \downarrow$), no desire that firms should become formal and disappear from the "tax base", and so on.

In the following section, the policy maker is the government and the optimal penalty for maximizing firm formalization is derived.

5.2.2 Effects from penalties on formalization

By analyzing $\theta^{\text{formalization}}$, we can study how the government policy parameter vis-a-vis informal firms affects formalization. Whereas $\theta^{\text{formalization}}$ increases in F, z, δ and ρ and decreases in A^F, there is an ambiguous effect with respect to $A^I h(x)$. This effect was observed in analyzing proposition 2 and is restated here:

Proposition 4. The effect of the penalty parameter on the threshold for formalization $\theta^{\text{formalization}}$ is U-shaped, first decreasing in x for small values of x and then increasing. The penalty parameter that minimizes $\theta^{\text{formalization}}$, i.e. that maximizes the amount of firms that formalize, is $x = 0$ when $A^I \leq \dfrac{A^F}{2}\dfrac{\rho}{\delta+\rho}$, in an intermediate range of A^I it is given by the x that solves $h(x) = \dfrac{A^F}{2A^I}\dfrac{\rho}{\delta+\rho}$ and it is $x = 1$ when $A^I \geq \dfrac{3\sqrt{3}A^F}{4}\dfrac{\rho}{\delta+\rho}$.

The policy maker can affect the incentive to formalize through the penalty on informal production. For small values of x, the threshold effect will dominate – formalization becomes more attractive when penalties are increased. This is seen in (11'), where the level of capital at which the firm optimally formalizes (the RHS) goes down. For large penalties, it is instead the case that the investment effect will dominate – firms will accumulate less capital and will therefore not be able to become formal. This is the LHS of 11'.

The penalty that maximizes the amount of firms that formalize is derived through the necessary condition for a minimum on $\theta^{\text{formalization}}$, i.e. $\partial\theta^{\text{formalization}}/\partial x = 0$. This condition gives $h(x) = \dfrac{A^F}{2A^I}\dfrac{\rho}{\delta+\rho}$ which, in turn, must lie between $h(0) = 1$ and $h(1) = \dfrac{2}{3\sqrt{3}}$. This gives proposition.4.

A restatement of proposition 4 is that a policy designed by a government to incentivize firms to become formal should be conducted with a "carrot and stick" approach: neither too mild nor too tough. The accumulation- and threshold effects will be balanced and the amount of firms that become formal is maximized.

As an illustration, assume that $A^F = 1$, $A^I = \frac{A^F}{2}$ and that $\delta = \rho$. Since $\frac{1}{4} < A^I < \frac{3\sqrt{3}}{8}$, we have an interior solution and we get $h(x) = \frac{1}{2}$. Using the $h(x)$-function (appendix 2), we get that $x^* \approx 0.68$ maximizes the amount of formalization. The graph below on the shape of $\theta^{\text{formalization}}$ shows that, for this particular illustration, lower penalties will result in much less formalization, whereas higher penalties do not affect the degree of formalization to the same extent.

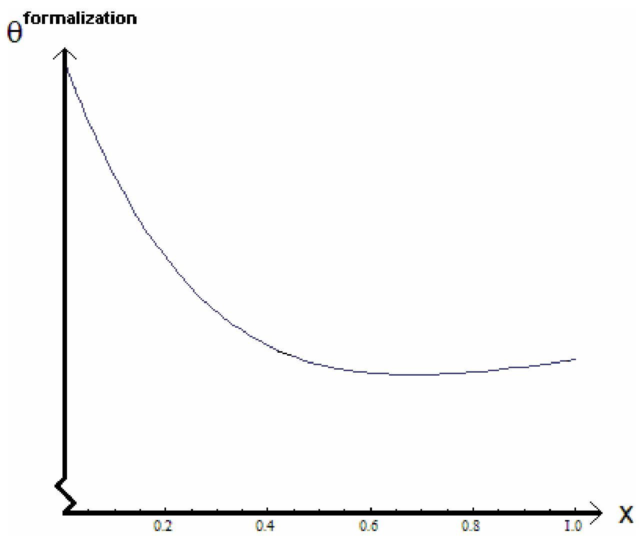

Figure 3. An illustration of the effects of penalties on the minimum ability threshold for formalization, $\theta^{\text{formalization}}$.

5.3 The formal sector productivity A^F

Although we have neither made explicit the formal sector productivity parameter A^F, nor have specified it as a policy parameter, it is worth pointing out that increases in A^F have two effects. First, investments increase (investment effect, LHS of 11). Second, the firm size at which formalization becomes beneficial goes down (threshold effect, RHS of 11). Obviously, decreasing A^F has the opposite effect.

The models by Loayza (1995) and Garcia Penalosa and Turnovsky (2005) include taxes and public goods as determinants of formal sector productivity. Higher taxes and less efficient public goods provisioning in the formal sector both act to increase informality. In the present paper, we can consider these policy parameters as potential determinants of A^F. The preceding paragraph then clarifies two channels through which investment in the informal sector and formalization is discouraged by higher taxation and less efficient public goods provisioning in the formal sector.

5.4 The aggregate of firms

What does the long-run firm size distribution predicted by the model look like? Does it resemble actual firm size distributions in developing economies?

In the long run, informal firms converge to a (firm-specific) size $k_{j,\infty}^{\text{informal}} = \dfrac{\theta_j A^I h(x)}{z\delta(\delta+\rho)}$ and formal firms to $k_{j,\infty}^{\text{formal}} = \dfrac{\theta_j A^F}{z\delta(\delta+\rho)}$. The firm with an ability parameter marginally lower than $\theta^{\text{formalization}}$ thus reaches a much smaller size than had the ability parameter been somewhat larger. The size of the firm size gap, i.e. $\dfrac{\theta^{\text{formalization}} A^F}{z\delta(\delta+\rho)} - \dfrac{\theta^{\text{formalization}} A^I h(x)}{z\delta(\delta+\rho)}$, is increasing in F, ρ, A^F and decreasing in $A^I h(x)$, z and δ. Together with the fact that $\theta^{\text{formalization}}$ increases in F and that $h(x)$ decreases in x, we can state the following proposition:

Proposition 5. In the long run, the model displays a low-end range, $0 < \theta_j \leq \theta^{\text{formalization}}$ of small informal firms and a high-end range, $\theta^{\text{formalization}} \leq \theta_j \leq 1$, of large formal firms. There is a "missing middle" in firm sizes, and the size of the gap is increasing in formalization costs F and penalties x.

Tybout (2000) documents firm size distributions for a number of developing economies and finds evidence of a "dual structure", with a large proportion of very small firms, a "missing middle" and then a few large firms. This contrasts with typical high-income countries. The author further argues that "small producers frequently operate partly or wholly outside the realm of government regulation" (Tybout 2000, page 15), discussing costs of dealing with the government as one explanation for the observed pattern. The present model shows how the profit maximization behavior of firms, with large costs of entry into formality, can generate such a "missing middle". It also delivers predictions about the size of this firm size gap. Figure 4 shows the long-run firm sizes as predicted by the model.

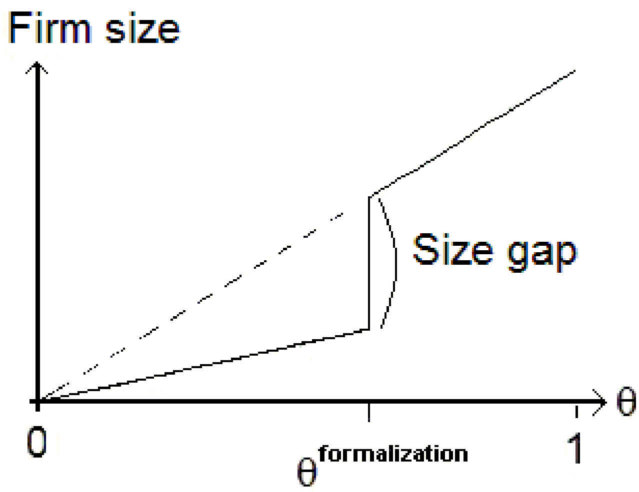

Figure 4. Long-term firm size distribution.

5.5 The dependence on initial capital, k_0

Heterogeneity between firms was introduced along an ability/productivity dimension, resulting in predictions about which firms that formalize. The focus of the poverty trap literature is instead on studying how the initial wealth distribution matters for the future wealth distribution. Translated into the present paper, the question is if differences in initial capital k_0 can explain differences in long-run firm sizes[20] . At face value, the answer to this question should be no: there are no explicit credit constraints in the model; therefore equally productive firms should converge to the same steady state in the long run (Banerjee, 2001; McKenzie and Woodruff, 2006).

Proposition 1 states that initial capital does play a role, however. Whenever k_0 is larger than $k_\infty^{\text{informal}}$, the maximum level of capital that can optimally be accumulated is larger than $k_\infty^{\text{formalization}} = \dfrac{A^i(\delta+\rho)+\delta A^f}{z\delta(\delta+\rho)(2\delta+\rho)}$. The maximum level of capital is also increasing in k_0. These statements were proven in appendix 1.

[20] In this section, I disregard the ability differences and assume all firms to be equal except for differences in k_0.

A high initial capital stock gives a firm an "initial cost advantage" in reaching a certain size. That is, it is (initially) less costly for a firm starting with a high k_0 to reach a certain (larger) capital level than if the firm had started with a low k_0. This effect vanishes over time because the initial capital depreciates at a rate δ. The way in which a firm starting with high k_0 takes advantage of the effect is to invest heavily in the beginning and then formalize early (at capital level k^F from 10). Early formalization makes the additional investment worthwhile because formalization profits come closer in time. A firm starting with a lower k_0 has no possibility of taking advantage of the effect of initial capital since reaching k^F would mean even higher initial investment losses and/or later formalization, thus implying a lower net present value of formality.

An alternative explanation to the fact that the cost advantage due to a high k_0 is only initial is the following: A firm that is informal forever converges to a capital stock $k_\infty^{\text{informal}}$ at which investments just compensate depreciated capital. For a firm that formalizes at some time T far into the future, it is costly to deviate to any large extent from $k_\infty^{\text{informal}}$ prior to formalization. Therefore, for a T that is large, investments will mimic the "ever-informal" case where depreciated capital is just replaced, as long as T is not close in time. This, in turn, means that the effect of the initial capital stock will vanish prior to T and, as a result, not affect the decision whether to formalize or not.

The result that k_0 is of importance resembles the poverty trap literature. A certain formalization level k^F that would be prohibitive for firms with k_0 smaller than $k_\infty^{\text{informal}}$ will still allow for a range of firms with sufficiently high initial capital to formalize and converge to k_∞^{formal}.

6 Extensions

This section discusses some modifications to the model in section 4. I abstract from firm heterogeneity in the presentation below to save on notation.

6.1 The effect of borrowing and savings constraints

So far, nothing hinders the firm from making losses while informal. These losses can come from three sources: early in the process due to investment costs exceeding production revenue, due to higher investment costs in anticipation of formalization, as well as in the instant when F is paid. In the formulation in (3), the only concern is that the net present value of revenue exceeds the net present value of costs.

It is beyond the scope of this paper to review the literature on credit.[21] Instead, the introduction of a no-borrowing constraint in this model is, somewhat innocently, motivated by the observation that most informal firms do not use any credit at all. In the Brazilian representative sample of small informal firms, 6% of the firms had used credit during the three months prior to the study and 17% of the firms had any debt outstanding whatsoever (SEBRAE, 2005). These numbers seem to be quite typical. In a study of six African countries, 2% of the micro firms (1-5 employees) and 7% of the small firms (6-25 employees) had received a loan in the year prior to the study. 16% of each group had debt in the informal finance sector (Bigsten et al., 2003).[22] Levy (1993) reports small percentages of Tanzanian and Sri Lankan micro firms (1-5 employees) with formal credit.

In two recent empirical papers estimating returns to capital, McKenzie and Woodruff (2006) and de Mel, McKenzie and Woodruff (2008) find evidence on binding credit constraints for small informal firms in Mexico and Sri Lanka.

In addition to borrowing constraints, savings constraints may be important in developing countries (Duflo and Banerjee, 2007; Dupas and Robinson, 2009). In the

[21] Banerjee (2001) reviews a number of "stylized facts" about credit markets in developing countries.

[22] As pointed out by the authors, not having credit does not mean that firms do not want it and 64% of the micro firms had either been denied credit or say they would be denied if they applied. It should also be pointed out that several studies report trade credit to be an important source of financing for small firms.

model presented here, this would mean that informal firms cannot save from current profits to pay the formalization fee. The impact of such a constraint is discussed below.

6.1.1 No borrowing to finance investment

First, assume that the firm cannot borrow to finance investment. Because the firm would like to invest at least an amount i^{informal} from the start, the no-borrowing constraint will clearly be binding for firms that start with a small initial capital stock (imagine k_0 close to zero, this gives a profit flow in the unconstrained case of $A^i k_0 - \frac{z}{2}\left(i^{\text{informal}}\right)^2 < 0$, and hence a need to borrow). Formally, the following constraint should be added to the problem formulation in (3):

$$A^i k_t - \frac{z}{2} i_t^2 \geq 0 \tag{15}$$

Adding this constraint will not affect the formalization decision. Because investment in the capital stock is perfectly divisible and the return on small investment amounts is very high, it will be optimal for the firm to forsake small current profits against future higher profits generated by a larger capital stock. The firm will invest as much as possible, i.e. $i_t = \sqrt{2A^i k_t / z}$, until the constraint in (15) is no longer binding. The firm "bootstraps" its way out of the borrowing constraint. The time it takes depends on k_0, but the long-run capital stock and the formalization decision will not be affected.

6.1.2 Limited borrowing to finance the formalization cost F

The introduction of a constraint with respect to the financing of F is more complicated.[23] A shortcut to studying the impact of a borrowing constraint in financing F is as follows: Assume that a fraction γ of formalization costs can be financed as

[23] The model solved in this paper assumes that there is no other asset than the firm's own capital stock in which to invest: no financial saving can be accumulated. Given the evidence on little access to savings devices in the informal sector, this assumption is not implausible. Introducing a no-borrowing constraint to finance F then implies that firms would have to save "in the mattress", i.e. at zero % interest rate, in anticipation of formalization. Introducing a savings control variable (s_t), we get that with zero interest, total savings before formalization should at least equal F, i.e.

before, i.e. by the firm's own financing. A fraction $(1-\gamma)$ of the formalization cost can be financed by borrowing, at the instant of formalization, at an interest rate of r. This results in a "per-period" interest payment of $(1-\gamma)\,rF$, from T and onwards. Assuming that the amount borrowed is rolled over indefinitely, the only change to the problem is in the total profit expression in (3), which changes to become

$$\int_0^T \left(A^i k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} dt + \int_T^\infty \left(A^f k_t - \frac{z}{2} i_t^2 - (1-\gamma)\,rF\right) e^{-\rho t} dt - \gamma F e^{-\rho T}.$$

The solution is only affected through condition (9), when solving for the optimal level at which to formalize. The formalization criterion becomes

$$k^F = \frac{\gamma\rho F + (1-\gamma)\,rF}{A^f - A^i}. \tag{10'}$$

If the firm borrows at a high interest rate r $(>\rho)$, the more difficult formalization will become. The smaller is γ, the less likely is formalization.

Expression 10' gives a tractable expression through which to analyze credit constraints. The entire analysis in section 5 remains unchanged, and we can consider the credit constraint as being represented by a larger value on the formalization cost F.

The representation above is admittedly a simplified way of introducing a credit constraint. In relation to the literature on poverty traps, one might imagine that the interest rate r at which the firm can borrow depends on initial capital/initial establishments, such that firms that started off at higher capital stocks represent less of a moral hazard risk to lenders. Such a specification would effectively mean that firms with a larger initial capital stock face lower effective formalization costs, and act as a source of a poverty trap.[24]

$\int_0^T s dt \geq F$. Solving the model with this constraint turns out to be very complicated. However, it is likely that the period leading up to formalization would imply a trade-off between investment and "saving in the mattress", implying less capital accumulated and, as result, a smaller possibility for formalization.

[24] That is, the combination of the non-convexity (F) and higher interest rates for firms with less initial capital could, in itself, generate a long-run distribution of capital that depends on k_0.

6.2 An alternative view on penalties

It is often argued that informal firms run a larger risk of detection if they grow, and therefore prefer to stay small. The model can be modified to investigate this argument. Instead of firms spending time on hiding, we now modify the original informal production function. By writing informal production as $A^i k_t \left(1 - \frac{\beta}{2} k_t\right)$, rather than $A^i k_t$, we explicitly recognize that as the firm grows, a larger fraction of output will be captured by the authorities (we think of penalties x as incorporated into the β-term). Although it seems likely that informal firms will now grow (even) less, the incentive to formalize is also stronger than before.

What does this modification to the problem in (3), stated in appendix 3, yield? Although the solution is somewhat complex, parametrizing and comparing it to the model in section 4 gives at hand that for small values of A^i (in comparison to A^f), the investment effect is dominating. That is, informal firms now simply cannot grow and will formalize to a lesser extent. However, for values of A^i close to A^f, the opposite becomes true. If informal firms are relatively productive in the original problem, the threshold effect dominates (proposition 2) and firms have little to gain from becoming formal. This effect is reversed with the new specification. That is, growth implies higher penalties and that the incentive to become formal is strengthened. For large values of A^i, the firm is then able to escape such penalties by formalizing.

7 Discussion

This paper sets up a dynamic model of profit maximizing informal firms to study investment, growth and possible entry into formality, in the face of a large formalization cost that has to be paid at one instant in time. In the model, there is a basic dynamic trade-off: On the one hand, there is an incentive to invest and grow to be able to reap the benefits from formalization. On the other hand, this may require too much investment at the early stages and prove too costly, and the firm may therefore choose to remain informal.

The model generates a number of predictions. When aggregating over firms that differ along an ability dimension, the long-run firm size distribution exhibits a range of small firms and a range of larger firms but also a "missing middle", much in line with actual firm size distributions observed in developing countries (Bigsten et al. 2004, Tybout, 2000).

The model is also broadly consistent with recent empirical evidence from McKenzie and Woodruff (2006). Using a representative small-firm sample from Mexico, these authors find no evidence of barriers to growth for the smallest entrepreneurial activities. However, McKenzie and Woodruff (2006) do find some support for non-convexities which are at place and act as barriers to growth for larger firms. In line with their evidence, the present model generates a firm distribution where there is a range of small firms, growing up to a certain size but remaining informal and unable to grow further due to the formalization cost. Some firms however, by investing a lot at early stages, manage to surpass this barrier to growth and can then grow further due to a higher productivity once formal.

In addition to predictions related to long-run firm sizes, the model also predicts that the firm size at formalization, as well as the timing of formalization, is firm-specific. In particular, firms run by more able entrepreneurs formalize at smaller firm sizes and earlier than do firms run by less able entrepreneurs.

With respect to policy, the paper offers two views. If the policy maker is a benevolent government, a policy designed to incentivize firms to become formal should be conducted with a "carrot and stick" approach: neither too mild nor too tough. Such a policy will make an accumulation effect and a threshold effect balance and maximize the amount of firms that become formal. The second interpretation of policy is one in which the government is absent and where the policy maker is rather "malevolent" police, corrupt bureaucrats, local mafias etc. With the latter view on policy, growth and possible formalization seem less likely to occur.

Finally, an interesting theoretical result in this paper is that an "informality trap" can result in a model with only a non-convexity but without a credit constraint. This contrasts with standard neoclassical theory, where non-convexities alone should not affect long-run allocations. Economic agents could simply borrow to overcome such hurdles (Banerjee, 2001; McKenzie and Woodruff, 2006). In the present model, it is the combination of adjustment costs of investment and the formalization cost that makes initial capital matter for long-run capital distributions.

Bibliography

[1] Antunes, A., de V. Cavalcanti, T, 2006. Start-up costs, limited enforcement and the hidden economy. European Economic Review 51, 203-224.

[2] Banerjee, A., 2001. Contracting constraints, credit markets and economic development. MIT Dept. of Economics Working Paper, 02-17.

[3] Banerjee, A., Duflo, E., 2007. The Economic Lives of the Poor. Journal of Economic Perspectives 21, 141-167.

[4] Banerjee, A., Newman, A., 1993. Occupational choice and the process of development. Journal of Political Economy 101, 274-298.

[5] Bigsten, A., Collier, P., Dercon, S., Fafchamps, M., Gauthier, B., Gunning, J., Oduro, A., Oostendorp, R., Pattillo, C., Söderbom, M., Teal, F., Zeufack. A., 2003. Credit Constraints in Manufacturing Enterprises in Africa. Journal of African Economies 12, 104–25.

[6] Bigsten, A., Kimuyu, P., Lundvall, K., 2004. What to Do with the Informal Sector? Development Policy Review 22, 701-715.

[7] Buera, F., 2008. A Dynamic Model of Entrepreneurship with Borrowing Constraints: Theory and Evidence. Manuscript, Dept. of Economics, Northwestern University.

[8] Chen, M., 2007. Rethinking the Informal Economy: Linkages with the Formal Economy and the Formal Regulatory Environment, in "Unlocking Human Potential: Concepts and Policies for Linking the Informal and Formal Sectors" (Guha-Khasnobis, Kanbur and London eds). Oxford University Press, 75-92.

[9] de Mel, S., McKenzie, D., Woodruff, C., 2008A. Who are the Microenterprise Owners?: Evidence from Sri Lanka on Tokman v. de Soto. World Bank: Policy Research Working Paper #4635.

[10] de Mel, S., McKenzie, D., Woodruff, C., 2008B. Returns to Capital in Microenterprises: Evidence from a Field Experiment. The Quarterly Journal of Economics 123,1329–1372.

[11] de Soto, H., 1989. The Other Path: The Economic Answer to Terrorism. New York: Harper Collins.

[12] Djankov, S., La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2002. The regulation of entry. The Quarterly Journal of Economics 117, 1-37.

[13] Dupas, P., Robinson, J., 2009. Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya. Manuscript, University of California.

[14] Galor, O., Zeira, J., 1993. Income Distribution and Macroeconomics. Review of Economic Studies 60, 35-52.

[15] García Penalosa, C., Turnovsky, S., 2005. Second-best optimal taxation of capital and labor in a developing economy. Journal of Public Economics 89, 1045-1074.

[16] Ghatak, M., Jiang, N., 2002. A Simple Model of Inequality, Occupational Choice and Development. Journal of Development Economics 69, 205-26.

[17] Harris, J., Todaro, M., 1970. Migration, Unemployment, and Development: A Two Sector Analysis. American Economic Review 60, 126–142.

[18] Harstad, B., Svensson, J., 2009. From Corruption and Lobbying to Economic Growth. (Revision of) CEPR Discussion Paper 5759.

[19] Hart, K., 1973. Informal Income Opportunities and Urban Employment in Ghana. Journal of Modern African Studies 11, 61-89.

[20] Hopenhayn, H., 1992.Entry, Exit, and Firm Dynamics in Long Run Equilibrium. Econometrica 60, 1127-1150.

[21] ILO., 1972. Employment, Incomes and Equality: A Strategy for Increasing Productive Employment in Kenya. Geneva: International Labour Office.

[22] INEGI - Instituto Nacional de Estadística, Geografía, e Informática, 2003. Encuesta nacional de micronegocios 2003.

[23] Jovanovic, B., 1982. Selection and the Evolution of Industry. Econometrica 50, 649-670.

[24] Levenson, A., Maloney, W., 1998. The Informal Sector, Firm Dynamics and Institutional Participation. World Bank: Policy Research Working Paper #1988.

[25] Levy, B., 1993. Obstacles to Developing Indigenous Small and Medium Enterprises: An Empirical Assessment. The World Bank Economic Review 7, 65-83.

[26] Lewis, W., 1954. Economic Development with Unlimited Supplies of Labour. Manchester School 22, 139-191.

[27] Loayza, N., 1995. The economics of the informal sector: a simple model and some empirical evidence from Latin America. Carnegie-Rochester Conference Series on Public Policy 45, 129-162.

[28] Lucas, R., 1978. On the Size Distribution of Firms. Bell Journal of Economics 9, 508-523.

[29] Maloney, W., 2004. Informality Revisited. World Development 32, 1-20.

[30] McKenzie, D., Woodruff, C., 2006. Do Entry Costs Provide an Empirical Basis for Poverty Traps? Evidence from Mexican Microenterprises. Economic Development and Cultural Change 55, 3-42.

[31] Melitz, M., 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Econometrica 71, 1695-1725.

[32] Rauch, J., 1991. Modelling the informal sector informally. Journal of Development Economics 35, 33-47.

[33] Schneider, F., 2002. Size and measurement of the informal economy in 110 countries. Workshop on Australian National tax centre.

[34] SEBRAE, 2005. Economia Informal Urbana 2003.

[35] Todaro, M., 1969. A Model of Labor Migration and Urban Unemployment in Less Developed Countries. American Economic Review 59, 138-48.

[36] Tokman, V., 1992. Beyond Regulation: The Informal Economy in Latin America. Boulder, CO: Lynne Rienner Publishers.

[37] Tybout, J., 2000. Manufacturing Firms in Developing Countries: How Well Do They Do, and Why? Journal of Economic Literature 38, 11-44.

[38] Woodruff, C., 2006. Self-employment: Engine of Growth or Self-help Safety Net? World Bank Studies in Development series.

[39] World Bank, 2009a. http://www.doingbusiness.org

[40] World Bank, 2009b. http://web.worldbank.org/wbsite/external/datastatistics

Appendix

A1: Proof of proposition 1

Define the auxiliary capital level $k_{\infty}^{\text{formalization}} = \frac{A^i(\delta+\rho)+\delta A^f}{z\delta(\delta+\rho)(2\delta+\rho)}$. This is the (hypothetical) level of capital that a firm would reach at $T=\infty$ if it followed the formalization investment path forever. Using this expression, we can write the capital stock a firm reaches at the time of formalization, which is the LHS of expression (11), as

$$k_{t=T}^{\text{formalization}} =$$
$$k_{\infty}^{\text{formalization}} + \left(k_0 - k_{\infty}^{\text{informal}}\right)e^{-\delta T} + \left(k_{\infty}^{\text{informal}} - k_{\infty}^{\text{formal}}\right)\frac{\delta e^{-(2\delta+\rho)T}}{2\delta+\rho}.$$

As long as $k_0 \leq k_{\infty}^{\text{informal}}$, $k_{t=T}^{\text{formalization}}$ is increasing in T and converges to $k_{\infty}^{\text{formalization}}$ as $T\to\infty$. By equating $k_{\infty}^{\text{formalization}}$ with k^F and solving for F, the formalization criterion in proposition 1 is obtained.

For $k_0 > k_{\infty}^{\text{informal}}$, the capital stock $k_{t=T}^{\text{formalization}}$ reaches a maximum value of $k_{\infty}^{\text{formalization}} + \frac{\delta+\rho}{2\delta+\rho}\frac{\left(k_0-k_{\infty}^{\text{informal}}\right)^{(2\delta+\rho)/(\delta+\rho)}}{\left(k_{\infty}^{\text{formal}}-k_{\infty}^{\text{informal}}\right)^{\delta/(\delta+\rho)}}$ at $T=\frac{1}{\delta+\rho}\ln\left(\frac{k_{\infty}^{\text{formal}}-k_{\infty}^{\text{informal}}}{k_0-k_{\infty}^{\text{informal}}}\right)$.

The capital stock that can be obtained is thus larger than $k_{\infty}^{\text{formalization}}$ and depends on the initial stock of capital. As a result, the firm can face a higher formalization fee and still optimally choose to formalize. Comparing this capital stock with k^F and solving for F, the second part of the proposition is obtained (with $G(k_0) \equiv \left(\frac{A^f-A^i}{\rho}\right)\frac{\delta+\rho}{2\delta+\rho}\frac{\left(k_0-k_{\infty}^{\text{informal}}\right)^{(2\delta+\rho)/(\delta+\rho)}}{\left(k_{\infty}^{\text{formal}}-k_{\infty}^{\text{informal}}\right)^{\delta/(\delta+\rho)}}$).

A2: Solving for the time allocated to hiding

The first-order condition $-\theta_j A^I + p'(l)x\theta_j A^I(1-l) + x\theta_j A^I(1-p(l)) = 0$ gives the following solution:

$$l^* = \frac{2-4x+5x^2-2\sqrt{(x-1)^2(1-2x+4x^2)}}{9x^2}$$
$$p^* = \sqrt{l^*}$$

$A_j^i = \theta_j A^I h\left(x\right)$ where $h\left(x\right) = 1 - l^* - x\left(1 - l^*\right)\left(1 - p^*\right).$

Except for $h\left(0\right) = 1$ and $h\left(1\right) = \dfrac{2}{3\sqrt{3}}$, we also have that $h'\left(x\right) < 0$ and $h''\left(x\right) > 0$

7.1 A3: Penalties that increase in k_t

Choose i_t, T to

$$\text{Max}\left[\int_0^T \left(A^i k_t \left(1 - \frac{\beta}{2} k_t\right) - \frac{z}{2} i_t^2\right) e^{-\rho t} dt + \int_T^\infty \left(A^f k_t - \frac{z}{2} i_t^2\right) e^{-\rho t} dt - F e^{-\rho T}\right]$$

s.t. $\dot{k}_t = i_t - \delta k_t$ and $k\left(0\right) = k_0$

The optimal investment path, prior to formalization, becomes:

$$i_t = \delta \frac{C_2}{C_1} + C_4\left(\delta + r_1\right) e^{r_1 t} + C_5\left(\delta + r_2\right) e^{r_2 t}$$
$$k_t = \frac{C_2}{C_1} + C_4 e^{r_1 t} + C_5 e^{r_2 t}$$

where

$$C_1 = \frac{\beta}{z} + \delta^2 + \delta\rho,\ C_2 = \frac{A_i}{z}$$
$$r_1 = \frac{\rho}{2} + \sqrt{\left(\frac{\rho}{2}\right)^2 + C_1},\ r_2 = \frac{\rho}{2} - \sqrt{\left(\frac{\rho}{2}\right)^2 + C_1}$$
$$C_3 = \left(zC_1\left(\left(\delta + r_1\right) e^{r_1 T} - \left(\delta + r_2\right) e^{r_2 T}\right)\right)^{-1}$$
$$C_4 = C_3\left(C_1 F - \delta C_2 z + e^{r_2 T}\left(C_2 - C_1 k_0\right)\left(\delta + r_2\right) z\right)$$
$$C_5 = C_3\left(-C_1 F + \delta C_2 z - e^{r_1 T}\left(C_2 - C_1 k_0\right)\left(\delta + r_1\right) z\right)$$

After formalization, the expressions in (4) apply.

Chapter 4

Compositional and dynamic Laffer effects in a model with constant returns to scale*

1 Introduction

There is a renewed interest in the dynamic effects of tax cuts. Methods for not only including "micro" behavioral effects but also dynamic "macro" effects of tax cuts in the US budget process are being discussed (Auerbach, 2006). In a recent paper, Mankiw and Weinzierl provide "back of the envelope" calculations comparing static and dynamic "scoring" for the neoclassical growth model. They argue that tax cuts can, through a new higher steady state level of capital and therefore a larger tax base, to a large extent pay for themselves (Mankiw and Weinzierl, 2006). Leeper and Yang (2008) in turn show that the specific financing scheme for such tax cuts matters.

This paper follows a different literature than the two papers above and studies effects from tax cuts in a model with constant returns to scale in broad capital. These models are different from the neoclassical growth model studied by Mankiw

* I thank Jonas Agell at the Department of Economics at Stockholm University for advice and support. In addition, thanks to Andreas Müller, John Hassler, Lars Johansson, Martin Flodén, Mats Persson, Timothy Kehoe, Victor Rios-Rull, Fabrizio Zilibotti and participants in seminars at the Department of Economics at Stockholm University, the IIES Macro Study Group and the 10th workshop on Dynamic Macroeconomics in Vigo, Spain. Finally, thanks to Christina Lönnblad for editorial assistance. Any remaining errors are mine.

and Weinzierl (2006) and Leeper and Yang (2008). Since they display "endogenous" long-run growth, the scope for dynamic effects is different.[1] Leeper and Yang (2008) state the issue in their section 2, page 3: "some doubt remains about whether a deficit-financed tax cut can actually be self-financing". The present paper sheds light on this question by carefully defining what "self-financing" means and conditions are derived for when such self-financing is possible.

I develop a tractable framework introducing human capital and a labor/leisure choice in the AK-model to make three main points. First, I further define "Laffer effects" in the constant returns models by dividing effects of tax cuts into dynamic and compositional effects. This is crucial when there is more than one factor of production. Second, simple analytical expressions for when tax cuts in AK-style models will fully finance themselves are provided. Third, I follow both the endogenous growth literature and Mankiw and Weinzierl and add a labor/leisure choice and study how the scope for self-financing tax cuts changes.

Having added leisure to the model, we have a framework with three incentive margins that, as a result of tax cuts, can create Laffer effects on their own or in combination. The three incentive margins are 1) dynamic effects of taxes on interest and growth rates, 2) compositional effects of taxes on production (an "uneven playing field") and 3) the labor/leisure choice. In a world with the first – dynamic – effect only, there is a direct revenue effect of a tax cut and an indirect effect of different interest and growth rates. The second – compositional – effect comes in when physical and human capital are taxed differently; the current tax base is then also affected by tax cuts, adding to the direct revenue effect and the growth effect. Adding the third margin – leisure – there is an additional effect on the tax base through a different labor/leisure choice after a tax cut and there is also an additional growth effect.

In this setup, I derive what combinations of tax rates on physical and human capital are required for a tax cut to be self-financing. The results suggest that dynamic and compositional distortions will need to be large if there are to be Laffer effects; less so, however, if the model contains a labor/leisure choice. I show that

[1] There are thus, broadly speaking, two strands of literature: a neoclassical growth literature and an "endogenous" growth literature. As the neoclassical and endogenous growth models have different long-run properties, the analysis of dynamic effects of tax cuts differ.

the margin opened up by the endogenous labor/leisure choice may be quantitatively important.

Regarding terminology and main scope, this paper follows the tradition of the "endogenous" growth literature and studies "Laffer effects" rather than "dynamic scoring". This means that we are interested in when tax cuts can fully finance themselves, maintaining government spending. I derive conditions for what starting point of tax rates is required for tax cuts to be fully self-financing. As shown by Agell and Persson (2001) and as further detailed here, "maintaining government spending" must be accurately defined and several cases arise. Specifically, I add one definition of Laffer effects to the definitions provided by these authors.

As the direct revenue effects of tax cuts are negative, government bonds function as the means of intertemporal financing. In the long run, government bonds must obey a transversality condition. This analysis differs from the analysis of the neoclassical model by Mankiw and Weinzierl (2006) where an atemporal government budget constraint is always obeyed through a reduction in transfers when distortionary taxes are reduced. Their study of "scoring" is therefore different from the study of Laffer effects. Leeper and Yang (2008) add government bonds obeying a transversality condition to the model and study different ways of financing reductions in taxes in the neoclassical model: future tax increases or lower government consumption/transfers. Differently from both these papers I use an endogenous growth setting. With permanent tax cuts I study if government spending (transfers and consumption) can be maintained.[2]

Much of the earlier literature on taxation in "endogenous" growth models has focused on growth effects of taxation in CRTS two-sector models with physical and human capital, e.g. Lucas (1988), King and Rebelo (1990), Rebelo (1991), Pecorino (1993, 1994), Stokey and Rebelo (1995) and Milesi-Ferretti and Roubini (1998a, 1998b). A key aspect of all these papers, as well as of the few studies of Laffer effects, is that in almost all specifications, the return to capital and the growth rate are affected by tax cuts. Milesi-Ferretti and Roubini (1998a, 1998b) clarify the role

[2] There is also a literature, related to both scoring and Laffer effects, comparing the level of present value government revenue along balanced growth paths for different sets of tax rates in calibrated endogenous growth models (Pecorino, 1995; Bianconi, 2000).

that different model assumptions have on growth responses from taxation for these two-sector models.[3]

Ireland (1994) and Bruce and Turnovsky (1999) study dynamic Laffer effects in the one-sector AK-model and Novales and Ruiz (2002) in a two-sector model. Agell and Persson (2001) clarify the role of different assumptions regarding "maintaining government spending" in explaining why Ireland and Bruce and Turnovsky get seemingly different results. The fact that government spending grows with the "old" growth rate even after tax cuts in both Ireland (1994) and Novales and Ruiz (2002) makes it less difficult to obtain dynamic Laffer effects.

This paper extends the study of Laffer effects from the one-sector AK models towards the two-sector models. For this purpose, I add human capital and a leisure choice to the one-sector AK-framework. For analytical tractability, I first add human capital and work out the effects and then, in a later section, add the leisure decision to the model. The framework has the considerable advantage of there being no transitional dynamics. The economy "jumps" from one growth path to another as a result of tax cuts, thereby facilitating the analysis of Laffer effects.[4,5]

Section 2 outlines the basic model. Laffer effects are defined in section 3. The conditions for Laffer effects are derived and discussed in section 4. In the model description until section 4, the first two effects of taxation, the dynamic effect and the compositional effect from above, are present in the analysis. Section 5 introduces the third effect of taxation by endogenizing the leisure decision and shows how this additional incentive margin affects the scope for Laffer effects. Section 6 summarizes

[3] Milesi-Ferretti and Roubini (1998a, 1998b) study the balanced growth path responses to taxation in a full catalogue of models that have been used in the literature; they investigate different specifications of leisure, the importance of human capital being a market- (taxed) or home (untaxed) activity and the different cases arising depending on what the human capital production function looks like.

[4] In order to analytically isolate the three incentive margins I impose (1) the restriction of one common production function for physical and human capital and (2) no restriction on deinvestment in either type of capital. The assumptions imply that a two-sector model with equal production functions for physical and human capital collapses into the one-sector model presented here. There will be no transitional dynamics since adjustments to tax changes are immediate. There are growth effects, though. To the best of my knowledge, the method used to solve for the level of leisure in this paper has not been presented before.

[5] Novales and Ruiz (2002) parametrize a version of the two-sector model with physical and human capital and use numerical methods to study Laffer effects when government spending grows with the "old" growth rate even after a tax cut.

and discusses the results. In the appendix, the relationship between the model and the general two-sector model with human and physical capital is discussed and the analytical solution in the leisure case is discussed. Finally, the level of leisure in a special case is solved for.

2 The model

I set up a perfect foresight and full commitment model with utility maximizing agents holding physical and human capital. Agents derive utility from consumption. The capital is rented out to firms and agents pay tax on the returns to their capital stocks. The government uses tax receipts to finance lump-sum transfers and government consumption. Having set up the model, I define Laffer effects in section 3 and in section 4 then ask: what combination of tax rates is required for the government to be able to reduce a tax rate but still maintain its spending paths?

2.1 Production and capital

The model used in this paper is a modified AK-model, a one-sector model with physical and human capital in the production function. It has constant returns to scale in physical capital K and human capital H altogether. The production function for physical as well as human capital is $F(K, H) = AK^{\alpha}H^{1-\alpha}$, i.e. Cobb-Douglas[6] with $0 < \alpha < 1$.

Output in the economy is used for consumption or for building physical and human capital stocks. It is then assumed that output can be directly used for both physical capital build-up and human capital build-up and that one unit of physical capital can be converted into one unit of human capital. It is also assumed that investment in physical capital and human capital can be negative and immediate. This implies that capital stocks can "jump" from one level to another; the aggregate

[6] Stokey and Rebelo (1995) use the more general CES production function studying growth effects from tax rates and conclude that the elasticities of substitution in production are relatively unimportant.

of physical and human capital cannot jump, however. In effect, there is thus one aggregate capital stock, defined in per-capita terms as $z_t = k_t + h_t$, where t is a time index.[7]

2.2 Representative agent optimization

Agents derive utility from consumption and have an infinite time horizon. Utility maximizing agents sell their physical and human capital to profit maximizing firms and receive factor returns. The agent also receives a lump-sum transfer from the government and returns from government bonds. Income is spent on consumption or invested in the assets of the economy, physical and human capital and government bonds. Income is also used to pay taxes on the returns on these assets. The government uses tax receipts to finance government consumption and lump-sum transfers to the agents. Depreciation rates are set at zero and there is no population growth.

Before solving the representative agent's optimal consumption path, the return to capital is derived. Firms rent physical and human capital from the agents in order to maximize profits with respect to inputs k_t and h_t. The per-capita production function is $f(k_t, h_t) = Ak_t^{\alpha}h_t^{1-\alpha}$. From the competitive equilibrium condition that $r_t = \partial f/\partial k_t$ and $w_t = \partial f/\partial h_t$, where r_t is the return on physical capital and w_t is the return on human capital, the standard arbitrage condition of equal after-tax returns on k_t and h_t becomes

$$r_t(1-\tau_k) = w_t(1-\tau_h) \tag{1}$$

where τ_k and τ_h are taxes on returns to physical and human capital, respectively.

Condition (1) allows us to define the agent's after tax return to capital (k, h as well as z) to become

$$\phi \equiv r_t(1-\tau_k) = A\alpha^{\alpha}(1-\alpha)^{1-\alpha}(1-\tau_k)^{\alpha}(1-\tau_h)^{1-\alpha}. \tag{2}$$

[7] The model is equivalent to a two-sector model with equal production functions for physical and human capital and no restrictions on deinvestment of k and h. As a result of these assumptions, I get a framework where responses to tax cuts are immediate. Absent transitional dynamics between the old and new growth paths, it is possible to decompose the effects of tax cuts into compositional and dynamic effects. In appendix 1, I discuss the relation between the model in this paper and the two-sector models. See Barro and Sala-i-Martin (1995) for a presentation of the model used here.

This return ϕ will also be the return paid by the government on the stock of government bonds, defined as b_t.[8] The agent's total wealth, defined as $W_t \equiv z_t + b_t$, thus earns the return ϕ.

The agent maximizes lifetime utility from consumption subject to the budget constraint, i.e.

$$\begin{aligned}
&\text{Max} \int_0^{\infty} U\left(c_t, G_t\right) e^{-\rho t} dt \\
&\text{s.t. } \dot{W}_t = \phi W_t + T_t - c_t \text{ and } W_0 = z_0 + b_0. \qquad (3)
\end{aligned}$$

There is also a transversality condition. $U\left(c_t, G_t\right)$ is the instantaneous utility function, c_t is private consumption, G_t is government consumption, ρ is the time preference factor of the agent, T_t are lump-sum transfers received from the government, z_0 is period-zero total capital and b_0 is period-zero government bonds.[9] Dotted variables are time derivatives. Time indices will normally be suppressed. The utility function is additively separable in private consumption c and government consumption G, $U\left(c, G\right) = u\left(c\right) + v(G)$ where $u\left(c\right)$ takes the Constant Inter-temporal Elasticity of Substitution (CIES) form,

$$u\left(c\right) = \frac{c^{1-\theta}}{1-\theta}, \qquad (4)$$

where θ is the inverse of the intertemporal elasticity of substitution. Attaching the dynamic Lagrange multiplier λ to the budget constraint in (3), using control c and state W, the first-order conditions of this problem are

$$\begin{aligned}
&u'\left(c\right) e^{-\rho t} = \lambda \\
&\lambda \phi = -\dot{\lambda} \qquad (5)
\end{aligned}$$

and the transversality condition is $lim_{t\to\infty} \lambda_t W_t = 0$.

Using the CIES utility function and conditions (5) gives the Euler equation

---

[8] The return from bonds is taxed with the physical capital tax τ_k, so that the pre-tax return to bonds R_t would be determined by $\phi = R_t(1-\tau_k)$.

[9] Regarding notation, I use capital letters for government spending variables, i.e. transfers T and government consumption G. I use W for per-capita wealth in order to not confuse it with the return to human capital, w. Small letters are used for all other stocks and flows. I use the capital letter A in the production function because of the resemblance with the AK-model.

$$\gamma \equiv \frac{\dot{c}}{c} = \frac{1}{\theta}(\phi - \rho). \tag{6}$$

The growth rate of consumption γ depends on the degree of intertemporal substitution $\sigma \equiv \theta^{-1}$, the after-tax return on capital ϕ (from 2) and the time preference factor ρ. The degree of intertemporal substitution has the usual interpretation: an agent with a low degree of intertemporal substitution prefers a stable consumption path and will not react to tax changes to any considerable extent. A low after-tax return ϕ will discourage investment and slow growth. So will a high degree of impatience (ρ high) of the agent, because future consumption flows are less valued. Taxes on physical and human capital affect the growth rate, through ϕ, to the same degree as physical and human capital affect total output. Reductions in τ_k and τ_h make investment more productive and hence increase the growth rate, $\partial\gamma/\partial\tau_k < 0$ and $\partial\gamma/\partial\tau_h < 0$.[10]

2.3 Composition effect, return to capital and total production

Since we will work with compositional as well as dynamic effects from tax cuts, we need to keep track of how tax changes affect the composition of the human-to-physical capital stocks and how this affects production and returns to capital. This section therefore discusses three variables that will be important in what follows: the agent's h/k-ratio, the private return to capital ϕ (from above) and the economy-wide return to capital.

From the above, we get the equilibrium h/k-ratio, derived from the arbitrage condition in (1),

$$\frac{h}{k} \equiv \Omega = \frac{1-\alpha}{\alpha}\frac{1-\tau_h}{1-\tau_k}. \tag{7}$$

A smaller α increases the h-to-k ratio, as does an increase in τ_k.[11]

[10] The transversality condition $lim_{t\to\infty}\lambda_t W_t = 0$ implies that the consumption growth rate must be smaller than the private return on capital; $\phi - \gamma > 0$. For values of θ above or equal to unity, this condition is always satisfied. Below unity we require $A\alpha^{\alpha}(1-\alpha)^{1-\alpha}(1-\tau_k)^{\alpha}(1-\tau_h)^{1-\alpha} < (1-\theta)^{-1}\rho$.

[11] As seen in (2), a higher τ_k reduces ϕ with the factor $(1-\tau_k)^{\alpha}$ and not $(1-\tau_k)$. Since h/k increases, physical capital k becomes more scarce and its return r goes up, partly but not fully compensating the effect of the tax cut on the private return and growth.

In interpreting Ω, note that maximizing output $Ak^{\alpha}h^{1-\alpha}$ subject to the constraint $h+k=z$ would yield a ratio $h/k=(1-\alpha)/\alpha$. The agent's h/k ratio, and therefore what is used for production, differs from this value as soon as $\tau_h \neq \tau_k$. A differentiated tax treatment of h and k thus adds the second effect discussed in the introduction, a compositional distortion in production, to the first effect, the dynamic distortion always present when capital is taxed. This compositional distortion from a differentiated capital taxation is important in the analysis of Laffer effects. The importance can be seen by defining the economy-wide return to capital which, using $h/k=\Omega$, is

$$\beta \equiv \frac{wh+rk}{h+k} = \frac{A\Omega^{1-\alpha}}{1+\Omega} = \frac{A\alpha^{\alpha}(1-\alpha)^{1-\alpha}(1-\tau_k)^{\alpha}(1-\tau_h)^{1-\alpha}}{\alpha(1-\tau_k)+(1-\alpha)(1-\tau_h)}. \quad (8)$$

It can be shown that the economy-wide return to capital β, with one tax rate given, is at its maximum when the second tax is set equal to the first tax. β is nothing else than the de facto production factor in the economy which we see by rewriting $Ak^{\alpha}h^{1-\alpha}$ as follows:

$$Ak^{\alpha}h^{1-\alpha} = \beta z. \quad (9)$$

Total production is therefore tax-dependent, a feature which is absent in standard one-sector AK-models. A reduction in the highest tax will decrease the compositional distortion and production will jump to a higher level, which will open up a new margin for Laffer effects.[12]

For future reference, also note that

$$\phi = \beta\left(1-\tau_{avg}\right), \quad (10)$$

where the "average tax rate" $\tau_{avg} \equiv \alpha\tau_k + (1-\alpha)\tau_h$ has been used. Whereas β is the economy-wide return to capital, the agents face the (lower) private return ϕ because taxes must be paid. Without taxes, the returns are equal.[13]

[12] There are no traditional dynamics but immediate adjustment in the h/k-ratio as a result of tax cuts. If we had a more general model with different production functions for physical and human capital, we would get transitional dynamics as a result of tax cuts but still, along a balanced growth path, a constant h/k-ratio. There is still a compositional reoptimization as a response to tax cuts. It would not be immediate, however. As will be discussed in detail in the following section, the tax cuts that will be considered are such that we reduce the highest tax and hence, increase GDP.

[13] I use the term "economy-wide" to refer to the pre-tax return to capital in the economy. It

2.4 Intertemporal constraints

Throughout the analysis of Laffer effects, the tools of analysis will be the consumption rule and the present value resource constraint. These will describe how the representative agent responds to tax changes and, as a result, what scope there is for Laffer effects. The consumption rule is derived by integrating the budget constraint

$$\dot{W} = \phi W + T - c. \tag{11}$$

The present value budget constraint, using initial total wealth $W_0 = z_0 + b_0$ and $c_t = c_0 e^{\gamma t}$ from (6) and applying the transversality condition, becomes

$$\frac{c_0}{\phi - \gamma} = z_0 + b_0 + \int_0^\infty T_t e^{-\phi t} dt \tag{12}$$

where c_0 is period-zero private consumption. This relationship says that the present value of consumption should be equal to initial assets plus the present value of transfers received from the government. We get the consumption rule by multiplying through by $(\phi - \gamma)$,

$$c_0 = (\phi - \gamma) \left(z_0 + b_0 + \int_0^\infty T_t e^{-\phi t} dt \right). \tag{13}$$

This consumption rule, which depends both on the transfer and the tax policy of the government, will be used to study how government consumption can vary with tax rates complying with the economy's resource constraint. In the resource constraint, the production of the economy is either consumed by the government or by the agents or added to the stocks of k and h, $Ak^\alpha h^{1-\alpha} = c + G + \dot{k} + \dot{h}$. With GDP written as βz (from 9), the resource constraint becomes

$$\dot{z} = \beta z - c - G, \tag{14}$$

is different, however, from the first-best (social) return as long as taxes are differentiated. In the AK-model, the private return would be $A(1 - \tau_k)$ and the economy-wide (and social) return is A. Expression (10) is the same relationship in a model with two types of capital and two tax rates.

and the present value resource constraint, using initial capital z_0 and $c_t = c_0 e^{\gamma t}$ and applying the transversality condition, becomes

$$\int_0^\infty G_t e^{-\beta t} dt + \frac{c_0}{\beta - \gamma} = z_0. \tag{15}$$

The constraint says that the present value of total consumption must equal initial resources. Note that it is the "economy-wide" discount rate β which is of importance for the use of resources, whereas it is the private return ϕ that determines the behavior of the agent in the consumption rule.

We have now derived the tools to study Laffer effects, i.e. the tools for studying how government spending reacts to tax cuts and if it will be possible to "maintain government spending" after tax cuts. In the following sections, different versions of expressions (13) and (15) will be differentiated with respect to tax rates in order to study Laffer effects. Before that, however, we need to rigorously define Laffer effects and what is meant by "maintaining government spending".

3 Definitions of Laffer effects

The three definitions of Laffer effects follow and extend the work by Agell and Persson (2001) where differences between earlier results on dynamic Laffer effects were clarified. It extends this work to a context where there is more than one factor of production and therefore not only dynamic but also compositional effects of taxes.[14]

Definition 1 follows the Agell and Persson definition whereas definitions 2 and 3 comprise two slightly different cases that collapse into one case in the basic AK framework. The difference between the three definitions is related to what is meant by "maintaining government spending". In the model presented so far, a balanced growth path exists where private consumption, capital stocks and government consumption and transfers all grow at the same rate. After a tax cut, the return to

[14] For this reason, I use the term "Laffer effect" instead of "dynamic Laffer effect".

private capital ϕ and the growth rate of consumption γ increase. We can then either allow for government consumption and transfers to adjust their growth rate to the new higher rate or they can maintain their pre tax cut growth rates.

We also need to distinguish between the case where we account for the jump in production as a result of tax cuts and the case where we do not. After a tax cut, because the "economy-wide" return β is tax-dependent, period-zero GDP discretely adjusts from $f_0(k_0^{\mathrm{pre}}, h_0^{\mathrm{pre}}) \equiv \beta^{\mathrm{pre}} z_0$ to $f_0(k_0^{\mathrm{post}}, h_0^{\mathrm{post}}) \equiv \beta^{\mathrm{post}} z_0$. Definition 2 does not take this discrete adjustment into consideration whereas definition 3 does. That is, in definition 2, because GDP discretely increases as a result of a tax cut, the transfer-to-GDP ratio goes down for a given period-zero transfer T_0 and we allow this to happen. This is why we will use the post tax cut GDP $f_0(k_0^{\mathrm{post}}, h_0^{\mathrm{post}})$ in the transfer-to-GDP ratio in definition 2 below. We do not require T_0 to adjust to maintain the original ratio. In definition 3, we want to maintain the original transfer-to-GDP ratio and therefore divide by the original GDP, $f_0(k_0^{\mathrm{pre}}, h_0^{\mathrm{pre}})$. In the basic AK framework, definitions 2 and 3 collapse into one case only.[15]

Definition 1 *Assume that the resource constraint $\int_0^\infty G_t e^{-\beta t} dt + c_0 (\beta - \gamma)^{-1} = z_0$ holds for some initial tax rates τ_k^{pre} and τ_h^{pre} and flows of government consumption $(G_t)_0^\infty$ and transfers $(T_t)_0^\infty$. If there is some lower set of tax rates $\tau_k^{post} \leq \tau_k^{pre}$ and $\tau_h^{post} \leq \tau_h^{pre}$, where at least one of the inequalities should be strict, that allows the government to maintain its transfer program $(T_t)_0^\infty$ and for some time $\Delta t > 0$ increase its consumption flow G_t and not decreasing it at any other time, there is a Laffer effect.*

In definition 1, government transfers T_t follow their pre tax cut path, even after the tax cut. This path of T_t will be taken to be $T_t = T_0^{\mathrm{pre}} e^{\gamma^{\mathrm{pre}} t}$ where γ^{pre} is the pre-tax cut growth rate of consumption, GDP and capital stocks and T_0^{pre} constitute the pre tax cut period-zero level of transfers. When implementing definition 1 in this paper, government consumption G_t will also grow at the old growth rate of private

[15] Regarding notation, the superindices "pre" and "post" refer to the values pre- and post- tax cut, respectively. The subindex t refers to time, a subindex 0 therefore means the value at time zero.

consumption, γ^{pre}, and we will ask the question whether period-zero government consumption G_0 can increase when a tax is reduced.[16]

Definition 2 *Assume that the resource constraint* $\int_0^\infty G_t e^{-\beta t} dt + c_0 (\beta - \gamma)^{-1} = z_0$ *holds for some initial tax rates* τ_k^{pre} *and* τ_h^{pre} *and flows of government consumption* $(G_t)_0^\infty$ *and transfers* $(T_t)_0^\infty$*. If there is some lower set of tax rates* $\tau_k^{post} \leq \tau_k^{pre}$ *and* $\tau_h^{post} \leq \tau_h^{pre}$*, where at least one of the inequalities should be strict, that allows the government to maintain its transfer to GDP ratio, i.e. at all times after the tax cut keep* $T_t / f_t(k, h) = T_0^{pre} / f_0(k_0^{post}, h_0^{post})$*, and for some time* $\Delta t > 0$ *increase its consumption to GDP ratio* $G_t / f_t(k, h)$ *to exceed* $G_0^{pre} / f_0(k_0^{post}, h_0^{post})$ *and not decreasing it at any other time, there is a Laffer effect.*

Using the more demanding definition 2**,** all government spending follows the new higher growth rate of private consumption and GDP, even after the tax cut. That is, $T_t = T_0^{\text{pre}} e^{\gamma^{\text{post}} t}$ and we will let G_t grow at the rate γ^{post} as well and we ask the question whether period zero government consumption G_0 can increase when a tax is reduced.

Definition 3 *Assume that the resource constraint* $\int_0^\infty G_t e^{-\beta t} dt + c_0 (\beta - \gamma)^{-1} = z_0$ *holds for some initial tax rates* τ_k^{pre} *and* τ_h^{pre} *and flows of government consumption* $(G_t)_0^\infty$ *and transfers* $(T_t)_0^\infty$*. If there is some lower set of tax rates* $\tau_k^{post} \leq \tau_k^{pre}$ *and* $\tau_h^{post} \leq \tau_h^{pre}$*, where at least one of the inequalities should be strict, that allows the government to maintain its initial transfer to GDP ratio at all times, i.e. keep* $T_t / f_t(k, h) = T_0^{pre} / f_0(k_0^{pre}, h_0^{pre})$*, and for some time* $\Delta t > 0$ *increase its consumption to GDP ratio* $G_t / f_t(k, h)$ *to exceed* $G_0^{pre} / f_0(k_0^{pre}, h_0^{pre})$ *and not decreasing it at any other time, there is a Laffer effect.*

Using the even more demanding definition 3, all government spending follows the new higher growth rate of private consumption, γ^{post}. In addition, the new

[16] That is, can G_t, as a response to a tax cut, shift up to a higher level and then continue to grow at its old growth rate but starting at this new higher level so that G_t is permanently on a higher level than before the tax cut?

period zero transfers T_0^{post} and therefore the whole path of transfers T_t has made a discrete adjustment to match the discrete adjustment in GDP, i.e. $T_0^{\text{post}}/T_0^{\text{pre}} = f_0(k_0^{\text{post}}, h_0^{\text{post}})/f_0(k_0^{\text{pre}}, h_0^{\text{pre}})$. We ask the question whether period zero government consumption G_0 can make a discrete adjustment that is larger than the adjustment in GDP and then grow at the rate γ^{post}.

Imagine a tax cut that increases the consumption growth rate from 2% to 3% and as a result of the tax cut, GDP experiences a 1% discrete jump from 1.00 to 1.01 in period zero.

With definition 1, transfers T_t should not jump in period zero and continue to grow with 2% and we ask whether period-zero G_t can jump to a higher level and then grow with 2%. With definition 2, transfers T_t should also not jump in period zero but then grow with 3% and we ask whether period-zero G_t can jump to a higher level and then grow with 3%. With definition 3, transfers T_t should jump up 1% in period zero and then grow with 3% and we ask whether period-zero G_t can jump more than 1% and then grow with 3%.

The scope for each type of Laffer effect is studied below.

4 Conditions to get Laffer effects

4.1 Pre tax cut setting

In order to study Laffer effects, we start out in a situation at time $t = 0$ with initial capital z_0, zero outstanding government debt ($b_0 = 0$) and government transfers and consumption equalling government revenue, $T_0 + G_0 = r^{\text{pre}}\tau_k^{\text{pre}}k_0^{\text{pre}} + w^{\text{pre}}\tau_h^{\text{pre}}h_0^{\text{pre}}$. Using the equilibrium expressions for r, w and h/k, this expression can be written as

$$T_0 + G_0 = \tau_{avg}^{\text{pre}}\beta^{\text{pre}}z_0. \tag{16}$$

Prior to a tax cut GDP, private consumption, capital stocks as well as government consumption and transfers all grow at the pre tax cut growth rate $\gamma^{\text{pre}} = \frac{1}{\theta}\left(\phi^{\text{pre}} - \rho\right)$. Taxes are then changed according to $\tau_k^{\text{post}} \leq \tau_k^{\text{pre}}$ and $\tau_h^{\text{post}} \leq \tau_h^{\text{pre}}$, where at least one of the inequalities should be strict. For the expressions derived below, the analysis is restricted to reducing one tax at a time, i.e. we either have $\left(\tau_k^{\text{post}} < \tau_k^{\text{pre}}, \tau_h^{\text{post}} = \tau_h^{\text{pre}}\right)$ or $\left(\tau_k^{\text{post}} = \tau_k^{\text{pre}}, \tau_h^{\text{post}} < \tau_h^{\text{pre}}\right)$. Moreover, we are naturally interested in decreasing the highest tax as this reduces the compositional distortion in production.

For the subsequent analysis, it will be useful to express the tax-derivatives of the consumption growth rate γ and the economy-wide return to capital β as functions of the tax-derivatives of the private return to capital ϕ. A few algebraic steps will show that

$$\frac{\partial \gamma}{\partial \tau_h} = \frac{1}{\theta} \frac{\partial \phi}{\partial \tau_h} \text{ and } \frac{\partial \beta}{\partial \tau_h} = \Gamma_h \frac{\partial \phi}{\partial \tau_h} \text{ where } \Gamma_h = \frac{\alpha \left(\tau_h - \tau_k\right)}{\left(1 - \tau_{avg}\right)^2} \tag{17}$$

$$\frac{\partial \gamma}{\partial \tau_k} = \frac{1}{\theta} \frac{\partial \phi}{\partial \tau_k} \text{ and } \frac{\partial \beta}{\partial \tau_k} = \Gamma_k \frac{\partial \phi}{\partial \tau_k} \text{ where } \Gamma_k = \frac{(1 - \alpha) \left(\tau_k - \tau_h\right)}{\left(1 - \tau_{avg}\right)^2}. \tag{18}$$

Γ_h and Γ_k are important factors in the Laffer effect analysis. They represent the second effect of taxation in the model, the impact of the compositional distortion from a differentiated tax treatment of h and k. In (17) and (18), we get that if $\tau_h = \tau_k$, both Γ_h and Γ_k are zero and β is not tax dependent. A non-zero value of either Γ_h or Γ_k means that we have a compositional distortion and that the maximum production capacity is not achieved (as GDP equals βz from 9). As discussed earlier, tax changes that affect β will therefore make available more/less resources in all periods and affect the possibility for Laffer effects (more resources when we reduce the highest tax which is what we are interested in). For future reference, also note that if Γ_h or Γ_k are larger than unity, we get a larger change in the economy-wide return β than in the private return to capital ϕ when taxes are changed.

4.2 Mathematical criterion

Let subindex i refer to either the physical capital or human capital tax. The criterion to get a Laffer effect is $\partial G_0 / \partial \tau_i < 0$ for definitions 1 and 2 Laffer effects and $\partial \left(G_0 / \text{GDP}\right) / \partial \tau_i < 0$ for definition 3 Laffer effects.

4.3 Laffer effect according to definition 1

Following definition 1, we will let T grow at its pre tax cut growth rate γ^{pre} and study scope for increased G. G is set to grow at the original growth rate γ^{pre} as well and it is therefore enough to study the impact on G_0, G in period zero. The consumption rule and the present value resource constraint, expressions (13) and (15), are repeated with these assumptions for T and G:

$$\frac{c_0}{\phi-\gamma} = z_0 + \frac{T_0}{\phi-\gamma^{\text{pre}}} \tag{19}$$

$$\frac{G_0}{\beta-\gamma^{\text{pre}}} = z_0 - \frac{c_0}{\beta-\gamma}. \tag{20}$$

I differentiate (19) and (20) with respect to either of the tax rates and study whether such a tax change makes the new G_0 comply with the condition for a Laffer effect, $\partial G_0/\partial\tau_i < 0$. A change in taxation will, through its effect on growth, the private discount rate and future value of transfers in the first constraint affect c_0. This change in c_0 then adds to the effects on the growth rate and on the economy-wide discount rate in the second constraint to give a total effect on G_0 such that the resource constraint is always fulfilled.[17] Note that $\partial/\partial\tau_i$ can mean a change in either tax rate, τ_h or τ_k. Differentiation of (20) and (19 gives[18]

$$\frac{\partial G_0}{\partial\tau_i} = -\frac{\partial c_0}{\partial\tau_i} - \frac{c_0}{\beta-\gamma}\frac{\partial\gamma}{\partial\tau_i} + z_0\frac{\partial\beta}{\partial\tau_i}, \tag{21}$$

where

$$\frac{\partial c_0}{\partial\tau_i} = \frac{\partial(\phi-\gamma)}{\partial\tau_i}\left(z_0 + \frac{T_0}{\phi-\gamma^{\text{pre}}}\right) + (\phi-\gamma)\frac{\partial}{\partial\tau_i}\left(z_0 + \frac{T_0}{\phi-\gamma^{\text{pre}}}\right). \tag{22}$$

In (21), tax changes will indirectly affect G_0 through their effect on c_0, and directly through the change in the growth rate of private consumption and through the effect on β. The second term in (21) is always positive as $\partial\gamma/\partial\tau_i < 0$; a higher growth rate of private consumption from tax cuts makes Laffer effects more difficult to achieve. The third term comes from the impact of a tax change on the economy-wide return to capital. A change in β, through a change in compositional distortions that affects output in all periods, changes the present value of given flows of lifetime private and government consumption and thereby the scope for

[17] That is; G_0 is residually calculated such that the resource constraint always holds.

[18] I differentiate with respect to a tax and then evaluate the derivative in the pre tax-cut point.

Laffer effects. Expression (22) is the standard consumption response in period-zero consumption through income and substitution effects (first term) and wealth effects of transfers (second term). The wealth effect of transfers plays a crucial role in the possibility to get Laffer effects.[19]

Summing up and rewriting (21) and (22) gives that $\partial G_0/\partial \tau_i$ is a sum of the growth effects ($\partial \gamma/\partial \tau_i$) on the two different present values of consumption and the effects through the different returns to capital. The condition to get a definition 1 Laffer effect, $\partial G_0/\partial \tau_i < 0$, becomes

$$\frac{\partial G_0}{\partial \tau_i} = \frac{\partial \gamma}{\partial \tau_i}\left(\frac{c_0}{\phi-\gamma} - \frac{c_0}{\beta-\gamma}\right) + \left(\frac{\partial \beta}{\partial \tau_i} - \frac{\partial \phi}{\partial \tau_i}\right)\left(\frac{c_0}{\phi-\gamma} - \frac{T_0}{\phi-\gamma}\right) < 0. \tag{23}$$

Using relationships (17) and (18) between changes in γ, β and ϕ gives

$$\frac{\partial G_0}{\partial \tau_i} = \left(\frac{1}{\phi-\gamma}\right)\frac{\partial \phi}{\partial \tau_i}\left(\frac{1}{\theta}\frac{\beta-\phi}{\beta-\gamma}c_0 + (\Gamma_i - 1)(c_0 - T_0)\right) < 0. \tag{24}$$

From (24), because $\partial \phi/\partial \tau_i < 0$ and $\phi - \gamma > 0$, there are two possible cases giving a Laffer effect:

Proposition 1 *There is a Laffer effect, $\partial G_0/\partial \tau_i < 0$, in the sense of definition (1), where $i = k$ or h, if*
$\frac{T_0}{c_0} > 1 - \frac{1}{1-\Gamma_i}\frac{\beta-\phi}{\beta-\gamma}$ or if $\Gamma_i > 1$.

The first part of proposition 1 is written to stress the importance of transfers and the wealth effect that results from tax cuts. It simplifies to the case of the AK-model when taxes are equal, i.e. when $\Gamma_i = 0$, meaning that there is only a dynamic and no compositional margin. The proposition then tells us how large a share of consumption that should be transfer-financed to get a *dynamic* Laffer effect.[20]

[19] From (22), c_0 is affected through the change in the portion of lifetime income consumed in the first period $(\phi - \gamma)$ and through the change in valuation of lifetime transfers $T_0/(\phi - \gamma^{\text{pre}})$. If the intertemporal elasticity of substitution θ^{-1} is less than unity, the income effect dominates the substitution effect and the first term is negative. The second term is the wealth effect and is always positive. It will act to reduce period-zero consumption when taxes are reduced because future transfers are worth less as a result of the tax cut. The wealth effect from transfers must be sufficiently large, i.e. the higher $\partial c_0/\partial \tau_i$, the more likely is a Laffer effect.

[20] See Agell and Persson (2001) for a full discussion.

The discussion of the criterion $\Gamma_i > 1$ is postponed until proposition 2, regarding definition 2 Laffer effects, has been derived.

4.4 Laffer effect according to definition 2

With government transfers and consumption following the (higher) growth rate of private consumption after a tax cut, the present value budget and resource constraints, (13) and (15), simplify to become

$$c_0 = z_0\,(\phi - \gamma) + T_0 \tag{25}$$
$$G_0 = z_0\,(\beta - \gamma) - c_0. \tag{26}$$

The condition to get a Laffer effect is, once more, $\partial G_0/\partial \tau_i < 0$. Differentiation of G_0 with respect to any tax gives

$$\frac{\partial G_0}{\partial \tau_i} = z_0 \left(\frac{\partial \beta}{\partial \tau_i} - \frac{\partial \phi}{\partial \tau_i} \right) = z_0 \frac{\partial \phi}{\partial \tau_i} \left(\Gamma_i - 1 \right). \tag{27}$$

Because $\partial \phi/\partial \tau_i < 0$, we arrive at the following proposition:

Proposition 2 *There is a Laffer effect, $\partial G_0/\partial \tau_i < 0$, in the sense of definition (2) if* $\Gamma_i > 1$.

4.5 Interpretation

As expected, definition 1 Laffer effects are the easiest to obtain. Government spending is only required to grow at the old growth rate, whereas for definition 2 government spending should grow at the new higher growth rate that follows from a tax cut. A positive Γ_i in the first part of proposition 1 reduces the requirement on the transfer/consumption ratio in order to obtain a Laffer effect. This is because a tax cut reduces the compositional distortion, thus helping the self-financing of a tax cut. The second part of proposition 1 is the same as proposition 2. The requirement for a definition 2 effect, $\Gamma_i > 1$, is now analyzed in more detail.

If we combine the intertemporal constraints (25) and (26), we get $T_0 + G_0 = (\beta - \phi)\, z_0$. Because $(\beta - \phi) = \beta\tau_{avg}$ (from 10), this is nothing but the time-zero budget constraint of the government from (16). Since all variables grow at the same rate, the dynamic constraint collapses to the static government budget constraint. The reason is that with the assumptions on c, T and G growing at the same growth rate, for the optimal solution also total capital z will grow at the same rate and no bonds will ever be issued. That is, if we are in a condition to get a dynamic Laffer effect, $\Gamma_i > 1$, no bonds are needed; G_0 will be residually determined to fulfill the present value (and static) resource constraint. If we are not in a condition to get a dynamic Laffer effect, $\Gamma_i < 1$, if bonds were to be issued they could not be recovered (a transversality condition would be violated) and there is no way for G_0 complying with $\partial G_0 / \partial \tau_i < 0$ to fulfill the present value resource constraint.

If we maximize static government tax revenue, $\beta\tau_{avg}$, keeping one tax constant (say τ_k), we get the condition that τ_h should fulfill $\Gamma_h = 1$ for maximum revenue. For $\Gamma_h > 1$, we are below maximum tax revenue. Therefore, the condition for definition 2 Laffer effects is the same as a static government revenue maximization problem. If τ_h is beyond the point where $\Gamma_h = 1$, we are at the wrong side of the τ_h-Laffer curve and can hence increase revenue by reducing τ_h. From figure 1 below, where $\beta\tau_{avg}$ is plotted for two different values of α, the physical capital tax is fixed at $\tau_k = 0.3$. The graph shows that a human capital tax above 0.75 would be needed to get a definition 2 Laffer effect. As seen in the graph, this result is not very sensitive to the value of α.

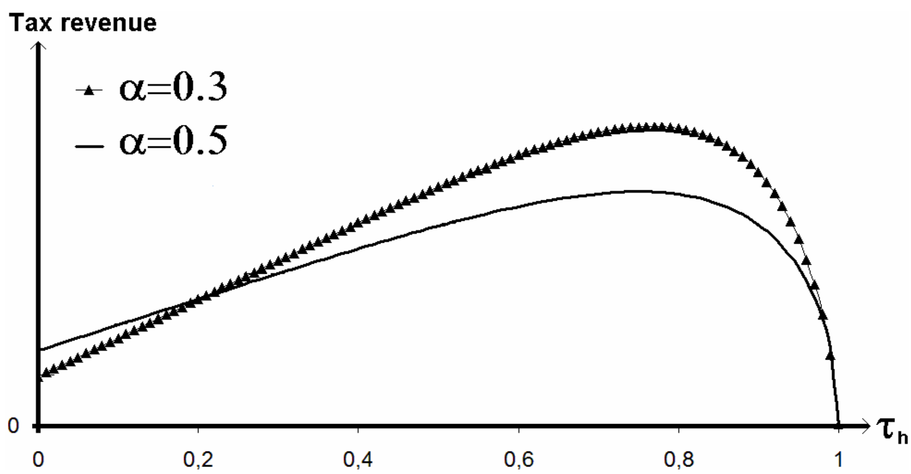

Figure 1. The τ_h-Laffer curve for two different values of α when $\tau_k = 0.3$. (Note that the shape of these curves does not depend on A.)

The analysis above tells us that there are two sources for Laffer effects, *compositional* and *dynamic*. In our model, where there are no transitional dynamics but immediate adjustment in the h/k-ratio, the compositional effect is static. As a result of a tax change, the agent immediately reoptimizes the h/k-ratio according to (7) and there is an immediate adjustment in the returns to capital and the consumption growth rate. Production βz will "jump" through the discrete adjustment in β and more resources are made available in all periods. This compositional effect makes it easier for the government to maintain spending and it is possible to get a definition 2 Laffer effect. The *dynamic* Laffer effect is captured by definition 1. Here, because of a less stringent requirement on spending and an increased growth rate of consumption and the capital stock, there is a true dynamic effect of tax cuts.

The interpretation of Laffer effects as compositional and dynamic is likely to carry over to the more general two-sector model with separate production functions for physical and human capital. The model used is a special case of this two-sector model; the same production function for both physical and human capital and immediate adjustment in the stocks of h and k has been assumed. These assumptions have allowed us to separate compositional from dynamic effects. In the general model, along a balanced growth path, the h/k-ratio will also be constant. A tax change will result in a period of transition where the ratio - or composition - readjusts to the new tax rates. A tax cut in these models also generates the growth effect, which is the source of the dynamic Laffer effects.

I state a final proposition regarding Laffer effects and then proceed to studying what the introduction of a labor/leisure choice implies for the analysis of Laffer effects.

4.6 Laffer effect according to definition 3

With definition 3, total government spending should increase as a *fraction* of GDP.[21] It is straightforward to show that this can never be possible. Rewrite (26) using (25) to get

[21] Increasing the total government spending $(G+T)$ to GDP ratio is equivalent to asking whether G can increase as a fraction of the new GDP, letting T increase to exactly preserve its GDP ratio.

$G_0 = z_0 (\beta - \phi) - T_0 = \beta \tau_{avg} z_0 - T_0.$

Division by GDP, βz_0, gives

$$\frac{G_0}{\text{GDP}} = \tau_{avg} - \frac{T_0}{\text{GDP}}.$$

We are interested in the sign of $\partial (G_0/\text{GDP}) / \partial \tau_i$. If we reduce either tax rate, the factor τ_{avg} will decrease. If the T_0/GDP-ratio is to remain intact, the left-hand side must then decrease as a result of the tax cut, i.e. $\partial (G_0/\text{GDP}) / \partial \tau_i > 0$. We are thus in a case where there are no dynamic or compositional margins from which resources for increased government consumption can be generated and we get the following result;

Proposition 3 *There can never be a Laffer effect, $\partial (G_0/\text{GDP}) / \partial \tau_i < 0$, in the sense of definition (3).*

5 Extending the framework: endogenous leisure

So far in this paper we have seen how capital taxation in general and an uneven taxation of factors of production in particular affect the scope for self-financing tax cuts. When a tax is reduced, dynamic and compositional margins are affected and there may be Laffer effects. There is symmetry between changes in τ_k and τ_h. In this section, I extend the model by introducing a labor/leisure choice and leisure in the agent's utility function. I follow most of the literature and model leisure as "raw-time", where human capital and leisure are bundled together and the human capital effectively supplied for production is $h(1-l)$ rather than h. Here, $(1-l)$ is the fraction of the unitary time endowment used for labor and l is leisure time.[22]

Before presenting the extended model, we can say something about what results to expect. First, we should expect the scope for Laffer effects to increase because we have a new margin of adjustment. As we will see, the growth rate in this model

[22] See Milesi-Ferretti and Roubini (1998a) for a discussion of different specifications of leisure and for a full discussion of the problem set-up and first-order conditions. The model used here is a special case of their model, the case when the production functions for physical and human capital are the same.

will be increasing in the level of labor time. Therefore, a tax cut that increases labor time adds a new dynamic margin which is indeed a new source of a *dynamic* Laffer effect. Second, there is also a new compositional effect. If labor time and therefore production increase in response to a tax cut, this also opens up for Laffer effects (although we also need to consider the general equilibrium response in the h/k-ratio).

The introduction of leisure should also break the symmetry between the two taxes. In particular, the agent will be faced with an intratemporal allocation decision between consumption and leisure. This decision will be directly affected by the tax on human capital, whereas the physical capital tax will only have indirect effects on the consumption/leisure decision.

A limitation in the analysis is that the method to integrate the budget and resource constraints will no longer be easily applicable for definition 1 Laffer effects. This is due to the fact that no constant-leisure level will exist other than in the long run when we have variables growing at different growth rates. An analytical condition for definition 1 Laffer effects with leisure will therefore not be provided. We can, however, get the intuition for definition 1 Laffer effects, discussing how a tax cut has affected the growth rate through the leisure level. For definition 2 Laffer effects, we will get a solution where leisure jumps from one constant level to another as taxes change. This means that consumption, capital stocks, government consumption and government transfers can all grow at the same rate and we can analyze the compositional effect of having introduced leisure[23,24].

5.1 The model with leisure

The model description is limited to what has changed from above. A fraction l of the agent's unitary time endowment will be removed from production. The remaining part of the time endowment, $(1-l)$, will be used in production so that effective

[23] I use a utility function that is consistent with a steady state with a constant leisure level as derived by King et al. (1988).

[24] The definitions of Laffer effects remain the same. For definition 2 effects, this still means that if there is such an effect, it will be compositional in nature. Leisure in the model may change the way the growth rate responds to tax cuts. This change in growth rate also applies to government spending, however.

human capital supplied in production is $h(1-l)$ and $w = \partial f/\partial(h(1-l))$ will be the return to effective human capital. We can proceed with the model setup from above, but h should be replaced with $h(1-l)$ in the production function and in the budget constraint.[25] I continue suppressing time indices on the variables and, in order to not introduce additional confounding notation, the same symbols w, r, γ, ϕ, β as above are used. As an example, w is still the return to human capital but its expression will be slightly different from above because of the introduction of leisure in the model.

The arbitrage condition in (1) now becomes

$$r(1-\tau_k) = w\,(1-l)\,(1-\tau_h). \tag{1'}$$

From this condition, we derive the h/k-ratio, which will still be $h/k = \Omega$ from above. It is unaffected by the introduction of leisure, but part of the human capital stock is no longer deployed.[26] Knowing h/k, we can derive the expressions for the private return ϕ and the economy-wide return β. These will look as in the no-leisure case, (2) and (10), but will now include a leisure component $(1-l)^{1-\alpha}$:

$$\phi \equiv A\alpha^{\alpha}(1-\alpha)^{1-\alpha}(1-\tau_k)^{\alpha}(1-\tau_h)^{1-\alpha}\,(1-l)^{1-\alpha}\,.$$
$$\beta = \frac{\phi}{1-\tau_{avg}}.$$

Leisure thus affects the returns to capital in the economy. The fact that not all human capital is deployed in production has a negative effect on the return to capital and, as we shall see, the growth rate. Changes in leisure, induced by tax cuts, thus affect the scope for Laffer effects. In particular, it seems likely that decreases in the leisure level from tax cuts, $\partial l/\partial\tau_i > 0$, will act as a new margin that increases the scope for self-financing tax cuts both through more human capital deployed in production and through a higher growth rate.[27] This growth effect is in addition

[25] When the representative agent's problem was solved in the non-leisure section, we first derived the non-arbitrage condition between k and h and then worked with the state variable W (or equivalently, z and b) in the budget constraint. With leisure, h is replaced by $h\,(1-l)$, and the return to capital in the budget constraint is instead $hw\,(1-l)\,(1-\tau_h) + kr(1-\tau_k)$.

[26] A fraction l of human capital h is no longer productive. The return on effective human capital $h(1-l)$ has increased and the return on k has decreased. The h/k-ratio that satisfies condition (1') remains intact. The fraction of human capital used in production, i.e. $h(1-l)/k$, has decreased, though, which is what we should expect.

[27] The total (general equilibrium) effect also needs to take the change in the h/k-ratio into account.

to the positive effect on the growth rate from the tax cut itself and thus constitutes an additional margin for definition 1, or *dynamic*, Laffer effects.

The utility function is additively separable in the private goods (c, l) and government consumption (G), $U(c, l, G) = u(c, l) + v(G)$. I follow Devereux and Love (1994) and Novales and Ruiz (2002) and use

$$u(c, l) = \frac{\left(c^{\eta} l^{1-\eta}\right)^{1-\theta}}{1-\theta}.$$

The first-order conditions with respect to consumption and leisure imply that the marginal rate of substitution should equate the relative price:

$$\frac{u_l}{u_c} = wh(1-\tau_h). \tag{28}$$

The human capital tax has a direct effect on this trade-off through its direct effect on the relative price of leisure. Both τ_h and τ_k also indirectly affect the trade-off through changes in w and h. We can see the full tax-dependence in the consumption/leisure trade-off by rewriting (28) using the utility function and the general equilibrium expressions for w and h:

$$\frac{c}{l(1-l)^{-\alpha}} = \frac{\xi(1-\tau_h)^{2-\alpha}(1-\tau_k)^{\alpha}}{\alpha(1-\tau_k) + (1-\alpha)(1-\tau_h)}. \tag{29}$$

Here, ξ is a function of the capital stock z and the non-tax parameters of the problem.[28] The right-hand side is unambiguously decreasing in τ_h and it is also decreasing in τ_k if $\tau_k > \tau_h$ which is the case when we study reductions in τ_k. Moreover, it can be shown that the size of the effect on the intratemporal margin from tax cuts is larger for changes in τ_h than in τ_k (because of the direct effect of τ_h). We should thus expect the largest jump in the leisure level from reductions in τ_h and, as a consequence, a larger scope for Laffer effects (compared to when τ_k is reduced).

The intertemporal conditions still need to be derived, both to get a second relationship between c and l from the consumption rule and to get the present value resource constraint from which we study Laffer effects. The Euler equation is once

[28] $\xi = Az\frac{\eta}{1-\eta}\alpha^{\alpha}(1-\alpha)^{2-\alpha}$.

again derived from first-order conditions similar to (5), where the utility function and ϕ now also include leisure. With $\sigma \equiv (1-\eta+\eta\theta)^{-1}$, the Euler equation becomes

$$\frac{\dot{c}}{c}-\frac{\dot{l}}{l}\sigma(1-\theta)(1-\eta)=\sigma(\phi-\rho). \tag{30}$$

With constant leisure, the Euler equation reduces to $\gamma \equiv \frac{\dot{c}}{c}=\sigma(\phi-\rho)$ and the method of integrating the budget and resource constraints remains valid. The Euler equation looks as before but the "intertemporal elasticity of substitution" σ now equals $(1-\eta+\eta\theta)^{-1}$.

5.2 Laffer effect according to definition 2 – with leisure

The expressions for the consumption rule and the present value resource constraint from (25) and (26) remain valid; they maintain their simple form because any adjustment in the growth rate also applies to government transfers T and government consumption G. The only difference is the factor $(1-l)^{1-\alpha}$ now present in γ, ϕ and β. The expressions are repeated here:

$$c_0=z_0(\phi-\gamma)+T_0 \tag{31}$$

$$G_0=z_0(\beta-\gamma)-c_0 \tag{32}$$

To get an explicit expression for Laffer effects, we first need to solve for c and l from the intratemporal and intertemporal relationships (equations 29 and 31). Implicitly, however, the procedure from the non-leisure section can be followed: (32) and (31) are differentiated in order to obtain the condition for definition 2 Laffer effects. That is, we calculate the derivative $\partial G_0/\partial\tau_i = z_0(\partial\beta/\partial\tau_i - \partial\phi/\partial\tau_i)$ as in (27), without first solving for the leisure level, explicitly recognizing that β and ϕ contain the factor $(1-l)^{1-\alpha}$, where l is tax-dependent. This means that the proposition will contain the leisure level itself, a variable for which we have not yet solved. It turns out that qualitative statements from the analysis can be made if we manage to determine how leisure reacts to tax changes, i.e. the sign of $\partial l/\partial\tau_i$.

From the criterion for a Laffer effect, $\partial G_0/\partial\tau_i<0$, we get the following proposition, using the auxiliary positive parameter Ψ_i:[29]

[29] $\Psi_k=\frac{1-\alpha}{\alpha}\frac{1-\tau_k}{1-\tau_{avg}}\tau_{avg}$ and $\Psi_h=\frac{1-\tau_h}{1-\tau_{avg}}\tau_{avg}$

Proposition 4 *In the case with leisure, there is a Laffer effect, $\partial G_0/\partial\tau_i < 0$, in the sense of definition (2), if* $\Gamma_i > 1 - \Psi_i \dfrac{\partial l/\partial\tau_i}{1-l}$.

The difference between propositions 4 and 2 is the term $\Psi_i \partial l/\partial\tau_i/(1-l)$. The scope for a definition 2 Laffer effect is larger or smaller in the model with leisure depending on the sign of the leisure derivative with respect to the tax rate. A positive tax derivative, $\partial l/\partial\tau_i > 0$, produces a larger change in β than in ϕ, as compared to the non-leisure case, thereby facilitating the desired $\partial\beta/\partial\tau_i - \partial\phi/\partial\tau_i < 0$. The opposite holds for a negative tax derivative, $\partial l/\partial\tau_i < 0$. The earlier requirement that the compositional distortion must be such that $\Gamma_i > 1$ in order to get a definition 2 Laffer effect is thus "relaxed" when there is leisure in the model (if $\partial l/\partial\tau_i > 0$).

5.3 Interpretation of Laffer effects, with leisure

The introduction of leisure into the model has opened up a new channel for dynamic as well as compositional Laffer effects. Although definition 1 Laffer effects are not studied in this section, we see that the labor/leisure level affects the growth rate. The private return $\phi \equiv A\alpha^\alpha(1-\alpha)^{1-\alpha}(1-\tau_k)^\alpha(1-\tau_h)^{1-\alpha}(1-l)^{1-\alpha}$ and therefore the growth rate, $\gamma = \sigma(\phi - \rho)$, will increase as either tax τ_i is reduced. If $\partial l/\partial\tau_i > 0$, there is an additional boost to the growth rate. This is a new dynamic effect. It is likely to make it easier for the government to maintain its spending path according to Laffer effect definition 1, as compared to the no-leisure case.

As for definition 2 effects, we once again get an expression for the compositional distortion required for there to be a Laffer effect from tax cuts. If $\partial l/\partial\tau_i > 0$, which is likely to be the case at least for the human capital tax, the agent will work more and production will increase. We start out in a situation with production βz where β is increasing in labor time $(1-l)$. This effect comes in addition to any effect through the h/k-ratio which, as previously, is determined by Ω. Thus, we have a new compositional margin that may increase production and therefore be a source of definition 2 Laffer effects. This is what is captured by proposition 4 above.

An analytical solution for the leisure level from (31) and (29) can only be obtained for certain values of α and the analysis is therefore only suggestive[30] . I have solved

[30] See appendix 2.

it for the special cases of $\alpha = 1/2$ and $\alpha = 1/3$. For a broad range of values on the other parameters I then go on to compare $\partial l/\partial \tau_h/(1-l)$ when (τ_h, τ_k)=$(0.5, 0.3)$ with $\partial l/\partial \tau_k/(1-l)$ when (τ_h, τ_k)=$(0.3, 0.5)$. I get that the following conditions are obeyed:

$$\frac{\partial l/\partial \tau_h}{1-l} >> \frac{\partial l/\partial \tau_k}{1-l} \text{ and } \frac{\partial l/\partial \tau_h}{1-l} > 0.$$

Changes in the human capital tax are thus likely to be more effective in creating Laffer effects. This applies to both the dynamic and the compositional part of the effect. Figure 2, which is explained in detail in the next paragraph, illustrates, for the case of $\alpha = 1/2$, the left- and right-hand sides of proposition 4 where the Γ_i-term, i.e. the left-hand side, should be larger than the leisure term, $1 - \Psi_i \partial l/\partial \tau_i/(1-l)$, for there to be a definition 2 Laffer effect.

Figure 2 displays Laffer effects according to propositions 2 and 4, for reductions in both τ_k and τ_h. The graph should be read by considering one tax rate only at a time.

First consider a reduction in τ_k. With $i = k$ in the graph, the horizontal axis is τ_k and the Γ_i-term is Γ_k. This Γ_k-term has been drawn for a human capital tax fixed at $\tau_h = 0.3$. The graph thus shows how high τ_k must be in order to get a compositional Laffer effect, from reducing τ_k when $\tau_h = 0.3$. According to the no-leisure case in proposition 2, there is a Laffer effect if $\Gamma_k > 1$. From the graph we see that τ_k must be at least 0.755 in order to get such an effect (intersection of Γ_i and "1").

In the leisure case it is instead the intersection of the leisure term, $1 - \Psi_k \dfrac{\partial l/\partial \tau_k}{1-l}$, with Γ_k, that matters. The leisure term is the downward-sloping thick concave line. It intersects Γ_k at a $\tau_k \approx 0.72$. Above this tax rate there is a Laffer effect. The effect of leisure, on the possibility to get a Laffer effect, is thus small, and large distortions are required to get Laffer effects in both the no-leisure and the leisure cases.

Now consider instead a reduction in τ_h. With $i = h$ in the graph, the horizontal axis is τ_h and the Γ_i-term is Γ_h. This Γ_h-term has been drawn for a physical capital tax fixed at $\tau_k = 0.3$. The graph thus shows how high τ_h must be in order to get a compositional Laffer effect, from reducing τ_h when $\tau_k = 0.3$. According to proposition 2, there is a Laffer effect if $\Gamma_h > 1$. From the graph we see that τ_h must be at least 0.755 in order to get such an effect (intersection of Γ_h and "1")[31] . In the leisure case it is instead the intersection of the leisure term, $1-\Psi_h\frac{\partial l/\partial\tau_h}{1-l}$, with Γ_h, that matters. The leisure term is now the dotted downward-sloping concave line. It intersects Γ_h at a $\tau_h \approx 0.50$. The effect of introducing leisure into the model, on the possibility to get a Laffer effect from human capital tax reductions, is thus large.[32]

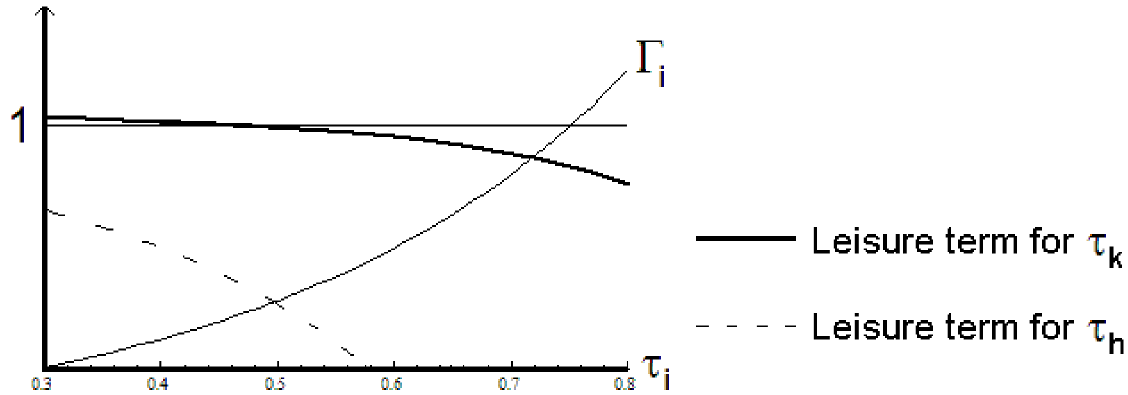

Figure 2. Requirements on τ_k to get a Laffer effect when τ_h is fixed at 0.3 and requirements on τ_h to get a Laffer effect when τ_k is fixed at 0.3, respectively (see text preceding graph).

I conclude the leisure section by stating that Laffer effects according to definition 3 are not possible in the case with leisure. The reasoning from proposition 3 remains unchanged.

Proposition 5 *In the case w*ith leisure, ther*e can never be a Laffer effect,* $\partial\left(G_0/\mathrm{GDP}\right)/\partial\tau_i<0$*, in the sense of definition (3).*

[31] This is the same as in the τ_k-case. The fact that $\alpha = 1/2$ is what allows us to display the two cases in one graph. If α takes a different value, the Γ_k and Γ_h-curves would not coincide.
[32] Parameters are: $\alpha = 0.5$, $A = 0.1$, $\theta = 2$, $\eta = 0.7$, $\rho = 0.02$ and half of government revenue goes to transfers.

6 Discussion

In the general version of the two-sector model, there is a separate sector accumulating human capital. An endogenous labor/leisure choice is also standard in such models. As a result of tax cuts, a period of transitional dynamics follows, during which stocks of human and physical capital allocated to each sector readjust. There is also an increase in the growth rate. In this paper, I have made simplifying assumptions regarding technology and the transition phase in order to separate different effects arising from tax cuts.

First, there are adjustments along a dynamic margin from the tax cut itself – the growth rate increases. This opens up for *dynamic* Laffer effects if we assume that government spending grows at its pre tax cut growth rate. With endogenous leisure, the growth rate also increases because more human capital is deployed, this is a second source of a *dynamic* Laffer effect.

Second, there are adjustments along a *compositional* margin if there is more than one factor of production. Tax cuts change the human capital to physical capital equilibrium composition. This affects production and is a source of Laffer effects. The level of leisure also changes as a result of the tax cut. This also changes production and is a second source of a *compositional* Laffer effect.

Propositions 1-5 summarize the results of this paper. They show that the possibility to get "self-financing" tax cuts - Laffer effects - depend on what is meant by "self-financing". It is indeed possible, for certain initial parameter combinations, to maintain (and even increase) period-zero government spending levels after a tax cut and then let spending grow with the new higher growth rate produced by the tax cut. As this is possible, it is then also possible to apply the less stringent requirement to maintain (and increase) period-zero government spending and let it grow only with its pre tax-cut growth rate.

Fulfilling the more stringent of these two requirements, where government spending grows with the new higher growth rate, depends on *compositional* effects from tax cuts and is, as stated above, facilitated when there is an endogenous leisure decision in the model. Fulfilling the less stringent requirement, where the growth rate of government spending does not adjust, is also helped by *dynamic* effects.

The Laffer effects in Ireland (1994) are dynamic effects. There is only one tax rate, no compositional adjustments, and any possible increase in government spending will be explained by *dynamic* effects. In Novales and Ruiz (2002), with two tax rates, leisure in the model and government spending growing with the pre tax-cut growth rate, possible Laffer effects are explained by a combination of *compositional* and *dynamic* Laffer effects. The present model separates and clarifies the reasons why Laffer effects are obtained in such a setting.

With an even stricter requirement, i.e. that the government spending to GDP ratio should increase to a higher level after the tax cut than before the tax cut, no such Laffer effects is possible. As this definition implies a jump in government spending to match adjustments in GDP, there are no further margins, compositional or dynamic, from which to get resources to increase government spending further.

Compositional and dynamic distortions need to be quite large in order to be able to maintain period-zero government spending levels and then let spending grow with the new higher growth rate. Fixing one tax rate at 30%, the other tax rate must be above 70% if a reduction in this higher tax is to be self-financing (for one parametrization of the model). If leisure is in the model, this requirement goes down. Through the direct effect on the consumption/leisure trade-off, a human capital tax cut produces a direct effect which is not present for a physical capital tax cut. Therefore, human capital tax cuts are a more likely source of Laffer effects when there is leisure in the model. For one parametrization, fixing the physical capital tax rate at 30% requires the human capital tax rate to be above 50% to get a Laffer effect (instead of above 75% when leisure is not in the model).

Bibliography

[1] Agell, J., Persson, M., 2001. On the analytics of the dynamic Laffer curve. Journal of Monetary Economics 48, 397-414.

[2] Auerbach, A., 2005. Dynamic Scoring: An Introduction to the Issues. American Economic Review 95, 421–425.

[3] Barro, R., Sala-i-Martin, X., 1995. Economic Growth. Cambridge, MA: MIT Press.

[4] Bianconi, M., 2000. The effects of alternative fiscal policies on the intertemporal government budget constraint. International Review of Economics and Finance 9, 31–52.

[5] Bruce, N., Turnovsky, S., 1999. Budget balance, welfare, and the growth rate: "dynamic scoring" of the long-run government budget. Journal of Money, Credit, and Banking 31, 162–186.

[6] Devereux, M., Love, D., 1994. The effects of factor taxation in a two-sector model of endogenous growth. The Canadian Journal of Economics 27, 509-536.

[7] Ireland, P., 1994. Supply-side economics and endogenous growth. Journal of Monetary Economics 33, 559–571.

[8] King, R., Plosser, C., Rebelo, S., 1988. Production, growth and business cycles. II. New directions. Journal of Monetary Economics 21, 309–341.

[9] King, R., Rebelo, S., 1990. Public policy and economic growth: developing neoclassical implications. Journal of Political Economy 98, 126–150.

[10] Leeper, E., Yang, S., 2008. Dynamic Scoring: Alternative financing schemes. Journal of Public Economics 92, 159-182.

[11] Lucas, R., 1988. On the mechanics of Economic Development. Journal of Monetary Economics 22, 3-42.

[12] Lucas, R., 1990. Supply-side economics: an analytical review. Oxford Economic Papers.42, 293–316.

[13] Mankiw, G., Weinzierl, M., 2006. Dynamic scoring: A back-of-the-envelope guide. Journal of Public Economics 90, 1415-1433.

[14] Milesi-Ferretti, G., Roubini, N., 1998a. On the taxation of human and physical capital in models of endogenous growth. Journal of Public Economics 70, 237–254.

[15] Milesi-Ferretti, G., Roubini, N., 1998b. Growth effects of income and consumption taxes. Journal of Money, Credit, and Banking 30, 721–744.

[16] Novales, A., Ruiz, J., 2002. Dynamic Laffer curves. Journal of Economic Dynamics & Control 27, 181-206.

[17] Pecorino, P., 1993. Tax structure and growth in a model with human capital. Journal of Public Economics 52, 251-271.

[18] Pecorino, P., 1994. The Growth Rate Effects of Tax Reform. Oxford Economic Papers 46, 492-501.

[19] Pecorino, P., 1995. Tax rates and tax revenues in a model of growth through human capital accumulation. Journal of Monetary Economics 36, 527–539.

[20] Rebelo, S., 1991. Long-Run Policy Analysis and Long-Run Growth. The Journal of Political Economy 99, 500–521.

[21] Stokey, N., Rebelo, S., 1995. Growth effects of flat-rate taxes. The Journal of Political Economy 103, 519–550.

Appendix

A1: Relationship to the general two-sector model

The model in this paper uses two assumptions regarding the production function and reversibility of investments that require some motivation. In discussing these assumptions and their implications the presentation is brief, for additional details see the work by Milesi-Ferretti and Roubini (1998a, 1998b) and papers referenced therein.[33] Barro and Sala-i-Martin (1995) also discuss the model used here. For simplicity, I discuss the model assumptions in a setting without leisure. The first assumption is that output and human capital are produced with the same production function. Second, there are no restrictions on deinvestment in the stocks of H and K.

For the first assumption, consider the general model where output is produced with the following production function, $Y = A_K (vK)^{\alpha_1} (uH)^{1-\alpha_1}$ and human capital is produced with $\dot{H} = A_H ((1-v) K)^{\alpha_2} ((1-u) H)^{1-\alpha_2}$. Physical capital input is divided according to vK used for final good production and $(1-v) K$ for human capital production and there is a similar division of human capital input. As discussed at some length in Barro and Sala-i-Martin (1995), this model is difficult to analyze in its general version. If $\alpha_1 = \alpha_2 \equiv \alpha$, however, things simplify a lot. From the first-order conditions, we get that when $\alpha_1 = \alpha_2$, the marginal impact of increasing the fraction v has the same impact on final good production relative to human capital production as an additional unit of u has on output production relative to human capital production. Therefore, $v = u$, so that irrespective of the global H/K-ratio, an equal fraction of each stock is deployed in final goods production and the rest in human capital production. As a consequence, the relative price of human capital produced to final good produced is unaffected by the global H/K-ratio and it is also unaffected by v (because u adjusts to al-

[33] Milesi-Ferretti and Roubini (1998a, 1998b) study growth responses from capital taxation in a catalogue of models using physical and human capital as factors of production and in which there is a different technology for producing human capital than for the final good. They use different versions of production functions in the two sectors; they separate the two cases where human capital is or is not a market activity and also elaborate on different leisure specifications.

In the terminology of Milesi-Ferretti and Roubini, the model in this paper falls under the category "raw-time" leisure, human capital is a market good and physical and human capital are produced with the same production function.

ways be equal to v). The relative price is therefore constant and being equal to $p = A_K/A_H$. Broad output can therefore be simplified to become $Q \equiv Y + p\dot{H} = A_K (vK)^{\alpha} (vH)^{1-\alpha} + pA_H ((1-v) K)^{\alpha} ((1-v) H)^{1-\alpha} = A_K K^{\alpha} H^{1-\alpha}$. This is the "broad" production function used in this paper.

The second assumption is that the model does not exhibit any transitional dynamics. The relative price of human to physical capital is constant and a unit of physical capital can be "deinvested" and instantaneously converted into a unit of human capital. Transitional dynamics could be included in the model by adding non-negativity constraints on capital accumulation. Outside the balanced growth H/K-ratio, the stock which is relatively abundant would remain constant for a finite time whereas the other stock would catch up until the equilibrium ratio is reached. Although there are interesting properties of this transition, it is not the objective of this paper and it would considerably complicate the analysis of Laffer effects.

The assumption of $\alpha_1 = \alpha_2$ means that human capital production is as intensive in physical capital input as is the production of output. A more realistic assumption is that human capital is relatively intensive in human capital input, $\alpha_2 < \alpha_1$. This is the Uzawa-Lucas model and in its extreme version $\alpha_2 = 0$ such that $\dot{H} = A_H (1-u) H$. Milesi-Ferretti and Roubini (1998a), King and Rebelo (1990) and Rebelo (1991) all discuss this model with respect to capital taxation. Growth along the balanced growth path will be driven by human capital accumulation and in the case of a taxed human capital sector, the growth rate is a function of τ_h but not of τ_k. Changes in τ_k will cause changes in the H/K-ratio used in producing output, but will not affect the growth rate. For Laffer effects, only the human capital tax has a dynamic effect. Presumably; in the intermediate model with $0 < \alpha_2 < \alpha_1$, the lower is α_2, the more important is τ_h for the possibility to obtain Laffer effects, although τ_k plays some role. In the case with $\alpha_1 = \alpha_2$, there is complete symmetry between the two taxes (absent leisure).

A2: Analytical solution with leisure

By combining the intertemporal relation (31) with the time-zero version of the intratemporal constraint (29), consumption can be eliminated to get one equation in one unknown[34] , the time-zero leisure level l_0:

$$l_0\left(C_1+C_2\right)=C_2+C_3\left(1-l_0\right)^{\alpha}.$$

This equation cannot be solved in the general case. Instead, I solve it for the special cases of $\alpha = 1/2$ and $\alpha = 1/3$. The leisure level is a non-trivial function of initial transfers T_0 to initial capital z_0 and all parameters including the tax rates[35] , $l\left(T_0/z_0, A, \theta, \eta, \rho, \tau_h, \tau_k\right)$. It can be differentiated, however, so that we get analytical expressions for the tax derivatives $\partial l/\partial \tau_h$ and $\partial l/\partial \tau_k$.

When $\partial l/\partial \tau_h$ is calculated I set $\tau_h > \tau_k$ and vice versa when computing $\partial l/\partial \tau_k$. This is done because we are interested in tax cuts where we decrease the highest tax. As an example, if we use τ_h=0.5 and τ_k=0.3 to evaluate l and $\partial l/\partial \tau_h$ to get the term $\partial l/\partial \tau_h/\left(1-l\right)$, this tax cut is compared to a situation where τ_k=0.5 and τ_h=0.3 to evaluate l and $\partial l/\partial \tau_k$ to get the term $\partial l/\partial \tau_k/\left(1-l\right)$.

As discussed in the main text, the intratemporal condition (29) is such that an increase in either tax makes the agent substitute consumption for leisure. In the intertemporal relationship (31), a tax cut will reduce the consumption level for a constant leisure level (for $\sigma < 1$). The full intertemporal effect also depends on the reaction of the leisure level to the tax cut. A non-ambiguous statement regarding the sign can therefore not be made. However, the relationship between c and l in (31) is symmetric in the tax rates, so that any difference in the response in l between changes in τ_h and τ_k will be due to the intratemporal constraint.

Having taken derivatives, I derive comparative statics of l and $\partial l/\partial \tau_i$ with respect to the variables (T_0/z_0,A, θ, η, ρ) varying one parameter at a time.[36] The main results can be summarized as follows

[34] The three constants are

$$C_1=\frac{\eta}{1-\eta}\frac{Az_0\alpha^{\alpha}(1-\alpha)^{2-\alpha}(1-\tau_h)^{2-\alpha}(1-\tau_k)^{\alpha}}{\alpha\left(1-\tau_k\right)+\left(1-\alpha\right)\left(1-\tau_h\right)}$$

$$C_2=A\left(1-\left(1+\eta\theta-\eta\right)^{-1}\right)\alpha^{\alpha}(1-\alpha)^{1-\alpha}(1-\tau_k)^{\alpha}(1-\tau_h)^{1-\alpha}$$

$$C_3=\frac{T_0}{z_0}+\rho\left(1+\eta\theta-\eta\right)^{-1}$$

[35] For the case of $\alpha = 1/2$;

$$l=\frac{C_2}{C_1+C_2}-\frac{C_3^2}{2\left(C_1+C_2\right)^2}+\frac{\left(C_3^4+4C_1^2+4C_1C_2\right)^{1/2}}{2\left(C_1+C_2\right)^2}$$

Because l is constant, the subindex 0 is dropped.

[36] Typical parameter values over which I have evaluated the expressions are $0.1 \leq A \leq 0.5$,

- As expected, the level of leisure l is decreasing in the preference for consumption η. It is also decreasing in A so that the substitution effect dominates and the agent works more when A increases.

- From the intertemporal constraint (31), period zero consumption is increasing in the time preference factor ρ and in T_0/z_0. The intratemporal c/l-ratio is not affected, however, meaning that leisure must also be increasing in ρ and T_0/z_0.

- The only ambiguous comparative static is with respect to θ as it will depend on the preference parameter η. For low values of η, the leisure level is decreasing in θ and for high values of η, leisure is increasing in θ.[37]

For the parametrizations of (τ_k, τ_h, T_0/z_0,A, θ, η, ρ) and for both values of α, we get the result that

$$\frac{\partial l/\partial \tau_h}{1-l} >> \frac{\partial l/\partial \tau_k}{1-l} \text{ and } \frac{\partial l/\partial \tau_h}{1-l} > 0.$$

In most cases also $\partial l/\partial \tau_k/\left(1-l\right)$ is positive.

At least for the parametrizations used it is thus possible to say that, in order to get a Laffer effect, a human capital tax cut helps a great deal more than a physical capital tax cut. The model delivers what the basic intuition tells us, i.e. that the direct effect on the price of leisure creates a larger effect on the leisure level from human capital tax cuts than from physical capital tax cuts. The lower level of leisure delivered by the tax cut then helps producing a definition 2 Laffer effect according to proposition 4.

$1.1 \leq \theta \leq 5$, $0.1 \leq \eta \leq 0.9$, $0 \leq \rho \leq 0.05$. In doing the comparative statics, T_0/z_0 can be no higher than $\tau_{avg}\beta$ which is the case when all government revenue is used for transfers. Choosing T_0/z_0 thus amounts to choosing the ratio of government revenue that is used for transfers and we let this ratio vary between 0.1 and 0.9. Furthermore, I set the lowest tax rate at 0.3 and increase the other tax up to 0.9. I check that there is positive growth and that the transversality condition holds, for all parameter combinations.

[37] From $\sigma = (1-\eta+\eta\theta)^{-1}$, we see that low values of η make σ very close to unity, substitution and income effects cancel out, and the fraction of lifetime wealth that is consumed in period zero is $(\phi - \gamma) \approx \sigma\rho$. The effect of a change in σ is then mainly to make the effective time discount factor larger, increasing consumption in period zero and also leisure (through the intratemporal constraint that has not changed). For large values of η, even small increases in θ will lead to large changes in σ so that the fraction of lifetime wealth consumed in period zero now decreases in σ. Period zero consumption will then decrease in σ and so will the leisure level. Because σ is inversely proportional to θ, the comparative statics with respect to θ follow.

MONOGRAPH SERIES

1. Michaely, Michael *The Theory of Commercial Policy: Trade and Protection*, 1973
2. Söderström, Hans Tson *Studies in the Microdynamics of Production and Productivity Change,* 1974
3. Hamilton, Carl B. *Project Analysis in the Rural Sector with Special Reference to the Evaluation of Labour Cost*, 1974
4. Nyberg, Lars and Staffan Viotti *A Control Systems Approach to Macroeconomic Theory and Policy in the Open Economy*, 1975
5. Myhrman, Johan *Monetary Policy in Open Economies*, 1975
6. Krauss, Melvyn *International Trade and Economic Welfare*, 1975
7. Wihlborg, Clas *Capital Market Integration and Monetary Policy under Different Exchange Rate Regimes*, 1976
8. Svensson, Lars E.O. *On Competitive Markets and Intertemporal Resources Allocation*, 1976
9. Yeats, Alexander J. *Trade Barriers Facing Developing Countries*, 1978
10. Calmfors, Lars *Prices, Wages and Employment in the Open Economy*, 1978
11. Kornai, János *Economics of Shortage*, Vols I and II, 1979
12. Flam, Harry *Growth, Allocation and Trade in Sweden. An Empirical Application of the Heckscher-Ohlin Theory*, 1981
13. Persson, Torsten *Studies of Alternative Exchange Rate Systems. An Intertemporal General Equilibrium Approach*, 1982
14. Erzan, Refik *Turkey's Comparative Advantage, Production and Trade Patterns in Manufactures. An Application of the Factor Proportions Hypothesis with Some Qualifications*, 1983
15. Horn af Rantzien, Henrik *Imperfect Competition in Models of Wage Formation and International Trade*, 1983
16. Nandakumar, Parameswar *Macroeconomic Effects of Supply Side Policies and Disturbances in Open Economies*, 1985
17. Sellin, Peter *Asset Pricing and Portfolio Choice with International Investment Barriers*, 1990
18. Werner, Ingrid *International Capital Markets: Controls, Taxes and Resources Allocation*, 1990
19. Svedberg, Peter *Poverty and Undernutrition in Sub-Saharan Africa: Theory, Evidence, Policy*, 1991
20. Nordström, Håkan *Studies in Trade Policy and Economic Growth*, 1992
21. Hassler, John, Lundvik, Petter, Persson, Torsten and Söderlind, Paul *The Swedish Business Cycle: Stylized facts over 130 years*, 1992
22. Lundvik, Petter *Business Cycles and Growth*, 1992
23. Söderlind, Paul *Essays in Exchange Rates, Business Cycles and Growth*, 1993

24. Hassler, John A.A. *Effects of Variations in Risk on Demand and Measures of Business Cycle Comovements*, 1994
25. Daltung, Sonja *Risk, Efficiency, and Regulation of Banks*, 1994
26. Lindberg, Hans *Exchange Rates: Target Zones, Interventions and Regime Collapses*, 1994
27. Stennek, Johan *Essays on Information-Processing and Competition*, 1994
28. Jonsson, Gunnar *Institutions and Incentives in Monetary and Fiscal Policy*, 1995
29. Dahlquist, Magnus *Essays on the Term Structure of Interest Rates and Monetary Policy*, 1995
30. Svensson, Jakob *Political Economy and Macroeconomics: On Foreign Aid and Development*, 1996
31. Blix, Mårten *Rational Expectations and Regime Shifts in Macroeconometrics*, 1997
32. Lagerlöf, Nils-Petter *Intergenerational Transfers and Altruism*, 1997
33. Klein, Paul *Papers on the Macroeconomics of Fiscal Policy*, 1997
34. Jonsson, Magnus *Studies in Business Cycles*, 1997
35. Persson, Lars *Asset Ownership in Imperfectly Competitive Markets*, 1998
36. Persson, Joakim *Essays on Economic Growth*, 1998
37. Domeij, David *Essays on Optimal Taxation and Indeterminacy*, 1998
38. Flodén, Martin *Essays on Dynamic Macroeconomics*, 1999
39. Tangerås, Thomas *Essays in Economics and Politics: Regulation, Elections and International Conflict*, 2000
40. Lidbom, Per Pettersson *Elections, Party Politics and Economic Policy*, 2000
41. Vestin, David *Essays on Monetary Policy*, 2001
42. Olofsgård, Anders *Essays on Interregional and International Political Economics*, 2001
43. Johansson, Åsa *Essays on Macroeconomic Fluctuations and Nominal Wage Rigidity*, 2002
44. Groth, Charlotta *Topics on Monetary Policy*, 2002
45. Gancia, Gino A. *Essays on Growth, Trade and Inequality*, 2003
46. Harstad, Bård *Organizing Cooperation: Bargaining, Voting and Control*, 2003
47. Kohlscheen, Emanuel *Essays on Debts and Constitutions*, 2004
48. Olovsson, Conny *Essays on Dynamic Macroeconomics*, 2004
49. Stavlöt, Ulrika *Essays on Culture and Trade*, 2005
50. Herzing, Mathias *Essays on Uncertainty and Escape in Trade Agreements*, 2005
51. Bonfiglioli, Alessandra *Essays on Financial Markets and Macroeconomics*, 2005

52. Pienaar, Natalie *Economic Applications of Product Quality Regulation in WTO Trade Agreements,* 2005
53. Song, Zheng *Essays on Dynamic Political Economy*, 2005
54. Eisensee, Thomas *Essays on Public Finance: Retirement Behavior and Disaster Relief*, 2005
55. Favara, Giovanni *Credit and Finance in the Macroeconomy*, 2006
56. Björkman, Martina *Essays on Empirical Development Economics: Education, Health and Gender*, 2006
57. Larsson, Anna *Real Effects of Monetary Regimes*, 2007
58. Prado, Jr., Jose Mauricio *Essays on Public Macroeconomic Policy*, 2007
59. Tonin, Mirco *Essays on Labor Market Structures and Policies*, 2007
60. Queijo von Heideken, Virgina *Essays on Monetary Policy and Asset Markets*, 2007
61. Finocchiaro, Daria *Essays on Macroeconomics*, 2007
62. Waisman, Gisela *Essays on Discrimination and Corruption*, 2008
63. Holte, Martin Bech *Essays on Incentives and Leadership*, 2008
64. Damsgaard, Erika Färnstrand *Essays on Technological Choice and Spillovers,* 2008
65. Fredriksson, Anders *Bureaucracy, Informality and Taxation: Essays in Development Economics and Public Finance*, 2009