

Essays on Macroeconomics and International Trade

by

Martí Mestieri

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

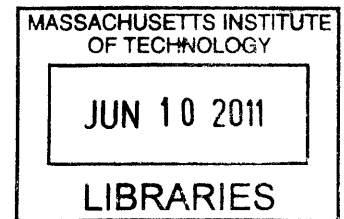
Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

ARCHIVES



© © Martí Mestieri. Some rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly
paper and electronic copies of this thesis document in whole or in part.

Author
Department of Economics
May 15, 2011

Certified by
Daron Acemoglu
Elizabeth and James Killian Professor of Economics
Thesis Supervisor

Certified by
Abhijit Vinayak Banerjee
Ford Foundation International Professor of Economics
Thesis Supervisor

Certified by
Robert M. Townsend
Elizabeth & James Killian Professor of Economics
Thesis Supervisor

Accepted by
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Chairman, Departmental Committee on Graduate Studies

Essays on Macroeconomics and International Trade

by

Martí Mestieri

Submitted to the Department of Economics
on May 15, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis focuses on the study of different aspects of income inequality across and within countries. In the first chapter, I study how the optimal provision of human capital is distorted in the presence of borrowing constraints and private information on talent and wealth. It shows that elitist, non-merit based, access to higher education can be constrained optimal in poor and unequal countries. The second chapter documents how the IT revolution has changed the patterns of North-South trade and analyzes its effects on wage inequality. It provides theoretical and empirical results on wage polarization and a changes in the pattern of specialization. Finally, the third chapter provides a framework for estimating technological diffusion across countries. The framework is applied to study the diffusion of major technologies across the world since the Industrial Revolution. It is shown that differences in technology diffusion in the last two hundred years can account for two thirds of current income per capita differences.

Thesis Supervisor: Daron Acemoglu
Title: Elizabeth and James Killian Professor of Economics

Thesis Supervisor: Abhijit Vinayak Banerjee
Title: Ford Foundation International Professor of Economics

Thesis Supervisor: Robert M. Townsend
Title: Elizabeth & James Killian Professor of Economics

Acknowledgments

I would like to thank my advisors, Daron Acemoglu, Abhijit Banerjee and Robert Townsend for their help and support. Arnaud Costinot has helped me greatly as well, specially for the chapter two of the thesis. I have had the pleasure to work with Sergi Basco and Diego Comín and I hope these collaborations continue in the future. Chapter 2 is written in collaboration with Sergi Basco and Chapter 3 is written in collaboration with Diego Comín.

Mar Reguant has helped me with her love and support throughout the Ph.D. and she is in great part responsible that I did not switch back to study quantum mechanics after passing the generals. I would like to thank my family for their unending love and support. I have had the pleasure of discussing my research ideas with many Ph.D. students at MIT. These discussions have not only been helpful in the chapters I present in the thesis (each chapter has a (probably incomplete) acknowledgment list) but for other research projects as well.

I would like to thank the professors at Universitat Autònoma de Barcelona who helped me to get started for this journey. Specially, I would like to thank Jordi Caballé, Inés Macho and David Pérez-Castrillo. I would to thank Teresa Garcia-Milà as well.

Finally, I would like to acknowledge the financial support of Fundació "La Caixa", Banco de España and Fundación Ramón Areces during my Ph.D.

Contents

1 Wealth Distribution and Human Capital: How Borrowing Constraints Shape Educational Systems	11
1.1 Introduction	12
1.2 Outline of the Paper	17
1.3 The Environment	21
1.4 First Best	23
1.5 Private Information without Borrowing Constraints	24
1.6 Private Information with Borrowing Constraints	26
1.6.1 The Mechanism Design Problem	26
1.6.2 Solution with School Fees Only	34
1.6.3 School Fees and Exams	37
1.7 Extensions	42
1.7.1 Yeomen farmers and general spillovers	43
1.7.2 Continuum of Schools	44
1.8 Appendix: Model with Negative Transfers	57
1.8.1 The mechanism design problem	57
1.8.2 Decentralization: A Fair Lottery Market on Wealth	58
1.8.3 School fees	58
1.9 Appendix: Proofs	58
1.10 Appendix: Optimal Test-Fee Schedule Problem	66
1.10.1 Optimal Control Formulation of the Inner Problem	68
1.10.2 Calculus of Variations Formulation of the Inner problem	70

1.10.3	Formulation of the Outer Problem	70
2	Heterogeneous Trade Costs and Wage Inequality: A Model of Two Globalizations	73
2.1	Introduction	74
2.2	Motivating Evidence and Related Literature	78
2.2.1	Motivating Evidence	78
2.2.2	Related Literature	80
2.3	Model	81
2.3.1	Baseline Model	82
2.3.2	Trade equilibrium	83
2.4	Main Results	84
2.4.1	The two Globalizations and their Complementarity	84
2.4.2	Two Souths and the Moving Band	88
2.5	Extensions	91
2.5.1	A Model with Endogenous Labor Supply	91
2.5.2	Technology Adoption	93
2.6	Concluding Remarks	94
2.7	Proofs and Auxiliary Propositions	101
2.8	Data Appendix	106
2.9	Tables	106
2.10	Figures	108
3	The Intensive Margin of Technology Adoption	117
3.1	Introduction	118
3.2	A one-sector growth model with extensive and intensive technology adoption	121
3.2.1	Preferences	122
3.2.2	Production	122
3.2.3	Factor demands, output, and optimal adoption	124
3.3	Diffusion of the new technology	128
3.3.1	Empirical application	129
3.3.2	Identification and estimation procedure	130

3.3.3	Non-homotheticities	132
3.4	Results	134
3.4.1	Estimated Intensive Margin	135
3.4.2	Development accounting	139
3.5	Conclusion	141
3.6	Appendix: Data	145
3.7	Appendix: Equilibrium and diffusion of the new technology	146
3.8	Tables and Figures	147

Chapter 1

Wealth Distribution and Human Capital: How Borrowing Constraints Shape Educational Systems*

Abstract

This paper provides a theory of how the wealth distribution of an economy affects the optimal design of its educational system. The model features two key ingredients. First, agents are heterogeneous both in their ability and wealth levels, neither of which is observable. Second, returns to schooling depend on the ability-composition of agents attending each school tier, for example, because of choices of common curricula. An educational system is characterized by an assignment rule of agents to schools and by endogenous sizes of tiers. I find that a benevolent planner seeking to maximize economic efficiency implements “elitist” educational systems in economies with poor, borrowing-constrained, agents. Compared to the first best, the optimal solution features (i) relatively low-ability, rich agents selecting into higher education and (ii) higher education schools with less capacity. The same qualitative results obtain when only two commonly used instruments are available to the planner: school fees and exams. In addition, I show that economies with relatively tighter borrowing constraints rely more extensively on exams, and that agents performing better on exams are rewarded with lower school fees.

*I thank Daron Acemoglu, Abhijit Banerjee and Robert Townsend for their guidance and support. I also thank Sergi Basco, Diego Comin, Arnaud Costinot, Fernando Duarte, Jonathan Goldberg, Pablo Kurlat, Guido Lorenzoni, Mónica Martínez-Bravo, Michael Peters, Mar Reguant, Jenny Simon, Iván Werning, Juan Pablo Xandri and participants at the MIT Macro Seminar for helpful comments and discussions. All errors are my own.

1.1 Introduction

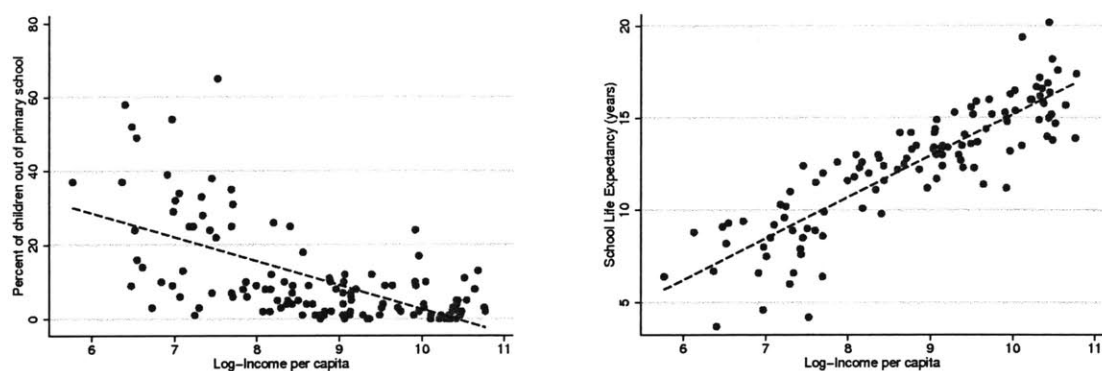
Educational systems in developing and rich countries differ in many respects. In particular, higher education in developing countries has lower attendance rates and lower educational achievement as measured by international tests. Moreover, access to higher education relies much more extensively on gate-keeping exams. A conventional view explaining these differences is that "...in many developing countries governments lack either the financial resources or the political will to meet their citizens' educational needs..."¹ These differences in the provision of human capital have lead many observers to emphasize the role of educational systems in developing countries as a means of generating and perpetuating the ruling elites (e.g., Engerman and Sokoloff (2000, 2002)).

This paper presents an alternative theory for which these same differences in educational systems occur even when governments seek to maximize aggregate welfare. The purpose of this theory is to emphasize that, when there are borrowing-constrained agents with private information on their valuation of education and wealth, there are economic forces pushing benevolent governments to design seemingly elitist educational systems. I illustrate this by showing that, even when education can be provided at *no* cost, a benevolent social planner implements a system in which higher education features (i) low attendance rates, (ii) a dampened ability-composition of agents attending higher education, and (iii) an allocation process for higher education that relies extensively on gate-keeping exams and rationing for poor agents in the form of lotteries.

The intuition for this result stems from the fact that access to education is a source of rents for agents. The combination of borrowing constraints and private information makes it difficult to separate true valuations from willingness to pay. Thus, in order to separate low-ability, unconstrained agents from high-ability but constrained agents, the educational system adopts additional screening mechanisms. This results in the usage of lotteries for poor people and extensive reliance on gate-keeping exams to access higher education (because it is less costly for high ability types to prepare exams). However, these are imperfect screening mechanisms. Therefore, in equilibrium, the ability-composition of agents that select into higher education is worse and the capacity of the higher education tier is reduced relative to an economy without borrowing-constrained agents.

The three stylized facts presented above, which emerge as the solution of the planner's problem in the presence of borrowing constraints, are well-documented features of educational systems in developing countries. First, there is ample evidence on more extensive use of gate-keeping exams in developing countries relative to richer countries, especially in Africa and Latin America, (Al-Samarrai and Peasgood, 1998; Kellaghan and Greaney, 1992;

¹This excerpt is from an article by Hillman and Jenkner prepared for the IMF publication "Economic Issues", <http://www.imf.org/external/pubs/ft/issues/issues33/index.htm>



(a) Percent of children not enrolled in primary school in 2005. (Source: UNESCO) (b) School life expectancy (in years) conditional on attending primary school. (Source: UNESCO)

Figure 1-1: Differences in School Enrollment

Kellaghan, 2004; Lockheed and Mete, 2007; Mete, 2004). For example, Kellaghan (2004) and Kellaghan and Greaney (1992) document that in most African countries, three (if not more) major examinations are required to complete secondary education.² Kellaghan emphasizes the role of examinations as gatekeepers and argues that this is reflected in the large numbers of students who fail exams and repeat their grade. This is consistent with UNESCO's data for 2005, which shows that the repetition rate at fifth grade (before accessing secondary education) in developing countries is 8.7% on average, versus 1.9% in rich countries.

Second, another difference between developing and rich countries is enrollment rates. Figure 1-1a shows that the fraction of children not enrolled in primary school is higher in low income countries. These differences in attendance rates are exacerbated as one moves forward in the education system. Figure 1-1b shows that the expected number of years children stay in school conditional on having some primary schooling is increasing in income per capita.³

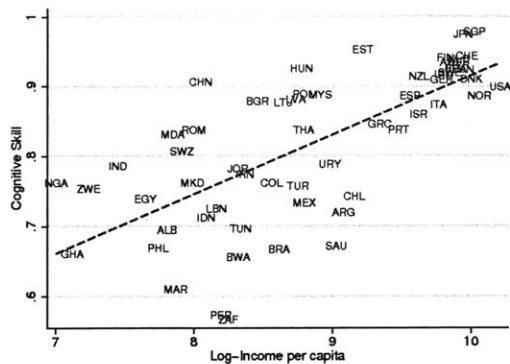
Third, differences in school quality are documented by Hanushek and Woessmann (2008, 2009) for a cross-section of countries. The authors construct an index based on results on a set of international tests to proxy for school quality and find significant cross-country differences.^{4,5} Figure 1-2a shows how their measure of education quality is positively correlated

²These examinations are typically at the end of primary schooling, after two or three years in secondary school and around the end of secondary school. Kellaghan and Greaney (1992) document that in Francophone African countries students tend to be subject to even more exams. In particular, additional examinations are administered during primary school and, also, a competitive examination, termed the *concours*, is used to select pupils for the next education level.

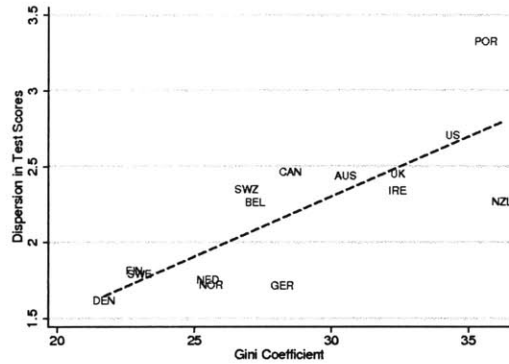
³The implicit assumption in this argument is that poorer countries have more borrowing constrained agents.

⁴Measures of school quality have been developed prior to these studies. For example, Hanushek and Kimko (2000) use a similar approach. Note that this concept of school quality differs from the approach of Barro and Lee (1993).

⁵Hanushek and Woessmann construct a cross-country comparable measure of acquired cognitive skills to proxy for education quality. The international student achievement tests they use include the following. Trends



(a) Cognitive Skill Score constructed from international tests and income per capita. (Source: Hanushek and Woessmann (2009))



(b) Education quality dispersion and inequality. Dispersion is the ratio 95th to 5th percentile of the average score of quantitative and prose in IALS. (Source: Nickell (2004))

Figure 1-2: Differences in School Quality

with income per capita. Nickell (2004) documents an additional correlation between wealth inequality and dispersion in school quality. Countries with a more unequal wealth distribution tend to have more dispersion in quality measures. This is shown in Figure 1-2b.⁶

Next, I discuss in more detail the main elements and results of the paper. The theory presented rests on two central elements. The first element is heterogeneity in agents' characteristics, ability and wealth, both of which are private information. The second element is the existence of complementarities in human capital formation across agents with the same level of education. A natural explanation for this complementarity is that the curriculum requirement of each education tier adjusts to students' ability.

In this context, an educational system is characterized by an assignment rule of agents to schools, and capacities of tiers. I characterize the educational system that a social planner would design and its decentralization under perfect capital markets and borrowing constraints. First, I show that in economies in which there are no borrowing constraints, private information alone does not prevent the educational system from being first best. In these economies, the educational system is meritocratic in the sense that agents are matched to different school tiers according to their ability. Moreover, the first best educational system can be decentralized through a market for schooling, even in the presence of private information.

in International Mathematics and Science Study, Programme for International Student Assessment and Progress in International Reading Literacy Study. In order to establish a baseline to compare the performance in different tests, the authors use the United States National Assessment of Education Progress. The reason is that this is the only test that has been administered consistently over a large period of time.

⁶Note that in this case the sample is limited to a particular international test in order to have a clear interpretation of the variance in the data. The method in Hanushek and Woessmann (2009) is not designed to generate comparable second moments.

Then, I turn to the study of the main object of interest of the paper: economies with borrowing constraints. Borrowing constraints generate a wedge between private valuation of education and ability to pay, as agents are constrained in the maximal transfer they can make. This distorts the matching of agents to schools because of the inability of agents to effectively signal their true valuations. I find that the optimal mechanism involves randomization in access to schooling for high-ability, poor agents, while high-ability, rich agents do not face any randomness in allocation. The capacity of the higher education tier is reduced relative to the economy without borrowing constraints. This is consistent with the evidence presented of low attendance rates in developing countries.

The comparative statics on wealth distribution show that in poorer countries the average ability of agents selecting into higher education is reduced. Thus, due to the complementarities in human capital formation, this endogenously reduces the human capital obtained in higher education in developing countries – which is consistent with worse performance in international tests. Changes in wealth dispersion have opposite effects depending on whether or not the median wealth type can afford higher education with certainty. If the original equilibrium features an allocation in which only agents above the median wealth can afford higher education without resorting to lotteries, an increase in wealth dispersion makes it optimal to restrict even further access to higher education, making the educational system more exclusive. Analogous comparative statics results obtain in the case that the social planner can only use school fees as instruments.⁷

Finally, I study an environment in which both school fees and a signaling technology (exams) can be used. There is a trade-off in using exams. They involve wasteful spending in order to be prepared, but allow for an additional screening mechanism because it is less costly for high ability agents to pass an exam. The optimal mechanism is such that agents that perform better in an exam are rewarded with a lower school fee. Thus, this mechanism resembles a scholarship scheme. The comparative statics on wealth distribution show that poorer countries use relatively more exams, and that exams are particularly used in the range in which there is more wealth inequality. These results fit with the evidence presented on extensive use of gate-keeping exams in developing countries.

Related Literature This paper emphasizes the role of asymmetric information and borrowing constraints to explain differences in the design of an educational system, and, ultimately, human capital provision. In this sense, while I focus on a different set of factors, this paper shares the approach of Banerjee (1997) and Esteban and Ray (2006) of focusing on asymmetric information and borrowing constraints to rationalize differences in provision of goods.

The paper relates to a rich and diverse literature on the determinants of human capital

⁷In fact, Section 1.6 shows that if the social planner cannot commit to exclude some high-ability poor agents from education once they reveal their type, the only credible instrument the planner can use are school fees.

acquisition. To the best of my knowledge, however, this is the first attempt to provide a theory of the design of an optimal educational system that focuses on the role of private information and borrowing constraints in matching of agents to schools.

Fernández and Galí (1999) and Fernández (1998) are the closest papers in terms of the framing of the problem. They study a matching problem with borrowing constraints and compare two alternative mechanisms (prices and exams). This paper differs from theirs in that it takes a mechanism design approach, thus endogenizing the usage of different instruments and the size of tiers. Moreover, this paper provides comparative statics results on the wealth distribution. Another important difference is that, in this paper, educational standards are set endogenously. In this respect, Costrell (1994) and Betts (1998) provide alternative theories on the determinants of educational standards, but they emphasize political economy reasons rather than private information and borrowing constraints.

The problem of allocating heterogeneous agents to schools studied in this paper can be interpreted as an extension of the assignment Roy's model (Sattinger, 1993), in which private information and borrowing constraints are introduced. With the exception of the aforementioned work of Fernández and Galí, the literature has typically analyzed other imperfections in the assignment process. For example, Legros and Newman (2007) and Durlauf and Seshadri (2003) study conditions under which monotone matching obtains in environments with non-transferabilities and endogenous coalition sizes. In this paper, the complementarity between ability and school tier and the fact that agents appropriate all the surplus from the match makes positive assortative matching efficient.⁸

The mechanism design problem considered in the presence of borrowing constraints constitutes a bi-dimensional screening problem. This type of problem has been studied in auction design by Che and Gale (1998, 2000) and Lewis and Sappington (2000, 2001).⁹ The problem analyzed in this paper differs in that the objective function of the principal is not to maximize profit but welfare, education is an indivisible good and there complementarities in payoffs across agents. These two features make the solution differ from these papers. For example there are bunching regions that would not be otherwise present. The conditions on the wealth distribution that I find for uniqueness of the solution using the first order approach in section 1.6 (MLRP and increasing hazard rate of the wealth distribution) are similar to the results derived in an auction setting without non-convexities in Che and Gale (2000) and Blackorby and Szalay (2007), respectively.

The optimal educational system implies a particular wage distribution and wage levels. The role of the educational system as a determinant of inequality, income per capita

⁸Indeed, this is only a convenient simplification of reality (and a common benchmark used in the literature). There are examples of reasonable technologies, as Kremer and Maskin (1996), that exhibit complementarities and fail to satisfy positive assortative matching in general.

⁹Condorelli (2009) and Che and Gale (2009) compare the performance of market and non-market mechanisms.

and growth has been studied by many authors. For example, Bénabou studies in a series of papers (Bénabou, 1993, 1996a,b) patterns of community formation and its implications for inequality and growth. In these papers he emphasizes the role of human capital formation and complementarities within types in the same community. This paper differs from Bénabou's in that the available mechanisms are endogenized and only a static problem is considered.¹⁰ Finally, this paper abstracts from the role of taxation and financing of education. This has been analyzed, for example, in De Fraja (2002), Fernández and Rogerson (2003) and Bénabou (2002).

The rest of the paper is organized as follows. Section 1.2 presents a detailed outline of the paper and summarizes the main results. Section 1.3 lays out the baseline model and Section 1.4 characterizes the first best educational system. Section 1.5 shows that under perfect capital markets, asymmetric information on agents' ability does not preclude the optimal mechanism to attain the first best educational system and how it can be decentralized. Section 1.6 studies how the educational system changes under borrowing constraints and presents the core results of the paper. Section 2.5 presents extensions of the baseline model to show the robustness of the results and Section 2.6 concludes.

1.2 Outline of the Paper

Section 1.3 lays out the baseline economic environment. As discussed before, agents are heterogeneous in both ability and initial wealth levels. These two characteristics are agents' private information. Agents obtain (linear) utility from a consumption good. The initial wealth endowment is in terms of this consumption good, so it can be consumed if desired. Additionally, final good is privately produced one-for-one with human capital and production cannot be observed by the planner.

Agents obtain human capital attending school. The human capital an agent obtains is jointly determined by its own ability, the school tier she attends and a spillover within the types that attend the same school tier. In the baseline model, these spillovers take the form of a "least common denominator", i.e., they are determined by the lowest ability type attending a school tier. A rationale for this is that curriculum requirements adjust to accommodate the lowest skill agent in a given school tier. Thus, this constitutes an endogenous margin of adjustment by which the curriculum taught at a given school tier, and ultimately, school quality, can differ from one economy to another. Moreover, there is a complementarity between ability and school tier: higher ability agents obtain relatively more human capital in higher tier schools. Finally, to focus solely on the matching problem of agents to schools, the

¹⁰Epple and Romano (1998), Fernández and Rogerson (1996, 1998) and Glomm and Ravikumar (1992) provide analysis of alternative (exogenously given) educational systems and their effect on growth, inequality, community formation and the public-private provision dichotomy.

baseline environment abstracts from any costs of school provision.

In the baseline model, there are only two school tiers: basic and higher education.¹¹ An educational system is characterized by (i) an allocation rule that maps types of agents to school tiers and (ii) school tier capacities. The linearity in the utility function implies that the social planner seeks efficiency and abstracts from any redistributive concerns when designing the educational system. In other words, the social planner implements the educational system that maximizes aggregate consumption and, thus, final good production and aggregate human capital.

I study the social planner problem and its decentralization in a variety of environments, which are summarized in Table 1.1. Section 1.4 characterizes the first best educational system. In this case, the allocation rule of agents to schools depends exclusively on ability. The mechanism is such that agents announce their ability type and have to make a (negative) transfer conditional on the announced type. I refer to negative transfers as school fees in Table 1.1. As discussed before, given the linearity in consumption of the objective function, the goal of the social planner is to achieve efficiency in human capital production (because this maximizes final good production). Given that the marginal cost of school provision is zero, the educational system can be decentralized by setting a negative transfer (tax) conditional on school attended, so that the spillover is internalized. These results are summarized in line (1) of Table 1.1.

Section 1.5 shows that in the absence of borrowing constraints, private information alone does not preclude the social planner from achieving the first best educational system. This result obtains because there is single-crossing in the human capital production function. As a result, high ability agents have higher valuation of higher education and, thus, a simple school fee can implement the first best educational system with private information. Moreover, the educational system can be decentralized. Similar to the first best case, in the decentralized equilibrium, the spillover in human capital is priced using school-contingent taxes. Thus, the allocation of agents to schools in an environment with private information coincides with the first best. This is, lines (1) and (2) in Table 1.1 implement the same allocation of agents to schools.

Section 1.6 contains the main results of the paper. It studies an economy with an extreme form of borrowing constraints: financial autarky. This introduces a potential wedge between private valuation of education and ability to pay, as agents can be constrained in the maximal transfer they can make. I characterize the optimal schooling system in the presence of borrowing constraints in Subsection 1.6.1. Agents announce their type (ability and wealth) and are assigned to schools according to a probability rule and a transfer conditional on their (bi-dimensional) type. I show that conditional on an ability level, richer agents are offered a

¹¹A more general form of the complementarity, a model with education costs and more than two tiers are introduced in Section 2.5.

Table 1.1: Summary of the Cases Considered in the Paper

Problem	Planner Implementation	Underlying Space	Decentralization (prices and taxes)
(1) Perfect Information	School Fees	$\mathcal{A} \times S$	✓
(2) Private Information	School Fees	$\mathcal{A} \times S$	✓
(3) Private Information with Borrowing Constraints			
(3.1) Constrained Optimum	Lotteries	$\mathcal{A} \times \Phi \times S$	+ Wealth Market
(3.2) No Commitment	School Fees	$\mathcal{A} \times \Phi \times S$	✓
(3.3) Private Outside Option	School Fees	$\mathcal{A} \times \Phi \times S$	✓
(3.4) No Comm. w/ Exams	Fees, Exams	$\mathcal{A} \times \Phi \times \mathcal{T} \times S$	+ Scholarships

higher probability of accessing higher education. The intuition for the use of lotteries is that they effectively allow to relax borrowing constraints. By offering a lower school fee with a corresponding lower probability of access to school, the social planner ensures that this lottery generates the same ex-ante payoff as a certainty-equivalent transfer. The constrained optimal mechanism has two important features compared to the first best: it admits agents with lower ability into higher education and it reduces the total mass of agents accessing higher education. Given the spillover in human capital formation across agents, this implies an endogenous degradation of the human capital obtained for all agents attending higher education.

In a comparative statics exercise, I show that the degradation in higher education quality due the selection of low ability types into higher education increases in poorer countries.¹² The flip side of this result is that, conditional on an ability and wealth level, the probability of accessing higher education is higher in poorer countries. Thus, this resonates with two features of educational systems in developing countries presented earlier: lower school quality and lower capacity of higher education schools. Changes in wealth dispersion have opposite effects depending on whether or not the median wealth type can afford higher education with certainty. If the original equilibrium features an allocation in which only agents above the median wealth can afford higher education without resorting to lotteries, an increase in wealth dispersion makes it optimal to restrict even further access to higher education, making the educational system more exclusive. The converse is true if the original equilibrium features agents with income below the median being able to afford higher education with certainty. Finally, I show that the constrained optimal mechanism can be decentralized through a market for wealth, in which agents play lotteries with each other over their wealth.

The optimal mechanism with borrowing constraints requires the commitment of the so-

¹²An economy is defined as being poorer than another if the wealth distribution of the poorer economy is MLRP-dominated by the richer.

cial planner to exclude from higher education some poor, high-ability agents that engage in lotteries. Note that once agents that select into lotteries to access higher education have revealed their types, the social planner would like to modify the allocation rule ex-post and allocate *all* these agents into higher education. The reason is that they have an ability (weakly) higher than the lowest ability of unconstrained (rich) agents that attend higher education. If the social planner cannot commit to exclude some of these agents from attending higher education once they have revealed their type, then the only credible mechanism that can be used are school fees (without lotteries).¹³ Subsection 1.6.2 shows that this environment delivers results that are qualitatively analogous to the environment with commitment. For example, the optimal educational system admits agents with lower ability into higher education institutions. The comparative statics results are also analogous to the case with commitment: there is endogenous degradation of human capital formation and less capacity of higher education institutions in poorer countries. Decentralization in this case does not require a market for wealth. A school-contingent tax in addition to prices of school suffices.

Finally, in this environment without commitment, Subsection 1.6.3 studies the design of the schooling system when the social planner has access to a signaling technology, which I interpret as exams. The usage of exams introduces a trade off: exams involve wasteful spending but allow for additional screening power because the cost of obtaining a particular score is decreasing in agents' ability. The optimal mechanism consists of a schedule of school fees and test scores contingent on agents' reported type. Agents that perform relatively better in an exam are rewarded with a lower school fee. Thus, the decentralization of the mechanism involves a set of taxes conditional on test scores that resemble a scholarship scheme.

When the exam technology is merely of a fail/pass-type, I show that the solution of the planner problem only makes use exams in sufficiently poor countries. The intuition for this result is simple. Once exams are put in place, all agents above a particular ability (regardless of whether they are rich or poor) take the exam. Thus, all agents that take the exam incur a wasteful spending. The benefit of using exams only comes through the additional mass of agents that access higher education thanks to exams. As a result, if borrowing constraints are not very severe, the additional screening power gained by using exams may be too costly to use. When the planner has access to a richer signaling technology, in which different test scores can be obtained, the comparative statics of the optimal schooling system on the wealth distribution show that, at any ability level, poorer countries use relatively more exams (i.e., require a higher score) and that the overall access cost to higher education (school fee plus exam cost) is lower in poorer countries. Similarly, for changes in the dispersion in the wealth

¹³If instead of no-commitment, agents have limited communication and cannot announce their types, then only a singleton can be used to allocate agents to higher education. If the planner has also limited communication, then only fees can be used.

distribution, I show that exams are relatively more used at the levels where inequality increases.

Section 2.5 relaxes some of the assumptions of the baseline model and shows that the main insights derived from the baseline model hold in more general set-ups. In Subsection 1.7.1, I allow for more general spillovers (for example, the case in which the spillover is determined by the average type attending a school) and a CES production function for the final good. Subsection 1.7.2 considers the case in which, within each school tier, there are a continuum of sub-tiers. This extension allows for more robust predictions on the mass of agents attending schools. Moreover, it provides a simple framework to study how the presence of an unregulated private provider of schooling may constrain the planner's problem. I show that the existence of a private unregulated sector undermines the capacity of the social planner to provide education to borrowing constrained agents.

1.3 The Environment

This section describes the fundamentals of the economies studied in the paper.

Endowments. The economy is populated by a unit mass of agents. Each agent is endowed with ability, $a \sim G(a)$, and initial wealth, $\phi \sim F(\phi)$. Initial wealth is distributed over the support $\Phi = [0, \bar{\phi}]$ with cumulative distribution function $F(\phi)$, and associated density $f(\phi)$. The upper bound on wealth, $\bar{\phi}$, can be either finite or infinite. Ability is distributed uniformly over a support $\mathcal{A} = [\underline{a}, \bar{a}] \subseteq [0, 1]$. Ability and wealth are uncorrelated across agents and are private information.

Technologies. In this economy, two technologies are used: a final good and a human capital production function. Final good is produced one-for-one with human capital H . Human capital is produced by the schooling system. There are S school tiers in the economy $s = \{0, 1, \dots, S - 1\}$, with associated capacities $c(s)$. School 0 provides the minimal, mandatory, level of education required for all the population, while schools $s > 0$ provide further education. The two main cases of analysis in the paper will be case of two and infinite schools. The marginal cost of school provision per student is $\kappa(s)$.

An agent with ability a attending school tier s obtains human capital $H(a, s)$, which is determined by the combination two different factors, $A(s)$ and $h(a, s)$, according to

$$H(a, s) = A(s)h(a, s). \tag{1.1}$$

The first factor is an intrinsic human capital production function $h(a, s)$ associated with each school tier. This production function is (weakly) increasing and concave in ability a and (weakly) increasing in school index s . There is complementarity between schools and ability.

Let $s > \tilde{s}$, then

$$\frac{\partial(h(a,s) - h(a,\tilde{s}))}{\partial a} \geq 0 \quad \text{for all } s, \tilde{s} \in S.$$

This means that high ability agents benefit relatively more from high index schools.

In addition to the intrinsic human capital production function, there is an spillover at each school tier level, $A(s)$. In the baseline model, this is modeled as an extreme complementarity between types

$$A(s) = \min_a \{a \in s\}. \quad (1.2)$$

This can arise because the social planner cannot commit to exclude from education students attending each education tier, and has to accommodate the curriculum level of each tier to the ability of its students. With this interpretation in mind, it is natural to have the lowest ability student attending a particular tier determining the spillover effect, because the curriculum level of each tier has to adjust to its “least common denominator.” This implies that if the spillover component, $A(s)$, differs across countries, the human capital obtained by an agent of ability a attending school tier s can differ across countries. It is in this sense that the model rationalizes differences in education quality across countries. In Section 1.7.1, I discuss how the results extend to more general type of spillovers, which include, among others, the mean type attending school s , rather than the minimum.

Preferences and Aggregate Welfare. An agent chooses actions so as to maximize her utility from consumption. Utility is linear in consumption and equals wage income, plus initial wealth and a possible lump-sum transfer from the government, minus any expenditures incurred to educate. Aggregate welfare is defined as the sum of the utilities of the agents in the economy. Throughout the paper, the social planner is assumed to be utilitarian and having as objective to maximize aggregate welfare. The linearity of the objective function implies that the social planner is concerned only by production efficiency and abstracts from any distributional consideration. Moreover, as the only input for production is human capital, this implies that the social planner objective is to maximize aggregate human capital and, thus, implement a schooling system based on efficiency considerations alone.

From section 1.4 to 1.6, I analyze the case in which there are two school tiers, $S = 2$. The main insights from the paper are obtained by this simple two-tier school model. As discussed before, the lower tier, $s = 0$, represents the basic or mandatory schooling, while $s = 1$ represents additional, non-compulsory education. Moreover, to highlight the frictions arising from private information, the cost of provision is assumed to be zero, $\kappa(s) = 0$. For the sake of brevity, I refer to $s = 1$ as higher education. Section 2.5 extends the model to allow for a richer production function, more general spillovers and a continuum of tiers within basic and higher education.

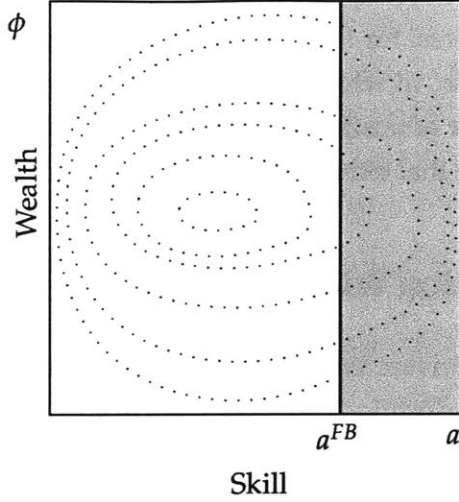


Figure 1-3: First Best Allocation. Agents are segregated by skill only.

1.4 First Best

In this section, I characterize the optimal schooling system when there is no private information and assignments to basic ($s = 0$) and higher education ($s = 1$) can be made contingent on types. Given the complementarity between skills and schools, segregation by skill is optimal. To see this, consider two agents with abilities a_1 and a_0 with $a_1 > a_0$. It cannot be optimal that agent 1 is in school 0 and agent 0 in school 1, as $H(a_1, 1) - H(a_0, 0) > H(a_1, 0) - H(a_0, 1)$. Given that it is mandatory to provide basic education to all agents in the economy, the problem of the social planner is to choose the lowest ability \tilde{a} in $s = 1$,

$$\max_{\tilde{a}} \int_{\tilde{a}}^{\tilde{a}} \tilde{a} h(a, 1) dG(a) + \int_{\underline{a}}^{\tilde{a}} \underline{a} h(a, 0) dG(a),$$

which has as implicit solution¹⁴

$$\underbrace{\int_{a^{FB}}^{\tilde{a}} h(a, s) da}_{\text{spillover benefit}} = \underbrace{a^{FB} h(a^{FB}, 1) - \underline{a} h(a^{FB}, 0)}_{\text{marginal output cost}}. \quad (1.3)$$

The social planner balances the spillover (quality) effect that is improved by increasing the ability of the marginal type attending school 1, with the reduction in the mass of agents that attend school 1. Note that the allocation is independent of agents' wealth, see Figure 1-3.

Conceivably, it could be the case that it would be optimal to have all agents attending higher education. For example, this could happen if the differences in ability in the popula-

¹⁴Sufficiency of the First Order Condition is shown in Appendix 2.7.

tion where low (imagine the extreme case in which everybody has the same ability). This is a pathological case of no interest for the discussion as, in this case, no theory of educational systems would be needed. So, in what follows, I shall focus the discussion on the cases in which both, basic and higher education coexist. Simple conditions that would ensure that both school tiers are used are that either $\underline{a} = 0$ or $h(a, 0) = 0$.

1.5 Private Information without Borrowing Constraints

In this section, I show that private information alone does not preclude the social planner from achieving the first best educational system.¹⁵ As in Fernández and Galí (1999), I assume perfect capital markets and that the interest rate paid by agents is constant and normalized to one. This market operates when the educational system is put in place, so that agents, if they desire to, can borrow and repay after production. Trades in this market cannot be monitored by the planner.

The planning problem now is constrained by the fact that the assignment to schools cannot be conditioned directly on agents' ability. The planner problem is divided in two stages. First, the social planner announces an assignment rule and school tier capacities. Second, the economy unfolds: conditional on the educational system chosen in the first stage, agents decide which school to attend (borrowing if necessary) and obtain human capital. Then, they supply their human capital in a competitive labor market, obtain a wage, repay debt (if any) and consume. Given the linearity of utility, the only concern of the social planner is to achieve efficiency by matching agents to schools. Thus, the attention is restricted to the design of the educational system.

Before proceeding, and to avoid the discussion of many cases, I shall make the expositional assumption that there are always "rich enough" agents in the economy.

Assumption 1 (*Expositional simplifying assumption*) *There is a positive mass of rich agents that can afford paying for extreme segregation,*

$$\bar{\phi} > \bar{a}h(\bar{a}, 1) - \underline{a}h(\bar{a}, 0).$$

This assumption implies that the spillover effect \bar{a} in school $s = 1$ is pinned down by the choice of unconstrained agents for any $a \in \mathcal{A}$.

In this environment, using school fees as an allocation device is enough to achieve the first best allocation. The reason is that agents' choices satisfy a single-crossing condition. Let ψ_s denote the school fee of attending school s , the indirect utility $u(a) \equiv \max_s H(a, s) - \psi_s$ is

¹⁵This result is analogous to Fernández and Galí (1999) for the case in which the spillover effect $A(s)$ is shut down, i.e., $A(s) = 1$ for all $s \in S$.

increasing in a . Moreover, the social planner cannot exclude agents from school 0 ($\psi_0 = 0$). As a result, the choice variable of the planner is to choose fee $\psi_1 \equiv \psi$ for school 1. The social planner problem can be written as

$$\max_{\psi} \int_{\underline{a}(\psi)}^{\bar{a}} \bar{a}(\psi) h(a, 1) dG(a) + \int_{\underline{a}}^{\bar{a}(\psi)} \underline{a} h(a, 0) dG(a), \quad (1.4)$$

where the marginal type obtaining higher education \bar{a} is implicitly defined by

$$\bar{a} h(\bar{a}, 1) - \underline{a} h(\bar{a}, 0) = \psi. \quad (1.5)$$

Note that the fees incurred by agents to attend higher education are just transfers and, as such, they are not wasted. The revenue that the social planner obtains from the transfers is redistributed ex-post back to agents. The specifics of the redistribution do not matter given the linearity of utility. For concreteness, I assume that revenue is redistributed back to agents in a lump-sum manner.

Proposition 1 *The marginal type obtaining education $\bar{a}(\psi)$ is strictly increasing in ψ for all $\psi \in \Psi = (\underline{a}(h(\underline{a}, 1) - h(\underline{a}, 0)), \bar{a}h(\bar{a}, 1) - \underline{a}h(\underline{a}, 0))$.*

The First Best schooling system can be implemented under private information by setting a school fee of $\psi^{PF} = H(a^{FB}, 1) - H(a^{FB}, 0)$ to agents attending $s = 1$, where a^{FB} is implicitly defined in (1.3).¹⁶

This proposition implies that there is a one-to-one mapping from the school fee ψ (in the relevant margin) to the marginal type selecting into higher education, $\bar{a}(\psi)$. Using this property, it follows that the First Best schooling system can be implemented by setting a school fee of $\psi^{PF} = H(a^{FB}, 1) - H(a^{FB}, 0)$. Section 2.5 shows that this result extends to the case of $S > 2$ schools. Indeed, if there was no complementarity within agents in higher education, it would be optimal to make everyone attend higher education.

Decentralization. Next, I discuss how the optimal educational system can be decentralized. One can imagine the social planner running a procurement auction for both basic and higher education tiers, and there being a competitive pool of firms willing to enter the market. Firms would undercut each other and this would result in firms willing to supply education at its marginal cost, which is zero in this case. A price of education equal to zero coincides with the optimal school fee for basic education. For higher education, the social planner would set a tax contingent on attending higher education equal to $H(a^{FB}, 1) - H(a^{FB}, 0)$. This ensures that the demand of schooling from the agents coincides with the first best allocation of agents to schools. Thus, this discussion shows how the first best educational system can be decentralized in the presence of private information.

¹⁶All proofs can be found in Appendix 2.7.

1.6 Private Information with Borrowing Constraints

This section analyzes an environment without perfect capital markets. The capital market imperfection studied is an extreme one: financial autarky. This is, capital markets are shut down entirely and agents have to self-finance their investments in education. Self-financing can be interpreted literally, or as the effective disposable wealth agents have to finance education after exhausting any potential way they have to obtain financing.

This section starts studying the mechanism design problem assuming that the social planner can commit to randomization. As I show below, this means that even though some agents reveal to be of high ability type to the social planner and “deserving” to attend higher education in the sense that agents with the same or even less ability attend higher education, it is optimal for the social planner *not* to let them obtain education. Then, Section 1.6.2 studies the case in which the social planner cannot commit to exclude from education all agents that signal to have high ability. I show that this is isomorphic to a world in which no communication is possible, and the optimal mechanism used to allocate agents to schools consist of just school fees. Finally, I allow for an exam technology to relax the commitment (or communication) problem. The exam technology gives an additional screening mechanism to the social planner and allows the planner to make the schooling allocation contingent on exam performance.

1.6.1 The Mechanism Design Problem

This subsection studies the design of the optimal educational system when agents are borrowing constrained. By the revelation principle, I restrict attention to direct-revelation mechanisms in which each type has an incentive to report private information truthfully. This constitutes a bi-dimensional screening problem, as both ability a and wealth ϕ are private information. The mechanism specifies a type-contingent transfer $t(a, \phi) \in \mathbb{R}$ and probability $\pi(a, \phi) \in [0, 1]$ of attending higher education. Formally, a mechanism is a mapping from the type space to the transfers and probability space, $\langle t, \pi \rangle : \mathcal{A} \times \Phi \rightarrow [-\infty, \infty] \times [0, 1]$. Any feasible mechanism has to satisfy participation and incentive compatibility constraints for all agents, in addition agents’ transfers cannot exceed their wealth (borrowing constraint) and the social planner has to satisfy a break-even constraint.

Let $w(a, \phi) = \pi(a, \phi)H(a, 1) + (1 - \pi(a, \phi))H(a, 0)$ denote the expected return or wage of an agent of type (a, ϕ) . With this notation at hand, the participation constraint can be stated as

$$w(a, \phi) + t(a, \phi) \geq 0, \quad \forall (a, \phi). \quad (1.6)$$

Note that the participation constraint is equivalent to the social planner being constrained to supply at least basic education to all agents in the economy. The incentive compatibility

constraint is

$$w(a, \phi) + t(a, \phi) \geq \pi(\bar{a}, \bar{\phi})H(a, 1) + (1 - \pi(\bar{a}, \bar{\phi}))H(a, 0) + t(\bar{a}, \bar{\phi}), \quad \forall(a, \phi) \text{ and } (\bar{a}, \bar{\phi}). \quad (1.7)$$

The borrowing constraint implies that

$$\phi + t(a, \phi) \geq 0 \quad \forall(a, \phi). \quad (1.8)$$

Finally, note that if the social planner could use negative transfers, it would like to do so to subsidize education and overcome the borrowing constraints. Thus, a budget constraint condition needs to be imposed. As the planner has no access to additional resources, a natural benchmark is that no subsidization is possible. That is, transfers are restricted to be negative,¹⁷

$$t(a, \phi) \in \mathbb{R}_-. \quad (1.10)$$

The social planner problem consists on finding the schedule of transfers t and associated probabilities π that maximize the objective

$$\int_{\underline{a}}^{\bar{a}} \int_0^{\bar{\phi}} [\pi(a, \phi)a^*h(a, 1) + (1 - \pi(a, \phi))\underline{a}h(a, 0)] dG(a)dF(\phi), \quad (1.11)$$

subject to (1.6), (1.7), (1.8) and (1.10), where a^* is defined by the lowest ability agent attending higher education, $a^* = \min_a \{a \in s = 1\}$.

Proposition 2 (Optimal Schedule) *The optimal mechanism featuring agents of ability $a \geq a^*$ in higher education takes the form of a higher education fee menu $t(\phi), \pi(\phi)$,*

$$t(a, \phi) = \begin{cases} -\psi & \text{for } \phi \geq \psi, a \geq a^*, \\ -\phi & \text{for } \phi < \psi, a \geq a^*, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \pi(a, \phi) = \begin{cases} 1 & \text{for } \phi \geq \psi, a \geq a^*, \\ \frac{\phi}{\psi} & \text{for } \phi < \psi, a \geq a^*, \\ 0 & \text{otherwise,} \end{cases}$$

with $\psi = a^*h(a^*, 1) - \underline{a}h(a^*, 0)$.

The formal proof of the result can be found in Appendix 2.7. Here I provide a sketch of the proof. Consider the case in which, due to spillovers in human capital production, the social planner wants to segregate agents in the two school tiers. In this case, the social planner has

¹⁷Section 2.5 and Appendix 1.8 analyze an alternative setup in which the social planner can do “simultaneous” redistribution and has only to break even on net,

$$\int_{\underline{a}}^{\bar{a}} \int_0^{\bar{\phi}} t(a, \phi) dG(a) dF(\phi) \leq 0, \quad (1.9)$$

and show that the same qualitative results hold.

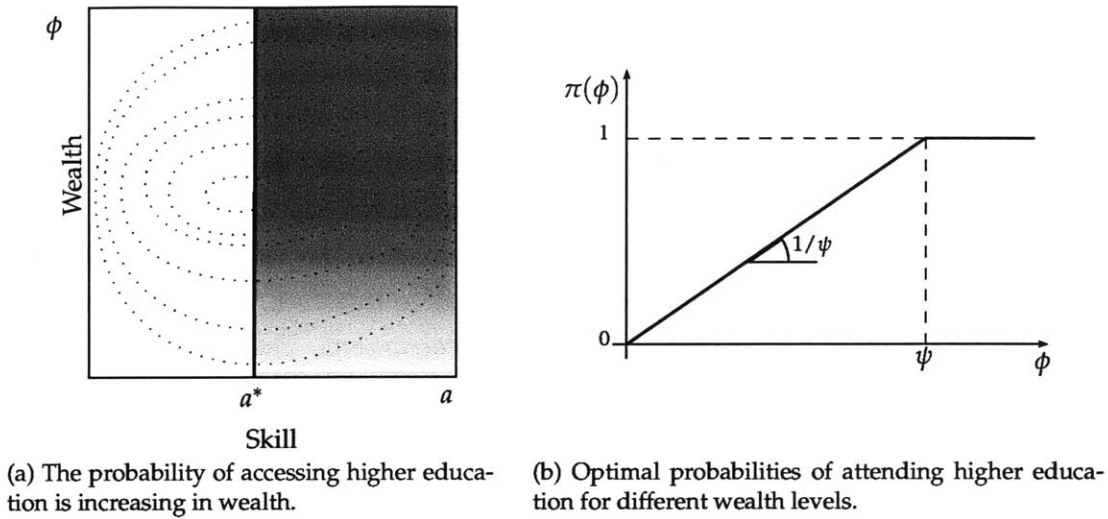


Figure 1-4: Constrained efficient allocation

to choose the spillover level of higher education (i.e., the threshold type a^*). For a given a^* , it would be optimal (in the first-best sense) to have all agents with $a \geq a^*$ attending higher education. However, borrowing constraints prevent doing so because the transfer level needed to separate agents according to their willingness to pay is too high for some agents (i.e., the borrowing constraint is binding for them). Note, however, that the constrained agents cannot afford any mechanism that involves paying a fee higher than their wealth. In other words, rich agents can always claim to be poor, but not the other way around. As a result, the optimal mechanism can offer to rich, unconstrained agents the “best deal”, which is to attend higher education with probability 1, provided that they are of ability greater than a^* . This deal does not affect credit-constrained agents because they cannot afford it. For the credit-constrained agents, then, the social planner provides a lottery that offers the highest possible probability to access higher education (provided agents are of ability higher than a^*). This lottery is decreasing in agents wealth to ensure incentive compatibility. In short, the fact that poorer agents cannot afford better lotteries compartmentalizes the problem into layers of wealth. The allocation patterns of the mechanism are represented graphically in Figure 1-4.

Proposition 2 allows to rewrite the original planner’s problem (1.11) as an essentially uni-dimensional optimization problem. For simplicity, I work with the negative of the transfers, which I refer to as fees. The objective function of the planner can be expressed now as

$$\max_{\psi} \int_{\psi}^{\bar{\phi}} \int_{a^*(\psi)}^{\bar{a}} \Delta w(a, a^*) dG(a) dF(\phi) + \int_0^{\psi} \int_{a^*(\psi)}^{\bar{a}} \pi(\phi, \psi) \Delta w(a, a^*) dG(a) dF(\phi), \quad (1.12)$$

subject to $\psi = a^*h(a^*, 1) - ah(a^*, 0)$, where I have used the notation $\Delta w(a, a^*) = a^*h(a, 1) - ah(a, 0)$. The case of interest is when the problem has an interior solution, i.e., it is optimal to have some agents with just basic education. In this case, the first order condition is¹⁸

$$(1 - F(\psi)) \left(-\frac{\partial a^*}{\partial \psi} \Delta w(a^*, a^*) g(a^*) + \int_{a^*}^{\bar{a}} \frac{\partial \Delta w}{\partial \psi} dG(a) \right) + \int_0^{\psi} \left(-\frac{\partial a^*}{\partial \psi} \frac{\phi}{\psi} \Delta w(a^*, a^*) g(a^*) + \int_{a^*}^{\bar{a}} \frac{\partial}{\partial \psi} \left(\frac{\phi}{\psi} \Delta w \right) dG(a) dF(\phi) \right) = 0. \quad (1.13)$$

From this expression, it is apparent that the optimal solution balances costs and benefits of raising tuition fees. The cost of raising the tuition fee is that it reduces the number of agents attending higher education. This effect is captured in the first and third terms of (1.13), as the threshold type a^* moves with ψ . Moreover, this appears in the decrease in the probability of accessing higher education for borrowing-constrained agents in the last term of (1.13) (i.e., ϕ/ψ being decreasing in ψ). The benefit of higher education fees comes through the spillover effect. By increasing ψ , the spillover effect increases and, thus, makes all agents that attend higher education obtain more human capital.

To investigate the effect of wealth distribution on the optimal schedule, condition (1.13) can be rewritten so that the effect of the wealth distribution is encapsulated in one term,

$$\frac{\frac{1}{\psi} \int_0^{\psi} \pi(\phi, \psi) dF(\phi)}{1 - F(\psi)} = \frac{\int_{a^*(\psi)}^{\bar{a}} \frac{\partial \Delta w}{\partial \psi} dG(a) - \frac{\partial a^*}{\partial \psi} \Delta w(a^*) g(a^*)}{\int_{a^*(\psi)}^{\bar{a}} \left(\Delta w - \psi \frac{\partial \Delta w}{\partial \psi} \right) dG(a) + \frac{\partial a^*}{\partial \psi} \psi \Delta w(a^*) g(a^*)}. \quad (1.14)$$

The left hand side of equation (1.14) contains all the influence of the wealth distribution on the optimal solution. If there were no borrowing constraints this term would be zero, and the optimal ψ would be given by equating the numerator of the right hand side to zero (which coincides with the first best first order condition, equation 1.3).

To gain intuition on how the wealth distribution affects the optimal fee ψ (i.e., the left hand side of (1.14)), suppose that no lotteries were used. In this case, the probability of attending higher education would be given by a step function, $\pi(\phi, \psi) = 0$ for $\phi < \psi$ and $\pi(\phi, \psi) = 1$ for $\phi \geq \psi$. This would simplify the left hand side to the hazard-rate of the wealth distribution. In this case what would only matter in making the trade-off between the mass of agents attending higher education and the spillover effect is the percent increase in the mass of agents that can afford education at the margin, $f(\psi)$, relative to the total mass of agents educating, $1 - F(\psi)$. Now, going back to the original formulation, rather than just having the density at ψ , thanks to the lottery, the mechanism can include some borrowing constrained agents with wealth $\phi < \psi$. Thus, the social planner considers a weighted average of the mass of agents that can attend higher education at different levels of

¹⁸The derivation of the first order condition and the proof of its sufficiency is shown in Appendix 2.7.

wealth. To see that, note that $\int_0^\psi \pi(\phi, \psi) dF(\phi)$ computes the total mass of constrained agents attending higher education, while the term $1/\psi$ in front, implies that the planner takes into account the average value of the integral relative to the total mass of unconstrained agents $1 - F(\psi)$.

The behavior of the left hand side term of (1.14) as a function of ψ can be, in general, non-monotonic. However, it is monotonically increasing for log-concave functions that have support in $[0, \infty)$ such as the Weibull, Exponential, Gamma (with shape parameter greater than one) and for some other distributions such as the Uniform and the Pareto distribution with well defined mean and variance (i.e., with shape parameter, greater than 2). Given that the left hand side term is monotonically increasing for the most usual distributions used to model wealth distributions (except for the Log-Normal, for which is non-monotonic), I restrict my attention to wealth distributions that generate an increasing left hand side term.

The right hand side of equation (1.14) is the ratio of the net marginal gain in output if all agents were unconstrained relative to the net gain in output due to spillover gains evaluated at the average-constrained agent.¹⁹ In Appendix 2.7, I show that under mild conditions on the intrinsic production function $h(a, s)$, the right hand side of (1.14) is strictly decreasing. This is intuitive, the numerator captures the net marginal gain from segregation (as in the first best), and it is decreasing in ψ because of the concavity in the human capital production function. On the contrary, the denominator is increasing in ψ because the quality of the higher education sector increases with the spillover. Thus, the right hand side of (1.14) is decreasing in ψ .

1.6.1.1 Comparative Statics on the Wealth Distribution

With the previous discussion at hand, I begin to study the main question of the paper: how optimal educational systems change with shifts in the wealth distribution? To do so, I introduce a one-dimensional ranking of wealth distributions that is amenable to our purposes.

Definition (Wealth abundance) Consider two wealth distributions, \tilde{F} and F , with associated densities \tilde{f} and f . A distribution \tilde{f} is more wealth abundant than f , denoted by $\tilde{f} \succ_w f$, if $\tilde{f}(\phi_1)f(\phi_0) \geq \tilde{f}(\phi_0)f(\phi_1)$ for all $\phi_1 > \phi_0$.

The interpretation of the notion of wealth abundance is intuitive. For non-vanishing values of the density, the wealth abundance condition can be written as

$$\frac{\tilde{f}(\phi_1)}{f(\phi_1)} \geq \frac{\tilde{f}(\phi_0)}{f(\phi_0)}.$$

¹⁹The denominator is always positive. This follows from the concavity of Δw , for which Appendix 2.7 provides sufficient conditions. A concave function satisfies the following property (c.f. Varian (1992)) $\Delta w(0) + \Delta w'(\psi)\psi \leq \Delta w(\psi)$. Note that if $\psi = 0$, $a^* = \underline{a} = 1$ and $\Delta w(0) = \underline{a}(h(a, 1) - h(a, 0)) \geq 0$. Using this result with $\Delta w'(\psi) > 0$, it follows that $\Delta w'(\psi)\psi \leq \Delta w(\psi)$ for all a .

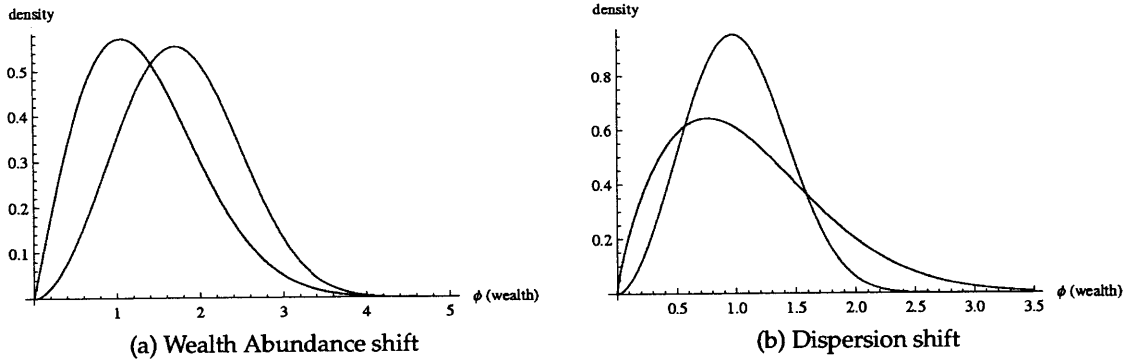


Figure 1-5: Wealth Abundance and Dispersion Shift of a Weibull Distribution

This means that if one economy is more wealth abundant than another, there are relatively more rich agents in this economy when comparing any arbitrary two wealth levels, ϕ_1 and ϕ_0 .²⁰ This notion of wealth abundance requires that the two distributions satisfy a Monotone Likelihood Ratio Property (MLRP). Figure 1-5 provides a graphical intuition. Having two distributions ranked according to MLRP implies both hazard rate and first order stochastic dominance between these two distributions.

Remark Consider two wealth distributions, $\tilde{F} \succ_w F$, then

$$\frac{\frac{1}{\psi} \int_0^\psi \phi f(\phi) d\phi}{1 - F(\psi)} \geq \frac{\frac{1}{\psi} \int_0^\psi \phi \tilde{f}(\phi) d\phi}{1 - \tilde{F}(\psi)}.$$

Proof: Rewrite the previous expression as $(\psi(1 - F(\psi))(1 - \tilde{F}(\psi)))^{-1} \int_0^\psi \phi (f(\phi)(1 - \tilde{F}(\psi)) - \tilde{f}(\phi)(1 - F(\psi))) d\phi$. Thus, a sufficient condition for the inequality condition to hold is that $f(\phi)(1 - \tilde{F}(\psi)) - \tilde{f}(\phi)(1 - F(\psi)) \geq 0$, where $\phi \leq \psi$. Now, I show that this sufficient condition is implied by the definition of wealth abundance. From the definition of wealth abundance, $\tilde{f}(\phi_1)f(\phi_0) \geq f(\phi_1)\tilde{f}(\phi_0)$ for all $\phi_1 > \phi_0$. Integrating both sides of the inequality from $\phi_1 = \psi \geq \phi_0$ up to $\bar{\phi}$,

$$\int_\psi^{\bar{\phi}} \tilde{f}(\phi_1)f(\phi_0) d\phi_1 \geq \int_\psi^{\bar{\phi}} f(\phi_1)\tilde{f}(\phi_0) d\phi_1,$$

this implies that $(1 - \tilde{F}(\psi))f(\phi_0) \geq (1 - F(\psi))\tilde{f}(\phi_0)$ for all $\phi_0 \leq \psi$. \square

With this definition at hand, I proceed to do comparative statics on the wealth distribution in terms of wealth abundance to derive the first main result of the paper.

Proposition 3 Consider a wealth abundance shift, $\tilde{F} \succ_w F$. The optimal maximal fee ψ under \tilde{F} is higher than under F .

The results follows immediately from the left hand side of equation (1.14) being increasing, the right hand side being decreasing and the observation that a Wealth Abundance shift only

²⁰This notion of abundance is analogous to the skill abundance notion in Costinot and Vogel (2010).

moves downwards the left hand side of equation (1.14).

Proposition 3 implies that more wealth abundant countries have higher ability agents in higher education because the threshold type a^* is strictly increasing in ψ . That is, agents accessing higher education in more wealth abundant economies have higher ability on average. Thus, the ability-composition of higher education is better in more wealth abundant economies. Moreover, the level of randomization at all wealth levels is smaller in the wealth abundant country $\pi_{\bar{f}}(\phi) \leq \pi_f(\phi)$ for all ϕ . This implies that it is *less likely* for borrowing constrained agents to access higher education in relatively more wealth abundant countries. Indeed, there are less agents constrained in wealth abundant economies, so the relative cost of not providing them the right type of education is relatively low.

Another consequence of this result is that the the schooling system in wealth abundant countries amplifies the dispersion in the earnings distribution compared to less wealth abundant countries. To see this, note that the equilibrium spillover level is increasing in wealth abundance. Thus, conditional on accessing higher education, wealth abundant countries generate more human capital for an agent of a given ability. However, access to higher education happens at higher levels of ability in more wealth abundant countries.

The previous discussion highlights another dimension of the effect of borrowing constraints: the mismatch of ability to schools. With borrowing constraints, there is an increasing mass of mismatched agents. That is, higher ability agents that have to attend basic education because of credit constraints. This results in a change in the ranking of earnings of agents relative to the first-best. As an economy becomes less wealth abundant, high-ability low-wealth agents tend to fall in the ranking at the expense of low-ability high-wealth agents who rise.

I now provide a comparison of the mass of agents that are being educated in environments with and without borrowing constraints.

Proposition 4 *There is a reduction in the mass of agents obtaining higher education when the economy transitions from no agents being effectively borrowing constrained to a (small) mass of agents being borrowing constrained.*

This result is intuitive. As the ability of agents to express their valuations of attending higher education is hindered, the social planner finds better to reduce the capacity of higher education. Note however that this is only true around the transition from no borrowing constrained agents to a small mass of borrowing constrained agents. The reason why the result does not hold for all levels of borrowing constraints is that benefits from the spillover are traded-off against the mass of agents attending higher education. With two school tiers, if borrowing constraints are very prevalent, it could be the case that it is better to reduce the spillover effect to be able to admit more students. In Section 1.7.2, I show that when there are more school layers, tighter results are obtained. In particular, I show that the result is true

at all levels of borrowing constraints for a positive measure of school tiers that contains the highest level school.²¹

A related question that can be investigated is how changes in wealth dispersion affect the educational system.

Definition (Wealth dispersion) Let ϕ_m denote the median wealth of an economy. A distribution with density \tilde{F} has more wealth dispersion, denoted by $\tilde{F} \succ_d F$, if and only if $\tilde{F}(\phi) \succ_w F(\phi)$ for $\phi > \phi_m$ and $F(\phi) \succ_w \tilde{F}(\phi)$ for $\phi < \phi_m$.

This definition captures the idea that there are more agents with extreme wealth values.

Applying Proposition 3, the following result follows.

Proposition 5 Consider a wealth dispersion shift $\tilde{F} \succ_d F$. If the optimal fee under F featured $\psi > \phi_m$, a wealth dispersion shift increases the optimal fee, while if $\psi < \phi_m$, a wealth dispersion shift reduces ψ .

This result shows that changes in wealth dispersion have opposite effects depending on whether the median wealth type can afford higher education with certainty. If the original equilibrium featured an allocation in which only agents above the median wealth could afford attending higher education with certainty, an increase in the wealth dispersion makes it optimal to set the education fee even higher. This makes the average type attending school 1 higher and increases the school quality. Indeed, the increase in school quality is at the expense of making it less likely for poor people to access it. This change in the educational system makes the earnings distribution more disperse than the original wealth distribution. In other words, the optimal educational system enhances inequality. The opposite is true if the original equilibrium featured agents with wealth below the mean. Upon a wealth dispersion shift, the educational system becomes more inclusive and, in fact, the optimal educational system tends to undo the increase in wealth dispersion, by making the ex-post earnings distribution less disperse under the new optimal schooling system.

1.6.1.2 Decentralization: A Fair Lottery Market on Wealth

In this subsection, I discuss how to decentralize the previous allocation.²² Similar to Becker et al. (2005) and Cole and Prescott (1997), I show that with a market for lotteries over income, the optimal allocation can be decentralized. Given the working assumption of zero marginal cost of provision, consider firms (schools) providing mandatory and higher education at

²¹When there is a cost of provision of schooling, this result in Proposition 4 is true when there is a large mass of borrowing constrained agents even with only two schools. The reason is that the social planner cannot reduce the price below the marginal cost of provision.

²²I thank Iván Werning for suggesting this discussion.

price 0.²³ Suppose that the social planner sets a school-contingent tax that has to be satisfied to attend school s such that

$$\tau(s) = \begin{cases} 0 & \text{for } s = 0, \\ \psi & \text{for } s = 1. \end{cases} \quad (1.15)$$

Now, consider a market that opens after the social planner announces the school-contingent taxes and before agents attend schools, in which fair lotteries l over wealth are traded. There is a continuum of these lotteries, indexed by $i \in [0, \psi]$. A lottery l_i delivers ψ with probability π_i and 0 with probability $1 - \pi_i$. There is a competitive market for each lottery.²⁴ Thus the price of lottery i , p_i , is given by the break even constraint (or the actuarially fair lottery), $p_i = \pi_i \psi$.

By the discussion on the previous section, the school-contingent tax $\tau(s)$ makes attending higher education attractive for all agents with ability $a \geq a^*$. Agents that are not borrowing constrained, do not derive any gain from participating in the lottery market. If they had to, they would purchase the lottery with corresponding price ψ , which returns wealth ψ with probability 1. However, agents that are borrowing constrained derive positive gains from participating in this market. Note that should they not participate, they would attend school 0 with probability one, while by participating in the lottery market they can attend school 1 with some positive probability.²⁵ Moreover, purchasing a lottery with a higher probability π_i and, hence, a higher price p_i is (weakly) better than purchasing a lottery with lower probability. Thus, constrained agents exhaust their initial wealth when purchasing lotteries. As a result, agents select into lotteries that have the same expected value of their initial wealth endowment. This discussion shows the following result.

Proposition 6 (Wealth Market) *The optimal schooling system can be decentralized with school-contingent taxes and a market for wealth.*

1.6.2 Solution with School Fees Only

In this section I show that in environments in which the social planner has no commitment or limited communication, no lotteries can be used.

Proposition 7 (Credible Mechanism) *Let the cost of reallocating agents to higher education be zero. If the social planner has no commitment, the only credible mechanism are school fees. Similarly, if the economy has limited communication such that no announcements can be made, the only feasible mechanism are school fees.*

²³Section 2.5 shows that an analogous result holds when there is a positive marginal cost of provision.

²⁴If each market is operated by more than one broker, the assumption is that each broker serves a positive mass of agents, so that there is no uncertainty on the returns of the lottery.

²⁵The formal argument is analogous to equation (1.7), $\pi_i \Delta w - p_i + w(a, 0) \geq w(a, 0)$, for all $a \geq a^*$.

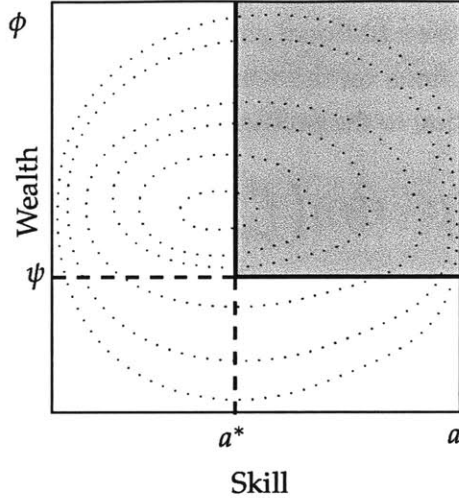


Figure 1-6: Allocation with school fees only.

Note that in the previous subsection, some borrowing constrained agents with $a \geq a^*$ (that truthfully report their type) are not allocated to higher education. This would not happen in a first-best world, and it requires a commitment from the social planner of not reallocating ex-post agents once they have announced their type.²⁶ In this environment, the social planner cannot use lotteries to relax the borrowing constraints, because agents anticipate that if any randomization is announced, it is not credible. Thus, only school fees (without lotteries) can credibly be used. For the case of limited communication, the result follows purely from the constraints imposed from the limited communication in the transfer space.

The problem of the social planner reduces to decide the fee ψ that it charges to attend higher education. As in the previous section, the optimal school choice of an individual of type (a, ϕ) is to choose $s = 1$ if $H(a, 1) - \psi \geq H(a, 0)$ and $\phi \geq \psi$. Otherwise, either because she does not have a high enough ability or enough wealth, she chooses $s = 0$. Figure 1-6 represents the region of agents that attend higher education in the type space.

The objective function of the Social Planner in the restricted problem is

$$\max_{\psi} \int_{\psi}^{\bar{\phi}} \int_{a^*(\psi)}^{\bar{a}} \Delta w(a, a^*(\psi)) dG(a) dF(\phi) + \int_0^{\bar{\phi}} \int_{\underline{a}}^{\bar{a}} \underline{a} h(a, 0) dG(a) dF(\phi), \quad (1.16)$$

where $a^*(\psi)$ is implicitly defined by $a^* h(a^*, 1) - \underline{a} h(a^*, 0) = \psi$, and the notation $\Delta w \equiv a^* h(a^*, 1) - \underline{a} h(a^*, 0)$ has been used. The interpretation of the objective function is that all agents obtain at least human capital $\underline{a} h(a, 0)$ while the mass of agents $(1 - F(\psi))(1 - G(a^*(\psi)))$

²⁶Note that the zero cost of reallocation is important. If there was an investment stage in which the number of "seats" (capacity) of each tier are decided and they could not be changed ex-post, this would suffice to ensure that lotteries are credible.

attending school 1 obtain an additional amount of human capital. The goal of the Social Planner is to precisely maximize the additional gain coming from higher education.

Under regularity assumptions on the wealth distribution F to be discussed below, the FOC gives a sufficient condition to the problem,

$$f(\psi) \int_{a^*(\psi)}^{\bar{a}} \Delta w dG(a) da = (1 - F(\psi)) \left(\int_{a^*(\psi)}^{\bar{a}} \frac{\partial \Delta w}{\partial \psi} dG(a) - \frac{\partial a^*}{\partial \psi} \Delta w(a^*, a^*) g(a^*) \right). \quad (1.17)$$

The left hand side term in equation (1.17) captures the costs that raising the tuition fee has in reducing the number of agents attending school 1. The right hand side captures the marginal benefit that increasing the tuition fee has on raising the spillover for agents in school 1.

The next proposition identifies sufficient conditions for the solutions implicitly defined by the first order condition (1.17) to be local maxima.

Proposition 8 *If the hazard rate of the wealth distribution,*

$$\frac{f(\phi)}{1 - F(\phi)},$$

is increasing, then the first order condition (1.17) has a unique solution that is the maximum of the planner's objective function (1.16).²⁷

Proposition (8) establishes a sufficient condition for the first order condition uniquely pinning down the optimal fee ψ . This requires the hazard rate of the wealth distribution to be increasing. Thus, any distribution with a log-concave density yields a unique solution. For the purposes of the paper, the empirically relevant distributions with log-concave density are the Beta, Weibull, Gamma and Exponential distributions.²⁸ Note, however, that there can be distributions that are not log-concave and have an increasing hazard rate.²⁹ Finally, footnote 27 identifies a relaxed sufficient condition that allows to include the Pareto distribution in the set of distributions for which the first order condition is sufficient.

²⁷ If the wealth distribution satisfies the relaxed condition

$$-\frac{d}{dp} \frac{f(p)}{\psi(p)} < \left(\frac{f(p)}{\psi(p)} \right)^2, \quad (1.18)$$

then, the solution of the first order condition (1.17) may not be unique but contains the solution that maximizes the planner's objective function (1.16). The interest on the relaxed formulation is that it accommodates Pareto distributions with well defined mean, i.e., with shape parameter greater than one. This can be readily verified by checking the condition directly. The hazard rate of a Pareto distribution with index α and support lower bound x_m , $1 - F(x) = \left(\frac{x_m}{x}\right)^\alpha$, $f(x) = \alpha \frac{x_m^\alpha}{x^{\alpha+1}}$, is equal to αx^{-1} . Thus, condition (1.18) is satisfied if and only if and only if $\alpha > 1$.

²⁸ Except for the Exponential distribution, these distributions are log-concave when they have a hump-shape. Bagnoli and Bergstrom (2005) provide a discussion of the parameter ranges in which these distributions have a hump-shape.

²⁹ Bagnoli and Bergstrom (2005) show these results and provide further examples. Note that the log-normal distribution has a non-monotonic hazard rate and proposition 8 does not apply.

1.6.2.1 Comparative Statics on the Wealth Distribution

I now go back to the study of the behavior of the first order condition of the constrained problem, (1.17), under changes in the wealth distribution. The discussion is brief, as this is essentially a particular case of equation (1.14) with a degenerate lottery. To simplify the discussion in the comparative statics exercise, I assume that the wealth distribution has increasing hazard rate. Rearranging, equation (1.17) can be written as

$$\frac{f(\psi)}{1 - F(\psi)} = \frac{\int_{a^*(\psi)}^{\bar{a}} \frac{\partial \Delta w}{\partial \psi} dG(a) - \frac{\partial a^*}{\partial \psi} \Delta w(a^*, a^*) g(a^*)}{\int_{a^*(\psi)}^{\bar{a}} \Delta w dG(a)}. \quad (1.19)$$

The left hand side of equation (1.19) is the hazard rate of the wealth distribution. This is a particular case of the discussion following the general case with lotteries, equation (1.14). The right hand side is the ability-average marginal return of increasing segregation in higher education divided by the ability-average of people who stay in higher education. Note that the right hand side is independent of the wealth distribution. Moreover, the numerator is decreasing in ψ while the denominator is increasing in ψ . Thus, the right hand side of (1.19) is decreasing in ψ . This discussion yields to the following result.

Proposition 9 *Consider two wealth distributions with $\tilde{F} \succ_w F$. The education fee ψ is higher under \tilde{F} than under F .*

There is a reduction in the mass of agents obtaining higher education when the economy transitions from no agents being effectively borrowing constrained to a (small) mass of agents being borrowing constrained.

The first part of the proposition shows that the comparative statics derived in the previous section hold when only school fees are used. However, in this case a weaker condition on the ranking of wealth distribution suffices to ensure the result: as long as there is dominance in terms of the hazard rate the result follows. The second part of Proposition 9 shows that the comparative statics on the mass of agents attending higher education is inherited as well. Similarly, changes in the wealth dispersion yield results analogous to the case with lotteries.

As a final remark, the decentralization of this schooling system is immediate. The planner sets school contingent taxes equal to the optimal fees.

1.6.3 School Fees and Exams

This section studies an environment in which the social planner has access to a signaling technology: exams. These can be used as an additional mechanism to screen agents that access higher education. The reason why the social planner may want to use them is that exams have screening power. This is, it is more costly for low ability agents to obtain a given

test score. However, exam preparation implies a waste of resources (e.g., tutoring). This introduces a trade-off between better screen capacity and wasteful spending. Obviously, without borrowing constraints it is never optimal to use exams, because incentives can be given perfectly with school fees, which do not convey wasteful spending.³⁰

The signal technology considered is similar to Fernández and Galí (1999). It is represented by the mapping $T : \mathcal{A} \times \Phi \rightarrow \mathcal{T}$ with $t(a, c)$ measuring the score generated by an agent of type a who spends resources c , and $\mathcal{T} \subseteq \mathbb{R}_+$. I shall be working with the associated cost function, $c(a, t)$, which is defined implicitly by $t(a, c(a, t)) = t$, for all $a \in \mathcal{A}$ and $t \in \mathcal{T}$. In this context t has the natural interpretation of a test level. The interpretation of the cost function is that an agent with ability a has to spend an amount $c(a, t)$ to obtain a test score t in the exam. I assume that $c_a < 0$, $c_t > 0$, $c_{at} \leq 0$, $c(a; t = 0) = 0$ (i.e., not taking the exam has zero cost).

I consider two different exam specifications. I begin by characterizing the optimal mechanism with a simple pass/fail exam in which the difficulty of passing is exogenously given. Then, I consider an exam technology in which there is a continuum of possible test scores and, thus, the difficulty of an exam to access higher education becomes endogenous. Moreover, more than one test score can give access to higher education (with different school fees associated to different scores). I show that that the qualitative results derived in the simple pass/fail exam hold in the general set-up.

A simple Pass/Fail Exam

The main insight of this section can be obtained by looking at a simple formulation in which the government has access to a very limited technology, a pass-fail exam, $\mathcal{T} = \{0, 1\}$. In this environment, school fees can be indexed by whether or not an agent passes the exam (i.e., invests in the signal technology) –if the social planner decides to use the signaling technology.

When does the social planner want to use exams? Suppose only a small mass $\varepsilon > 0$ of agents is constrained in an allocation that uses only fees. In this case, it is not optimal for the social planner to use exams in the assignment mechanism. To see this, suppose that agents can access education by either paying a fee ψ_0 and not taking the exam ($t=0$) or by paying a ψ_1 and passing the exam ($t = 1$). With this mechanism at hand, all agents with ability $a \geq \tilde{a}$, with \tilde{a} defined by $\psi_0 = \psi_1 + c(\tilde{a}, t)$ prefer to take the exam. Note that the mechanism cannot give incentives along the wealth dimension and thus *all* agents with ability $a \geq \tilde{a}$ take the

³⁰The assumption that the all exam preparation is wasteful spending is an extreme one. It could be the case that agents learn by preparing an exam. In this sense, the exam component that the model is capturing is the resources that are spent to do well in an exam that are orthogonal to knowledge acquisition. Fernández (1998) argues that this constitutes a sizable part of exam preparation. Other researchers, such as Bishop (1997), have argued that exams can be beneficial because they are coordination devices. I abstract from this feature as well.

exam. As a result, conditional on a ψ_0 (which pins down the spillover level), the additional gain in output of using exams comes from the additional mass of poor agents with wealth between ψ_0 and $\psi_1 + c(a, t)$ that can access school by taking the exam but would otherwise be excluded,

$$\int_{\bar{a}}^{\bar{a}} \int_{\psi_1 + c(a, t)}^{\psi_0} \Delta w dF(\phi) dG(a). \quad (1.20)$$

The cost of using exams is the wasteful spending incurred by all agents with $a \geq \bar{a}$

$$\int_{\bar{a}}^{\bar{a}} \int_{\psi_1 + c(a, t)}^{\bar{\phi}} c(a, 1) dF(\phi) dG(a). \quad (1.21)$$

Thus, output gains are smaller than costs of introducing exams whenever the mass of additional agents that select into higher education relative to the original mass, $\int_{\bar{a}}^{\bar{a}} \int_{\psi_1 + c(a, t)}^{\psi_0} dF(\phi) / (1 - \psi_0)$, is small. This discussion shows the following proposition.

Proposition 10 *Consider a test technology $\mathcal{T} = \{0, 1\}$ and a family of wealth distributions that can be ranked according to the wealth abundance criterion. Then, there exists a threshold distribution F^* such that for all $F \succ_w F^*$ the optimal allocation mechanism does not use exams, i.e., $a^* = \bar{a}$. Moreover, if $F_1 \succ_w F_0$, then $a_1^* \leq a_0^*$ and $\bar{a}_1 \leq \bar{a}_0$.*

Proposition 10 states the two key results of this section. First, sufficiently wealth abundant economies do not use exams. Second, the less wealth abundant an economy is, the more it relies on exams to allocate agents to higher education. To be more precise, the ability range in which exams are used, $[\bar{a}, \bar{a}]$, increases. Moreover, the comparative statics on the threshold type attending higher education are the same as in the previous sections (propositions 3 and 9). Thus, less wealth abundant economies have a worse selection of agents into higher education in terms of ability.

These results resonate with the empirical evidence presented in the Introduction. Developing countries make relatively more extensive use of gate-keeping exams to complete basic education and access higher levels of education. This is often coupled with tutoring to prepare exams, especially in Asia, Africa, Latin America and Eastern Europe (Bray, 2000).

The pattern of selection of types that attend higher education is pictured in Figure 1-7. In general, it can be the case that $\bar{a} > a^*$. This resonates as well with practices of access to higher education. For example, in India, access to prestigious higher education institutions can be done through two different paths. Access for the general body of students is through an exam requirement and a tuition fee. But, in addition, there is access through school fees only, known as management quotas.³¹ A similar finding is documented for Tanzania by Al-Samarrai and Peasgood (1998).

³¹I thank Abhijit Banerjee for pointing out this example to me.

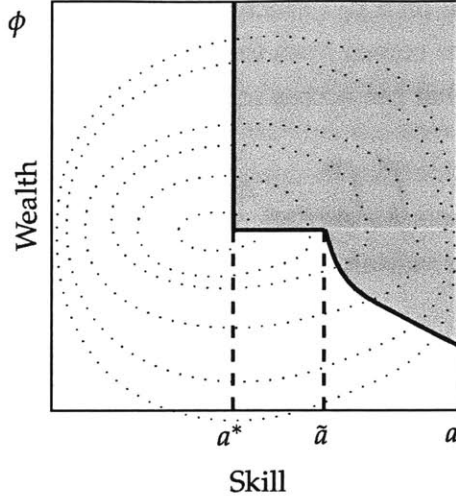


Figure 1-7: Allocation with school fees and a pass/fail exam.

Full-Blown Exam Technology and Fellowship Schemes

I now consider the case in which the social planner has access to a continuum of exam technologies, $\mathcal{T} = [0, \infty)$. That is, the planner can ask an agent to obtain any score $t \in \mathcal{T}$. I show that in this environment the qualitative results of Proposition 10 hold. More specifically, it is still optimal to reduce school quality and resort more extensively in exams in less wealth abundant economies. Before proceeding, the following assumption on costs is made.

Assumption 2 *The associated cost function to the exam technology takes the form $c(t, a) = c_1(t)c_2(a)$, where c_1 and c_2 are positive, twice-continuously differentiable functions. Moreover, $c_2(a)$ is log-convex.*

Using the revelation principle, I look for a schedule $\{\psi(a), t(a)\}$. The social planner problem is

$$\max_{\psi(a), t(a)} \int_{a^*}^{\bar{a}} \int_{\psi+c(t,a)}^{\bar{\phi}} (\Delta H(a, a^*) - \kappa - c(a, t)) dG(a) dF(\phi), \quad (1.22)$$

where $\Delta H(a, a^*) \equiv a^*h(a, 1) - ah(a, 0)$, subject to

$$\Delta H(a, a^*) - \psi(a) - c(t(a), a) \geq 0 \quad \text{for all } a \in [a^*, 1], \quad (1.23)$$

$$\Delta H(a, a^*) - \psi(a) - c(t(a), a) \geq \Delta H(a, a^*) - \psi(\hat{a}) - c(t(\hat{a}), a), \quad (1.24)$$

for all $a, \hat{a} \in [a^*, 1]$, where (1.23) and (1.24) are the participation and incentive compatibility constraints. I adopt a first order approach to solve the problem. I first discuss the sufficient conditions for implementability and then discuss the optimization part. The set of sufficient conditions for implementability under the first order approach are as in a standard screening problem (e.g., Bolton and Dewatripont (2005)). The incentive compatibility constraints are

satisfied if and only if there is local incentive compatibility

$$\psi'(a) + c_t(t, a)t'(a) = 0, \quad (1.25)$$

and monotonicity,

$$t'(a) \geq 0 \quad \text{and} \quad \psi'(a) \leq 0. \quad (1.26)$$

Note that the implementation problem is essentially one dimensional, because incentives are provided along the ability dimension only. Calculating the difference of the second order condition with the total derivative of the local incentive compatibility constraint only involves $t'(a)$. However, differently than the standard screening problem, one cannot get rid of t and ψ in the objective function by integrating by parts. Finally, before proceeding to the optimization stage, using the complementarity in the intrinsic human capital technology, the participation constraint can be substituted by the condition that the marginal type accessing higher education has to be indifferent,

$$\Delta H(a^*, a^*) - \psi(a^*) - c(t(a^*), a^*) = 0. \quad (1.27)$$

The Social Planner problem needs to be solved in two steps because the optimization problem cannot be written exactly as an optimal control or calculus of variations problem. This comes from the complementarity between the spillover and the intrinsic human capital production function. That is, in addition to the initial boundary condition for a^* to lie on the curve defined by (1.27), any change in a^* affects the value at any point of the integrand of the objective function (1.22). Thus, first I solve an inner problem in which the spillover term a^* that enters in $\Delta H(a, a^*)$ is kept fixed at \hat{a}^* . This problem is readily amenable to optimal control techniques. Note that at this step, I obtain an optimal value for a^* (\hat{a}^*) coming from the initial condition problem. Then, I solve an outer problem for \hat{a}^* under the constraint that $a^* = \hat{a}^*$.

Appendix 1.10 characterizes the optimal solution and the comparative statics in the wealth distribution. The results are summarized below.

Proposition 11 *Consider a family of log-concave wealth distributions. Let $\tilde{F} \succ_w F$, then (i) the total cost of education $\psi(a) + c(a, t(a))$ at all ability levels $a \in [a^*, \bar{a}]$ is greater under \tilde{F} , (ii) the test level $t(a)$ at all ability levels $a \in [a^*, \bar{a}]$ is smaller under \tilde{F} , (iii) the threshold type a^* is greater under \tilde{F} and (iv) the optimal fee-test schedule does not have any bunching region.*

These results generalize the ones obtained in Proposition 10 for the simple pass/fail exam technology. Poorer economies rely more on exams to access higher education. The total cost of education holding ability constant is increasing with a wealth abundance shift. This means that the total cost of education (exam plus transfer) in poor countries is less than in

rich countries. However, point (ii) of Proposition 11 shows that the test level required to an agent of a given ability is higher in less wealth abundant countries. These two observations imply that the ratio of exam cost to total cost

$$\frac{c(a, t(a))}{c(a, t(a)) + \psi(a)}$$

is decreasing upon wealth abundance shifts. Thus, exam expenditure relative to total school expenditure is higher in poorer countries. This result relates to the stylized fact discussed in the Introduction that developing countries rely more on exams than rich countries. Finally, as in the previous comparative statics results, point (iii) shows that the ability-composition of agents selecting into higher education is worse in poorer countries.

A corollary of Proposition 11 is the following. Consider a wealth distribution such that higher education is provided to some agents with wealth below the median. Let $\tilde{F} \succ_d F$, then the total cost of education $\psi(a) + c(a, t(a))$ decreases for agents below the median wealth relative to those above, but test requirements increase. The difference in test scores required to access schools increases for agents with wealth below the median relative to those above it. That is, exam requirements (i.e., test levels) increase in the range in which the mass of borrowing constrained agents rises. This comes as no surprise, exams are more intensively used in the region where inequality changes are more pronounced, which is precisely where their screening power has a higher relative benefit.

Decentralization. The previous mechanism gives incentives by rewarding (high ability) agents that obtain high test scores with low school fees. In a decentralized equilibrium, (as the marginal cost of provision is zero) this is implemented with a tax contingent on school and test performance, so that agents that obtain better grades pay a lower tax. Thus, this very much resembles the usage of scholarship schemes. The previous discussion implies that poorer countries rely more in scholarship-like schedules to implement the optimal solution.

1.7 Extensions

The goal of this section is to show that the results highlighted in the baseline model hold in more general environments. In Subsection 1.7.1, I consider a yeomen-farmer like economy, in which each agent produces a differentiated intermediate good, and the final consumption good is a CES composite of all intermediates. Then, I discuss how the results extend to more general type of spillovers.

Subsection 1.7.2 analyzes an environment in which, within each education tier, there is a continuum of sub-tiers, so that a school can be tailored to each ability level. Even though this extension mutes the endogenous quality degradation margin, it provides a useful benchmark to analyze how the mass of mismatched agents and school capacities change at differ-

ent wealth levels. I use this simplified environment to show how the planner uses cross-subsidization within schools to increase access of borrowing constrained agents to schools and to show how the presence of an unregulated private sector provider may hamper its ability to do so.

1.7.1 Yeomen farmers and general spillovers

Consider the following extension of the baseline model. Each agent produces a differentiated intermediate good with her human capital. The final good is produced as an aggregator of intermediates with elasticity ε ,

$$Y = \left(\int y(i)^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \quad \varepsilon > 1.$$

Thus, the technology specification in the baseline model is a limiting case in which all intermediates are perfectly substitutable ($\varepsilon \rightarrow \infty$).

Markets are competitive, and payments to factors of production are made according to marginal productivity.³² Thus, an agent with ability a that attends school s , earns

$$w(a, s) = Y^{1/\varepsilon} H(a, s)^{\frac{\varepsilon-1}{\varepsilon}} = Y^{1/\varepsilon} A(s)^{\frac{\varepsilon-1}{\varepsilon}} h(a, s)^{\frac{\varepsilon-1}{\varepsilon}}.$$

This payoff structure resembles the structure of Bénabou (1996b) in that there is a “local” and “global” externality.

Next, I show that the same qualitative results hold in this generalized set-up. Define $\tilde{h}(a, s) = h(a, s)^{\varepsilon/(\varepsilon-1)}$. Note that the complementarity between a and s is preserved as $\varepsilon > 1$. Analogously, define $\tilde{A}(s) = (\min_s \{a \in s\})^{\varepsilon/(\varepsilon-1)}$. Let $a^* = \min_s \{a \in s\}$. The derivative of $\tilde{A}(s)$ with respect to a^* is positive. The argument that maximizes the value of $Y(a^*)$ coincides with $Y(a^*)^{(\varepsilon-1)/\varepsilon}$. Thus, the problem of finding the argmax of $Y(a^*)^{(\varepsilon-1)/\varepsilon}$ is isomorphic to the baseline model replacing $\tilde{h}(a, s)$ for $h(a, s)$, and $\tilde{A}(s)$ for $A(s)$. As a result, the same methods and results derived in the linear technology case apply to this general set-up.³³

Now, I discuss how the results extend in two different alternative specifications of the spillover. The main lead theme of both specifications is to relax the “least common denominator” specification to address the concerns that (i) agents may learn even if a curriculum is tailored for higher ability agents (Duflo et al. (2008) offer evidence along these lines), (ii) there are other forces that can generate the complementarity in human capital production.

A simple extension that generalizes the results presented in the baseline model without

³²I maintain the assumption that the social planner cannot manipulate the production of goods, it can only choose the educational system structure.

³³Note, however, that in this case the sufficiency conditions derived for the baseline model need to be adjusted by the presence of the additional factor $\varepsilon/\varepsilon - 1$.

adding any complexity to the problem is to allow the spillover to be a convex combination of the highest and lowest ability agent in a particular school,

$$A(s) = \alpha \min_s \{a \in s\} + (1 - \alpha) \max_s \{a \in s\}, \quad \alpha \in [0, 1]. \quad (1.28)$$

Note that in this formulation, a reduction on the threshold type attending higher education would affect both the spillovers at the basic and higher education. However, it is immediate to check that the complementarity in the intrinsic human capital h is enough to ensure that the results derived in the baseline hold with a spillover as in (1.28).

One might argue that specification (1.28) is blind to whether most of agents are close to the maximum or the minimum, and that this is likely to matter. One can further generalize the spillover to

$$A(s) = \left(\iint a^{\frac{\sigma-1}{\sigma}} dZ(a, \phi|s) \right)^{\frac{\sigma}{\sigma-1}}, \quad (1.29)$$

where $Z(a, \phi|s)$ is the joint ability-wealth distribution of types selecting into school s . This spillover spans the range from a Leontief as $1/\sigma$ grows to infinity to a “best shot” as $1/\sigma$ goes to minus infinity. Note that as $\sigma \rightarrow \infty$ the spillover becomes the average type attending the school. The results derived in the baseline section can be extended to a spillover with this specification (1.29) if basic education is modeled as an outside option with its value normalized to zero for all agents (i.e., $h(a, 0) = 0$). In this case, it is immediate to verify that analogous results to the baseline follow. Intuitively, when the spillover is of the “best shot” no segregation is optimal, while when it is Leontief, some segregation is always optimal. An intermediate value of $1/\sigma$, gives an intermediate level of segregation between these two extremes.

If the outside option of basic education is not normalized to zero then the problem becomes more complicated and some additional structure to the human capital production function and the distribution of types is needed in order to have clear comparative statics. This comes from the fact that the difference in spillovers $A(1)h(a, 1) - A(0)h(a, 0)$ may be either concave or convex on changes in the marginal type obtaining education when there are borrowing constraints.

1.7.2 Continuum of Schools

This section analyzes the optimal schooling system when within the two broad curricula of basic and higher education, there are finer curriculum options. For example, higher education can be subdivided in associate, bachelor, master and Ph.D. degrees. Even further, it may be the case that different schools can have some margin to adapt their curricula. For example, to the extent that educational institutions have some discretion in setting their own standards, curriculum requirements may vary to some extent among institutions in the same

education tier. I capture this richer environment by allowing for different sub-tiers in both mandatory and higher education. These can be interpreted as either a finer partition across different educational tiers or differences at school level. As it turns out, the most convenient formulation is to allow for a continuum of sub-tiers within each educational tier.

The goal of this exercise is two-fold. First, it allows to investigate to which extent the reduction in capacity of higher education schools and the mismatch of ability to schools persists even when there is a school tailored to each ability level. Second, it provides a natural framework to analyze competition between an education sector regulated by the social planner and an unregulated sector.³⁴

Compared to the baseline environment, in this section I allow for a marginal cost of provision of higher education κ . More importantly, I allow the planner to cross-subsidize across schools. That is, the budget constraint of transfers has to break even on net, and not school by school. That is, as opposed to the baseline model, the planner has the ability to do instantaneous redistribution, as in equation (1.9).

1.7.2.1 First Best characterization and no Borrowing Constraints

This section shows that the result derived in the two-tier case which stated that private information alone does not prevent the social planner from implementing the first best allocation system generalizes to more school tiers. The intuition for the result is the same: given the complementarities between endogenous school quality and types, higher ability agents are willing to pay more for higher quality education. This section proceeds as follows. I begin by characterizing the first best. Then, I show how a mechanism can be designed to implement the first best educational system when private information is present.

Given the complementarity of a and s , there is full segregation in the first best. That is, the first best features one school for each type a . To see this, suppose to the contrary that two types of different ability $a_l < a_h$ with associated mass m_{a_l} and m_{a_h} attend the same school tier. Then, by segregating them, output can be increased as a_h benefit from a higher spillover, without reducing the output of a_l . Types with $a \leq a^*$ attend school 0, where a^* is defined by

$$a^*(h(a^*, 1) - h(a^*, 0)) = \kappa.$$

Next, I show that the first best educational system can be achieved with private information using a school fee schedule. This is the continuous case counterpart to Proposition 1.

³⁴As opposed to the two-tier framework, this framework allows to isolate the effect of competition. In the two-tier system, additional provision from the unregulated sector generates a mechanical force towards having gains from an additional private provider just because it allows more segregation. Moreover, a two-tier system easily runs into the problem of multiple equilibria.

The problem of the social planner is

$$\max_{\psi(a)} \int_{\underline{a}}^{a^*} H(a,0)da + \int_{a^*}^{\bar{a}} (H(a,1) - \kappa)da$$

subject to

$$\begin{aligned} ah(a, \mathbf{1}(a)) - \psi(a) &\geq 0 \quad \text{for all } a \in [0, 1], \\ ah(a, \mathbf{1}(a)) - \psi(a) &\geq \hat{a}h(a, \mathbf{1}(\hat{a})) - \psi(\hat{a}) \end{aligned}$$

for all $a, \hat{a} \in [0, 1]$. And the break even constraint,

$$\int_{\underline{a}}^{\bar{a}} (\psi(a) - \kappa)da \geq 0.$$

The fee schedule that solves the problem is,

$$\psi(a) = \begin{cases} \int_{\underline{a}}^a h(a,0)da & \text{for } a \leq a^*, \\ \int_{a^*}^a h(a,1)da + \int_{\underline{a}}^{a^*} h(a,0)da & \text{for } a > a^*. \end{cases} \quad (1.30)$$

where it is used the normalization that $h(a = \underline{a}, 0) = 0$. Figure 1-8a provides a graphical representation of the result. Note that the solution features standard properties from screening mechanisms. "All rents" are extracted for the lowest ability type $a = 0$ and consumption is increasing in ability, $c'(a) = ah_a(a, \cdot)$. Moreover, it can be verified that the break even constraint is satisfied with inequality.

Finally, this solution coincides with the First Best because there is complete segregation in skill. The following proposition summarizes the previous discussion.

Proposition 12 *The first best educational system features full-segregation. The optimal mechanism with private information can implement the first best educational system.*

Before proceeding note that there can be other price schedules that implement the first best if $h(a,0) > 0$. However, these are constrained to have the same slope as (1.30). The reason is simple, at any other slope, there would be bunching of some types and this cannot be optimal by the previous argument that segregation is always optimal. This is stated in the next remark.

Remark Any transfer that implements truth-telling revelation of ability has to have slope in ability given (almost everywhere) by either the slope of $\int_0^a h(a,0)da$ for $s = 0$ or $\int_{a^*}^a h(a,1)da$ for $s = 1$.

1.7.2.2 Decentralization without borrowing constraints

In this section I show that there exists an equilibrium that decentralizes the optimal schooling system. Define a competitive schooling equilibrium as a supply of schools S , pricing function $p : S \rightarrow \mathbb{R}_+$ and agents' choices $c : \mathcal{A} \rightarrow S$ such that (i) agents' school choices maximize utility at the stated prices, (ii) firms maximize profits, (iii) markets clear.

First, I show that it exists an equilibrium in which prices coincide with fees set by the social planner (1.30). The reason why sub-tiers can exist is that, despite using the same intrinsic human capital production function, different schools offer different spillovers at different prices. Thus, school prices p are indexed by both the intrinsic technology s and the spillover level $a \in \mathcal{A}$. Consider the following price schedule for $p(a, s)$

$$\begin{cases} p(a, 0) = \int_0^a h(a, 0) da & \text{for } a < \bar{a}, \\ p(a, 1) = \int_{\bar{a}}^a h(a, 1) da + \kappa & \text{for } a \geq \bar{a}, \end{cases} \quad (1.31)$$

where κ is the marginal cost of school provision and \bar{a} is implicitly defined by $\bar{a}h(\bar{a}, 1) - \kappa = \bar{a}h(\bar{a}, 0) - p(\bar{a}, 0)$. Note that this price schedule, conditional on s , is convex in a . Agents utility maximization,

$$\max_{\hat{a}, s} \{\hat{a}h(\hat{a}, s) - p(\hat{a}, s), 0\} \quad (1.32)$$

can be solved sequentially. First, find the optimal demand of \hat{a} in each school tier, and then comparing utility at the optimal $\hat{a}(s)$ from attending basic and higher education (or not educating at all). The solution to (1.32) is $\hat{a} = a$. Thus an agent with ability $a < \bar{a}$ chooses $s = 0$ and an agent with $a > \bar{a}$ chooses $s = 1$, an agent with $a = \bar{a}$ is indifferent, and no agent decides not to educate.

Given this price schedule, education provision for any (a, s) is provided at no loss, as $p(a, 0) \geq 0$ and $p(a, 1) \geq \kappa$. In fact, schools make positive profits in equilibrium and no entrant can attract agents by offering lower prices. To see this, consider an entrant that provides (a, s) at a price lower than $p(a, s)$. At a lower price, agents with lower ability would purchase the schooling good (a, s) , but this would result in a spillover effect of lower ability than a . Thus, no price lower than $p(a, s)$ can be credibly offered. The question is that with positive profits, there will always be schools willing to enter the market. One possibility to discipline the model is to have a given measure of potential entrants, so that schools per se become infinitesimal. Another is to have an additional type specific school production input in fixed supply that is needed to provide education, so that as more and more firms enter a particular school market, the price of the input goes up. This would pin down the number of schools entering the market. The details of how to pin down the number of firms is inessential for purposes of this discussion. In either case, the relevant part is that at the stated prices, the demand for schooling at all ability levels is positive and is met by a supply.

Thus, there is market clearing.

Remark Any price schedule that perfectly separates agents in terms of ability has to have slope in ability given (almost everywhere) by either the slope of $\int_0^a h(a,0)da$ for $s = 0$ or $\int_a^1 h(a,1)da$ for $s = 1$.

The equilibrium proposed as it stands does not coincide with the first best solution. In this equilibrium there is more provision of higher education than in the first best, as the lowest ability type attending higher education is

$$ah(a,1) - \kappa = ah(a,0) - \int_0^a h(a,0)da, \quad (1.33)$$

while in the social planner solution is $ah(a,1) - \kappa = ah(a,0)$. This makes clear that a tax contingent on attending school tier $s = 1$ of value $\tau = \int_0^a h(a,0)da$ implements the same assignment of agents to schools as in the first best.³⁵

1.7.2.3 Private Information with Borrowing Constraints

I study the mechanism design problem when there are borrowing constrained agents. The following result greatly simplifies the analysis.

Remark (No bunching) Whenever there exists an agent of ability a that can afford segregation, it is optimal to offer a school tailored for agents of ability a .

This result comes from the extreme complementarity in the spillover, which implies that segregation of high-ability, unconstrained agents does not reduce the utility of high-ability constrained agents. Segregation has the advantage of increasing output and increasing revenue of the social planner (which implies a reduction on the schooling fee across-the-board). An important corollary of this remark is that it is not optimal to use lotteries. The reason is simple: given the Leontief spillover, this would involve bunching types. Thus, announcements on the wealth dimension are not relevant to the social planner because there is no added value on randomization.

With this results at hand, the mechanism to be used can be rewritten as follows. Let $\psi(a)$ denote the fee that agents of ability a pay if they announce to be of ability a . Then, given an announcement (a, ϕ) the fee that any agent with $\phi > \psi(a)$ has to pay is $\psi(a)$, while if the

³⁵Note that if $h(\underline{a}, 0) = 0$, this is the only equilibrium that features full segregation. If $h(\underline{a}, 0) > 0$, there could be other equilibria that decentralize the first best, which would be a translation $p(a, s) + \beta$ for some $0 \leq \beta \leq \underline{a}h(\underline{a}, 0)$.

agent is constrained the fee is just ϕ . The problem of the social planner can be written as

$$\begin{aligned} \max_{\psi(\underline{a}), \mathbf{1}(a)} \quad & \int_{\underline{a}}^{\bar{a}} \int_{\psi(a)}^{\bar{a}} (H(a, \mathbf{1}(a)) - \mathbf{1}(a)\kappa) dF(\phi) dG(a) + \\ & \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\psi(a)} (\psi^{-1}(\phi)h(a, \mathbf{1}(\psi^{-1}(\phi))) - \mathbf{1}(\psi^{-1}(\phi))\kappa) dF(\phi) dG(a), \end{aligned}$$

subject to the participation and incentive compatibility constraints,

$$\begin{aligned} ah(a, \mathbf{1}(a)) - \psi(a) &\geq 0 \quad \text{for all } a \in [\underline{a}, \bar{a}], \\ ah(a, \mathbf{1}(a)) - \psi(a) &\geq \hat{a}h(a, \mathbf{1}(\hat{a})) - \psi(\hat{a}), \end{aligned}$$

for all $a, \hat{a} \in [\underline{a}, \bar{a}]$ that are unconstrained, and the break-even constraint

$$\begin{aligned} \int_{\underline{a}}^{\bar{a}} \left[(1 - F(\psi(a)))\psi(a) + \int_{\underline{a}}^{\psi(a)} \phi dF(\phi) \right] dG(a) \\ \geq \kappa(1 - F(\psi(a^*))) \int_{a^*}^{\bar{a}} dG(a). \end{aligned}$$

From the results in the previous section, a fee-schedule that achieves full segregation has to have the slope of equation (1.30). In this case, the presence of borrowing constraints imply that it is optimal to reduce as much as possible fees (conditional on the break even constraint being binding). Thus, the level of the fee schedule $\psi(a)$ is going to be reduced as much as possible (up to the break-even constraint being binding) without changing the slope. More specifically, the solution of the problem is given by

$$\psi(a) = \begin{cases} \int_0^a h(a, 0) da - \mathcal{C} & \text{for } a \leq a^*, \\ \int_{a^*}^a h(a, 1) da + \int_0^{a^*} h(a, 0) da - \mathcal{C} & \text{for } a > a^*. \end{cases} \quad (1.34)$$

with $a^*h(a^*, 1) - \kappa = a^*h(\tilde{a}, 0)$, where \mathcal{C} is pinned down by the break even constraint

$$\int_{\underline{a}}^{\bar{a}} \left[(1 - F(\psi(a)))\psi(a) + \int_0^{\psi(a)} \phi f(\phi) d\phi \right] da = \kappa(1 - F(\psi(a^*))(1 - G(a^*)). \quad (1.35)$$

Remark on impossibility of decentralization. The decentralization that can be achieved is exactly the same as in the case without borrowing constraints and it is omitted from the discussion. Note that this differs from the optimal mechanism, as the price level in the decentralized equilibrium is too high. The reason is that while the planner uses revenue from schools to reduce the price level, private providers cannot do that. Thus the social planner educational

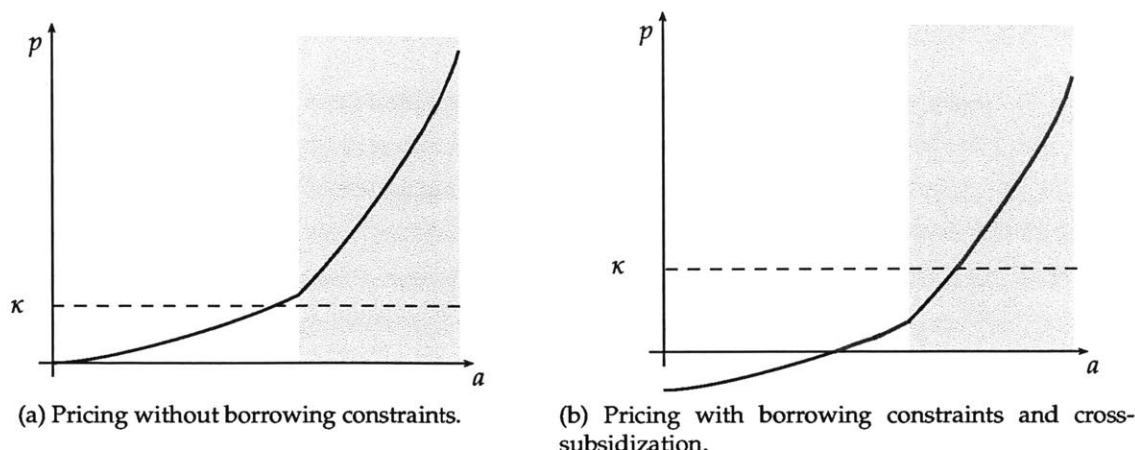


Figure 1-8: Optimal pricing with a continuum of schools and two technologies

system cannot be decentralized in this case.³⁶

1.7.2.4 Comparative Statics on the Wealth Distribution

This section analyzes how changes in wealth abundance and dispersion affect the mismatch of agents to schools and school tier capacity. Note that the extreme assumption of the continuum of tiers washes out the endogenous deterioration of quality arising from changes in the wealth distribution.

Proposition 13 (*Mismatch of agents to school*) Let $\tilde{F} \succ_w F$, the percentage of agents mismatched is higher at all levels of schooling under F . Let $\tilde{F} \succ_a F$, the percentage of agents with wealth above the median that are mismatched decreases relative to agents below the median.

This results are immediately interpretable, and generalize those of the baseline model. In countries that are relatively poor, there are more agents mismatched in the sense that they attend a school tailored for lower ability agents at all levels of education. In more unequal countries, the mismatch is specially aggravated at low levels of education. The logic of the proof is simple. Upon a wealth shift, planner's revenue increases and this makes the fee level to go down -this makes borrowing constraints less severe. The fraction of agents mismatched in school a is

$$\frac{f(\psi(a))}{1 - F(\psi(a))} \frac{\int_a^{\bar{a}} dG(a)}{g(a)}, \quad (1.36)$$

which decreases as well upon a wealth abundance shift. Thus, both effects go in the same direction and the result follows. A similar reasoning applies for changes in wealth dispersion.

³⁶Appendix 1.8 shows that the same result holds in the two school case if simultaneous redistribution is allowed.

In this case, given the convexity of school fees, revenue of the social planner can increase. In this case, this would reduce the fee level across the board. For agents above the median wealth this amplifies the decrease in the fraction of mismatched agents. However, for agents below the median wealth the effect would be ambiguous because of the dispersion shift and the reduction in fee levels going on opposite directions.

A related question that this model is better suited to answer than the baseline model is how does the mass of agents attending each school-tier change with changes in the wealth distribution.

Proposition 14 (*Mass of agents attending school*) Consider $\tilde{F} \succ_w F$, then the school capacity of all schools with index $a > a_m$ increases under \tilde{F} . Consider $\tilde{F} \succ_d F$, then the school capacity of all schools with index $a > a_m$ increases under \tilde{F} , the effect at the bottom is ambiguous.

This result implies that top schools (those with index greater than a_m) in poorer and unequal societies feature less capacity than in richer countries. The exact value of a_m depends on the specifics of the wealth distribution. It suffices to note that this result always holds for \bar{a} . Then, by a continuity argument, it holds in some neighborhood of \bar{a} . The results in Proposition 14 are the analogous counterparts of the message provided in the baseline model that higher education provision is reduced in poor, unequal economies. The continuum of schools case allows to identify that the reduction in capacity is localized at the top schools within the higher education tier.

1.7.2.5 Optimal Mechanism with (Unregulated) Private Schools Outside Option

Finally, I discuss the case in which the social planner has to design the optimal mechanism facing a new additional constraint. There is a mass of private agents that have access to the schooling technology and can escape from the regulation of the social planner. Thus, these private agents are free to provide education at any sub-tier that they find profitable to. This set-up is meant to be a first pass exercise in understanding possible interactions between public and private education, with the caveat that public education needs not to coincide with the social planner's optimal educational system as it may be subject to additional constraints not modeled here.

The point I want to illustrate is simple: private provision undermines the capacity of the social planner to cross-subsidize education. The reason is that private provision competes with the planner only on the profitable segments of the market, which are precisely the source the planner uses to provide subsidization. Put shortly, private firms cream-skim the market for education. To illustrate the point, suppose that the social planner tried to implement the optimal mechanism described in (1.34). At all school levels in which $p(a, 0) \geq 0$ and $p(a, 1) \geq \kappa$, private provision occurs, because at the stated fees private firms make positive

profits. These are the regions depicted in red in Figure 1-8b. Consider for now the extreme case in which, *ceteris paribus*, agents prefer to attend private schools. In this case, all the sources of positive revenue from the social planner would disappear and the social planner's budget constraint would not be satisfied. Thus, the conjectured equilibrium ceases to be an equilibrium in the presence of private schools. In this case, the planner, would have to increase the fee level to ensure some positive revenue. However, this would backfire because all the schools yielding positive profit would be captured by the private sector. As a result, the only equilibrium that would survive is one in which there is only the private provision, as in (1.31). In this case, there would not be any scope for school cross-subsidization.

On the opposite extreme, one can consider the case in which, *ceteris paribus* agents prefer to attend public schools. By a similar argument, one can show that the (constrained) efficient mechanism analyzed in the previous section can be implemented in this case. Presumably, a realistic benchmark is somewhere between these two opposite poles. The point to take away is that if private providers coexist with schooling provided by the planner, then private school provision puts limits to-cross school subsidization. This suggests that the cream-skimming of the sources of cross-subsidization by the private providers in poor countries may hamper the capability of the social planner to cross-subsidize education, ultimately reducing the effective level of education that can be provided to credit constrained agents.

Bibliography

- Al-Samarrai, S. and Peasgood, T. (1998). Educational attainments and household characteristics in tanzania. *Economics of Education Review*, 17(4):395–417.
- Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory*, 26(2):445–469.
- Banerjee, A. V. (1997). A theory of misgovernance. *The Quarterly Journal of Economics*, 112(4):1289–1332.
- Barro, R. J. and Lee, J.-W. (1993). International comparisons of educational attainment. *Journal of Monetary Economics*, 32(3):363–394.
- Becker, G. S., Murphy, K. M., and Werning, I. (2005). The equilibrium distribution of income and the market for status. *Journal of Political Economy*, 113(2):282–310.
- Bénabou, R. (1993). Workings of a city: Location, education, and production. *The Quarterly Journal of Economics*, 108(3):619–52.
- Bénabou, R. (1996a). Equity and efficiency in human capital investment: The local connection. *Review of Economic Studies*, 63(2):237–64.
- Bénabou, R. (1996b). Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance. *American Economic Review*, 86(3):584–609.
- Bénabou, R. (2002). Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica*, 70(2):481–517.
- Betts, J. R. (1998). The impact of educational standards on the level and distribution of earnings. *American Economic Review*, 88(1):266–75.
- Bishop, J. H. (1997). The effect of national standards and curriculum-based exams on achievement. *American Economic Review*, 87(2):260–64.

- Blackorby, C. and Szalay, D. (2007). Multidimensional screening, affiliation, and full separation. The Warwick Economics Research Paper Series (TWERPS) 802, University of Warwick, Department of Economics.
- Bolton, P. and Dewatripont, M. (2005). *Contract Theory*. MIT Press.
- Bray, M. (2000). The shadow education system: private tutoring and its implications for planners. Technical report, UNESCO, International Institute for Educational Planning.
- Chachuat, B. (2007). Nonlinear and Dynamic Optimization: From Theory to Practice. url: <http://lawwww.epfl.ch/page4234.html>.
- Che, Y.-K. and Gale, I. (1998). Standard auctions with financially constrained bidders. *Review of Economic Studies*, 65(1):1–21.
- Che, Y.-K. and Gale, I. (2000). The optimal mechanism for selling to a budget-constrained buyer. *Journal of Economic Theory*, 92(2):198–233.
- Che, Y.-K. and Gale, I. (2009). Market versus non-market assignment of ownership. Discussion Papers 0607-05, Columbia University, Department of Economics.
- Cole, H. L. and Prescott, E. C. (1997). Valuation equilibrium with clubs. *Journal of Economic Theory*, 74(1):19–39.
- Condorelli, D. (2009). Market and non-market mechanisms for the optimal allocation of scarce resources. Discussion Papers 1483, Northwestern University, Center for Mathematical Studies in Economics and Management Science.
- Costinot, A. and Vogel, J. (2010). Matching and inequality in the world economy. *Journal of Political Economy*, 118(4):747–786.
- Costrell, R. M. (1994). A simple model of educational standards. *American Economic Review*, 84(4):956–71.
- De Fraja, G. (2002). The design of optimal education policies. *Review of Economic Studies*, 69(2):437–66.
- Duflo, E., Dupas, P., and Kremer, M. (2008). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. NBER Working Papers 14475, National Bureau of Economic Research, Inc.
- Durlauf, S. N. and Seshadri, A. (2003). Is assortative matching efficient? *Economic Theory*, 21(2):475–493.

- Engerman, S. L. and Sokoloff, K. L. (2000). Institutions, factor endowments, and paths of development in the new world. *Journal of Economic Perspectives*, 14(3):217–232.
- Engerman, S. L. and Sokoloff, K. L. (2002). Factor endowments, inequality, and paths of development among new world economics. Working Paper 9259, National Bureau of Economic Research.
- Epple, D. and Romano, R. E. (1998). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, 88(1):33–62.
- Esteban, J. and Ray, D. (2006). Inequality, lobbying, and resource allocation. *American Economic Review*, 96(1):257–279.
- Fernández, R. (1998). Education and borrowing constraints: Tests vs. prices. CEPR Discussion Papers 1913, C.E.P.R. Discussion Papers.
- Fernández, R. and Galí, J. (1999). To each according to . . . ? markets, tournaments, and the matching problem with borrowing constraints. *Review of Economic Studies*, 66(4):799–824.
- Fernández, R. and Rogerson, R. (1996). Income distribution, communities, and the quality of public education. *The Quarterly Journal of Economics*, 111(1):135–64.
- Fernández, R. and Rogerson, R. (1998). Public education and income distribution: A dynamic quantitative evaluation of education-finance reform. *American Economic Review*, 88(4):813–33.
- Fernández, R. and Rogerson, R. (2003). Equity and resources: An analysis of education finance systems. *Journal of Political Economy*, 111(4):858–897.
- Glomm, G. and Ravikumar, B. (1992). Public versus private investment in human capital endogenous growth and income inequality. *Journal of Political Economy*, 100(4):813–34.
- Hanushek, E. A. and Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, 90(5):1184–1208.
- Hanushek, E. A. and Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3):607–68.
- Hanushek, E. A. and Woessmann, L. (2009). Do better schools lead to more growth? cognitive skills, economic outcomes, and causation. NBER Working Papers 14633, National Bureau of Economic Research, Inc.
- Kellaghan, T. (2004). Public examinations, national and international assessments, and educational policy. Mimeo, Educational Research Centre St Patrick s College.

- Kellaghan, T. and Greaney, V. (1992). Using examinations to improve education: a study of fourteen african countries. World Bank Technical Papers 165, World Bank.
- Kremer, M. and Maskin, E. (1996). Wage inequality and segregation by skill. NBER Working Papers 5718, National Bureau of Economic Research, Inc.
- Legros, P. and Newman, A. F. (2007). Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities. *Econometrica*, 75(4):1073–1102.
- Lewis, T. R. and Sappington, D. E. M. (2000). Contracting with wealth-constrained agents. *International Economic Review*, 41(3):743–67.
- Lewis, T. R. and Sappington, D. E. M. (2001). Optimal contracting with private knowledge of wealth and ability. *Review of Economic Studies*, 68(1):21–44.
- Lockheed, M. and Mete, C. (2007). Tunisia: Strong central policies for gender equity. In *Exclusion, Gender and Schooling: Case Studies from the Developing World*, chapter 8, pages 205–230. Center for Global Development, Washington, D.C.
- Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. John Wesley & Sons.
- Mete, C. (2004). The inequality implications of highly selective promotion practices. *Economics of Education Review*, 23(3):301–314.
- Nickell, S. (2004). Poverty and worklessness in britain. *Economic Journal*, 114(494):C1–C25.
- Sattinger, M. (1993). Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31(2):831–80.
- Varian, H. R. (1992). *Microeconomic Analysis, Third Edition*. W. W. Norton & Company, 3rd edition.

1.8 Appendix: Model with Negative Transfers

1.8.1 The mechanism design problem

In this appendix, I show how the same qualitative results of the baseline model hold when the planner can make use of negative transfers subject to a global break-even constraint, (1.9). First, I state the counterpart of Proposition 2 for the optimal mechanism.

Proposition 15 (*Optimal Schedule with negative transfers*) *If transfers are unrestricted, i.e., $t(a, \phi) \in \mathbb{R}$, the optimal transfer schedule under unrestricted transfers $t^N(a, \phi)$ is a translation of the restricted schedule $t^N(a, \phi) = t(a, \phi) - k$ with*

$$k = \int_{\underline{a}}^{\bar{a}} \int_0^{\bar{\phi}} t(a, \phi) dG(a) dF(\phi). \quad (1.37)$$

*The structure of the optimal probability remains unaltered with $\psi^N = a^*h(a^*, 1) + k - \underline{a}h(a^*, 0)$.*

Proposition 15 highlights that the social planner effectively provides cross-subsidization between agents. This is, the social planner anticipates the revenue from transfers of rich agents to reduce the level of all transfers (so that incentives are preserved)

Moreover, Proposition 15 allows to separate the problem in two stages. First, one can define a “virtual” fee and solve for the optimal mechanism ignoring the break-even constraint, (1.9). Once the mechanism is obtained, there exists a one-to-one transformation to the “real” fees, provided that a sufficient condition for uniqueness in the first stage of the problem is met.

By construction, the virtual transfer is defined as

$$\psi^v = \psi - \int_{a^*(\psi)}^{\bar{a}} dG(a) \left(\int_0^{\psi} \phi dF(\phi) + 1 - F(\psi) \right). \quad (1.38)$$

The problem can be solved as follows. First find the solution of the restricted problem (i.e., $t(a, \phi) \in \mathbb{R}_+$ with virtual transfers, i.e., find the virtual transfers that maximize (1.12). By definition, these coincide with the solution to the restricted problem. Then, given ψ^v , find the optimal transfer ψ that solves (1.38). This second stage does not need to have unique solution.³⁷ This shows that the same efficient allocation can be sustained by diverse educational systems if the social planner can cross-subsidize agents. As the only relevant margin for efficient allocation is the virtual fee, which determines the marginal type attending higher education a^* , two seemingly different alternative mechanisms $\{\psi_h(\psi^v), \psi_l(\psi^v)\}$ with $\psi_h > \psi_l$ can coexist. In the mechanism with ψ_h , the discount from the virtual fee k is high, but the probability of attending higher education of borrowing constrained is low. On the contrary, for ψ_l the discount k is low, but the probability of attending higher education of borrowing constrained agents is relatively higher. In any event, is important to emphasize that because the same virtual fee is implemented, the allocation of agents to schools is the same. Thus, the

³⁷This can be seen by taking the derivative

$$\frac{\partial \psi^v}{\partial \psi} = 1 + \frac{\partial a^*}{\partial \psi} g(a^*) \left(\int_0^{\psi} \phi dF(\phi) + 1 - F(\psi) \right) + \int_{a^*(\psi)}^{\bar{a}} dG(a) f(\psi) (1 - \psi).$$

For $\psi = 0$ the derivative is unambiguously positive, while for $\psi \rightarrow \infty$ it may be negative.

main object of interest remains to be the solution the constrained problem, i.e., the virtual transfer ψ^v .

1.8.2 Decentralization: A Fair Lottery Market on Wealth

In this subsection, I discuss how to generalize the decentralization to this relaxed environment. Following the discussion in the main text, the social planner sets a school-contingent tax that has to be satisfied to attend school s such that

$$\tau(s) = \begin{cases} -k & \text{for } s = 0, \\ \psi^v - k & \text{for } s = 1. \end{cases} \quad (1.39)$$

Now, consider a market that opens after the social planner announces the school-contingent taxes and before agents attend schools, in which fair lotteries l over wealth are traded. There is a continuum of these lotteries, indexed by $i \in [0, \phi^v - k]$. A lottery l_i delivers $\psi^v - k$ with probability π_i and 0 with probability $1 - \pi_i$. There is a competitive market for each lottery. Thus the price of lottery i , p_i is given by the break even constraint (or the actuarially fair lottery), $p_i = \pi_i(\psi^v - k)$.

The rest of the discussion to show how the decentralization is achieved mimics the main text and is omitted.

1.8.3 School fees

The environment where the only credible mechanism are school fees can be solved in a similar fashion as the full mechanism design problem. First, derive the solution of the restricted problem, in which transfers being restricted to be negative $t(a, \phi) \in \mathbb{R}_-$. Then, characterize the unrestricted problem, in which $t(a, \phi) \in \mathbb{R}$.

Define the virtual fee as

$$\psi^v = \psi - (1 - F(\psi)) \int_{a^*(\psi)}^{\bar{a}} dG(a). \quad (1.40)$$

Solve the restricted planner's problem (1.16) for the virtual price. Once it is found, use the one-to-one positive relationship defined by (1.40) to determine the optimal fee. Note that this result allows us to focus on ψ^v for the comparative statics with the wealth distribution, as ψ inherits the shifts in ψ^v .

1.9 Appendix: Proofs

Proof of Proposition 1 The proof relies on the implicit function theorem. Define $I(a, \psi) = ah(a, 1) - \underline{a}h(a, 0) - \psi$. Denoting derivatives with subindexes, it can be verified that $I_a = h(a, 1) + ah_a(a, 1) - \underline{a}h_a(a, 0) > 0$ (because of the complementarity of the intrinsic human capital production function and the fact that $a \geq \underline{a}$) and $I_\psi = -1$. The implicit function theorem states that $da/d\psi = -I_\psi/I_a > 0$. \square

Proof of Proposition 2 and 15 Start considering the environment in which transfers are restricted to be negative, $t(a, \phi) \in \mathbb{R}_-$. There are two cases to distinguish. The first case is

the no-segregation, in which all agents attend higher education. In this case $t(a, \phi) = 0$ and $\phi(a, \phi) = 1$ for all $a \in [\underline{a}, \bar{a}]$ and $\phi \in [0, \bar{\phi}]$ implements the desired allocation. The second case is the segregation case, in which some agents are excluded from higher education. In what follows, let $\bar{a} = \min_a \{a \in s = 1\}$. The proof is presented in a series of lemmas.

Lemma 1 *Conditional on a given \bar{a} and wealth level ϕ , it is optimal to maximize the amount of agents with $a \geq \bar{a}$ that attend $s = 1$.*

Proof By contradiction. Suppose that there exists a mechanism $\langle \hat{t}, \hat{\pi} \rangle$ that implements the same allocation (respecting constraints (1.6), (1.7) and (1.8)) as the original mechanism $\langle t, \pi \rangle$, except for agents with wealth $\bar{\phi}$, in which

$$\begin{aligned} & \int_{\bar{a}}^{\bar{a}} \hat{\pi}(a, \bar{\phi}) \bar{a} h(a, 1) dG(a) + \int_{\underline{a}}^{\bar{a}} (1 - \hat{\pi}(a, \bar{\phi})) \underline{a} h(a, 0) dG(a) \\ & > \int_{\bar{a}}^{\bar{a}} \pi(a, \bar{\phi}) \bar{a} h(a, 1) dG(a) + \int_{\underline{a}}^{\bar{a}} (1 - \pi(a, \bar{\phi})) \underline{a} h(a, 0) dG(a). \end{aligned}$$

If this is true, then $\langle t, \pi \rangle$ does not maximize the objective function (1.11), a contradiction. \square

Lemma 2 *Consider a mechanism that achieves truth-telling in the wealth dimension. The transfer schedule as a function of the ability reported for a given wealth level ϕ is a step function with a jump at \bar{a} .*

Proof Consider first agents with $a < \bar{a}$. They are allocated in $s = 0$ with probability one. Otherwise, the spillover effect for $s = 1$ would not be \bar{a} . Thus, conditional on being allocated in school $s = 0$ the agent reports the ability that conditional on its wealth maximizes his utility, this is, maximizes the transfer $t(a, \phi)$. Thus, for $a \leq \bar{a}$, $t(a, \phi) = t(\phi)$. Moreover, given the single-crossing property of the intrinsic human capital production function, for this allocation to be incentive compatible it has to be the case that

$$w(\bar{a}, 0) + t(\phi) = \pi(\bar{a}, \phi) w(\bar{a}, 1) + (1 - \pi(\bar{a}, \phi)) w(\bar{a}, 0) + t(\bar{a}, \phi), \quad (1.41)$$

because otherwise agents with $a < \bar{a}$ would choose to report ability \bar{a} .

Next consider the case of agents with $\bar{a} \geq a$. Given $(t(\bar{a}, \phi), \pi(\bar{a}, \phi))$ and the single-crossing property, it is clear that they choose to report, at least, to be of type \bar{a} . If $t(\bar{a}, \phi) \leq t(a, \phi)$ and $\pi(\bar{a}, \phi) = \pi(a, \phi)$ for $a > \bar{a}$, agents weakly prefer to report type \bar{a} , as the payoff from attending $s = 1$ remains constant but the transfer schedule may not. From Lemma 1, it follows that, $\pi(\bar{a}, \phi) = \pi(a, \phi)$ for $a > \bar{a}$, as otherwise output can be increased. To see this, suppose that $\pi(\bar{a}, \phi) \leq \pi(a, \phi)$ and $t(\bar{a}, \phi) \leq t(a, \phi)$ for $a > \bar{a}$. Then, consider the alternative mechanism such that the highest probability and the lowest transfer are preserved, $\bar{\pi}(\bar{a}, \phi) = \pi(a, \phi)$ and $t(\bar{a}, \phi) = \bar{t}(a, \phi)$. This alternative mechanism increases the value of the objective function and implements the same allocation of agents to schools. Thus, this analysis shows that it is optimal to set $(t(a, \phi), \pi(a, \phi)) = (t(\bar{a}, \phi), \pi(\bar{a}, \phi))$ for all $a > \bar{a}$. \square

Lemma 3 *All agents with $a < \bar{a}$ attend school $s = 0$ with probability one, the associated transfer is at most zero irrespective of agents' wealth and ability.*

Proof The first claim is already shown in the proof of Lemma 2. For the second claim, consider the poorest, lowest ability agent in the economy, $(\underline{a}, \phi = 0)$. Given that he is borrowing constrained, the maximal transfers he can afford is $t = 0$. Thus, to satisfy the participation constraint (1.6) the maximal transfer for an agent of type $t(\underline{a}, 0) = 0$. Note that this can be negative. But, given that it is optimal to set $\pi(a, \phi) = 0$ for all $a < \bar{a}$ irrespective of the wealth level, agents with $a < \bar{a}$ can always report being of type $(a, \phi = 0)$ and thus, only, one level of transfers is implemented in equilibrium. \square

Lemma 4 *Conditional on a , agents' payoff is maximal for agents with $(a, \bar{\phi})$, in particular for $a \geq \bar{a}$. This is implemented by a mechanism in which these agents pay the highest transfers and receive the highest probability of attending $s = 1$.*

Proof The intuition for the result is as follows. Agents with $\phi = \bar{\phi}$ that are the "least" borrowing constrained in the economy. Thus, from the incentive compatibility constraint, (1.7), it follows that these agents can select the highest return announcement. As a result, the truth-telling mechanism that induces agents to report their true wealth has to be "expensive" enough for poorer agents not to be able to imitate them, and has to offer an attractive enough reward, for rich agents being willing to self-select, hence the high probability. The next lines make this intuitive reasoning more precise. Consider the announcement made by a type $(a, \bar{\phi})$ with $a \geq \bar{a}$, which has associated transfers and probabilities, $(t(a, \bar{\phi}), \pi(a, \bar{\phi}))$. By construction, from the incentive compatibility constraint (1.7), denoting the expected wage of reporting truthfully by $w(a, \phi)$, it is the case that $w(a, \phi) + t(a, \phi) \geq w(a, \hat{a}, \hat{\phi}) - t(\hat{a}, \hat{\phi})$. Note however that for types with $\phi < \bar{\phi}$, if $t(a, \phi) > \phi$ it is not possible to pretend that they are richer than they actually are. Thus, given the operating assumption that it is optimal to have some segregation, it has to be the case that the fee paid by agents with $(a, \bar{\phi})$ and $a \geq \bar{a}$ is weakly higher than for agents with $\phi < \bar{\phi}$ and $a \geq \bar{a}$. As a result, given the incentive compatibility constraint, wealthy agents have to be compensated to report a higher wealth by a higher expected wage, which can only be achieved by a (weakly) higher probability π of accessing school 1. \square

Lemma 5 *Given a threshold \bar{a} , it is optimal to allocate agents with $\phi \geq \hat{\phi}$ and $a \geq \bar{a}$, where $\bar{a}h(a, 1) - \underline{a}h(a, 0) = \hat{\phi}$ to school 1 with probability 1.*

Proof Note first that $\hat{\phi}$ is to be readily interpreted as $t(a, \phi) = -\hat{\phi}$ for all $\phi \geq \hat{\phi}$ and $a \geq \bar{a}$. From Lemma (1) it is immediate to check that if this allocation is implementable it is optimal provided that it does not distort the inframarginal allocation of types with $\phi < \hat{\phi}$. But, by construction, agents with $\phi < \hat{\phi}$ cannot afford signaling themselves as having wealth $\hat{\phi}$. To satisfy the incentive compatibility, it has to be ensured that the expected return of reporting (a, ϕ) with $\phi < \hat{\phi}$ and $a \geq \bar{a}$, $w(a, \phi) + t(a, \phi)$ is less or equal to $H(a, 1) + \hat{\phi}$. Note that this restriction does not impose additional constraints. The reason is simple: $t(a, \phi) = \hat{\phi}$ is the minimal transfer consistent with threshold \bar{a} . (And note again that by Lemma 1 it would not be optimal to set a probability lower than one). \square

Lemma 6 *Given a threshold \bar{a} , the optimal mechanism for agents with $\phi < \hat{\phi}$ and $a \geq \bar{a}$ is*

$$t(a, \phi) = -\phi, \quad \pi(a, \phi) = \frac{\phi}{\hat{\phi}} \quad (1.42)$$

Proof The optimal schedule comes from maximizing the mass of agents with $a \geq \bar{a}$ of a particular wealth level that attends $s = 1$. Once this is done, it remains to be checked that truth-telling is optimal.

Consider agents with wealth $\phi < \hat{\phi}$. Note that these agents are borrowing constrained, and thus cannot announce to be of type $(a, \hat{\phi})$ for $a \geq \bar{a}$. Using Lemma 2, the transfer-probability pair that maximizes attendance of agents above \bar{a} to school 1 has to satisfy the condition at the boundary \bar{a} ,

$$\pi(t(\phi), \hat{\phi})\bar{a}h(\bar{a}, 1) + (1 - \pi(\phi, \hat{\phi}))\underline{a}h(\bar{a}, 0) + t(\phi) = \underline{a}h(\bar{a}, 0) \quad (1.43)$$

First, note that it is suboptimal to set $t(\phi) > -\phi$. Suppose, to the contrary, that the optimal school fee is less than ϕ . As $\pi(t(\phi), \hat{\phi})$ is a strictly increasing function of t , this implies that by setting $t(\phi) < \phi$ the mass of agents attending school 1 can be increased by setting $t(\phi) = \phi$. Second, setting $t(\phi) = \phi$ does not affect the incentives of agents with wealth strictly lower than ϕ . For agents with wealth above ϕ , from equation (1.43) it can be verified that they are indifferent (or strictly prefer if they have $\phi > \hat{\phi}$) between the transfer-probability designed for them and this alternative. From this analysis, equation (1.42) follows. \square

The previous series of lemmas show that the optimal mechanism takes the form of a menu of prices, as stated in the main proposition. I now discuss the case of unrestricted transfers, $t(a, \phi) \in \mathbb{R}$. In this case, it is clear that it is always weakly better to set the transfers so that the borrowing constraint of the social planner breaks-even. (If there are agents borrowing constrained is strictly better). From the incentive compatibility conditions, equation (1.7), it is clear that the only relevant object as far transfers are concerned for agents when considering to deviate from truth-telling is the difference between transfers, $t(a, \phi) - t(a^\dagger, \phi^\dagger)$. Thus, letting k denote the revenue (in negative terms) from the transfers when they where constrained to be negative,

$$k = \int_{\underline{a}}^{\bar{a}} \int_0^{\hat{\phi}} t(a, \phi) dG(a) dF(\phi), \quad (1.44)$$

it is immediate to verify that a lump-sum decrease in the transfer schedule of the type $t(a, \phi) - k$, does not modify any of agents' decisions. So, the threshold \bar{a} is still implemented. Yet, now agents are effectively less borrowing constrained, and thus the mass of agents that effectively can attend higher education increases. \square

Derivation of equation 1.14 and proof of sufficiency of FOC The derivation of expression (1.14) comes from taking the derivative of the objective function with respect to ψ . Note, moreover, that for the first term in (1.14) the integration with respect to wealth is independent of ability. Using the Leibniz rule, this can be written as

$$\begin{aligned} & -f(\psi) \int_{a^*}^{\bar{a}} \Delta w dG(a) + (1 - F(\psi)) \int_{a^*}^{\bar{a}} \frac{\partial \Delta w}{\partial \psi} dG(a) - (1 - F(\psi)) \frac{\partial a^*}{\partial \psi} (\Delta w|_{a=a^*}) g(a^*) + \\ & + f(\psi) \int_{a^*}^{\bar{a}} \Delta w dG(a) + \int_0^\psi \int_{a^*}^{\bar{a}} \frac{\partial}{\partial \psi} \left(\frac{\phi}{\psi} \Delta w \right) dG(a) dF(\phi) - \int_0^\psi \frac{\partial a^*}{\partial \psi} \left(\frac{\phi}{\psi} \Delta w|_{a=a^*} \right) g(a^*) dF(\phi). \end{aligned}$$

The terms appearing in the first line come from the derivative of the first term of the objective function, and the terms on the second line, correspond to the derivative of the second term. Note that by the incentive compatibility constraints, $\frac{\phi}{\psi}\Delta w|_{a=a^*} - \phi = 0$ for all $\phi \leq \psi$. Next, take the derivative of the integrand of the second term of the second line

$$\frac{\partial}{\partial \psi} \left(\frac{\phi}{\psi} \Delta w \right) = -\frac{\pi(\phi, \psi)}{\psi} \left(\Delta w - \psi \frac{\partial \Delta w}{\partial \psi} \right).$$

This allows to rewrite the integrand that contains this derivative as

$$\int_0^\psi \int_{a^*}^{\bar{a}} \frac{\pi(\phi, \psi)}{\psi} \left(\Delta w - \psi \frac{\partial \Delta w}{\partial \psi} \right) dG(a) dF(\phi) = 0.$$

Noting that the second integral can be expressed as

$$\int_0^\psi \frac{\pi(\phi, \psi)}{\psi} dF(\phi) \int_{a^*}^{\bar{a}} \left(\Delta w - \psi \frac{\partial \Delta w}{\partial \psi} \right) dG(a),$$

the result stated in the main text, equation (1.14), follows.

For the concavity of the objective function, I provide a sufficient condition for concavity. Instead of showing that the first order condition is decreasing, I analyze the stronger condition that the first order condition multiplied by an increasing function is still decreasing. The increasing function chosen is $1/(1 - F(\phi))$. Thus, I want to show that the following function is decreasing in ψ :

$$\begin{aligned} & \int_{a^*(\psi)}^{\bar{a}} \frac{\partial \Delta w}{\partial \psi} dG(a) - \frac{\partial a^*}{\partial \psi} \Delta w(a^*) g(a^*) \\ & - \frac{\frac{1}{\psi} \int_0^\psi \pi(\phi, \psi) dF(\phi)}{1 - F(\psi)} \left(\frac{\partial a^*}{\partial \psi} \psi \Delta w(a^*) g(a^*) + \int_{a^*(\psi)}^{\bar{a}} \left(\Delta w - \psi \frac{\partial \Delta w}{\partial \psi} \right) dG(a) \right) \end{aligned} \quad (1.45)$$

In fact, as the next proof shows, this stronger condition ensures uniqueness of the solution as well. Before proceeding, there is a need to introduce an intermediate result regarding the concavity of $a^*(\psi)$. \square

Proposition 16 (*Concavity of $a^*(\psi)$*) *The marginal type obtaining education $\bar{a}(\psi)$ is concave in $\psi \in \Psi$ if and only if $2\partial h(a, 1)/\partial a + \underline{a}|\partial^2 h(a, 0)/\partial a^2| \geq a|\partial^2 h(a, 1)/\partial a^2|$ for all $a \in [\underline{a}, \bar{a}]$.*

Proof The proof relies on the implicit function theorem. Define $I(a, \psi) = ah(a, 1) - \underline{a}h(a, 0) - \psi$. Denoting derivatives with subindexes, it can be verified that $I_a = h(a, 1) + ah_a(a, 1) - \underline{a}h_a(a, 0) > 0$ (because of the complementarity of the intrinsic human capital production function and the fact that $a \geq \underline{a}$, $I_\psi = -1$, $I_{a\psi} = 0$ and $I_{aa} = 2h_a(a, 1) + ah_{aa}(a, 1) - \underline{a}h_{aa}(a, 0)$). The implicit function theorem implies the second derivative that $d^2 a/d\psi^2 = -\left(I_{\psi\psi} + 2I_{a\psi}da/d\psi + I_{aa}(da/d\psi)^2\right)/I_a < 0$, if and only if $I_{aa} > 0$. The sufficient condition mentioned in the next paragraph is obtained from ignoring the term $2h_a(a, 1)$ in the expression of I_{aa} . In this case, the condition for $I_{aa} \geq 0$ is

$$\frac{a}{\underline{a}} h_{aa}(a^*, 1) \geq h_{aa}(a^*, 0),$$

as $a \geq \underline{a}$, the sufficient condition stated in the text follows. \square

Proposition 16 identifies the conditions under which the threshold type \bar{a} is concave in school fee ψ . This condition is trivially satisfied by intrinsic production functions of the type $h(a, s) = a\hat{h}(s)$, as in this case $h_{aa}(a, s) = 0$. For more general human capital functions whether this condition is satisfied depends on the shape of the production function and the support of the ability distribution. For example if $|\partial^2 h(a, 1)/\partial a^2|/|\partial^2 h(a, 0)/\partial a^2| \leq \underline{a}/\bar{a}$. I proceed making the assumption that $a^*(\psi)$ is concave.

Assumption 3 $a^*(\psi)$ is concave for $\psi \in \Psi$. This is,

$$2\partial h(a, 1)/\partial a + \underline{a}|\partial^2 h(a, 0)/\partial a^2| \geq a|\partial^2 h(a, 1)/\partial a^2| \text{ for all } a \in [\underline{a}, \bar{a}].$$

Consider terms in the first line of (1.45), direct differentiation shows that it is decreasing in ψ ,

$$\begin{aligned} \frac{d}{d\psi} \left(\int_{a^*(\psi)}^{\bar{a}} \frac{\partial \Delta w}{\partial \psi} dG(a) - \frac{\partial a^*}{\partial \psi} \Delta w(a^*) g(a^*) \right) &= \int_{a^*(\psi)}^{\bar{a}} g(a) \left(\frac{\partial^2 \Delta w}{\partial \psi^2} \right) da - \frac{\partial a^*}{\partial \psi} \left(\frac{\partial \Delta w}{\partial \psi} \right) \Big|_{a=a^*} \\ &\quad - \frac{\partial^2 a^*}{\partial \psi^2} \Delta w(a^*) g(a^*) \\ &\quad - \frac{\partial a^*}{\partial \psi} \frac{\partial \Delta w(a^*)}{\partial \psi} g(a^*) < 0. \end{aligned} \quad (1.46)$$

To obtain the result that the derivative is decreasing, note that the integrand in the first term of right hand side of (1.46) is equal to $\frac{\partial^2 a^*}{\partial \psi^2} h(a, 1)$. Thus the difference between the first and third terms is negative because $h(a, 1)$ is weakly increasing in a , $h(a, 1) \geq ah(a, 1) - \underline{a}h(a, 0)$ (recall that $\bar{a} \leq 1$) and the ability distribution is uniform. Note, moreover, that $\frac{\partial g(a^*)}{\partial \psi} = 0$ because of the assumption that the ability distribution is uniform.

Consider the term in parenthesis in the second line of (1.45). Applying the Leibniz rule,

$$\begin{aligned} \frac{d}{d\psi} \left(\int_{a^*(\psi)}^{\bar{a}} g(a) \left(\Delta w - \psi \frac{\partial \Delta w}{\partial \psi} \right) dG(a) + \frac{\partial a^*}{\partial \psi} \psi \Delta w(a^*) g(a^*) \right) &= \\ - \int_{a^*(\psi)}^{\bar{a}} \frac{\partial^2 \Delta w}{\partial \psi^2} g(a) da - a^*(\psi) \left(\Delta w - \psi \frac{\partial \Delta w}{\partial \psi} \right) \Big|_{a=a^*} g(a^*) &\quad (1.47) \\ + \frac{\partial^2 a^*}{\partial \psi^2} \psi \Delta w(a^*) g(a^*) + \frac{\partial a^*}{\partial \psi} \Delta w(a^*) g(a^*) + \frac{\partial a^*}{\partial \psi} \psi \frac{\partial \Delta w(a^*)}{\partial \psi} g(a^*) &> 0. \end{aligned}$$

The result that the derivative is increasing follows from an analogous argument to the one used in the previous derivative, equation (1.46).

Finally, note that the reminder term, in (1.45),

$$\frac{\frac{1}{\psi} \int_0^\psi \pi(\phi, \psi) dF(\phi)}{1 - F(\psi)},$$

is increasing by assumption in the text. This analysis shows that (1.45) is decreasing, and,

hence, the objective function is globally concave. \square

Analysis of equation 1.14 Consider the right hand side of (1.14). From previous proof, equations (1.46) and (1.47) show that the numerator of the right hand side is decreasing and the denominator is increasing. Hence, the right hand side is decreasing.

For the left hand side, the main text already discusses under which conditions it is increasing, which is the working assumption in any case. As the LHS is increasing and the RHS is decreasing, they can cross almost once. Indeed, this is the case if the solution is assumed to be interior, which is the environment of interest discussed in the main text. \square

Proof of Proposition 3 The results follows immediately for the virtual fee ψ^v . The LHS of equation (1.12) being increasing, the RHS, decreasing and the property that a Wealth Abundance shift the whole LHS curve downwards. The RHS remains unaffected. As both the LHS and RHS are continuous and monotone the result follows. \square

Proof of Proposition 4 The mass of agents attending higher education is given by

$$\int_{a^*(\psi)}^{\bar{a}} dG(a) \left(\int_0^{\psi} dF(\phi) + 1 - F(\psi) \right).$$

In the limiting case in which no agent is borrowing constrained, the mass of agents attending school reduces to

$$\int_{a^*(\psi)}^{\bar{a}} dG(a),$$

as $f(\phi) < \varepsilon$ for a small $\varepsilon > 0$ in the range $\phi \in [0, \psi]$. The first order condition of the objective function, (1.4), for the case without borrowing constraints can be written as

$$\frac{d}{d\psi} \ln \left(\int_{\bar{a}(\psi)}^{\bar{a}} \Delta w dG(a) \right) = 0. \quad (1.48)$$

Let ψ^{unc} denote the solution to the problem without borrowing constraints, (1.48). When borrowing constraints start to bind, assuming

$$\frac{\frac{1}{\psi} \int_0^{\psi} \pi(\phi, \psi) dF(\phi)}{1 - F(\psi)} \leq \varepsilon$$

for some “small” $\varepsilon > 0$ in the relevant range, the first order condition becomes approximately

$$\frac{d}{d\psi} \ln \left(\int_{a^*(\psi)}^{\bar{a}} \Delta w dG(a) \right) = \varepsilon. \quad (1.49)$$

Denoting the solution to the problem by ψ^{bc} , it follows that $\psi^{bc} \simeq \psi^{unc} - \delta(\varepsilon)$ for some $\delta(\varepsilon) > 0$, where the dependence on ε is carried over to emphasize the dependence in the approximation of the solution. Now, approximating the integral

$$\int_{a^*(\psi^{bc})}^{\bar{a}} dG(a) \simeq \int_{\bar{a}(\psi^{unc})}^{\bar{a}} dG(a) - \frac{\partial \bar{a}}{\partial \psi} \delta(\varepsilon) g(\bar{a}), \quad (1.50)$$

the difference in the mass of agents attending higher education with borrowing constraints minus the mass without reduces to

$$(1 - F(\psi^{unc} - \delta)) \left(\int_{\bar{a}(\psi^{unc})}^{\bar{a}} dG(a) - \frac{\partial \bar{a}}{\partial \psi^{unc}} \delta(\varepsilon) g(\bar{a}) - \int_{\bar{a}(\psi^{unc})}^{\bar{a}} dG(a) \right) - F(\psi^{unc} - \delta) \int_{\bar{a}(\psi^{unc})}^{\bar{a}} dG(a), \quad (1.51)$$

where I have used the fact that borrowing constraints “start to bind” and thus $\int_0^{\psi^{unc} - \delta} \phi f(\phi) \ll (1 - F(\psi^{unc} - \delta))$. Expression (1.51) is unambiguously negative. This shows the result stated in the proposition. \square

Proof of Proposition 8 I first show that the relaxed condition (which implies increasing hazard rate) implies local concavity at the solution of the first order condition. Take the derivative of (1.17) with respect to ψ . Before analyzing the sign of the LHS of the FOC (1.17), it is convenient to take the derivative of the first term RHS of the FOC. Its derivative with respect to ψ is unambiguously negative,

$$-f'(\psi) \int_{a^*(\psi)}^{\bar{a}} \frac{d\Delta w}{d\psi} dG(a) - (1 - F(\psi)) g(a^*) \frac{da^*}{d\psi} \Big|_{a=a^*} + (1 - F(\psi)) \int_{a^*(\psi)}^{\bar{a}} \frac{d^2 \Delta w}{d\psi^2} dG(a) da < 0.$$

The analysis of the derivative of the second term is analogous to the previous proposition and is omitted. The derivative of the LHS in the FOC (1.17) is

$$-f'(\psi) \int_{a^*(\psi)}^{\bar{a}} \Delta w dG(a) - f(\psi) \int_{a^*(\psi)}^{\bar{a}} \frac{dw}{d\psi} dG(a) + f(\psi) \frac{da^*}{d\psi} \frac{dw}{d\psi} g(a^*). \quad (1.52)$$

The sum of the second term and third terms is negative for an analogous reason as in the previous proposition. This first term has an ambiguous sign, as $f'()$ can be either positive or negative. Take the expression of the derivative of the first term of the FOC, equation (1.52) and use that, at the optimum ψ , the FOC (1.17) is satisfied to rewrite (the ambiguous part of) equation (1.52) as

$$\left(-f'(\psi) - 2 \frac{f'(\psi)^2}{1 - F(\psi)} \right) \int_{a^*(\psi)}^{\bar{a}} \Delta w dG(a).$$

As the integrand is always positive, the sign of this expression is the sign of the term multiplying the integral, $\left(-f'(\psi) - 2 \frac{f'(\psi)^2}{1 - F(\psi)} \right)$. The sign of this first term coincides with $(1 - F(\psi))''(1 - F(\psi)) - 2f'(p)^2$, which in turn implies the relaxed condition.

To ensure uniqueness of the solution with an increasing the hazard rate, divide through the first order condition by $1 - F(\psi)$ and proceed exactly as in the previous proposition. Namely show that the without the hazard rate is decreasing in ψ and the term that is multiplying the hazard rate is increasing. Then a sufficient condition for uniqueness is an increasing hazard rate. \square

Proof of Proposition 9 The proof is analogous to the proof of Proposition 4 and it is omitted.

Proof of Proposition 10 The analysis in the main text identifies that the exams may not be used in wealth abundant countries. The problem of interest is the comparative statics

when they are used, which can be written as

$$\max_{a^*, \bar{a}} (1 - F(\psi_0)) \int_{a^*}^{\bar{a}} \Delta w dG(a) + \int_{\bar{a}}^{\bar{a}} dG(a) \int_{\psi_0 - c(\bar{a}, 1) + c(a, 1)}^{\bar{\phi}} dF(\phi) (\Delta w - c(a, 1)) \quad (1.53)$$

The problem can be thought as being solved in a “telescopic” way. That is, solve for the optimal $\bar{a}(a^*)$ and then solve for the optimal a^* . The FOC for \bar{a} is

$$\frac{c(\bar{a}, 1)}{c_a(\bar{a}, 1)} = \int_{\bar{a}}^{\bar{a}} \frac{f(\psi_0)}{1 - F(\psi_0)} (\Delta w - c(t, a)) dG(a). \quad (1.54)$$

Note that the left hand side is increasing in a because of the concavity assumption $c_{aa} \leq 0$. The right hand side is decreasing in \bar{a} , increasing in a^* and a wealth abundance shift. To see this last point, note that one can rewrite $\int_{\bar{a}}^{\bar{a}} f(\psi_0 - c(\bar{a}, 1) + c(a, 1)) / (1 - F(\psi_0))$ as $\int_{\phi_1}^{\phi_2} f(\phi) d\phi / (1 - F(\phi_2))$, and thus one can apply directly the result from the remark. Thus, \bar{a} is increasing in a^* and decreasing in a hazard rate shift. Rewrite the objective function with $\psi_0(a^*)$ and $\bar{a}(a^*)$. Then, the first order condition can be written as,

$$\begin{aligned} 0 = & -\frac{f(\psi_0)}{1 - F(\psi_0)} \psi_0'(a^*) \int_{a^*}^{\bar{a}} \Delta w dG(a) + \int_{a^*}^{\bar{a}} \Delta w dG(a) \\ & - \Delta w(a^*) - \bar{a}'(a^*) \int_{\psi_0 - c(\bar{a}, 1)}^{\bar{\phi}} (\Delta w(\bar{a}) - c(\bar{a}, 1)) \frac{dF(\phi)}{1 - F(\psi_0)} \\ & \int_{\bar{a}(a^*)}^{\bar{a}} \frac{dG(a)}{1 - F(\psi_0)} \left(-(\psi_0'(a^*) - c_a(\bar{a}, 1) \bar{a}'(a^*)) f(\psi_1 + c(a, 1)) (\Delta w - c(a, 1)) + \int_{\psi_1 + c(a, t)}^{\bar{a}} \frac{\partial \Delta w}{\partial a^*} dF(\phi) \right) \end{aligned} \quad (1.55)$$

One can use an analysis analogous to Proposition 3 to show that the first order condition is decreasing provided that the same regularity condition on $\int_{\phi_1}^{\phi_2} f(\phi) d\phi / (1 - F(\phi_2))$ of the mechanism design problem applies here to show that the first order condition is decreasing. Using the MLRP property of the wealth abundance definition as in the Remark in page 31 it follows that a^* is decreasing in a wealth abundance shift that reduces $\int_a^b f(\phi) d\phi / (1 - F(\phi))$. Thus, \bar{a} decreases with a wealth abundance shift as both the direct effect in (1.54) and the effect through a^* in (1.55) go in the same direction. \square

The omitted proofs and those corresponding to section 1.7.2 are discussed in the main text and the formal proof is omitted.

1.10 Appendix: Optimal Test-Fee Schedule Problem

This appendix presents a general solution to the optimal test-fee schedule that encompasses the results in Section 1.6 and an analogous exam problem with a continuum of schools, as in Subsection 1.7.2. (The latter is not discussed in the main text.)

Consider a payoff structure in the objective function of the type

$$\int_{a^*}^{\bar{a}} (w(a, a^*) - \kappa - c(t, a)) (1 - F(p(a) + c(t, a))) dG(a), \quad (1.56)$$

subject to three constraints. (Note the use of p instead of ψ). First, the incentive compatibility constraint

$$\zeta(a) - \dot{p}(a) - c_t(t, a)\dot{t} = 0, \quad (1.57)$$

where the operator dot stands for the total derivative with respect to a , and the subindex t , for the partial derivative with respect to t . Second, the possible levels for a^* have to belong to the family of curves of the type

$$g(a^*) \equiv w(a^*, a^*) - p(a^*) - c(a^*, t(a^*)) = 0. \quad (1.58)$$

Third, the monotonicity constraints,

$$p'(a) \leq 0, \quad t'(a) \geq 0 \quad \forall a \in [a^*, \bar{a}]. \quad (1.59)$$

The problem at hand is to

$$\max_{\{a^*, p(a), t(a)\}} \int_{a^*}^{\bar{a}} (w(a, a^*) - \kappa - c(t, a))(1 - F(p(a) + c(t, a)))dG(a) \quad (1.60)$$

subject to (1.57), (1.58) and (1.59). This problem almost fits the standard formulation of optimal control/calculus of variation. The only difference is that the initial condition a^* enters directly through $w(\cdot, a^*)$ the objective. In order to solve the problem fully, I proceed in two steps. First, conditional on a threshold \hat{a}^* on $w(\cdot, \hat{a}^*)$, I solve an inner optimization problem and find the optimal p , t (and boundary conditions) conditional on \hat{a}^* . This inner problem is formulated as an optimal control problem in subsection 1.10.1 (and subsection 1.10.2 shows the equivalence with a more intuitive formulation using calculus of variations). Then, the outer problem simply consists on a pointwise maximization of the objective with respect to \hat{a}^* subject to $\hat{a}^* = a^*$.

The following two lemmas simplify the analysis.

Lemma 7 *Any optimal solution features $t(a^*) = 0$.*

Proof By contradiction. Suppose the opposite, $(p(a^*), t(a^*))$ with $t(a^*) > 0$. Now, consider an alternative plan with $\tilde{t}(a^*)$ and $\tilde{p}(a^*) = p(a^*) + c(a^*, t(a^*))$ (note that by assumption $c(a, 0) = 0$). By construction, constraint (1.58) is satisfied. Yet, the objective function (1.56) increases under this alternative plan. A contradiction. \square

Lemma 8 *It is not optimal to set $t(a) = 0$ for $a \in (a^*, a^* + \varepsilon)$ with $\varepsilon > 0$.*

Proof I show that for every a within a radius ε exists a positive exam level that improves upon a zero test level. Suppose that the optimal solution features $t(a) = 0$. Consider the alternative policy of $t(a) = \delta > 0$. Use a first order approximation to write $p(a) \simeq p(a^*) + \dot{p}(a^*)\varepsilon$ and $c(a, t) \simeq c_t(a, 0)\eta$. From equation (1.20), the difference in output from this change in policy is proportional to

$$\varepsilon f(p(a^*))w(a^*, a^*)(-\dot{p}(a^*)\varepsilon - c_t(a^*, 0)\eta). \quad (1.61)$$

From equation (1.21), the wasteful spending is

$$\varepsilon f(p(a^*))c_t(a^*, 0)\eta. \quad (1.62)$$

Thus, selecting a η such that

$$\eta < \frac{(-\dot{p}(a^*))\varepsilon w(a^*, a^*)}{1 + w(a^*, a^*)}, \quad (1.63)$$

increases the objective function without violating any constraint. \square

1.10.1 Optimal Control Formulation of the Inner Problem

Define the state variable $x(a) = p(a) + c(a, t)$, and the control variable as $t(a)$. The incentive compatibility condition (1.57) and the boundary condition can be written as

$$\dot{x} = \zeta(a) + c_a(t, a), \quad (1.64)$$

$$0 = w(\hat{a}^*) - x(a^*) + p_0 \equiv g(\hat{a}^*, x(a^*), p_0) \quad (1.65)$$

Note that at this inner stage of the problem \hat{a}^* is taken as given, and it will be optimized over in the outer problem (subject to $a^* = \hat{a}^*$). As it is usually done in this types of problems, I proceed by ignoring the monotonicity constraints (1.59) and verifying that they hold ex-post. This allows to express the problem in a simpler manner and avoid discussing ironing and bunching procedures. Moreover, as it will become apparent, the same properties of the solution emphasized in the text arise when using monotonicity constraints in an optimal control problem.

Define the following Hamiltonian,

$$\mathcal{H} = \int_{a^*}^{\bar{a}} [(w(a, \hat{a}^*) - \kappa - c(t, a))(1 - F(x)) + \lambda_1(a) (\zeta(a) + c_a(t, a))] dG(a) + \lambda_2 g(x(a^*), \hat{a}^*, p_0).$$

The necessary conditions for an optimum are³⁸

$$\dot{x}(a) = \zeta(a) + c_a(t, a), \quad (1.66)$$

$$\dot{\lambda}(a) = f(x)(w(a, a^*) - \kappa - c(t, a)), \quad (1.67)$$

$$0 = -c_t(t, a)(1 - F(x)) + \lambda_1(a)c_{at}(t, a), \quad (1.68)$$

and the boundary conditions

$$\lambda_1(a^*) = -\lambda_2, \quad (1.69)$$

$$\lambda_1(\bar{a}) = 0. \quad (1.70)$$

Equations (1.66) and (1.67) form a system of differential equations in x and λ , intermediated through the control t , (1.68). This system is somewhat complicated by the fact that boundary conditions are given at opposite ends. In any case, the system of equations (1.66) to (1.68), the boundary conditions (1.69) and (1.70), and the constraint (1.65) characterize the solution of the problem (there are 2 differential equations with two boundary conditions to pin down λ_1 and x , and 2 additional equations to pin down t and λ_2).

I now proceed to manipulate the system of differential equations in order to investigate how the optimal solution depends on the wealth distribution. Similar to the cases analyzed in the main text, the key element is the hazard ratio of the wealth distribution. Rearranging

³⁸These can be found, for example, in Chachuat (2007) or Luenberger (1969)

(1.68),

$$\lambda_1(a) = \frac{c_t(t, a)}{c_{at}(t, a)}(1 - F(x)),$$

and taking the total derivative with respect to a ,

$$\dot{\lambda}(a) = \frac{d}{da} \left(\frac{c_t(t, a)}{c_{at}(t, a)} \right) (1 - F(x)) - \frac{c_t(t, a)}{c_{at}(t, a)} f(x) \dot{x}. \quad (1.71)$$

This expression makes clear how the assumption that $c(a, t) \equiv c_1(t)c_2(a)$ greatly simplifies the analysis: in this case c/c_a is independent of t . For example, equation (1.67) can be written now as

$$c_1(t) = \frac{1}{c_2(a)} \left(w(a) - \kappa - \frac{\dot{\lambda}(a)}{f(x)} \right). \quad (1.72)$$

Using (1.72) and (1.71) in (1.66), omitting dependence from a , and denoting c_2 by just c ,

$$\dot{x} = \zeta + \frac{\dot{c}}{c}(w - k) - \frac{\dot{c}}{cf(x)} \left[\frac{d}{da} \left(\frac{c}{\dot{c}} \right) (1 - F(x)) - \dot{x} f(x) \frac{\dot{c}}{c} \right], \quad (1.73)$$

which results into

$$\frac{1 - F(x)}{f(x)} = \frac{1}{\frac{\dot{c}(a)}{c(a)} \frac{d}{da} \left(\frac{c(a)}{\dot{c}(a)} \right)} \left(\zeta(a) + \frac{\dot{c}(a)}{c(a)} (w(a) - k) \right). \quad (1.74)$$

It can be verified by direct derivation that log-convexity of c is sufficient to guarantee that

$$\frac{\dot{c}(a)}{c(a)} \frac{d}{da} \left(\frac{c(a)}{\dot{c}(a)} \right) > 0.$$

Moreover, log-convexity implies that the terms in brackets in (1.72) is strictly increasing in a . The assumptions made in the main text guarantee that the right hand side is increasing in a and the left hand side decreasing. For the two school case, Section 1.6, $\zeta(a) = 0$, and then it follows from log-convexity that the right hand side is increasing. For the case in Section 1.7.2, a sufficient condition for the left hand side to be increasing is that $\zeta(a)$ grows at a faster rate than \dot{c}/c . Thus, x is an increasing function of a .

Consider a family of wealth distributions $f(x; s)$ that can be ranked in their hazard-rate according to an index $s \in \mathbb{R}$, so that the hazard rate is decreasing in s , (i.e., high s are relatively wealth abundant countries). It is immediate to verify that

$$\frac{\partial x}{\partial s} > 0. \quad (1.75)$$

That is, more wealth abundant societies use higher x at each level of a . An analogous argument can be done for changes in the wealth dispersion. Now, integrating equation (1.66)

$$x = \int \dot{x} da = \int_{a^*}^a \zeta(a) da + \int_{a^*}^a \dot{c}_2(a) c_1(t(a)) da, \quad (1.76)$$

it is immediate to verify that any increase in $x(a)$ has to be accompanied with a decrease in

$t(a)$, as $\dot{c}_2 < 0$.

For the terminal condition, combine equation (1.69) and (1.68) to find that

$$(1 - F(x(a^*))) = -\lambda_2 \frac{\dot{c}_2(a^*)}{c_2(a^*)}. \quad (1.77)$$

The left hand side is decreasing in a while the right hand side is increasing. This makes apparent that a more wealth abundant country chooses a higher a^* .

1.10.2 Calculus of Variations Formulation of the Inner problem

In this subsection, I briefly show how a calculus of variations approach in which explicitly the two functions over which the problem is optimized are p and t yields the same set of necessary conditions. Construct the following lagrangean

$$\mathcal{L} = \int_{a^*}^{\bar{a}} [(w(a, a^*) - \kappa - c(t, a))(1 - F(p(a) + c(t, a))) - \lambda_1(a)(\zeta(a) - \dot{p}(a) - c_t(t, a)\dot{t})] dG(a)$$

Now, the problem under consideration is

$$\max_{t(a), p(a), \lambda_1(a)} \mathcal{L} \quad \text{s.t.} \quad w(a^*) - p(a^*) = 0, \quad (1.78)$$

where I have used the result of Lemma 7 to simplify constraint (1.58). The necessary conditions for the solution are the Euler-Lagrange equations

$$\frac{d}{da} \mathcal{L}_x = \mathcal{L}_x, \quad (1.79)$$

where $x = \{t(a), p(a), \lambda_1(a)\}$. These are

$$\dot{\lambda}(a) = f(p(a) + c(a, t))(w(a, a^*) - \kappa - c(a, t)) \quad (1.80)$$

$$\dot{\lambda}(a) + \frac{c_{at}(t, a)}{c_t(t, a)} \lambda(a) = 1 - F(p(a) + c(a, t)) + f(p(a) + c(a, t))(w(a, a^*) - \kappa - c) \quad (1.81)$$

$$0 = \zeta(a) - \dot{p}(a) - c_t(t, a)\dot{t}. \quad (1.82)$$

Note that (1.80) is equivalent to (1.67). Combining (1.80) and (1.81) one obtains (1.68), while (1.82) is merely the incentive compatibility constraint as (1.66). Thus the set of equations are equivalent.

1.10.3 Formulation of the Outer Problem

Once the solution of the inner problem has been found, it remains to ensure that the optimal a^* has been selected. This can be done by point-wise optimization

$$\max_{\hat{a}^*} \int_{a^*}^{\bar{a}} [(w(a, \hat{a}^*) - \kappa - c(t(a, \hat{a}^*), a))(1 - F(x(a, \hat{a}^*)))] dG(a) \quad \text{s.t.} \quad a^* = \hat{a}^*. \quad (1.83)$$

The dependence of x and t with respect to a^* is through w (see equation (1.74), for instance). The first order condition is

$$\int_{a^*}^{\bar{a}} \frac{\partial(w(a, \hat{a}^*) - c(t(a, \hat{a}^*), a))}{\partial \hat{a}^*} (1 - F(x)) dG(a) + \lambda = \quad (1.84)$$

$$\int_{a^*}^{\bar{a}} (w(a, a^*) - \kappa - c(t, a)) f(x) \frac{\partial x(a, a^*)}{\partial a^*} dG(a),$$

where λ is the Lagrange multiplier on the constraint. Using the specifics of the model of interest (note that the problem with the continuum of schools only has the inner problem to be solved), $w(a, a^*) = a^*h(a, 1) - \underline{h}(a, 0)$, it follows that

$$\frac{\partial^2 w(a, a^*)}{\partial a^{*2}} = 0.$$

If the hazard rate of the wealth distribution is concave, the first term in the first line of equation (1.84) is decreasing in a^* . To ensure that the first order yields a maximum, it is necessary to impose more structure on the wealth distribution. A sufficient condition is that the wealth distribution is log-concave, see Bagnoli and Bergstrom (2005). In this case, the the second line of equation (1.84) is increasing in \hat{a}^* , as $1 - F(x)$ and the hazard rates are ensured to be concave. (Note that x depends on the inverse hazard rate, from equation 1.74).

Chapter 2

Heterogeneous Trade Costs and Wage Inequality: A Model of Two Globalizations*

with Sergi Basco

Abstract

We develop a model for analyzing the distributional effects of two phases of globalization and their interdependencies. We distinguish between (i) a First Globalization, characterized by trade liberalizations during the 1980s, which mainly increased trade in low skill-intensive goods and (ii) a Second Globalization, characterized by a reduction in communication costs due to the IT revolution, which raised trade in more skill-intensive goods during the 1990s. We consider a North-South trade economy in which the North is skill abundant. A freely traded final good is produced in the North using high-skill services and a bundle of inputs. Inputs differ on the intensity of middle and low-skill workers required to be produced, and are subject to heterogeneous trade costs. In the North, we find that wage inequality increases in the First Globalization. During the Second Globalization, the relative wage of high- to middle-skill workers increases, while the relative wage of middle- to low-skill is hump-shaped. In the South, we find that wage inequality increases in both. We find a complementarity between the two globalizations. The decline in the relative wage of northern middle- to low-skill workers is delayed by the extent of trade in the First Globalization. Finally, we show how asymmetric participation in the Second Globalization of two southern countries generates a discontinuous pattern of specialization. The southern country participating in the Second Globalization specializes in the least and most skill-intensive traded inputs and wage inequality rises in this country.

*This paper is joint work with Sergi Basco. A previous version of this paper circulated under the title "A Model of Two Globalizations: implications for wage inequality and the pattern of specialization." We thank Daron Acemoglu, Pol Antràs and Arnaud Costinot for their guidance. We also thank James Anderson, David Autor, Abhijit Banerjee, Ricardo Caballero, Gene Grossman, Pablo Kurlat, Kiminori Matsuyama, Sahar Parsa, Mar Reguant, Robert Townsend and seminar participants at MIT and Banco de España for useful comments and suggestions. David Autor and David Dorn generously shared their code and data with us.

2.1 Introduction

The skill content of North-South trade has changed over the last decades. Figure 2-1 documents the evolution of the share of southern exports in industries with skill intensity above the average skill intensity of U.S. industries.¹ During the 1980s, southern exports increased relatively more in industries with skill intensity below the U.S. average. In the 1990s, this pattern reversed and southern exports grew disproportionately more in industries above the U.S. average skill intensity.

These changes in southern exports can be exemplified by the bilateral trade of the United States with Chile and India. Figure 2-2 shows that in the 1980s, Chilean exports increased in below-average skill-intensive industries. During the 1990s, Indian exports rose in above-average skill-intensive industries. We interpret these differential increases in North-South trade as reductions in different trade costs. Chile underwent a dramatic trade liberalization in the late 1970s and 1980s,² while India has benefited from offshoring of information technology (IT) industries and services in the 1990s.³

Our empirical analysis suggests that the trade patterns described for India and Chile hold more broadly. Trade liberalizations in the 1980s increased northern imports in low-skill-intensive industries, whereas the IT revolution increased northern imports in middle-skill-intensive industries. We label the expansion of trade in low-skill goods in the 1980s as the *First Globalization* and the increase in trade in more skill-intensive goods during the 1990s, the *Second Globalization*. This suggestive evidence will guide our comparative statics exercises.

This paper analyzes these two phases of the globalization process and their complementarity. We investigate the effects of the Second Globalization on wage inequality, the pattern of specialization and how these effects change with the extent of First Globalization trade.

Our first main result highlights the complementarity between the two globalizations in the North. We find that the relative wage of high- to middle-skill workers increases throughout the Second Globalization. On the contrary, the relative wage of middle- to low-skill workers is hump-shaped. The decline in the relative wage of middle- to low-skill workers together with the increase in the relative wage of high-skill workers has been termed “wage

¹A southern country is defined as having less than half of 2000 U.S. GDP per capita PPP adjusted. Skill intensity is constructed from U.S. census data, based on educational attainment of workers in different jobs. The average skill intensity of U.S. imports is the average of the skill intensity of each industry at 3-NAICS level weighted by the value of the U.S. imports in this industry. We only have consistent world trade data flows from Feenstra database for the period 1984-2000.

²This pattern is not specific for Chile. Goldberg and Pavcnik (2007) document that trade liberalizations during the 1970s and 1980s in several emerging countries (e.g. Mexico, Colombia and Morocco) were biased towards low-skill-intensive industries.

³India is one of the countries which has benefited the most from this new wave of offshoring. Trefler (2006) documents that India hosted the highest number of new IT services projects (around 19% of the world total) and call centers (around 12% of the world total) in 2003 and 2004.

polarization.”⁴ Thus, the equilibrium wage distribution tends to wage polarization as the Second Globalization progresses. Our complementarity result states that wage polarization is delayed by the extent of First Globalization trade.

Our second main result starts from the observation that some southern countries lack the minimum stock of specific capital needed to benefit from the IT revolution. This lack may lead to an asymmetric participation in Second Globalization trade, as Figure 2-3 suggests for India and Pakistan. To account for this asymmetric participation, we extend the model to two southern countries, with only one participating in the Second Globalization. We show that this asymmetric participation generates a discontinuous pattern of specialization. The country participating in the Second Globalization specializes in the least skill-intensive First Globalization goods, in addition to Second Globalization goods. The other southern country specializes in the relatively higher skill-intensive goods of the First Globalization. Wage inequality increases in the former and decreases in the latter.

Section 2.2 provides suggestive evidence consistent with a First Globalization characterized by trade liberalizations affecting low-skill industries and a Second one driven by a fall in communication costs, mainly affecting middle-skill industries. We show that U.S. tariff reductions were only skill-biased during the First Globalization. In particular, they were biased toward low-skill-intensive industries. For the Second Globalization, we use a Routine Task Intensity (RTI) index as a proxy for offshorability and show that high levels of RTI (and, thus, offshoring) are associated with middle-skill industries.⁵ Next, we look at world exports to the United States and show that countries with lower communication costs export relatively more in skill-intensive industries. We document a negative correlation between changes in U.S. trade openness and changes in the U.S. wage bill in low-skill industries during the First Globalization and middle-skill industries in the Second. Finally, we show that the skill content of southern exports increased in industries with skill requirement above and below U.S. average in both globalizations.

Our model features a North-South trade economy. A freely traded final good is produced in the North by combining a bundle of inputs and high-skill labor.⁶ This bundle is assembled using a continuum of inputs, which are produced by combining middle- and low-skill labor in different proportions. Thus, this model can be thought of as an offshoring decision by

⁴See Autor and Dorn (2009) and the references therein.

⁵This measure is closely related to impersonal services, which Blinder (2006) emphasizes as a distinctive element of Second Globalization trade. The use of the RTI is motivated by the observation that goods that can be electronically delivered (or monitored) are fairly standardized and follow determined procedures. The Index is taken from Autor and Dorn (2009), who also link the IT revolution with the loss of middle-skill jobs in northern industries. In this paper, we focus on the role of offshoring of jobs to the South rather than the substitution of jobs by computers emphasized in Autor and Dorn.

⁶Our results do not rely on the assumption that the final good is only produced in the North. In a working paper version available on request we allow for final good production in both countries and obtain similar results.

northern firms. We assume that the North is skill-abundant and that there are heterogeneous trade costs across different inputs.

Throughout the paper and for ease of exposition, we label the inputs traded in the First Globalization as *intermediates*, and those traded in the Second as *tasks*.⁷ We frame the First Globalization as an increase in the set of traded low-skill *intermediates* and the Second Globalization as an increase in the set of traded middle-skill *task* traded. Consistent with the evidence on the evolution of the skill content of southern exports presented in Section 2 (see Figures 2-4 and 2-5), we assume that the skill requirement of the marginal traded input increases in both globalizations.⁸

Section 2.4 presents the main results of the model. In the First Globalization, the set of low-skill-intensive inputs imported from the South increases. As in Wood (1995) and Feenstra and Hanson (1996), since the intermediates imported by the North are below its mean skill intensity, the relative demand for middle-skill workers increases, thereby raising their relative wage. Moreover, the wage of high-skill workers relative to both middle- and low-skill agents increases because more intermediates can be bought at cheaper prices and demand for northern intermediates decreases. In the Second Globalization the set of traded tasks increases. The relative wage of high-skill workers increases for the same reasons as in the First. The relative wage of middle- to low-skill workers exhibits a humped pattern. The reason is that the relative demand of northern middle-skill workers declines when the marginal task being offshored to the South is above the skill intensity of the mean input produced in the North after the First Globalization. Thus, the equilibrium tends to wage polarization in the North.

We find a complementarity between trade in the First and the Second Globalizations. Wage polarization is delayed by the extent of trade in the First Globalization. A larger set of intermediates traded during the First Globalization implies a higher skill intensity of the mean input produced in the North. Thus, more trade in the First Globalization allows a larger set of tasks to be offshored during the Second Globalization before the relative wage of northern middle-skill workers starts to decline. This complementarity highlights the importance of having a unified view of the First and the Second Globalizations, which is one of the novelties of our framework. We provide empirical evidence consistent with this prediction. We find a positive relationship between trade openness before the onset of the IT revolution, which we assume to be in 1990, and changes in lower-tail northern wage inequality in the

⁷Note that these definitions do not change with the endogenous equilibrium objects. In particular, we assume that there exists a continuum of inputs indexed with $z \in [0, 1]$. We label *intermediates* the inputs with skill intensity $z < \bar{z}$ and *tasks* inputs with skill intensity $z > \bar{z}$, where \bar{z} is exogenous.

⁸An increase in the skill content of southern exports below and above average is consistent with the U-shape of the share of southern exports in industries with skill intensity above average documented in Figure 2-1. These findings are robust to use the measure of skill-intensity based on educational attainment of workers from U.S. census. These figures are available upon request.

1990s.

In the South, relative wages increase both in the First and the Second Globalizations. The intuition is analogous to Wood (1995) and Feenstra and Hanson (1996): the marginal input being offshored is relatively more skill-intensive, which raises the relative demand of middle-skill labor and the relative wage throughout the globalization process.

Subsection 2.4.2 introduces a second southern country to study how asymmetric participation in the Second Globalization affects the pattern of specialization and wage inequality in the South. We want to capture the notion that participation in the Second Globalization is constrained by the stock of specific capital required to benefit from the IT revolution (e.g. knowledge, institutions and infrastructure). To embody this idea in our model, we assume that only one southern country participates in the Second Globalization. In equilibrium, this country exports tasks to the North and the relative wage of low-skill workers in this country decreases. Thus, this country gains comparative advantage in the least skill-intensive intermediates, which generates a discontinuous pattern of specialization. The most and least skill-intensive traded inputs are produced by this southern country. As the Second Globalization progresses, the equilibrium tends to complete specialization. One South produces tasks and the other, intermediates. The distributional implications are that the relative wage of middle-skill workers increases in the South participating in the Second Globalization, while it declines in the other. We provide evidence for the changes in the pattern of specialization amongst southern countries consistent with the predictions of our model. We show that southern countries with a high stock of IT technologies tend to increase exports in industries with levels of RTI above average, Second Globalization goods, and decrease exports in industries with skill intensity below-average, First Globalization goods.

Section 2.5 presents two extensions. Subsection 2.5.1 allows for endogenous supply decisions. It shows that the comparative statics for wages derived in the baseline model hold in this extended version. Moreover, we find that the mass of northern agents selecting into middle-skill jobs increases during the First Globalization and eventually shrinks during the Second. Finally, we have emphasized the role of the IT revolution in allowing firms to participate in the Second Globalization. However, the Second Globalization is also an outcome of the adoption of new technologies that replace middle and low-skill jobs (Autor et al., 1998). Subsection 2.5.2 shows that the adoption of a new technology needed to benefit from Second Globalization trade is delayed by the extent of trade in the First Globalization.

The rest of the paper proceeds as follows. In Section 2.2, we present the motivating evidence and discuss the related literature. Section 2.3 presents the baseline model and Section 2.4 derives the main results of the paper. The two extensions of the model are presented in Section 2.5. Section 2.6 concludes. Proofs are in the appendix.

2.2 Motivating Evidence and Related Literature

2.2.1 Motivating Evidence

The premise of our analysis is that trade costs have changed differentially across sectors of different skill intensity over time. The overall evidence in this section paints a picture of two different phases of globalization: a First Globalization, driven by a decline in tariffs which increased trade in low-skill-intensive industries and affected the relative demand of low-skill workers; and a Second Globalization characterized by a fall in communication costs, which raised trade in intermediate skill-intensive industries and affected the relative demand of middle-skill workers. Finally, we show that the skill content of southern exports increased in both globalizations.

We present our findings in three steps. First, we document heterogeneous changes in trade costs across skill and over time. Second, we relate these changes to trade flows and the relative demand of skill. Third, we report the evolution of the skill content of southern exports in industries with skill requirement below and above U.S. average.

The first piece of evidence on heterogeneous changes in trade costs comes from analyzing changes in U.S. tariffs and transportation costs over time. Our data is disaggregated at 3-digit NAICS level and we use education attainment from U.S. census as proxy for skill intensity. First, we find that tariff reductions between 1978 and 1988 were concentrated in low-skill industries (Figure 2-8). This result is similar to Haskel and Slaughter (2003), who use non-production workers to proxy for skill. Second, we show that changes in U.S. tariffs were not significantly different from zero at any level of skill intensity during the 1990s (Figure 2-9).⁹ Third, we show that changes in transportation costs were not statistically different from zero at any level of skill, neither in the 1980s, nor in the 1990s (Figure 2-10). This evidence is consistent with Hummels (2007) findings for the ad-valorem shipping cost not having changed much since the 1950s. There is a growing literature emphasizing that there are more dimensions in trade costs than tariffs and transportation costs, e.g. Hummels (2007). However, data along other dimensions of trade costs are difficult to obtain, specially for non-tariff barriers (Anderson and van Wincoop, 2004).

Our second piece of evidence on heterogeneous changes in trade costs aims at capturing some of the effects of the IT revolution. During the 1990s, new services such as telephone operators or data entry keyers started to be offshored (Trefler, 2006). The standard measures of trade costs are less relevant for this new trade pattern. A common characteristic of these

⁹Goldberg and Pavcnik (2007) document that trade liberalizations in the 1970s and 1980s were biased towards low skill-intensive industries. Amongst others, they cite Hanson and Harrison (1999) and Robertson (2000, 2004) documenting this pattern for Mexico, Currie and Harrison (1997) for Morocco and Attanasio et al. (2004) for Colombia. This is also consistent with the evidence presented for Chile in figure 2-2. Therefore, the overall evidence points towards a differential effect of trade liberalizations on low-skill-intensive industries during the 1980s, but not in the 1990s.

services is that they are standardized and follow tight and determined procedures. In order to capture this feature, we use the Routine Task Intensity (RTI) index from Autor and Dorn (2009) as a proxy for “offshorability.” The argument is that a reduction in communication costs (IT revolution) makes jobs with high RTI index more likely to be offshored.¹⁰ Grossman and Rossi-Hansberg (2006) and Ebenstein et al. (2009) also relate routine indices with offshoring. Figure 2-6 reports anecdotal evidence pointing that higher RTI jobs are performed by middle-skill workers. It suggests that there exists an inverse U-shape relationship between skill intensity and RTI, which we find when aggregating at the industry level (Figure 2-7).

Next, we investigate how trade flows relate to changes in the trade costs described above. More specifically, we analyze the relationship between U.S. imports and tariffs and communication costs for different levels of industry skill-intensity. We use Internet adoption at country level as a proxy for communication costs. The reason is that goods that become offshorable can be electronically delivered (e.g., data entry keyers) or require intense usage of IT (e.g., call centers). We run the following regression for 1990 and 2000,

$$X_{ic} = \alpha + \beta \tau_i + \gamma \text{Internet}_c \cdot \text{Skill Intensity}_j + \delta_j + \delta_c + \varepsilon_{ic}, \quad (2.1)$$

where X_{ic} are exports of product i from country c to the United States, τ_i is U.S. tariff on product i , Internet_c is the fraction of the population with access to Internet in country c , Skill Intensity_j is the average skill of industry j and δ_j and δ_c represent industry and country fixed effects, respectively.¹¹

Columns 1 and 2 of panel A in Table 2.1 show a negative, significant correlation between tariffs and U.S. imports in 1990. This correlation is larger when the sample is restricted to southern countries (columns 6 and 7). Columns 1, 2, 6 and 7 of panel B report not significant coefficients on tariffs for year 2000. In contrast, the coefficient on the interaction between Internet and Skill Intensity is positive and significant. The coefficient is larger when restricting the sample to the South, pointing at a differential effect of the IT revolution on poor countries.¹² Note that our results abstract from trade in services, because these data

¹⁰Each job is assigned a routine and a manual score, and this index is the log ratio of the two. Therefore, tasks with high RTI imply a high routine and a low manual score. The RTI index assigns a value of “routine intensity” to a representative set of 332 occupations in the U.S. census. See Autor and Dorn (2009) for further details. Note that the findings reported in Blinder and Krueger (2009) are not inconsistent with the use of the RTI index as proxy for offshorability. The reason is that Blinder and Krueger only consider a measure of routine that abstracts from the manual content of a job.

¹¹U.S. imports are from Feenstra’s data base and U.S. tariffs come from Romalis. Romalis’ tariff data starts in 1989, thus we can only test for the last years of our First Globalization. Our measure of industry is a 3-digit NAICS and of product is a 6-digit HS. There are no data for Internet adoption in 1990 (presumably it was negligible for most of the sample).

¹²In this sample the highest skill intensity level is 11.4, which roughly coincides with skill level associated with the largest RTI. As robustness checks we added additional controls. One could think that our Internet adoption measure could be a proxy for other country variables such as country wealth, human capital levels

are not available. We think this lack of data on services underestimates our results because offshored services are RTI intensive and, as argued by Markusen (2006) and Markusen and Strand (2008), require above-average skills to be produced.

We investigate how the relative demand for skill relates to trade flows over time. We report how changes in the wage bill paid by different U.S. industries are correlated with changes in U.S. trade. Column 1 in Table 2.2 reports the results of regressing the change in U.S. wage bill during the 1980s on the interaction of average industry skill with the change in trade openness.¹³ The coefficient is positive and significant. Column 2 shows that the coefficient on the interaction term is not significant for the 1990s. Yet, when we add a quadratic term, the coefficients become significant, as shown in column 3. These regression coefficients imply a U-shape pattern for the response of wage bill to increases in trade. These results suggest that the relationship between trade and demand for skill has changed over time.

Finally, Figure 2-4 and Figure 2-5 report the skill content of southern exports in industries with skill requirement below and above U.S. average, respectively. These figures point to an increase in the skill content of southern exports in both the First and the Second Globalization. This suggestive evidence will motivate our comparative statics exercises in Section 2.4.¹⁴

2.2.2 Related Literature

The evidence presented above points towards two phases of globalization characterized by changes in trade costs affecting industries of different skill-intensity. Motivated by this evidence, our paper focuses on the effects of heterogeneous changes in trade costs on wage inequality and the pattern of specialization. To the best of our knowledge, our work is the first attempt to provide a unified view of the globalization process and its effect on wage inequality, both across North-South trade and between different southern countries.

This paper relates to a rich and diverse literature on international trade, wage inequality and the patterns of specialization. Our First Globalization comparative statics results are

and financial development and therefore their interaction with skill intensity could be relevant. Our coefficients of interest remained significant and with similar values to the baseline regression when adding these additional covariates. A second robustness check is to control for the potential endogeneity of our measure of skill intensity. In order to address this concern, we instrument skill intensity with the RTI index. Our exclusion restriction is that RTI only affects exports through the level of skill intensity of an industry. Columns 5 and 10 report the previous regressions using this instrumental variable approach. The coefficients of interest remain significant and with the expected sign.

¹³We use the U.S. wage bill data for 1980-1996 constructed in Autor et al. (1998). Our trade openness measure is the share of exports plus imports over GDP from the Penn World Tables.

¹⁴To perform this exercise we hold the skill requirement of an industry fixed at the level of the United States. Then we compute the weighted average of the skill embodied in southern exports. Both figures use nonproduction workers data from BLS. We find similar qualitative results when using our skill-intensity index constructed using U.S. census data from IPUMS as skill requirement measure (see appendix for more details). These findings are robust to consider only U.S. imports.

related to standard Heckscher-Ohlin models and the work of Wood (1995) and Feenstra and Hanson (1996). Feenstra and Hanson provide a rationale for increasing inequality in both the North and the South. They analyze the effect of capital inflows to the South in the context of a free trade equilibrium. These capital inflows reduce the unit cost of production in the South, allowing the South to produce more (relatively) skill-intensive goods at the margin. The underlying mechanism of our comparative statics for the First Globalization is similar. However, their comparative statics exercise is different from ours, as we focus on changes in trade costs. Another important difference is that our framework, by assuming heterogeneous trade costs, allows us to study the two waves of globalization and their interdependence.

Our analysis of the Second Globalization bears upon the literature on offshoring, outsourcing and wage inequality. It includes, among others, Antràs et al. (2006a,b), Dinopoulos et al. (2009), Grossman and Rossi-Hansberg (2008), Markusen and Strand (2008) and Zhu and Trefler (2005). Our paper shares the emphasis on middle-skill agents as in Antràs et al. (2006b). They focus on team problem solving. In contrast, we consider a segmented production process with firms supplying inputs, which enables us to distinguish the effects of different changes in trade costs on wage inequality. Grossman and Rossi-Hansberg (2008) consider the effect of heterogeneous transportation costs. However, they assume that tasks can be so perfectly partitioned that a fall in trade costs only affects one type of labor.

Anderson (2009), Costinot and Vogel (2009), Grossman and Maggi (2000) and Ohnsorge and Trefler (2007) among others study the role of sorting for wage inequality and the pattern of specialization. They emphasize the difference between North-South and North-North trade, from which we abstract. However, they ignore the differential effect of heterogeneous changes in trade costs across sectors of different skill-intensity.¹⁵

2.3 Model

In this section we present a simple model to study the effects of the two phases of globalization. A freely traded final good is produced in the North using high-skill labor and a bundle of inputs, which are produced by middle and low-skill workers.¹⁶ Inputs are subject to heterogeneous trade costs, which enable us to frame our two globalizations in a tractable manner. Our baseline model abstracts from endogenous labor supply decisions. It consists

¹⁵Our paper is also related to a broader literature on technology and wage inequality, for example, Acemoglu (2003), Blum (2008) and Yeaple (2005). We briefly discuss the incentives to adopt technologies within our two globalizations framework. Finally, labor economists have documented large changes in U.S. wage inequality, which our findings relate to. This vast literature includes Katz and Murphy (1992), Acemoglu (1999), Autor et al. (2003), Autor et al. (2008) and Autor and Dorn (2009) among others.

¹⁶The assumption that the final good is traded at no cost is not crucial for our results. Our comparative statics results rely on heterogeneous changes in trade costs of middle and low-skill industries. Changes in trade cost of the final good do not affect relative demand of middle and low-skill workers because of the homotheticity of the final good production function.

of three exogenous types in the North and two in the South. Subsection 2.5.1 shows that the results derived for the baseline case hold when there is a continuum of types and each type endogenously selects into one occupation.¹⁷ This section presents and characterizes the baseline model and Section 2.4 derives the main results of the paper.

2.3.1 Baseline Model

We consider a competitive world economy consisting of two countries, the North, N , and South, S . Each country is populated by a mass one of agents, which cannot migrate. Each agent is endowed with one unit of labor that inelastically supplies to the market. Northern agents can be divided between low, middle and high-skill types. The fraction of each type is $1 - \theta^N$, $\theta^N(1 - \varphi)$ and $\theta^N\varphi$, respectively. Our interpretation is that a fraction θ^N has basic education and a fraction φ obtains further education. We assume that $\varphi = 0$ in the South. Thus, southern population can be divided between a fraction $1 - \theta^S$ of low and a fraction θ^S of middle-skill agents. Finally, we assume that the South is relatively abundant in low-skill labor, i.e., $(1 - \theta^S)/\theta^S > (1 - \theta^N)/\theta^N(1 - \varphi)$.

All agents have the same utility function, $u(c)$, where c is final good consumption. The final good is produced by combining a bundle of inputs B and high-skill services h ,

$$Y = h^\alpha B^{1-\alpha}. \quad (2.2)$$

This h can be thought of as headquarter services, which are provided by high skill agents. The bundle is made by assembling a continuum of inputs, $I(z)$, with $z \in [0, 1]$,

$$B = \int_0^1 \ln I(z) dz. \quad (2.3)$$

Each input is produced using a Cobb-Douglas production function

$$I(z) = A \left(\frac{m(z)}{z} \right)^z \left(\frac{l(z)}{1-z} \right)^{1-z} \quad \text{for } z \in [0, 1], \quad (2.4)$$

where A denotes a productivity factor and $m(z)$ and $l(z)$ denote middle and low-skill workers employed in the production of input z , respectively.¹⁸ Note that z parametrizes the skill-intensity required to produce each input. The higher is z , the more intensive the input in middle-skill labor is.

¹⁷In a working paper version available upon request, we analyze a similar model with three types in both North and South and derive qualitatively analogous results. In this version, both North and South produce final good.

¹⁸We assume that $A > \frac{e}{1-\alpha} (\theta^S(1 - \theta^N))^{-1}$, where e is the Neper number, to ensure that the bundle is positive. This condition is implied by (2.4) and (2.5).

The final good is assumed to be freely traded and we normalize its price to one throughout the paper. Inputs are subject to heterogeneous iceberg costs. For one unit of input z to arrive at home, $\tau(z)$ units must be purchased abroad.

Definition A *competitive equilibrium* is a set of prices $p^i(z)$ for each input z and country $i \in \{N, S\}$, a price for the final good $p_f (\equiv 1)$, a wage for low-skill workers w_l^i , a wage for middle-skill workers w_m^i , a wage for northern high-skill workers w_h^N , an allocation of low-skill $l^i(z)$ and middle-skill $m^i(z)$ labor across inputs producers and a consumption choice c^i for each agent in country i such that agents maximize their utility given their income, firms maximize profits and all markets clear.

2.3.2 Trade equilibrium

We characterize the competitive equilibrium for a given trade cost function $\tau(z)$. Consider the problem of the final good producer,

$$\max_{\{h, l^i(z)\}} h^\alpha \left(\int_0^1 \ln l^i(z) dz \right)^{1-\alpha} - w_h h - \int_0^1 p^i(z) l^i(z) dz. \quad (2.5)$$

Demands for high-skill services and each input z are

$$\alpha Y = w_h h, \quad (2.6)$$

$$(1 - \alpha) Y = \int_0^1 p^i(z) l^i(z) dz. \quad (2.7)$$

Consider the problem of an input producer in country i ,

$$\max_{\{m^i(z), l^i(z)\}} p^i(z) A \left(\frac{m^i(z)}{z} \right)^z \left(\frac{l^i(z)}{1-z} \right)^{1-z} - w_m^i m^i(z) - w_l^i l^i(z).$$

The labor demands of a producer of input z in country i are given by

$$z p^i(z) l^i(z) = w_m^i m^i(z), \quad (2.8)$$

$$(1 - z) p^i(z) l^i(z) = w_l^i l^i(z). \quad (2.9)$$

Using labor market clearing, we can integrate labor demands across all input producers

in each country to obtain the following implicit expressions for wages

$$w_h = \alpha \frac{Y}{\theta^N \varphi}, \quad (2.10)$$

$$\int_0^1 m^N(z) dz = (1 - \alpha) \frac{Y}{B} \int_0^1 \left(\mathbb{1}_d^N(z) + \frac{\mathbb{1}_x^N(z)}{\tau(z)} \right) \frac{z}{w_m^N} dz = \theta^N (1 - \varphi), \quad (2.11)$$

$$\int_0^1 l^N(z) dz = (1 - \alpha) \frac{Y}{B} \int_0^1 \left(\mathbb{1}_d^N(z) + \frac{\mathbb{1}_x^N(z)}{\tau(z)} \right) \frac{(1 - z)}{w_i^N} dz = 1 - \theta^N, \quad (2.12)$$

$$\int_0^1 m^S(z) dz = (1 - \alpha) \frac{Y}{B} \int_0^1 \frac{\mathbb{1}_x^S(z)}{\tau(z)} \frac{z}{w_m^S} dz = \theta^S, \quad (2.13)$$

$$\int_0^1 l^S(z) dz = (1 - \alpha) \frac{Y}{B} \int_0^1 \frac{\mathbb{1}_x^S(z)}{\tau(z)} \frac{(1 - z)}{w_i^S} dz = 1 - \theta^S, \quad (2.14)$$

where $\mathbb{1}_d^i(z)$ and $\mathbb{1}_x^i(z)$ are indicator functions for each input z being produced in country i for domestic consumption and for exporting, respectively. The rest of the equilibrium outcomes can be fully characterized as follows. A price $p^i(z)$ for each input follows from the demand function (2.7). The allocation of middle-skill labor $m^i(z)$ is determined by (2.8) and the allocation of low skill labor $l^i(z)$ is given by (2.9). Using that optimality in consumers' behavior requires that all their income should be consumed, we can derive the optimal level of consumption. Finally, given that the final good is freely traded, trade balance results from the value of exports of inputs and final good being equal to the value of imports in a country.

2.4 Main Results

In this section we present the two main results of the paper. Subsection 2.4.1 derives the distributional consequences of the First and the Second Globalizations and derives the complementarity between the two. Subsection 2.4.2 extends the baseline model by dividing the original South in two different southern countries which open differently to trade in the Second Globalization. Our second main result states how the patterns of specialization and wage inequality depend on the differential participation in the Second Globalization.

2.4.1 The two Globalizations and their Complementarity

This section performs comparative statics for the two globalizations on relative wages. Then, we present our first main result (complementarity), which shows that trade in First Globalization delays the emergence of wage polarization in the North during the Second Globalization.

2.4.1.1 Comparative Statics for the First Globalization

Section 2.2 characterized the First Globalization as a decrease in the trade costs of the least skill-intensive inputs. To study its effects in a parsimonious way, we assume that trade is only possible in inputs with an index lower than z_l .¹⁹ In other words, trade costs are

$$\tau(z) = \begin{cases} 1 & \text{for } z \leq z_l, \\ \infty & \text{otherwise.} \end{cases} \quad (2.15)$$

We define First Globalization as an increase in the set of traded intermediates. This implies that the skill requirement of the marginal traded intermediate increases with globalization. Figure 2-4 shows that the skill content of southern imports has increased in the 1980s in industries with skill requirement below U.S. average.²⁰ Therefore, the comparative statics exercise we are interested in is an increase in z_l .

Assumption 4 $z_l < z^*(\theta_N, \theta_S) < 1$, where $z^*(\theta_N, \theta_S)$ is implicitly defined as

$$\left(\frac{1 - z^{*2} \theta_S}{z^{*2} \theta_N} \right)^{z^*} \left(\frac{(1 - z^*)^2}{1 - (1 - z^*)^2} \frac{1 - \theta_S}{1 - \theta_N} \right)^{1 - z^*} = 1. \quad (2.16)$$

The threshold $z^*(\theta_N, \theta_S)$ is an implicit function of the relative skill abundance of both countries. This assumption implies that all traded inputs are produced in the South. The North exports the freely traded final good to ensure balanced trade.

Proposition 17 (First Globalization) *The First Globalization equilibrium features an increase of the relative wage of middle-skill workers in the North and the South. The relative wage of northern high-skill workers increases.*

The proof follows from using equations (2.10) to (2.14) and the trade cost structure (2.15). The relative wages of middle- to low-skill workers are

$$\frac{w_m^N}{w_l^N} = \frac{1 - \theta^N}{\theta^N} \frac{1 + z_l}{1 - z_l}, \quad \frac{w_m^S}{w_l^S} = \frac{1 - \theta^S}{\theta^S} \frac{z_l^2}{1 - (1 - z_l)^2}.$$

By inspection, the relative wages are increasing in z_l . The relative wage of high-skill is increasing because B increases with z_l and demand of northern inputs decreases.

¹⁹The threshold z_l can be endogenized as an equilibrium outcome in a model with constant iceberg costs $\tau(z) = \tau$ on intermediates. The reason is that the South has comparative advantage in low-skill-intensive inputs. In this case, our First Globalization comparative statics exercise (i.e., an increase in z_l) could be endogenously obtained as a decrease in τ .

²⁰Our qualitative results would hold if we allowed for $\tau(z) = 1$ for $z \in [z, z_l]$, with $z > 0$. The key assumption is that an increase in the set of traded inputs in the First Globalization translates into an increase in the relative demand of middle-skill labor in the South.

Note that the relative wages of middle- to low-skill workers consists of two parts. The term $(1 - \theta^i)/\theta^i$ corresponds to the relative supply (of low-skill agents), while the term containing z corresponds to the relative demand. Therefore, our First Globalization comparative statics represents a shift in the relative demand curves, while keeping the relative supply fixed.²¹ This comparative statics is summarized in figure 2-11.

The relative wage of middle- to low-skill workers in the North increases because it off-shores the least skill-intensive inputs. As a result, the relative demand of middle-skill workers increases, thereby increasing the relative wage. The relative wage of middle- to low-skill workers also increases in the South. The reason is that an increase in traded intermediates (i.e., an increase in z_i) translates into a larger relative demand of middle-skill jobs. This result is similar to Wood (1995) or Feenstra and Hanson (1996).²²

2.4.1.2 Comparative Statics for the Second Globalization and Complementarity Result

Based on our results in Section 2.2, we characterize the Second Globalization as an increase in traded middle-skill-intensive inputs. We argued that the reduction in communication costs was the driver of the Second Globalization and it mainly affected trade in middle-skill-intensive industries. Thus, we add to the set of traded intermediates a new set of tradeable tasks. Given that the nature of trade costs driving the First and the Second Globalizations is different, it is natural to allow for the two sets to be possibly disjoint. We frame this observation in the following trade cost structure

$$\tau(z) = \begin{cases} 1 & \text{for } z \leq z_i \text{ and } \bar{z} \leq z \leq z_{II}, \\ \infty & \text{otherwise,} \end{cases} \quad (2.17)$$

where $0 \leq z_i \leq \bar{z} \leq z_{II} < 1$. This is, in addition to the First Globalization trade in intermediates $z \in [0, z_i]$, we now allow for trade in more skill-intensive tasks $z \in [\bar{z}, z_{II}]$.

We formally define the Second Globalization as an increase in z_{II} . Thus, the comparative statics exercise that we perform is to increase the set of traded inputs with skill intensity above \bar{z} by increasing z_{II} .²³

Allowing for the sets of First and Second Globalization traded inputs to be disjoint enables us to have a natural measure of depth of the First Globalization. Other formulations that do not rely on disjoint sets are possible and deliver similar insights. The two key as-

²¹Subsection 2.5.1 endogenizes labor supply decisions and shows that the comparative statics on wages remain unchanged.

²²Despite the mixed evidence presented in Goldberg and Pavcnik (2007) for inequality in the South, conventional wisdom seems to point to an increase in southern inequality as the right prediction to have. Our model is consistent with this prediction.

²³We increase z_{II} instead of decreasing \bar{z} because the skill content of southern exports in industries above U.S. average increased in the 1990s, see Figure 2-5.

sumptions are (i) trade in the Second Globalization affects more skill-intensive industries than in the First and (ii) the set of inputs that can be traded during the First and the Second Globalizations increases by incorporating inputs that are relatively more skill-intensive. These two assumptions are borne out by the data, as discussed above.

Assumption 5 $z_{II} < z^*(\theta_N, \theta_S)$, where $z^*(\theta_N, \theta_S)$ is implicitly defined in equation (2.16).

Assumption 5 ensures that in equilibrium South produces all traded inputs.

Proposition 18 (Second Globalization) *In the Second Globalization equilibrium, the relative wage of middle- to low-skill workers in the North has an inverse U-shape pattern. It increases in z_{II} for $z_{II} < \tilde{z}_{II}(z_I, \bar{z})$ and decreases thereafter. The relative wage of high-skill workers in the North and the relative wage of middle- to low-skill workers in the South increase in z_{II} .²⁴*

The intuition for the comparative statics of the relative wage of middle-skill workers in the North is as follows. Assume that $\bar{z} = 1/2$ and note that, to a first order approximation (for small z_I), the threshold $\tilde{z}_{II}(z_I, \bar{z} = 1/2)$ is the arithmetic mean of the skill intensity of inputs produced in the North after the First Globalization, i.e., $\tilde{z}_{II}(z_I) = \frac{1+z_I}{2}$. Therefore, when North offshores tasks below the skill requirement of the average input produced domestically, the relative demand of middle-skill workers increases, raising the relative wage. Conversely, the relative wage decreases when the tasks being offshored require a skill intensity higher than the skill of the average input. The relative wage of high-skill workers increases because the size of the bundle of inputs increases with the set of tradeable inputs. These results imply that the equilibrium tends to wage polarization: the relative wage of high- to middle-skill workers increases and the relative wage of middle- to low-skill workers eventually decreases. The relative wage in the South increases in the Second Globalization. The reason is that the marginal input being offshored is more skill-intensive, which raises the relative demand of middle-skill workers.

Our results for the wage distribution in the North are consistent with the 90/50 and 50/10 measures of U.S. wage inequality in the last three decades. Namely, the 90/50 measure has steadily increased, and the 50/10 increased during the 1980s, flattening and, eventually declining thereafter. This pattern is consistent with our model. Figure 2-12 reports the predictions of our model and data on U.S. wage inequality used in Autor et al. (2008).

Proposition 19 (Complementarity in the North) *The threshold $\tilde{z}_{II}(z_I, \bar{z})$ below which the relative wage of middle-skill workers in the North increases in z_{II} is increasing in z_I .*

Figure 2-13 summarizes the results in Proposition 19. Consider the extreme case in which the First Globalization did not happen, i.e., $z_I = 0$. The mean skill intensity of northern inputs

²⁴All the remaining proofs can be found in Appendix 2.7.

is $\tilde{z}_H(0) = 1/2$. Thus, the relative wage decreases from the onset of the Second Globalization. Consider now the case in which there has been some First Globalization, i.e., $z_1 > 0$. In this case, the mean skill is larger ($\tilde{z}_H(z_1) > 1/2$), implying that the relative wage increases in the first stages of the Second Globalization ($z_H < \tilde{z}_H$), to decrease thereafter. This interdependence brings about the importance of taking into account the First Globalization to predict the effects of the Second. There is a complementarity between trade in the First and the Second Globalizations: northern wage polarization is delayed by the extent of trade in the First Globalization.

We provide suggestive evidence consistent with the complementarity result in Figure 2-14 and Table 2.3. Proposition 19 states that the deeper the First Globalization is, the higher the relative wage of middle- to low-skill workers (our 50/10 measure) rises. Figure 2-14 shows a positive relationship between trade openness in 1990 and changes in lower-tail (50/10) northern wage inequality in the 1990s for the countries in the Luxembourg Income Study plus Japan and Spain. Table 2.3 reports the coefficients of regressing changes on wage inequality on trade openness for the whole sample and for G-7 countries. The coefficients of both regressions are positive and significant. The coefficient is larger when the sample is restricted to G-7 countries. These findings remain when controlling for income per capita (columns 2 and 4).

2.4.2 Two Souths and the Moving Band

In this subsection, we investigate how the existence of different southern countries which asymmetrically participate in Second Globalization trade affects their pattern of specialization and wage inequality. As pointed out before, a key difference between the First and the Second Globalizations is that, while the First is driven by tariff reductions, the Second is driven by the IT revolution. Arguably, a trade liberalization is a policy relatively easier to implement than building the specific capital needed to benefit from the IT revolution.²⁵ Thus, it is reasonable to expect that not all southern countries can equally engage in Second Globalization trade. To account for this heterogeneity within our framework, we consider an extension in which two identical Souths, Southeast and Southwest, open asymmetrically to trade during the Second Globalization. More specifically, we assume that the two Souths open to trade in the First Globalization, but only Southeast participates in the Second.²⁶

The equilibrium in the First Globalization is simple. Due to the symmetry of the two southern countries, all competitive equilibria feature the same wage schedule in both Souths. Appendix 2.7 contains the formal proof. We now turn to the characterization of the equilib-

²⁵In policy circles, trade liberalizations can be categorized as “first-generation” reforms. On the other hand, building the stock of technology and creating the institutional features needed to benefit from the IT revolution would be considered “second-generation” reforms, which take a longer time to be completed.

²⁶We maintain Assumption 5.

rium in the Second Globalization.

Proposition 20 (*Pattern of Specialization*) *In the Second Globalization, Southeast exports tasks $z \in [\bar{z}, z_{II}]$ and intermediates $z \leq \check{z}_I(z_I, z_{II})$. Southwest exports intermediates $z \in [\check{z}_I(z_I, z_{II}), z_I]$, with $0 \leq \check{z}_I(z_I, z_{II}) < z_I$.*

The reason for this result is that when Southeast starts offshoring tasks, its relative wage of low-skill workers decreases (these tasks are more skill-intensive than the intermediates offshored in the First Globalization). This gives Southeast comparative advantage in the least skill-intensive intermediates. As a result, in addition to tasks ($z \in [\bar{z}, z_{II}]$), Southeast also produces the least skill-intensive intermediates ($z \in [0, \check{z}_I]$).

Proposition 21 (*Moving Band*) *The threshold $\check{z}_I(z_I, z_{II})$ is increasing in z_I and decreasing in z_{II} in the relevant range.*

An implication of Proposition 21 is that the equilibrium tends to complete specialization as the Second Globalization progresses (i.e., z_{II} increases). As the set of traded tasks increases, the labor demand in Southeast increases, raising wages. Thus, the range of intermediates in which Southwest has comparative advantage increases. Wages in Southeast rise and eventually reach a point in which Southeast is only able to produce tasks (i.e., \check{z}_I goes to zero). Therefore, the band of intermediates produced in Southeast shrinks with the progress of the Second Globalization. In this sense, we have a moving band of intermediates in which Southeast has comparative advantage.

In 2000, Internet access in India was twice as large as in Pakistan.²⁷ If we take Internet access as a proxy for IT usage, this difference suggests an asymmetric participation in the Second Globalization for India and Pakistan. Our model predicts India specializing in middle skill-intensive industries and Pakistan specializing in less skill-intensive industries. Figure 2-3 shows that Indian and Pakistani exports to the United States are consistent with this prediction. India increase exports in industries above the average skill requirement, and decreased exports in industries below. We analyze whether these results extend to a larger set of countries and run the following regression

$$\Delta X_{iz} = \beta \Delta \text{Internet}_i \delta_{SG(z)} + \gamma \Delta \text{Internet}_i \delta_{FG(z)} + \epsilon_{iz},$$

where ΔX_{iz} denotes changes in exports from a southern country i to the United States in industry z between 2000 and 1990, $\Delta \text{Internet}_i$ is Internet adoption in country i in 2000, $\delta_{SG(z)}$ is an indicator for industry z participating in Second Globalization trade and $\delta_{FG(z)}$ is an

²⁷We use the number of Internet users per 100 inhabitants and the International Internet Bandwidth measured in bits per person from the World Development Indicators (World Bank).

indicator for First Globalization industries.²⁸ The prediction of our model is that as a southern country participates more in the Second Globalization, it increases its exports in Second Globalization goods, $\beta > 0$, and reduces its exports in First Globalization goods, $\gamma < 0$.

Column 1 of Table 2.4 reports the coefficients of our baseline regression. The interaction between Internet adoption and Second Globalization industries is positive and significant and the interaction between Internet adoption and First Globalization industries is negative and also significant. In column 2 we reduce the number of industries participating in the Second Globalization by raising the RTI threshold from the 50th to the 66th percentile. The sign and significance of the coefficients remain the same.

Therefore, the evidence presented in Table 2.4 is consistent with the prediction of the model. As southern countries participate more in the Second Globalization, which we proxy as an increase in Internet adoption, they export more Second Globalization goods and less First Globalization goods.

Proposition 22 *The relative wage of middle-skill workers is increasing in Southeast and (weakly) decreasing in Southwest in z_{II} .*

The intuition for this result is similar to Proposition 18. Southwest increases the production of intermediates below the mean skill of its domestic production, raising the relative wage of low-skill workers. The converse happens with Southeast. The set of exported tasks increases, while the band of exported intermediates decreases. As a result, the relative demand for middle-skill labor rises, thereby increasing its relative wage.

Proposition 22 highlights how gains from the Second Globalization may not be equally shared between different types of workers across southern countries. Some studies suggest that there is low labor mobility within southern countries. For example, Munshi and Rosenzweig (2009) document low labor mobility in rural India, even though inequality has risen in recent years.²⁹ If we assume low labor mobility within countries, our model can be applied to different regions of the same country. Then, this model could explain why inequality has increased in Bengaluru, an Indian city specialized in Second Globalization exports, and declined in Bhopa, a city which has not benefited from Second Globalization trade.

This section provided a tractable framework to study how differential access to trade generates changes in the pattern of specialization and wage inequality in otherwise identical southern countries. In our model, we assumed that the source of differential access to

²⁸There are no data for Internet adoption in 1990 and it was presumably negligible for most of sample. $\delta_{FG(z)} = (1 - \delta_{SG(z)}) * \delta_{low(z)}$, where $\delta_{low(z)}$ is a dummy for industry z taking value of one for industries below the average skill requirement. $\delta_{SG(z)}$ is a dummy taking value of one for industry z with the RTI index above the 50th and 66th percentile of the distribution in columns 1 and 2, respectively. Note that high levels of RTI are associated with middle-skill-intensive industries.

²⁹Paweenawat and Townsend (2009) document a similar pattern for Thailand and show that wages are not equalized across different Thai regions. Candelaria et al. (2009) document a similar fact for China: inequality in coastal regions has increased, while it has remained fairly constant in inland regions.

trade comes from the necessity of building an IT specific capital to benefit from the Second Globalization. We think of this infrastructure as being inherently more difficult to create and manage than tariff reductions. Therefore, our globalization approach provides a rationale for asymmetric participation within southern countries. This asymmetric participation generates a discontinuous pattern of specialization for the country (or region) participating in the Second Globalization. It leads to increasing wage inequality in this country (or region), while reducing it in the one not participating.

2.5 Extensions

This section relaxes some of the assumptions of the baseline model. Subsection 2.5.1 endogenizes labor supply decisions and shows that all comparative statics results for relative wages hold in this generalized set-up. Moreover, we find that the mass of northern agents selecting into middle-skill jobs endogenously expands in the First Globalization and eventually shrinks during the Second Globalization. The converse is true for low-skill agents. Subsection 2.5.2 endogenizes technology adoption and shows that the adoption of IT related technologies is delayed by First Globalization trade.

2.5.1 A Model with Endogenous Labor Supply

We extend the baseline model to allow agents to self select in any of the occupations of the economy. Let j be the index of an agent. As in the previous sections, we assume that $j \in [0, 1]$.

If agent j chooses to be employed in a low, middle or high-skill job, this agent can supply one, $s^i(j)$ and $s^i(j)^{1+\varepsilon}$ units of labor in country i , respectively. ε is some small number greater than zero. Note that wages described in Section 2.3 should now be interpreted as wages per unit of effective labor. To avoid a taxonomical analysis, we assume that functions $s^i(j)$ are strictly increasing.

North and South only differ on $s^i(j)$, where $s^N(j)$ first order stochastically dominates $s^S(j)$.³⁰ Note that there is a single-crossing property built-in s^i . If an agent j with skill $s^i(j)$ chooses to be employed as a high-skill worker, another agent j' , with $j < j'$ will also work as high-skill worker. Therefore, there exists a cutoff level of skill \bar{s}^i , such that all agents with $s^i > \bar{s}^i$ choose to work as high-skill workers. A similar reasoning applies for the middle to low decision.

The agent \bar{j} in country i who is indifferent between being employed in a middle or low-skill job verifies that $s^i(\bar{j})w_m^i = w_l^i$. Similarly, the agent \bar{J} who is indifferent between being employed in a high or middle skill job in the North verifies $s^N(\bar{J})^\varepsilon w_h^N = w_m^N$. It is conve-

³⁰Formally, this is $\frac{\int_0^J s^N(j) dj}{\int_0^1 s^N(j) dj} \leq \frac{\int_0^J s^S(j) dj}{\int_0^1 s^S(j) dj} \quad \forall J \in [0, 1]$.

nient to choose a functional form for $s^N(j)$ to obtain analytic solutions. For tractability, we specialize $s^N(j) = j$ in what follows.

Proposition 23 *(First Globalization)* In the First Globalization equilibrium, the mass of agents selecting into middle and high-skill jobs increases with z_1 in the North. The relative wage of middle- to low-skill workers and the relative wage of high-skill workers in the North increase with z_1 . In the South, the mass of middle-skill workers and its relative wage increase with z_1 .

The intuition for the results in Proposition 23 is that an increase in the set of tradeable intermediates increases the relative demand of middle-skill workers in both North and South. Therefore, the mass of agents selecting into middle-skill jobs increases in both countries. However, these changes in the supply of skills do not offset the primary demand forces, and the comparative statics for relative wages is analogous to section 2.4. The return on high-skill labor increases with trade because it increases the set of intermediates that can be purchased in the South at a cheaper price, while the demand for low and middle-skill workers declines in the North.

Proposition 24 *(Second Globalization)* In the Second Globalization equilibrium, the mass of northern middle-skill workers increases for $z_{II} < \tilde{z}_{II}(z_I, \bar{z})$ and decreases thereafter, where $\tilde{z}_{II}(z_I, \bar{z})$ is defined in Proposition 18. The mass of high-skill workers increases with z_{II} . The mass of low-skill workers decreases for $z_{II} < \tilde{z}_{II}(z_I, \bar{z}) + \eta(z_I, \bar{z})$, with $\eta > 0$ and increases thereafter. The relative wage of high-skill workers increases with z_{II} and the relative wage of middle- to low-skill workers increases for $z_{II} < \tilde{z}_{II}(z_I, \bar{z}) + \eta(z_I, \bar{z})$ and decreases thereafter. In the South, the mass of middle-skill workers and its relative wage increase with z_{II} .

An implication of Proposition 24 is that wage polarization emerges during the Second Globalization. Compared to the exogenous labor supply case, wage polarization is delayed when agents can endogenously select into occupations. This delay is intuitive because in the endogenous supply case there is an extra margin of adjustment. An additional insight from this exercise is to show the endogenous responses of the masses of agents selecting into each occupation. Consistent with the labor literature (e.g. Autor and Dorn, 2009), the mass of middle-skill workers in the North eventually shrinks with Second Globalization trade. In contrast, the mass of agents selecting low-skill jobs eventually expands with Second Globalization trade.

The results in this subsection suggest that from the point of view of the North, the First Globalization gave incentives to select into middle-skill jobs. In this sense, trade complemented middle-skills during the First Globalization in the North. However, this complementarity effect diminishes and it is eventually overturned as the Second Globalization progresses and more skill-intensive tasks are offshored to the South. In addition to a reduction

in the relative wage of middle-skill workers, this generates a reduction in the mass of northern middle-skill agents. For the South, trade complements skills in the First and the Second Globalizations.

Finally, the results for the Two Souths stated in Subsection 2.4.2, and for the Technology Adoption extension presented below (Subsection 2.5.2) apply in this extension of the model. The reason is that the relative wage of middle- to low-skill workers behaves in the same manner as in the baseline model and its behavior is the main driver of the results.

2.5.2 Technology Adoption

Our baseline model emphasized the role of the IT revolution for the Second Globalization, as argued by Blinder (2006). However, the Second Globalization is also the result of adopting new technologies which replace middle and low-skill jobs (Autor et al., 2003). It is therefore natural to investigate the effect of different types of trade on the incentives of firms to adopt new technologies.

Consider an extension of our baseline model in which agents can freely choose between two technologies. There is an Old Technology which can only benefit from trade in First Globalization intermediates. There is a New Technology which uses more skill-intensive inputs (e.g., computerization) and benefits from Second Globalization trade. This New Technology only uses tasks as inputs. More precisely,

$$\text{Old Technology : } h^\alpha \left(\int_0^1 \ln I(z) dz \right)^{1-\alpha},$$

$$\text{New Technology : } h^\alpha \left(\int_{\bar{z}}^1 \ln I(z) dz \right)^{1-\alpha}.$$

We assume that high-skill agents choose the production technology. Note that the optimal technology is the one that maximizes the bundle of inputs. These bundles in equilibrium are

$$B^{Old} = \int_0^{z_I} \ln(\tilde{p}^S(z))^{-1} dz + \int_{z_I}^1 \ln(\tilde{p}^N(z))^{-1} dz, \quad (2.18)$$

$$B^{New} = \int_{\bar{z}}^{z_{II}} \ln(\tilde{p}^S(z))^{-1} dz + \int_{z_{II}}^1 \ln(\tilde{p}^N(z))^{-1} dz, \quad (2.19)$$

where $\tilde{p}^i(z) = \frac{B}{(1-\alpha)Y} p^i(z)$ denotes a renormalized price in country i .

Proposition 25 *Let $\hat{z}_{II}(z_I)$ denote the threshold above which the New Technology starts to be adopted. The threshold $\hat{z}_{II}(z_I)$ is (weakly) increasing in z_I .*

There are two economic forces driving Proposition 25. First, the Old Technology benefits from the First Globalization by replacing northern intermediates by cheaper southern

intermediates. Second, the prices of tasks increase with First Globalization. As a result, the relative profitability of the Old Technology increases with trade in the First Globalization.

This simple technology choice model suggests that the effect of trade on technology adoption depends on the “type of trade” and the “type of technology”. The First Globalization complements Old Technology and the Second Globalization complements New Technology. However, First Globalization delays the adoption of New Technology and thus can be seen as a substitute for adoption of New Technology. If we think of New Technology as computerization, along the lines of Autor et al. (2003), Proposition 25 can be interpreted as saying that there is no dichotomy between offshoring of services and computerization.³¹

2.6 Concluding Remarks

In this paper, we provided a unified view of two phases of the globalization process and analyzed the interdependencies that arise. We distinguished between (i) the First Globalization, characterized by trade liberalizations, which mainly raised trade in low-skill-intensive goods and (ii) the Second Globalization, characterized by a reduction in communication costs, which increased trade in middle-skill-intensive goods.

We considered a North-South trade economy in which the North is skill-abundant. A final good is produced in the North employing high-skill agents and assembling a bundle of inputs. Inputs are produced combining middle and low-skill labor in different proportions and can be purchased in the North or the South.

First, we analyzed the distributional effects of the globalization process for the North and the South. We found a non-monotonic effect of trade in the Second Globalization on the relative wage of middle- to low-skill workers in the North. We showed that the relative wage of northern middle skill workers increases if the inputs being offshored are below the skill intensity of the average input produced in the North. Therefore, in our setup, the relative wage of middle-skill workers rises during the First Globalization and can decrease during the Second. A complementarity between the two globalizations arises because the threshold below which the relative wage rises during the Second Globalization increases with the set of traded intermediates in the First. Moreover, the relative wage of high-skill workers is increasing throughout the globalization process. Thus, as the Second Globalization progresses, the equilibrium tends to wage polarization. The complementarity result implies that

³¹The labor literature has proposed the computerization hypothesis to account for the differential loss of middle-skill jobs and wage polarization (Autor et al., 2003). In our model, computerization is needed to take advantage of the Second Globalization. Therefore, computerization leads to offshoring and, consequently, to the loss of middle-skill jobs and wage polarization in the North. This extension implies that trade in First Globalization delays computerization. This may be a rationale for why middling and wage polarization appear sooner in the relatively less open northern countries (e.g., the United States) than in relatively more open countries (e.g., the United Kingdom).

wage polarization is delayed by First Globalization trade. We provided empirical evidence consistent with this result. We found a positive relationship between trade openness before the onset of the IT revolution, which we assume to be in 1990, and increases in lower-tail northern wage inequality in the 1990s.

Second, we divided the original South in two identical southern countries and assumed that only one of the two southern countries could open to the Second Globalization. This exercise was meant to capture the notion that participation in the Second Globalization requires a stock of specific capital, which some southern countries may lack of. At the impact, the country that participates in the Second Globalization gains comparative advantage in the least skill-intensive intermediates. This generates a discontinuous pattern of specialization. As the Second Globalization progresses, the set of First Globalization intermediates in which this country has comparative advantage shrinks. Eventually, complete specialization is reached. The relative wage of middle-skill workers in the southern country that opens to the Second Globalization increases, while it decreases in the other. We provided evidence consistent with the prediction of the model. We showed that as southern countries raise their Internet adoption, they increase exports in RTI-intensive industries (Second Globalization goods) and decrease exports in low-skill-intensive industries (First Globalization goods).

Finally, we considered two extensions. First, we allowed for endogenous labor supply choices and showed that the comparative statics for relative wages hold in this generalized set-up. Moreover, we showed that the mass of northern agents selecting into middle-skill jobs increases during the First Globalization and eventually declines during the Second, while the converse is true for the mass of agents selecting into low-skill jobs. In the second extension, based on the assumption that the Second Globalization requires a new technology more intensive in skilled inputs, we showed that trade in the First Globalization delays trade in the Second through its effect on relative prices.

Our results depart from the findings of some of the previous literature on the distributional effects of the globalization process. For instance, Fujita and Thisse (2006) use a model with monopolistic competition and technological externalities to analyze the effect of trade on wage inequality. They find that the process of globalization can be detrimental to both skilled and unskilled workers in the North. The reason is that, in their framework, there exists a pecuniary externality that makes the price of the bundle of goods purchased in the North increase with globalization. In contrast, our model does not feature this externality and we find that the price of the bundle is decreasing with globalization, allowing northern high-skill agents to benefit from globalization.

Bibliography

- Acemoglu, D. (1998). Why Do New Technologies Complement Skills? Directed Technical Change And Wage Inequality. *The Quarterly Journal of Economics*, 113(4):1055–1089.
- Acemoglu, D. (1999). Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence. *American Economic Review*, 89(5):1259–1278.
- Acemoglu, D. (2003). Patterns of Skill Premia. *Review of Economic Studies*, 70(2):199–230.
- Anderson, J. E. (2009). Globalization and Income Distribution: A Specific Factors Continuum Approach. NBER Working Papers 14643, National Bureau of Economic Research, Inc.
- Anderson, J. E. and van Wincoop, E. (2004). Trade Costs. *Journal of Economic Literature*, 42(3):691–751.
- Antràs, P., Garicano, L., and Rossi-Hansberg, E. (2006a). Offshoring in a Knowledge Economy. *The Quarterly Journal of Economics*, 121(1):31–77.
- Antràs, P., Garicano, L., and Rossi-Hansberg, E. (2006b). Organizing Offshoring: Middle Managers and Communication Costs. NBER Working Papers 12196, National Bureau of Economic Research, Inc.
- Attanasio, O., Goldberg, P. K., and Pavcnik, N. (2004). Trade reforms and wage inequality in Colombia. *Journal of Development Economics*, 74(2):331–366.
- Autor, D. and Dorn, D. (2009). Inequality and Specialization: The Growth of Low-Skill Service Jobs in the United States. Working papers, MIT.
- Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *The Review of Economics and Statistics*, 90(2):300–323.
- Autor, D. H., Katz, L. F., and Krueger, A. B. (1998). Computing Inequality: Have Computers Changed The Labor Market? *The Quarterly Journal of Economics*, 113(4):1169–1213.
- Autor, D. H., Levy, F., and Murnane, R. J. (2003). The Skill Content Of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333.

- Blinder, A. S. (2006). Offshoring: The Next Industrial Revolution? *Foreign Affairs*.
- Blinder, A. S. and Krueger, A. B. (2009). Alternative Measures of Offshorability: A Survey Approach. Working Papers 1169, Princeton University, Department of Economics, Center for Economic Policy Studies.
- Blum, B. S. (2008). Trade, technology, and the rise of the service sector: The effects on US wage inequality. *Journal of International Economics*, 74(2):441–458.
- Candelaria, C., Daly, M., and Hale, G. (2009). Beyond Kuznets: persistent regional inequality in China. Working Paper 2009-07, Federal Reserve Bank of San Francisco.
- Costinot, A. and Vogel, J. (2009). Matching and Inequality in the World Economy. Working papers, MIT.
- Currie, J. and Harrison, A. E. (1997). Sharing the Costs: The Impact of Trade Reform on Capital and Labor in Morocco. *Journal of Labor Economics*, 15(3):S44–71.
- Dinopoulos, E., Syropoulos, C., and Xu, B. (2009). Intraindustry Trade and Wage-Income Inequality. Working paper, University of Florida.
- Ebenstein, A., Harrison, A., McMillan, M., and Phillips, S. (2009). Estimating the impact of trade and offshoring on american workers using the current population surveys. NBER Working Papers 15107, National Bureau of Economic Research, Inc.
- Feenstra, R. C. and Hanson, G. H. (1996). Foreign Investment, Outsourcing and Relative Wages. In Robert Feenstra, G. G. and Irwin, D., editors, *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bagwati*. MIT Press.
- Fujita, M. and Thisse, J.-F. (2006). Globalization And the Evolution Of The Supply Chain: Who Gains And Who Loses? *International Economic Review*, 47(3):811–836.
- Goldberg, P. K. and Pavcnik, N. (2007). Distributional Effects of Globalization in Developing Countries. *Journal of Economic Literature*, 45(1):39–82.
- Grossman, G. M. and Maggi, G. (2000). Diversity and Trade. *American Economic Review*, 90(5):1255–1275.
- Grossman, G. M. and Rossi-Hansberg, E. (2006). The rise of offshoring: it's not wine for cloth anymore. *Proceedings*, pages 59–102.
- Grossman, G. M. and Rossi-Hansberg, E. (2008). Trading Tasks: A Simple Theory of Offshoring. *American Economic Review*, 98(5):1978–97.

- Hanson, G. H. and Harrison, A. (1999). Trade liberalization and wage inequality in Mexico. *Industrial and Labor Relations Review*, 52(2):271–288.
- Haskel, J. E. and Slaughter, M. J. (2003). Have Falling Tariffs and Transportation Costs Raised US Wage Inequality? *Review of International Economics*, 11(4):630–650.
- Hummels, D. (2007). Transportation Costs and International Trade in the Second Era of Globalization. *Journal of Economic Perspectives*, 21(3):131–154.
- Katz, L. F. and Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78.
- Markusen, J. (2006). Offshoring of Business Services in Small Open Economies: Toward a General-Equilibrium Modeling Approach. *Brookings Trade Forum 2005: Offshoring White-Collar Work*.
- Markusen, J. and Strand, B. (2008). Offshoring of Business Services in Small Open Economies: Toward a General-Equilibrium Modeling Approach. *Journal of Industry, Competition and Trade*, 8(3):231–246.
- Munshi, K. and Rosenzweig, M. (2009). Why is Mobility in India so Low? Social Insurance, Inequality, and Growth. Working Paper 14850, National Bureau of Economic Research.
- Ohnsorge, F. and Trefler, D. (2007). Sorting It Out: International Trade with Heterogeneous Workers. *Journal of Political Economy*, 115(5):868–892.
- Paweenawat, A. and Townsend, R. M. (2009). Villages as Small Open Economies. Working paper, MIT.
- Robertson, R. (2000). Trade Liberalisation and Wage Inequality: Lessons from the Mexican Experience. *The World Economy*, 23(6):827–849.
- Robertson, R. (2004). Relative prices and wage inequality: evidence from Mexico. *Journal of International Economics*, 64(2):387–409.
- Topalova, P. (2005). Trade Liberalization, Poverty, and Inequality: Evidence from Indian Districts. NBER Working Papers 11614, National Bureau of Economic Research, Inc.
- Trefler, D. (2006). Service Offshoring: Threats and Opportunities. *Brookings Trade Forum 2005: Offshoring White-Collar Work*.
- Wood, A. (1995). How Trade Hurt Unskilled Workers. *The Journal of Economic Perspectives*, 9(3):57–80.

Yeaple, S. R. (2005). A simple model of firm heterogeneity, international trade, and wages. *Journal of International Economics*, 65(1):1–20.

Zhu, S. C. and Trefler, D. (2005). Trade and inequality in developing countries: a general equilibrium analysis. *Journal of International Economics*, 65(1):21–48.

2.7 Proofs and Auxiliary Propositions

Proof of Proposition 18 Using equations (2.11) to (2.14) and the trade cost structure, (2.17), the relative wages of middle skill workers are

$$\frac{w_h^N}{w_m^N} = \frac{1 - \varphi}{\varphi} \frac{\alpha}{1 - \alpha} \frac{2B}{1 - z_l^2 - z_{ll}^2 + \bar{z}^2} \quad (2.20)$$

$$\frac{w_m^N}{w_l^N} = \frac{1 - z_l^2 - z_{ll}^2 + \bar{z}^2}{(1 - z_l)^2 + (1 - z_{ll})^2 - (1 - \bar{z})^2} \frac{1 - \theta^N}{\theta^N (1 - \varphi)} \quad (2.21)$$

$$\frac{w_m^S}{w_l^S} = \frac{z_l^2 + z_{ll}^2 - \bar{z}^2}{1 - (1 - z_l)^2 - (1 - z_{ll})^2 + (1 - \bar{z})^2} \frac{1 - \theta^S}{\theta^S} \quad (2.22)$$

The relative wage of high skill workers in the North increases with z_{ll} because the bundle B increases with z_{ll} and the denominator decreases with z_{ll} . Note that the same reasoning applies for the relative wage of high to low skill workers. Taking the partial derivative of the relative wage of middle skill workers (2.21) in the North, we find that it is increasing in z_{ll} as long as $z_{ll} < \bar{z}_{ll}(z_l) \equiv 1 + \bar{z} - z_l - \sqrt{2(\bar{z} - z_l)(1 - z_l)}$. For the relative wage in the South, the sign of the partial derivative with respect to z_{ll} is always positive. \square

Proof of Proposition 19 Direct differentiation of $\bar{z}_{ll}(z_l)$ yields to

$$\frac{1 - 2z_l + \bar{z}}{\sqrt{2(\bar{z} - z_l)(1 - z_l)}} - 1.$$

Note that this expression is increasing in z_l and decreasing in \bar{z} , thus, a lower bound on it is $\frac{2}{\sqrt{2}} - 1$, which is positive. \square

Proposition 26 *In the two Souths model of Subsection 2.4.2, all competitive equilibria in the First Globalization have the same wage schedule for both Souths.*

Proof First, note that the price function in country i is a geometric mean of the middle and low skill wages. The price schedule in a country i , $p^i(z)$, is strictly monotone in z (if the wages of middle and low skill agents are different). Thus the price functions can cross at most once.

We proof the result by contradiction. Note that in autarky, the price of intermediates were the same in both Souths (because both are identical) and that as a result of opening to trade, the prices in the South strictly increase if there is positive demand from the North in any good. Suppose that North demands the set of goods χ_1 to Southeast and χ_2 to Southwest, where we are allowing for some traded intermediates $z \notin \chi_1 \cap \chi_2$. Note that intermediate $z = 0$ has to be produced *only* by one country, because otherwise the price in both countries would be the same and by single crossing we cannot have an equilibrium. Suppose Southeast

produces it. To have an equilibrium we must have the prices crossing in the relevant range. This means that $w_l^{\text{Southeast}} < w_l^{\text{Southwest}}$ and that $w_m^{\text{Southeast}} > w_m^{\text{Southwest}}$. In other words, the relative demand of middle skill workers in Southeast is higher than in Southwest. This implies that intermediates with low index z (low means below the threshold at which the two prices cross) are cheaper in Southeast and yet there is more demand of them in Southwest. This is a contradiction, unless both prices are equal, which implies that wages are equal in Southeast and Southwest. \square

Proof of Proposition 20 For algebraic convenience we normalize the population size of each southern country to one. Let $f(z)$ denote the fraction of each intermediate z produced by Southeast in the range $z \in [0, z_I]$. Thus, Southwest produces the remaining fraction $1 - f(z)$. Prices in both Souths will generically coincide if and only if wages of middle skill and low skill workers are equalized in equilibrium. Denoting $E_f[z] = \int_0^{z_I} z f(z) dz$, equalization of middle skill wages implies that $E_f[z] = \frac{1}{2} \int_0^{z_I} z dz$. Equalization of low skill wages implies that $E_f[z] = 1 - \frac{1}{2} \int_0^{z_I} (1 - z) dz$. These two conditions cannot be satisfied at the same time, and thus, the price schedule will be different in Southeast and Southwest.

By an analogous reasoning of proposition 26, prices can cross at most once. Thus, there is a threshold equilibrium. Denote by \check{z}_I the threshold intermediate. We show the result by contradiction. Suppose that Southwest produces $z \in [0, \check{z}_I]$ and that Southeast produces $z \in [\check{z}_I, z_I] \cup [\bar{z}, z_{II}]$. This can be an equilibrium if and only if $w_l^{\text{Southwest}} < w_l^{\text{Southeast}}$ and $w_m^{\text{Southwest}} > w_m^{\text{Southeast}}$. These conditions on wages imply

$$\begin{aligned} 0 &< z_I \left(1 - \frac{z_I}{2}\right) - 2\check{z}_I \left(1 - \frac{\check{z}_I}{2}\right) + z_{II} \left(1 - \frac{z_{II}}{2}\right) - \bar{z}_I \left(1 - \frac{\bar{z}_I}{2}\right), \\ 0 &< 2\check{z}_I^2 - z_I^2 - z_{II}^2 + \bar{z}_I^2, \end{aligned}$$

which cannot be satisfied simultaneously. Thus this cannot be an equilibrium. \square

Proof of Proposition 21 The threshold \check{z}_I can be expressed implicitly as the solution to the problem $p_{\text{Southeast}}(z) = p_{\text{Southwest}}(z)$ for some $z \in [0, z_I]$, where if the inequality is not satisfied, then either 0 or z_I is the solution, depending on whether the price schedule of Southeast is above or below the pricing schedule of Southwest for $z \in [0, z_I]$. Using that in order to have an equilibrium middle skill wages are higher in Southeast and low skill wages are lower in Southeast, we have that the geometric average with parameter z

$$\left(\frac{w_m^{\text{Southeast}}}{w_m^{\text{Southwest}}}\right)^z \left(\frac{w_l^{\text{Southeast}}}{w_l^{\text{Southwest}}}\right)^{1-z} \quad (2.23)$$

will be exactly one by some z between zero and one. Consider an interior solution for z . Inspection of the explicit equation (2.23) shows that both the ratios of middle skill and low

skill wages in Southeast to Southwest are decreasing in z_I and increasing in z_{II} and \check{z}_I . As result, and using implicit derivation, it follows that in this range $\check{z}_I(z_I, z_{II})$ is increasing in z_I and decreasing in z_{II} . Letting $A \equiv \frac{w_m^{\text{Southeast}}}{w_m^{\text{Southwest}}}$ and $B \equiv \frac{w_I^{\text{Southeast}}}{w_I^{\text{Southwest}}}$, the expression for the implicit derivatives of \check{z}_I becomes, after some manipulation,

$$\frac{\partial \check{z}_I}{\partial z_i} \left[\ln A - \ln B + \frac{\check{z}_I}{A} \frac{\partial A}{\partial \check{z}_I} + \frac{(1 - \check{z}_I)}{B} \frac{\partial B}{\partial \check{z}_I} \right] = -\frac{\check{z}_I}{A} \frac{\partial A}{\partial z_i} - \frac{(1 - \check{z}_I)}{B} \frac{\partial B}{\partial z_i}, \quad (2.24)$$

where $i = \{I, II\}$. The sign of the term in brackets in the left hand side is positive for all i and the term on the right hand side is positive for z_I and negative for z_{II} . Thus, the sign of the derivative of the threshold \check{z}_I with respect to z_i is unambiguous. \square

Proof of Proposition 22 The relative wages in Southeast and Southwest are proportional to

$$\frac{w_m^{\text{Southeast}}}{w_I^{\text{Southeast}}} \propto \frac{\check{z}_I^2 + z_{II}^2 - \bar{z}^2}{1 - (1 - \check{z}_I)^2 - (1 - z_{II})^2 + (1 - \bar{z})^2}, \quad \frac{w_m^{\text{Southwest}}}{w_I^{\text{Southwest}}} \propto \frac{z_I^2 - \check{z}_I^2}{(1 - \check{z}_I)^2 - (1 - z_I)^2}.$$

From proposition 21, the relative wage in Southwest is decreasing in z_{II} for the range in which there is an interior solution for \check{z}_I and is constant otherwise. For the relative wage in Southeast, if $\check{z}_I = 0$, it is immediate to check that the relative wage is increasing in z_{II} . If $\check{z}_I > 0$, we first show that a sufficient condition for the relative wage being increasing is that $|\partial \check{z}_I / \partial z_{II}| < 1$. If this is the case, the change induced in \check{z}_I by an infinitesimal change ε in z_{II} is bounded below by $z_I - \varepsilon$. Algebraic manipulation shows that as long as $z_{II} > \check{z}_I$ (which is true by assumption), the relative wage is increasing in z_{II} .

To show that $|\partial \check{z}_I / \partial z_{II}| < 1$, we show that an upper bound of this derivative is less than one,

$$\frac{\frac{\check{z}_I}{A} \frac{\partial A}{\partial z_{II}} + \frac{(1 - \check{z}_I)}{B} \frac{\partial B}{\partial z_{II}}}{\frac{\check{z}_I}{A} \frac{\partial A}{\partial \check{z}_I} + \frac{(1 - \check{z}_I)}{B} \frac{\partial B}{\partial \check{z}_I}} < 1.$$

This condition reduces to

$$\frac{-2\check{z}_I z_I + z_I + \check{z}_I}{(z_I - \check{z}_I)(z_I + \check{z}_I - 2)(z_I + \check{z}_I)} + \frac{\check{z}_I(z_{II} - \check{z}_I)}{\check{z}_I^2 + z_{II}^2 - \bar{z}^2} + \frac{(\check{z}_I - 1)(\check{z}_I - z_{II})}{(\check{z}_I - 2)\check{z}_I + (z_{II} - \bar{z})(z_{II} + \bar{z} - 2)} < 0,$$

which is true given that $0 < \check{z}_I < z_I < \bar{z} \leq z_{II} < 1$. \square

Proof of Proposition 23 The indifference conditions can be rewritten as

$$\bar{j}^\varepsilon \frac{\alpha}{1-\alpha} \frac{(2+\varepsilon)B}{1-\bar{j}^{2+\varepsilon}} = \frac{1-z_I^2-z_{II}^2+\bar{z}^2}{\bar{j}^2-\bar{j}^{N^2}} \quad (2.25)$$

$$\bar{j}^N \frac{1-z_I^2-z_{II}^2+\bar{z}^2}{\bar{j}^2-\bar{j}^{N^2}} = \frac{(1-z_I)^2+(1-z_{II})^2-(1-\bar{z})^2}{2\bar{j}^N}, \quad (2.26)$$

$$\bar{s}^S(\bar{j}^S) \frac{z_I^2+z_{II}^2-\bar{z}^2}{\int_{\bar{j}^S}^1 s^S(j) dj} = \frac{1-(1-z_I)^2-(1-z_{II})^2+(1-\bar{z})^2}{\bar{j}^S}. \quad (2.27)$$

Consider the case for the South. Equation (2.27) can be rewritten as

$$\frac{\int_{\bar{j}^S}^1 s^S(j) dj}{2\bar{j}^S \bar{s}^S(\bar{j}^S)} = \frac{\theta^S}{(1-\theta^S)} \frac{w_m^S}{w_I^S}, \quad (2.28)$$

where the expression for the wages corresponds to section 2.4. Thus, the right hand side of equation (2.28) is increasing in z_I . The left hand side of (2.28) is decreasing in \bar{j}^S . Therefore, \bar{j}^S is decreasing in z_I . Note that the relative wage of a middle skill agent can be written as

$$\frac{w_m^S}{w_I^S} = \frac{s^S(j)}{s^S(\bar{j}^S(z_I))}. \quad (2.29)$$

Thus, the relative wage in the South increases with z_I .

Consider the case for the North. Given that ε is a small positive number, we assume that $2+\varepsilon \approx 2$. Under this simplifying assumption, we find

$$\bar{j}^2 = \frac{(1+A)C}{1+(1+A)C'}, \quad (2.30)$$

$$\bar{j}^{N^2} = \frac{AC}{1+(1+A)C'}, \quad (2.31)$$

where $A = \frac{(1-z_I)^2+(1-z_{II})^2-(1-\bar{z})^2}{2(1-z_I^2-z_{II}^2+\bar{z}^2)}$ and $C = \frac{(1-\alpha)}{\alpha} \frac{1-z_I^2-z_{II}^2+\bar{z}^2}{2B}$. Note that optimality on offshoring requires B being increasing in z_I . In the First Globalization $z_{II} = \bar{z}$, thus, A and C (and AC) are decreasing in z_I . Therefore, \bar{j} and \bar{j}^N are decreasing in z_I . Finally, note that the size of middle agents is

$$\frac{\bar{j}}{\bar{j}^N} = \sqrt{\frac{1+A}{A}}, \quad (2.32)$$

which increases in the First Globalization. Finally, relative wages are

$$\frac{w_h^N}{w_m^N} = \frac{j^{1+\epsilon}}{\bar{j}^\epsilon(z_l)}, \quad (2.33)$$

$$\frac{w_m^N}{w_l^N} = \frac{j}{\bar{j}^N(z_l)}, \quad (2.34)$$

which are increasing in z_l . \square

Proof of Proposition 24 For the South, the same reasoning as in proposition 23 applies. For the North, the comparative statics is the same as in proposition 23, while A is decreasing. However, when $z_{II} > \bar{z}_{II}(z_l)$, A increases. From equation (2.32), it follows that the mass of middle skill workers declines. The comparative statics for the mass of high skill workers does not depend on A , but on AC , which is unambiguously decreasing in z_{II} . From equation (2.30), this implies that the threshold \bar{j} is decreasing in z_{II} . From equation (2.33), this implies that the relative wage of high skill agents is increasing. The threshold \bar{j}^N is implicitly defined by equation (2.31). Taking the total derivative of (2.31) with respect to z_{II} , we can isolate $d\bar{j}^N/dz_{II}$. Evaluating this derivative at $z_{II} = \bar{z}$ and $z_{II} = 1$, shows that the derivative takes negative and positive values, respectively. Moreover, it is immediate to check that the derivative is continuous and monotone. Intuitively, monotonicity follows from the derivatives of A and C being monotone. Thus, by the Bolzano theorem, we know that there is a unique threshold for z_{II} , above which $d\bar{j}^N/dz_{II} > 0$. Note that this threshold is above \bar{z}_{II} (defined in proposition 18) because $\partial A/\partial z_{II}|_{z_{II}=\bar{z}_{II}} = 0$, and from the implicit derivation of equation (2.31), it follows that $d\bar{j}^N/dz_{II}|_{z_{II}=\bar{z}_{II}} < 0$. \square

Proof of Proposition 25 Define $\Delta B(z_l, z_{II}) \equiv B^{New} - B^{Old}$, the difference in profits between the two technologies,

$$\Delta B(z_l, z_{II}) = \int_z^{z_{II}} \ln \left(\frac{\bar{p}^S(z)}{\bar{p}^N(z)} \right)^{-1} dz - \left(\int_0^{z_l} \ln(\bar{p}^S(z))^{-1} dz + \int_{z_l}^{\bar{z}} \ln(\bar{p}^N(z))^{-1} dz \right). \quad (2.35)$$

The first term in (2.35) summarizes the relative benefit of adopting the New Technology, whereas the second captures the additional benefit of using the Old Technology. The equation $\Delta B(z_l, z_{II}) = 0$ implicitly defines the threshold $\hat{z}_{II}(z_l)$ above which the New Technology starts to be adopted. The partial derivative of equation (2.35) with respect to z_{II} is positive, because $p^N(z) \geq p^S(z)$ in the trade region. The partial derivative of equation (2.35) with respect to z_l is negative. The first term decreases in z_l and the second term (in parenthesis) increases. The result for the first term comes directly from differentiation of prices. To obtain

the sign of the second term, note that by Leibniz's rule, we have that the partial derivative is

$$\ln\left(\frac{\tilde{p}^N(z_i)}{\tilde{p}^S(z_i)}\right) + \int_0^{z_i} \frac{\partial}{\partial z_i} \ln(\tilde{p}^S(z))^{-1} dz + \int_{z_i}^{\bar{z}} \frac{\partial}{\partial z_i} \ln(\tilde{p}^N(z))^{-1} dz. \quad (2.36)$$

The first term in (2.36) is non-negative as long as $p^N/p^S \geq 1$ for traded goods, which is assumed to be true to derive the equilibrium. The second and third terms can be expressed as

$$\frac{(1 - \bar{z} + 2z_i)(1 - \bar{z})}{1 - z_i^2}, \quad (2.37)$$

which is positive. Therefore, using the implicit function theorem it follows that $\hat{z}_H(z_i)$ is increasing in z_i . \square

2.8 Data Appendix

World bilateral trade flows are taken from Feenstra database. We obtain U.S. tariff data at industry level for the period 1978-1988 from Feenstra database. Feenstra's data is available from <http://cid.econ.ucdavis.edu/>. For the period 1990-2000, we use Romalis' database, available at <http://faculty.chicagobooth.edu/john.romalis/more/>. Transportation costs are cost of insurance and freight over customs import value from Feenstra database. Data on U.S. imports comes from Feenstra database.

We construct a skill intensity index by using 5 percent U.S. census data from IPUMS. The skill intensity variable is constructed assigning a score to each level of education reported in the US Census, using the variable educ99. We average across industries by same NAICS and across occupations when noted in the main text.

We take the routine-intensity index (RTI) from Autor and Dorn (2009). Roughly speaking, using the Dictionary of Tasks each task can be divided into three characteristics (abstract, routine and manual) and it is assigned a score for each of the three entries. The RTI index represents the importance of the routine part for each task. See Autor and Dorn (2009) for further discussion.

Internet measures are obtained from the World Development Indicators (WDI), available from the World Bank. For the robustness checks, the financial development measure is domestic credit to private sector over GDP. Human capital is the fraction of the labor force with secondary education. Both measures are obtained from the World Development Indicators (WDI).

2.9 Tables

Table 2.1: Trade Costs and Pattern of Specialization

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Whole Sample					South				
	OLS	OLS	OLS	OLS	2SLS	OLS	OLS	OLS	OLS	2SLS
<i>Panel A: Dependent Variable is U.S. Imports in 1990</i>										
Tariff	-6.77 (1.35)	-6.77 (2.31)				-8.2 (1.99)	-8.2 (2.56)			
Observations	26397	26397				12642	12642			
<i>Panel B: Dependent Variable is U.S. Imports in 2000</i>										
Tariff	-3.46 (3.63)	-3.46 (2.47)	-4.07 (3.62)	-4.07 (2.43)	-4.05 (3.62)	-1.29 (2.58)	-1.29 (2.88)	-1.49 (2.56)	-1.49 (2.89)	-1.48 (2.57)
Internet · Skill Intensity			6.3 (1.65)	6.3 (5.03)	6.09 (.93)			9.59 (3.19)	9.59 (2.66)	8.07 (1.64)
Observations	262303	262303	261961	261961	261961	126891	126891	126549	126549	126549
Std. Error	Robust	Cluster	Robust	Cluster	Robust	Robust	Cluster	Robust	Cluster	Robust

Standard errors are clustered by country. A southern country is defined as having less than half of 2000 U.S. GDP per capita adjusted for PPP from the Penn World Tables. RTI index is used as instrument of Skill Intensity in the first stage regressions, which are omitted. Regressions include country and industry fixed effects. Dependent variable is U.S. Imports from Feenstra's NBER Dataset. Tariff is U.S. Tariffs at HS6 level from Romalis' Dataset. Skill intensity is mean level of education from U.S. Census for industry. Internet is the fraction of population with access to Internet in 2000. See Appendix 2.8 for detailed data definitions and sources.

Table 2.2: Change in Trade Openness and Wage Bill in the U.S.

	(1)	(2)	(3)
	Δ Wage Bill 80-90	Δ Wage Bill 90-96	
Δ Trade Openness 80-90 \cdot Skill Int.	1.65 (.43)		
Δ Trade Openness 90-96 \cdot Skill Int.		.29 (.17)	-4.62 (2.17)
Δ Trade Openness 90-96 \cdot Skill Int. ²			.23 (.10)
Observations	118	118	118

Robust Standard Errors in parenthesis. Δ Trade Openness is the change in the share of exports and imports over GDP from the Penn World Tables. Wage bill data at industry level at 3-digit NAICS comes from Autor et al. (1998). Skill intensity is the mean level of education from U.S. Census by industry.

2.10 Figures

Table 2.3: Complementarity in the North

	(1)	(2)	(3)	(4)
	Whole Sample		G-7 Countries	
	<i>Dep. Var.: Change in 50/10 wage in the 1990s</i>			
Trade Openness 1990	.18 (.05)	.17 (.06)	.24 (.07)	.30 (.13)
Income per capita		-.01 (.01)		.01 (.02)
Observations	19	19	7	7

Robust Standard Errors in parenthesis. The change in 50/10 wage inequality comes from the LIS data, to which we add Japan and Spain from the OECD. We restrict the LIS sample to countries that have more than 50% of U.S. income per capita. The LIS data are taken from rounds V and III. Trade Openness is the share of exports and imports over GDP from the Penn World Tables in 1990. The income per capita data are taken from Penn World Tables for 1995.

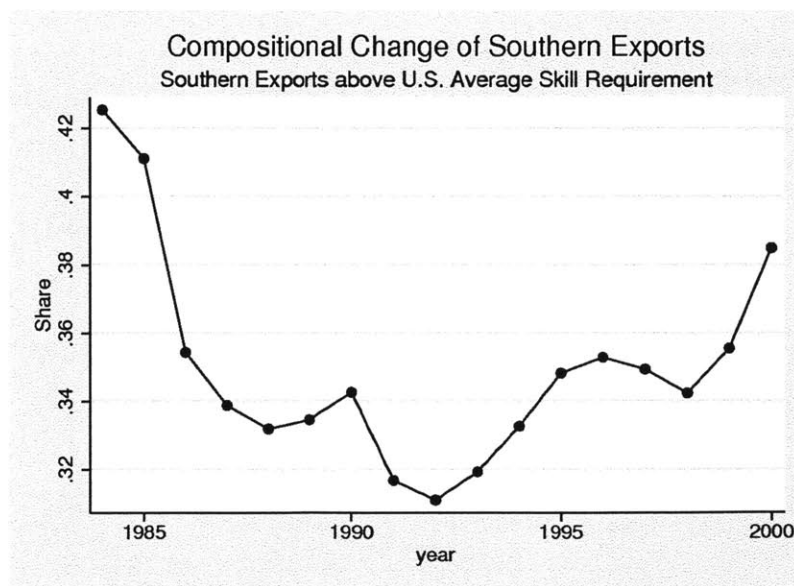


Figure 2-1: Changes in southern exports to the North in industries with above average U.S. skill intensity. The mean skill intensity of U.S. industries is measured using educational attainment in U.S. Census. North is defined as having more than 50 percent of U.S. GDP per capita (PPP adjusted). Source: Feenstra World Trade Database.

Table 2.4: Testing the Moving Band

	(1)	(2)
	Dep. Var. is	Δ Imports
$\Delta Internet \delta_{SG(z)}$.053 (.015)	.091 (.025)
$\Delta Internet \delta_{FG(z)}$	-.049 (.013)	-.047 (.012)
Obs.	1705	1705
R ²	0.04	0.06

Standard errors are clustered by country. Dependent variable is change in U.S. southern imports between 1990 and 2000. U.S. Imports data are from Feenstra's NBER Dataset. Internet and Δ Internet is the fraction of the population with access to Internet in 2000. There are no data for Internet adoption in 1990 and it was presumably negligible for most of the sample. $\delta_{FG(z)} = (1 - \delta_{SG(z)}) * \delta_{low(z)}$, where $\delta_{low(z)}$ is a dummy for industry z taking value of one for industries below the average skill requirement. $\delta_{SG(z)}$ is a dummy taking value of one for industry z with the RTI index above the 50th and 66th percentile of the distribution in columns 1 and 2, respectively. A southern country is defined as having less than 50 percent of U.S. GDP per capita (PPP adjusted). See Appendix B for detailed data definitions and sources.

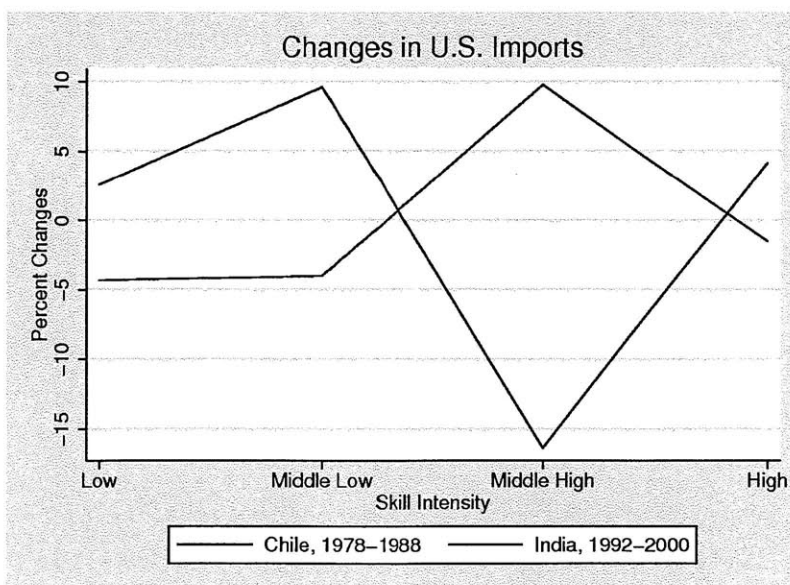


Figure 2-2: Changes in U.S. imports from India and Chile for different skill levels. For comparability with our tariff data, we can only consider the period 1978-1988 for the First Globalization. For India, the series starts in 1992 to dampen the effect of the trade liberalization in 1991, documented in Topalova (2005) among others. Source: Feenstra U.S. Imports Database, Skill Intensity constructed from U.S. Census.

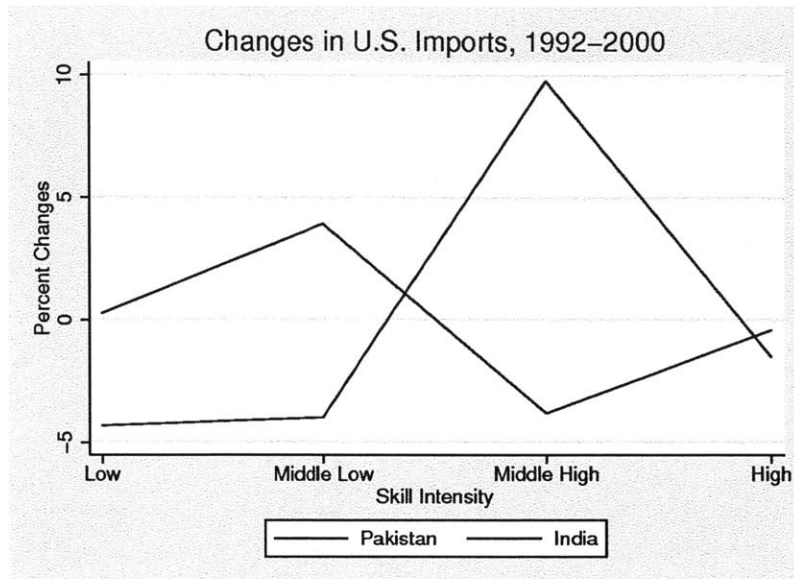


Figure 2-3: Changes in U.S. imports from India and Pakistan for different skill levels. Source: Feenstra U.S. Imports Database.

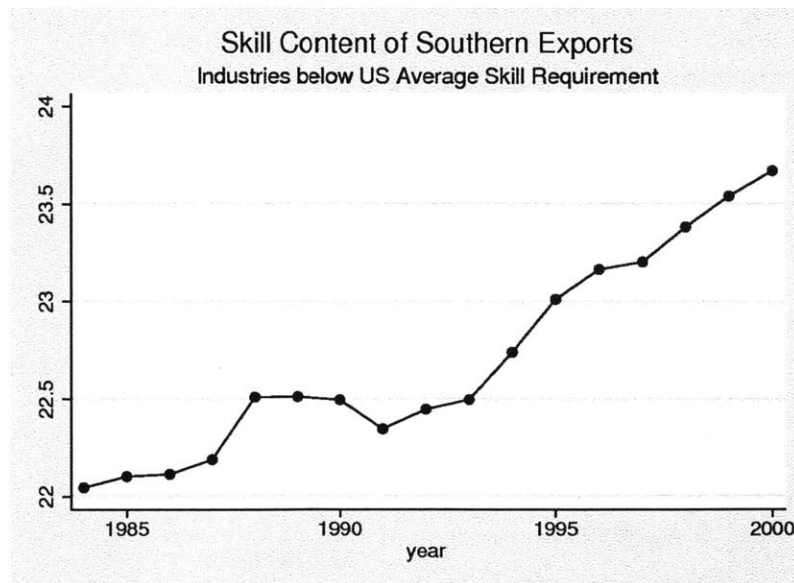


Figure 2-4: Skill Content of Southern Exports in industries with skill requirement below U.S. average. The skill content is the weighted average of the skill embodied in southern exports. The skill intensity of U.S. industries is measured using educational attainment in U.S. Census. North is defined as having more than 50 percent of U.S. GDP per capita (PPP adjusted). Source: Feenstra World Trade Database.

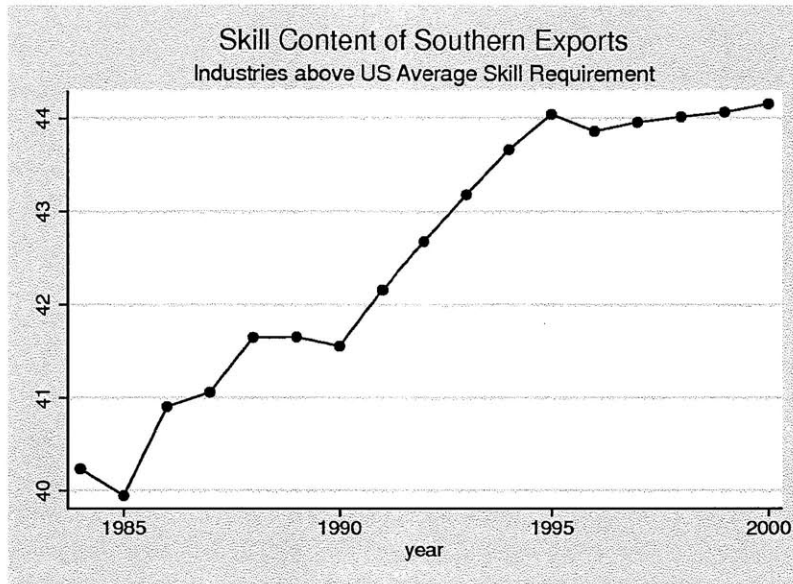


Figure 2-5: Skill Content of Southern Exports in industries with skill requirement above U.S. average. The skill content is the weighted average of the skill embodied in southern exports. The skill intensity of U.S. industries is measured using educational attainment in U.S. Census. North is defined as having more than 50 percent of U.S. GDP per capita (PPP adjusted). Source: Feenstra World Trade Database.

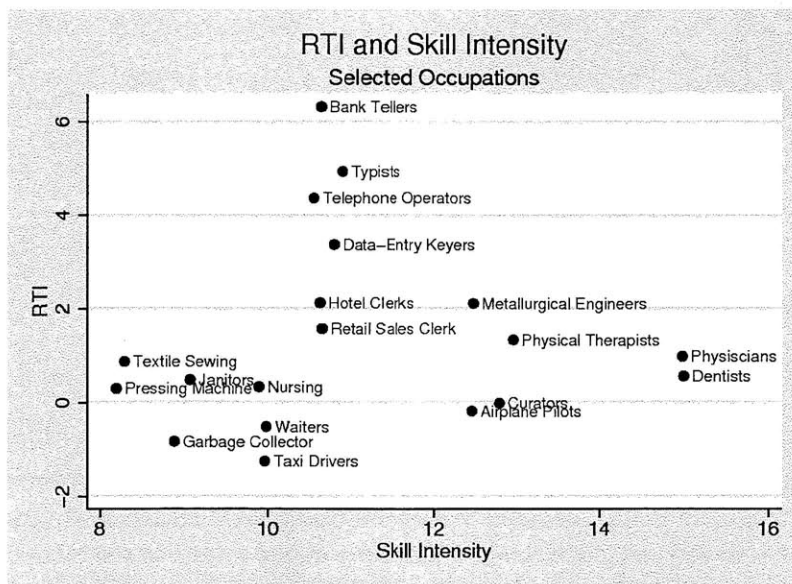


Figure 2-6: Examples of RTI for Selected Occupations. Source: Autor and Dorn (2009).

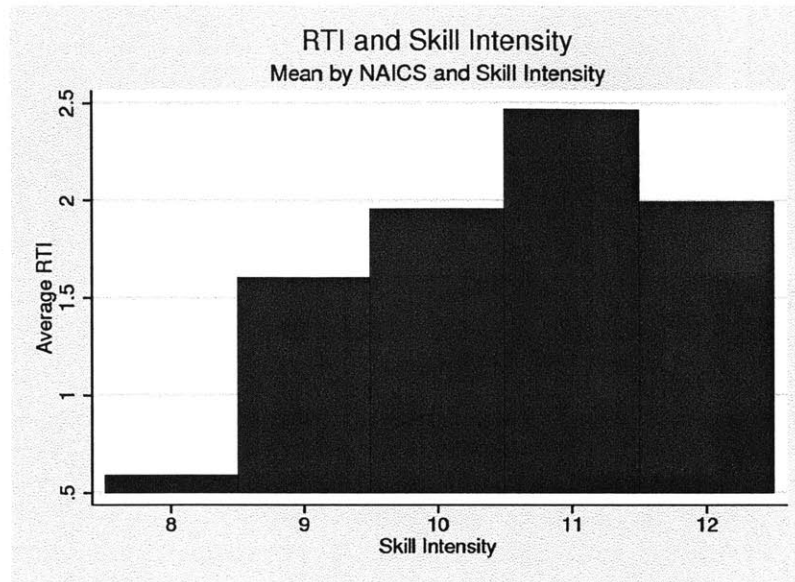


Figure 2-7: Average Skill Intensity by U.S. Industry. Source: Autor and Dorn (2009).

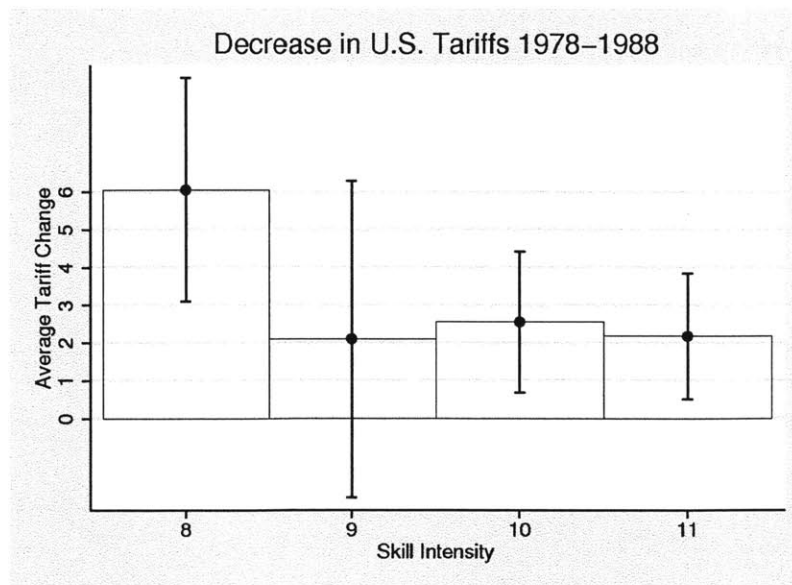


Figure 2-8: Changes in U.S. Tariffs by Skill in the First Globalization. (Two Std. Dev. bars). Source: Feenstra tariff data.

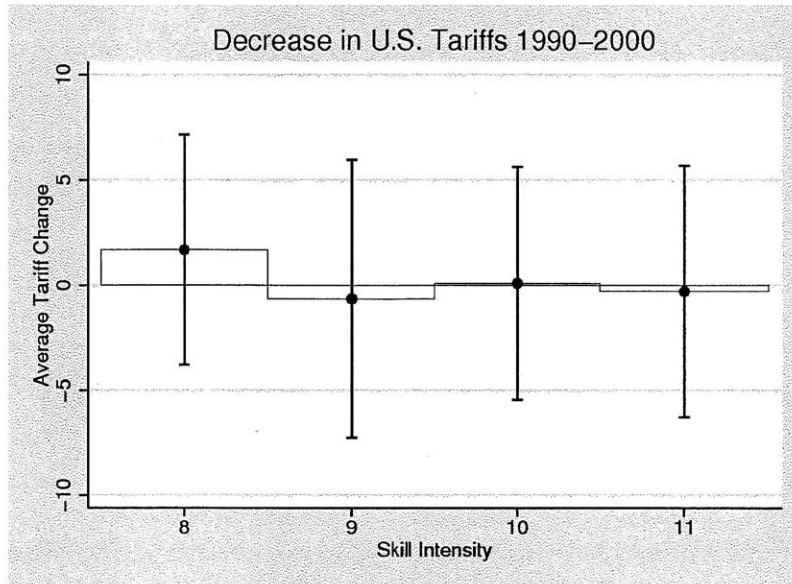


Figure 2-9: Changes in U.S. Tariffs by Skill in the Second Globalization. (One Std. Dev. bars). Source: Romalis tariff data.

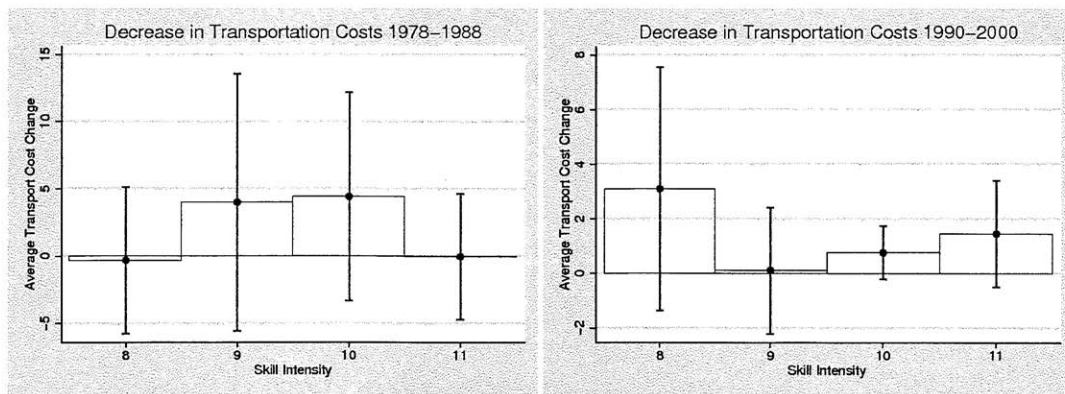


Figure 2-10: Changes in U.S. Transportation Costs (Insurance and Freight). One Std. Dev. bars are shown. Source: Feenstra database.

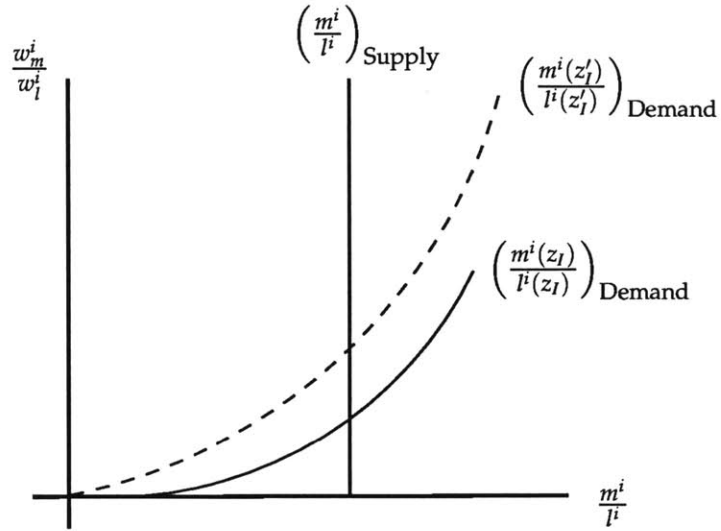


Figure 2-11: Relative Supply and Demand in the First Globalization for $z_I' > z_I$.

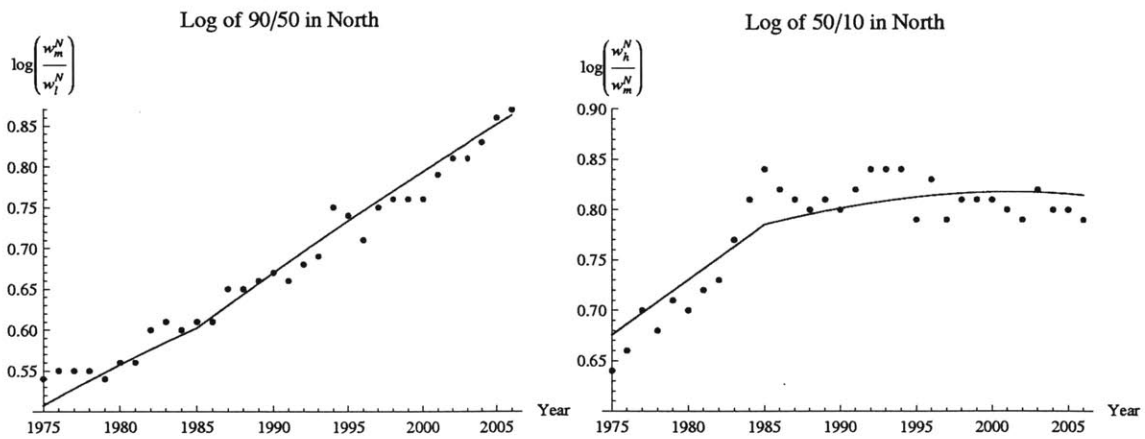


Figure 2-12: Model versus Autor et al. (2008) estimates. We take the structural change to be 1987, as in Autor et al. The First Globalization is a shift from $z_I = .15$ to $z_I = .21$. The Second Globalization is a shift from $z_{II} = .5$ to $z_{II} = .64$.

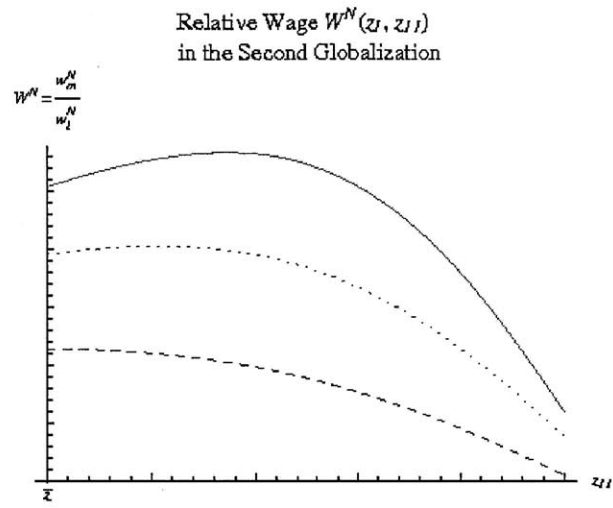


Figure 2-13: Interdependence in the North. This plot assumes $\bar{z} = 1/2$. The dashed line is for $z_I = 0$, dotted for $z_I = .2$ and regular line for $z_I = .3$.

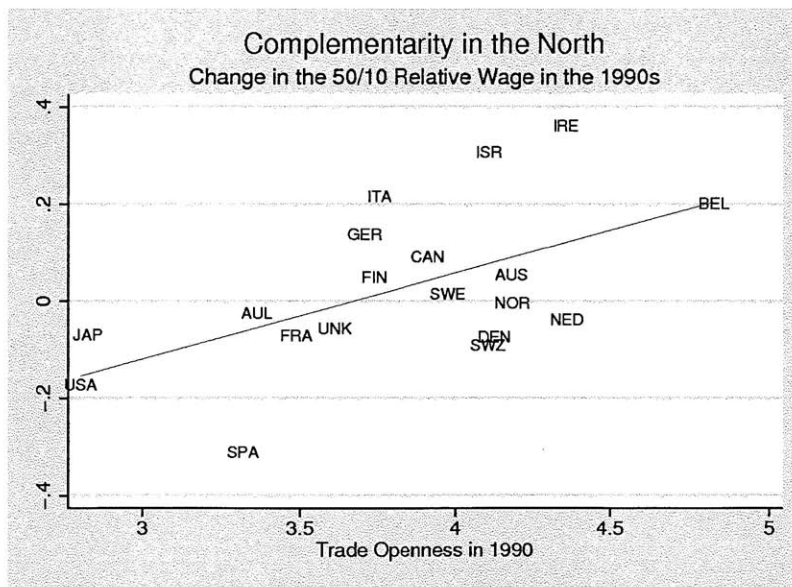


Figure 2-14: Scatter plot of changes in 50/10 wage inequality versus First Globalization trade. Wage inequality comes from the LIS data, to which we add Japan and Spain from OECD. We restrict LIS sample to countries that have more than 50% of U.S. income per capita. The LIS data are taken from rounds V and III. Trade Openness is share of exports and imports over GDP from the Penn World Tables in 1990.

Chapter 3

The Intensive Margin of Technology Adoption*

with Diego Comin

Abstract

We present a tractable model for analyzing the relationship between economic growth and the intensive and extensive margins of technology adoption. The “extensive” margin refers to the timing of a country’s adoption of a new technology; the “intensive” margin refers to how many units are adopted (for a given size economy). At the aggregate level, our model is isomorphic to a neoclassical growth model, while at the microeconomic level it features adoption of firms at the extensive and the intensive margin. Based on a data set of 15 technologies and 166 countries our estimations of the model yield four main findings: *(i)* there are large cross-country differences in the intensive margin of adoption; *(ii)* differences in the intensive margin vary substantially across technologies; *(iii)* the cross-country dispersion of adoption lags has declined over time while the cross-country dispersion in the intensive margin has not; *(iv)* the cross-country variation in the intensive margin of adoption accounts for more than 40% of the variation in income per capita.

*We would like to thank Mar Reguant and participants at the NBER Macroeconomics across Time and Space 2010 and the Growth and Development conference in Santa Barbara 2010 for comments and suggestions. Comin would like to thank the NSF (Grants # SES-0517910 and SBE-738101) for their financial assistance. Mestieri thanks the Bank of Spain for its financial support. The views expressed in this paper solely reflect those of the authors and not necessarily those of the National Bureau of Economic Research.

3.1 Introduction

Either as a fundamental or as a channel that amplifies other fundamentals (e.g., institutions), technology is surely important to understanding why there are rich and poor countries. However, several factors have limited economists' ability to assess its importance quantitatively. Broadly speaking, the goal of this paper is to overcome some of these difficulties.

One significant limitation in efforts to assess the role of technology has been the lack of direct measures of technology. Traditionally, technology diffusion has been measured as the share of producers who adopt a given technology (Griliches, 1957, Mansfield, 1961 and Gort and Klepper, 1982). Computing this measure requires firm-level data which are hard to collect for a large number of years, countries, and technologies. Consequently, economists have been unable to measure technology diffusion comprehensively.

Comin and Hobijn (2004) and Comin, Hobijn and Rovito (2006) present a new approach to measuring technology. They measure either the number of units of capital in a country that embody the new technology (e.g., the number of telephones) or the amount of output produced with the new technology (e.g., the tons of steel produced in blast oxygen furnaces). These measures have two advantages over traditional measures of technology diffusion. First, they just require data at the country level. Thus, it is easier to collect them for a large number of countries, technologies and years.¹ Second, they capture the number of units of the technology adopted by each adopter.²

As with any new data set, these new technology measures introduce the challenge of finding ways to extract information relevant to modeling the technology diffusion process. That is, they present the challenge of mapping the data into dimensions that we can interpret through the lens of our models.

Consider Figure 1 for an example of one technology in Comin, Hobijn and Rovito (2006). Figure 1 reports the number of land line phone calls normalized by total output for the United States, Australia, Japan, Malawi, Pakistan and Burkina Faso. These curves roughly appear to be the graphic result of plotting a single curve and then shifting it both horizontally and vertically. The hypothesis that this apparent graphic result reflects the actual process of technology adoption across countries was broadly confirmed in formal tests conducted in Comin and Hobijn (2010). Assuming this characterization of technology adoption, we can fully describe cross-country differences in technology dynamics if we know what drives the horizontal and vertical shifts in the diffusion curves. Section 1 develops a model based on Comin and Hobijn (2010) that provides a micro-foundation for these shifts.

A technology, in our model, is a group of production methods used to make an interme-

¹The CHAT data set described in Comin and Hobijn (2009) contains information about the diffusion of 104 technologies in 166 countries over the last 200 years.

²As shown by Comin, Hobijn and Rovito (2008), technology measures that include this dimension do not diffuse following a logistic curve which is characteristic of traditional measures (Griliches, 1957).

diate good or provide a service. We consider two aspects of technology adoption, which we call the “extensive” and “intensive” margins. The extensive margin of technology adoption gauges how long it takes a country to adopt a technology. Adopting a production method requires incurring in a fixed investment. The timing of this investment determines the lag with which production methods arrive in a country. In section 2, we show that the horizontal shifts seen in Figure 1 measure adoption lags. We call this lag the *extensive margin* of adoption.

Once a technology has been introduced, the *intensive margin* of adoption captures how many units of the good embodying it are demanded relative to aggregate demand. The intensive margin is determined by the productivity and price of goods that embody the technology and the cost that individual producers face in learning how to use it. Other things equal, these variables produce vertical shifts in the evolution of observable measures of technology adoption such as displayed in Figure 1. We call the vertical shift the intensive margin.

In this paper we pay particular attention to the intensive margin. To be clear, our goal is not to assess how important a particular factor may be in affecting the intensive margin of technology adoption. Instead, we just intend to understand how important cross-country differences in this margin are in explaining cross-country differences in productivity. Answering this question does not require taking a strong stand on the nature of the drivers of technology adoption.

Little has been known about how significant a role the intensive margin of technology adoption plays in determining overall productivity performance. Clark’s (1988) classic study on spinning-machine spindles documents large cross-country differences in the number of spindles that each worker operated circa 1900 and argues that this factor was a major contributor to differences in productivity. However, observing more units of a new technology in rich countries is not sufficient to establish the importance of the intensive margin for aggregate productivity, since it could just reflect reverse causation. In other words, higher aggregate demand could lead to the adoption of more units of technology per worker.

Filtering out the effect of aggregate demand on observable measures of technology is a key challenge that any attempt to assess the importance of the intensive margin needs to confront. We follow two different approaches to deal with this issue. First, we use our model predictions to pin down the income elasticity of our technology measures. On a balanced growth path, the income elasticity of demand for the goods embodying technology must equal one. Using this restriction we can filter out the effect of aggregate demand on technology adoption and then use the model to formalize the intuitions described above and identify the intensive and extensive margins of adoption.

Our second approach relaxes the restrictions of a balanced growth path (despite its appeal over long periods we study) to check the robustness of our findings. Here we use the

time series dimension of our panel to estimate the income elasticity of demand for goods embodying a technology. Because of the need for long time series to carry out this exercise and because we want to reduce possible biases in the estimates of the intensive margin,³ we proceed in two steps. First, we estimate the income elasticity of demand for goods embodying a technology using only U.S. data, and then we impose this estimate on the other countries to estimate the intensive adoption margin and the adoption lags for each technology-country pair.

We use data for 15 technologies and 166 countries, as in Comin, Hobijn, and Rovito (2006). Our data cover major technologies related to transportation, telecommunication, information technology, health care, steel production, and electricity. We obtain precise and plausible estimates of the adoption lags for two thirds of the 1294 technology-country pairs for which we have sufficient data.

Our exploration of the intensive margin of adoption, complementing Comin and Hobijn's (2010) analysis of the extensive margin, delivers four main findings. First, the magnitude of cross-country differences in the intensive margin of adoption and adoption lags are large. On average, they are 20% larger than the cross-country dispersion in per capita income. Second, there are significant differences across technologies in the cross-country dispersion in the intensive margin. For example, the dispersion in electricity and passenger rail represents 40% of the dispersion in per capita income, while in blast oxygen steel represents 170% of the dispersion in income. Third, a variance decomposition reveals that 33% of the variation in the intensive margin can be attributed to cross-technology variation, 43% can be attributed to cross-country variation, and the remaining 23% is not explained by these two factors. Fourth, the cross-country dispersion in adoption lags has declined monotonically over time. Specifically, for every decade later that a technology has been invented, the dispersion has been two years smaller. In contrast, we do not observe any cross country convergence in the intensive margin of adoption.

Our model is similar at the aggregate level to the neoclassical growth model, except that in our model the level of total factor productivity (TFP) is endogenous. In particular, TFP depends on both the intensive and extensive margins of technology adoption. We use this result to assess the magnitude of the cross-country differences in labor productivity that our estimated differences in the intensive margin generate. We find that differences in the intensive margin of adoption account for 44% of cross-country differences in income per capita. As our model makes clear, these effects could be fundamentally driven by differences in the costs individual producers face in adopting the new technologies or by differences in the overall efficiency of the economy that affect the intensity of adoption.

Finally, we show that the empirical results obtained in the baseline model hold when we

³See section 2.3 for more details.

allow for non-homothetic production functions that do not pin down the income elasticity of demand for goods embodying a new technology. For instance, the intensive margin still accounts for 52% of the cross-country variation in income per capita.

This paper is related to three strands of the literature. First, macroeconomic models of technology adoption (e.g., Parente and Prescott, 1994, and Basu and Weil, 1998) have tried to understand the role of technology in determining TFP. However, these studies have used an abstract concept of technology that is hard to match with data. Second, the applied microeconomic technology diffusion literature (Griliches, 1957, Mansfield, 1961, and Gort and Klepper, 1982, among others) has focused on estimating diffusion curves for technologies in different countries. However, these studies have only been able to investigate a relatively small number of technologies and countries. Moreover, the diffusion curves are purely statistical descriptions, not embedded in an aggregate model. Hence, it is difficult to use them to explore the aggregate implications of the findings.

Finally, the closest reference to this paper is Comin and Hobijn (2010). This paper differs from it in at least three important ways. First, our model provides a micro-foundation for the intensive margin of adoption as well as for the extensive margin. Second, in our empirical analysis we estimate and analyze the intensive margin of adoption. Third, we explore the robustness of our findings about the two margins of adoption by relaxing the balanced growth restriction and allowing the income elasticity of the demand for the technology to be endogenous.

The paper is divided into four sections. Section 3.2 sets out a one-sector neoclassical growth model featuring intensive and extensive margins of adoption. Section 3.3 describes the diffusion patterns of technology under the balanced growth path assumption, derives structural equations that can be estimated from the data, and explains how the margins of adoption are identified. Section 3.4 presents the results of the estimation, and Section 3.5 concludes.

3.2 A one-sector growth model with extensive and intensive technology adoption

We next present a one-sector growth model with endogenous technology adoption at the extensive and intensive margins. The model maps the adoption margins into the time-path of observable measures of technology diffusion and illustrates how each adoption margin affects endogenous TFP differentials. In what follows, we omit the time subscript, t , where obvious.

3.2.1 Preferences

A measure one of households populates the economy. They inelastically supply one unit of labor every instant, at the real wage rate W , and derive the following utility from their consumption flow

$$U = \int_0^{\infty} e^{-\rho t} \ln(C_t) dt. \quad (3.1)$$

Here C_t denotes per capita consumption and ρ is the discount rate. We further assume that capital markets are perfectly competitive and that consumers can borrow and lend at the real rate \tilde{r} .

3.2.2 Production

Technology:

Each instant, a new production method appears exogenously. We call these production methods, technology vintages or simply vintages. Production methods are capital embodied. The set of vintages available at time t is given by $\bar{V} = (-\infty, t]$. The productivity embodied in new vintages grows at a rate γ across vintages, such that

$$Z_v = Z_0 e^{\gamma v}. \quad (3.2)$$

Note that Z_v is constant over time. This characterizes the evolution of the world technology frontier. We shall choose the normalization parameter Z_0 such that vintage \underline{v} has productivity $Z_{\underline{v}}$.⁴

A country does not necessarily use all the capital vintages that are available in the world because, as we discuss below, making them available for production is costly. The set of vintages actually used is given by $V = (-\infty, t - D]$. Here $D \geq 0$ denotes the *adoption lag*. That is, the amount of time between when the best technology in use in the country became available and when it was introduced in the country.

In order to map the model into the data, we introduce the concept of technology. A technology is a set of production methods used to produce closely related intermediates. To simplify the exposition, we consider only two technologies: an old one, denoted by o , and a new one, denoted by n . The old technology consists of the production methods introduced up till a fixed time \underline{v} , such that the set of vintages associated with the old technology is $V_o = (-\infty, \underline{v}]$. The new technology consists of the newest production methods, invented after \underline{v} , such that it covers $V_n = (\underline{v}, t - D]$.

Output:

⁴This implies that $Z_0 = Z_{\underline{v}} e^{-\gamma \underline{v}}$.

The output associated to a technology τ , Y_τ , is given by:

$$Y_\tau = \left(\int_{V_\tau} Y_v^{\frac{1}{\mu}} dv \right)^\mu, \quad \tau \in \{0, n\}, \quad (3.3)$$

where Y_v denotes the intermediate output produced using technology vintage v . Final output Y , is produced competitively with the following production function:

$$Y = \left(\int_{-\infty}^{t-D} Y_v^{\frac{1}{\mu}} dv \right)^\mu = \left(\sum_{\tau \in \{0, n\}} Y_\tau^{\frac{1}{\mu}} \right)^\mu.$$

Once a technology vintage v is brought to the country, producers can find distinct ways to use it. Because each application developed solves a new problem, the larger the number of applications developed, N_v , the more efficient the production of intermediate service v is. In other words, there are efficiency gains from developing more applications. Each application yields a differentiated output, Y_{vi} . Differentiated outputs are produced monopolistically. A competitive producer then aggregates these outputs in the form of intermediate v , Y_v , as follows:⁵

$$Y_v = N_v^{-(\mu-\mu')} \left(\int_0^{N_v} Y_{vi}^{\frac{1}{\mu}} di \right)^\mu, \quad \text{with } \mu > \mu' > 1. \quad (3.4)$$

Output Y_{vi} is produced by combining labor and capital, K_{vi} , that embodies production method, v , as follows:

$$Y_{vi} = Z_v L_{vi}^{1-\alpha} K_{vi}^\alpha, \quad (3.5)$$

Capital goods production and taxes:

Capital goods are produced by monopolistic competitors. Each of them holds the patent of the capital good used for a particular production method v . It takes one unit of final output to produce one unit of capital of any vintage. This production process is assumed to be fully reversible. For simplicity, we assume that there is no physical depreciation of capital. The capital goods suppliers rent out their capital goods at the rental rate $\phi_R R_v$. R_v is the price received by the capital goods producer, while the wedge ϕ_R captures a tax on the price of capital that the government rebates back to the consumers with a lump sum transfer. ϕ_R is constant across vintages and over time. Below we show that ϕ_R can capture a wide range of institutional distortions that affect the efficiency of the economy.

Technology adoption costs:

There are two types of adoption costs. The cost of bringing to the country the production method associated with a capital vintage, Γ_{vt}^e , and the cost incurred by each individual pro-

⁵This specification is similar to Benassy (1996). The assumption that $\mu > \mu'$ ensures that the profits of a individual producer decline with N_v .

ducer to find a distinct application of a production method that is already available, Γ_{vt}^i . We define the former as the extensive and the latter as the intensive adoption costs. Both of these are sunk costs. The extensive cost of adopting vintage v at time t is given by (3.6) while the intensive cost of adoption is given by (3.7).

$$\Gamma_{vt}^e = \frac{\alpha}{\epsilon} \bar{\Psi} (1 + b_e) \left(\frac{Z_v}{Z_t} \right)^{\frac{1+\vartheta}{\mu-1}} \left(\frac{Z_t}{A_t} \right)^{\frac{1}{\mu-1}} Y_t, \text{ where } \vartheta > 0 \quad (3.6)$$

$$\Gamma_{vt}^i = \frac{\mu - 1}{\mu} \bar{\Psi} (1 + b_i) \left(\frac{Z_v}{Z_t} \right)^{\frac{1}{\mu-1}} \left(\frac{Z_t}{A_t} \right)^{\frac{1}{\mu-1}} Y_t. \quad (3.7)$$

In these expressions, A_t is the aggregate level of TFP to be defined below, b_e , b_i , and $\bar{\Psi}$ are constants. The parameters b_e and b_i reflect barriers to adoption for the agent that adapts the technology to the idiosyncrasies of the country or for individual producers that find a profitable use for the technology. $\bar{\Psi}$ is the steady state stock market capitalization to GDP ratio and is included for normalization purposes. The term (Z_v/Z_t) captures the idea that it is more costly to adopt technologies the higher is their productivity relative to the productivity of the frontier technology. The last two terms capture that the cost of adoption is increasing in the market size. We choose these formulations because, just like the adoption cost function in Parente and Prescott (1994), they yield an aggregate balanced growth path.⁶

3.2.3 Factor demands, output, and optimal adoption

The demand for the output produced with vintage v is:

$$Y_v = Y (P_v)^{-\frac{\mu}{\mu-1}}, \text{ where } P = \left(\int_{v \in V} P_v^{-\frac{1}{\mu-1}} dv \right)^{-(\mu-1)}. \quad (3.8)$$

We use the final good as the numeraire good throughout our analysis and normalize its price to $P = 1$. The demand faced by the i^{th} producer of differentiated output associated to vintage v is:

$$Y_{vi} = Y_v \left(\frac{P_{vi}}{P_v} \right)^{-\frac{\mu}{\mu-1}} N_v^{-\frac{\mu-\mu'}{\mu-1}}, \text{ where } P_v = N_v^{\frac{\mu-\mu'}{\mu-1}} \left(\int_0^{N_v} P_{vi}^{-\frac{1}{\mu-1}} di \right)^{-(\mu-1)}. \quad (3.9)$$

Note that all producers of differentiated outputs associated to a given vintage face the same demand and have access to the same technology. As a result, they will charge the same

⁶It could of course be the case that the linearity in the adoption cost function is violated for some particular technology for some particular country, without necessarily violating balanced growth, but to the extent that we are documenting adoption lags across many technologies this is perhaps not so critical.

price which is given by a constant markup, μ , times the marginal cost of production:

$$P_{vi} = \mu \left[\frac{1}{Z_v} \left(\frac{\phi_R R_v}{\alpha} \right)^\alpha \left(\frac{W}{1-\alpha} \right)^{1-\alpha} \right], \text{ for } i \in [0, N_v]$$

where R_v is the rental price of a unit of capital that embodies vintage v , ϕ_R is a tax on capital, and W is the wage rate. From (3.9), this implies that

$$P_v = N_v^{-(\mu'-1)} P_{vi}.$$

The revenue share of capital is α and labor exhaust the remaining revenue. This implies that the total demand faced by the producer of the capital good that embodies vintage v is:

$$\begin{aligned} K_v &= \int_0^{N_v} K_{vi} di = \frac{\alpha P_v Y_v}{\phi_R R_v} \\ &= Y \left(\frac{Z_v}{\mu} \right)^{\frac{1}{\mu-1}} N_v^{\frac{\mu'-1}{\mu-1}} \left(\frac{(1-\alpha)}{W} \right)^{\frac{1-\alpha}{\mu-1}} \left(\frac{\alpha}{\phi_R R_v} \right)^\epsilon, \text{ where } \epsilon \equiv 1 + \frac{\alpha}{\mu-1}. \end{aligned}$$

The supplier of each capital good takes as given the number of differentiated output producers but recognizes that the rental price he charges for the capital good, R_v , affects the price of the output associated with the capital good and, therefore, its demand, Y_v . The price elasticity of demand she faces, ϵ , is constant. As a result, the profit maximizing rental price equals a constant markup times the marginal production cost of a unit of capital, which we assume is equal to a unit of final output.

Because of the durability of capital and the reversibility of its production process, the per-period marginal production cost of capital is the user-cost of capital. Thus, the rental price that maximizes the profits accrued by the capital good producer is

$$R_v = R = \frac{\epsilon}{\epsilon-1} \tilde{\gamma}, \quad (3.10)$$

where $\frac{\epsilon}{\epsilon-1}$ is the constant gross markup factor.

Aggregate representation:

Our model has the following aggregate representation of production:

$$Y = AK^\alpha L^{1-\alpha}, \text{ where } K \equiv \int_{-\infty}^t K_v dv, L \equiv \int_{-\infty}^t L_v dv \quad (3.11)$$

Aggregate TFP, A , can be expressed as

$$A = \left[\int_{-\infty}^{t-D} \left(N_v^{\mu'-1} Z_v \right)^{\frac{1}{\mu-1}} dv \right]^{\mu-1} \quad (3.12)$$

Optimal adoption:

The flow profits accrued by producers of differentiated outputs associated with vintage v are equal to

$$\pi_{vi} = \frac{\mu - 1}{\mu} P_{vi} Y_{vi} = \frac{\mu - 1}{\mu} Y \left(\frac{Z_v}{A} \right)^{\frac{1}{\mu-1}} N_v^{-\frac{\mu-\mu'}{\mu-1}}$$

The market value of each differentiated output producer equals the present discounted value of the flow profits. That is,

$$M_{vi,t} = \int_t^\infty e^{-\int_t^s \tilde{r}_{s'} ds'} \pi_{vis} ds = \frac{\mu - 1}{\mu} \left(\frac{Z_v}{Z_t} \right)^{\frac{1}{\mu-1}} \left(\frac{Z_t}{A_t} \right)^{\frac{1}{\mu-1}} N_v^{-\frac{\mu-\mu'}{\mu-1}} \Psi_t Y_t, \quad (3.13)$$

where

$$\Psi_t = \left(\frac{\mu - 1}{\mu} + \frac{\alpha}{\epsilon} \right) \int_t^\infty e^{-\int_t^s \tilde{r}_{s'} ds'} \left(\frac{A_t}{A_s} \right)^{\frac{1}{\mu-1}} \left(\frac{Y_s}{Y_t} \right) \left(\frac{N_v(s)}{N_v(t)} \right)^{-\frac{\mu-\mu'}{\mu-1}} ds \quad (3.14)$$

is the stock market capitalization to GDP ratio.

Optimal adoption implies that, every instant, the value of becoming a user of a technology vintage v does not exceed the intensive cost of adoption. That is, for all vintages, v , that are adopted at time t

$$\Gamma_v^i \geq M_{vi}. \quad (3.15)$$

Thus, in equilibrium

$$N_v = \left(\frac{\Psi_t}{\bar{\Psi}_t(1 + b_i)} \right)^{\frac{\mu-1}{\mu-\mu'}}. \quad (3.16)$$

Given N_v , the flow profits that the capital goods producer of vintage v earns are equal to

$$\pi_v = \frac{\alpha}{\epsilon \phi_R} P_v Y_v = \frac{\alpha}{\epsilon} N_v^{\frac{\mu'-1}{\mu-1}} \left(\frac{Z_v}{A} \right)^{\frac{1}{\mu-1}} Y \quad (3.17)$$

The market value of each capital goods supplier equals the present discounted value of the flow profits. That is,

$$M_{v,t} = \int_t^\infty e^{-\int_t^s \tilde{r}_{s'} ds'} \pi_{vs} ds = \frac{\alpha}{\epsilon \phi_R} N_v^{\frac{\mu'-1}{\mu-1}} \left(\frac{Z_v}{Z_t} \right)^{\frac{1}{\mu-1}} \left(\frac{Z_t}{A_t} \right)^{\frac{1}{\mu-1}} \Psi_t Y_t. \quad (3.18)$$

Optimal adoption implies that, every instant, all the vintages for which the value of the firm that produces the capital good is at least as large as the adoption cost will be adopted. That is, for all vintages, v , that are adopted at time t

$$\Gamma_v^e \geq M_v \quad (3.19)$$

This holds with equality for the best vintage adopted if there is a positive adoption lag.

The adoption lag that results from this condition equals

$$D_v = \max \left\{ \frac{\mu - 1}{\gamma \theta} \left[\ln(1 + b_e) + \ln \phi_R - \frac{\mu' - 1}{\mu - 1} \ln N_v + (\ln \bar{\Psi} - \ln \Psi) \right], 0 \right\} \equiv D$$

and is constant across vintages, v .⁷ The lag with which new vintages are adopted is increasing in the adoption costs, b_e , and the tax wedge, ϕ_R , is decreasing in N_v and in the deviation of the stock market to output ratio from its steady state level. As shown in equation (3.16), the number of producers that develop distinct uses for technology vintage v , N_v , declines with the intensive cost of adoption, b_i , and increases with the deviation of the stock market to output ratio from the balance growth level.⁸

Conversely, there are several significant factors that do not influence the adoption decisions. First, given the specifications of the production function and the costs of adoption, the market size symmetrically affect the benefits and costs of adoption at both the intensive and extensive margins. Hence, variation in market size does not affect the timing of adoption, D , and the number of producers that use a new vintage, N_v . By the same token, the adoption margins are not affected by the productivity of technology at time zero, Z_0 . Second, since on the balanced growth path $\Psi = \bar{\Psi}$, the steady-state adoption lags and number of producers do not depend on the stock market to output ratio.

These observations together with equation (3.12) help us understand what drives aggregate TFP in this model. Three factors can drive cross-country differences in TFP: The adoption lag, the number of producers that adopt each technology vintage, and the normalized productivity level of the initial vintage. Note that, Z_0 affects directly aggregate TFP but, as mentioned above, has no effect through D or N_v . The costs of adopting new technologies affect TFP because they influence the range of technologies available for production and how many different applications are developed. Finally, the tax wedge (and other related frictions) only affect aggregate TFP through their effect on the adoption decisions.

⁷In what follows we focus on the interior case where $\Gamma_{\tau} \leq M_{\tau}$.

⁸Note that ϕ_R does not affect the intensity of adoption as measured by the number of producers that adopt a new vintage. That is the case because, from the perspective of the potential producers of differentiated outputs, ϕ_R only affects aggregate demand. Aggregate demand, in turn, has a symmetric effect on the costs and benefits of adopting the vintage for the differentiated output producers. Instead, corporate income taxes (or expropriation risk) also affect the profit margin net of taxes. This asymmetric effect would affect the number of producers that adopt the new vintage.

3.3 Diffusion of the new technology

We define the equilibrium of this economy in Appendix 3.7. In what follows, we focus on the balanced growth path of the economy.⁹ Along the balanced growth path, adoption lags, D , are constant, the number of adopters that adopt each vintage once it is available in the country is constant and equal to N , and the economy grows at a constant rate equal to $\gamma/(1-\alpha)$.¹⁰

So far, we have derived expressions for output and capital at the vintage level. However, because of the nature of available data, we are interested in the total demand for capital goods and the output produced with the production methods that make up the new technology $\tau = n$. We can express output produced with technology τ in the following Cobb-Douglas form

$$Y_\tau = A_\tau K_\tau^\alpha L_\tau^{1-\alpha}, \quad (3.20)$$

where

$$K_\tau \equiv \int_{v \in V_\tau} K_v dv, \quad L_\tau \equiv \int_{v \in V_\tau} L_v dv, \quad (3.21)$$

$$\text{and } A_\tau = \left(\int_{V_\tau} \left(N_v^{\mu'-1} Z_v \right)^{\frac{1}{\mu-1}} dv \right)^{\mu-1}. \quad (3.22)$$

Substituting in for Z_v , and recognizing that, along the balanced growth path, the adoption lag and the number of differentiated producers are constant and equal to D and N_n , respectively, we can express the endogenous level of TFP for technology $\tau = n$ at time t as

$$A_n = \left(\frac{\mu-1}{\gamma} \right)^{\mu-1} \underbrace{N_n^{\mu'-1}}_{\text{intensity of adoption}} \underbrace{Z_v}_{\text{embodiment effect}} \underbrace{\left[1 - e^{-\frac{\gamma}{\mu-1}(t-D-v)} \right]^{\mu-1}}_{\text{variety effect}}. \quad (3.23)$$

The path of the new technology TFP is driven by the adoption margins. First, there are efficiency gains from the number of producers that adopt a given vintage. This affects the level of technology through the ‘intensity of adoption’ term in (3.23). The trend in TFP is driven by the economy-wide adoption of new, more productive, vintages. The adoption lag determines the best vintage adopted and affects the level of TFP through the ‘embodiment effect’ term in (3.23). Finally, adoption lags also drive the curvature of A_n at a given moment in time. The marginal productivity gain from adopting new vintages decreases as more vintages are adopted. Because the adoption lag determines how far in the adoption process a

⁹The transitional dynamics of the model are similar to the one described in the working paper version of Comin and Hobjin (2009).

¹⁰Of course, D and N could differ across countries.

country is at time t , they affect the evolution of the slope through the ‘variety effect’. Graphically, this effect is captured by horizontal shifts in the path of A_n as adoption lags vary. We shall use this result to identify the adoption lags in the data.

These properties of the path for the level of TFP for technology τ affects the output and capital associated with the technology through its effect on the marginal cost of production of the technology-specific output measured by the price P_τ .

$$P_\tau = \frac{\mu}{A_\tau} \left(\frac{W}{1-\alpha} \right)^{1-\alpha} \left(\frac{\phi_R R}{\alpha} \right)^\alpha. \quad (3.24)$$

3.3.1 Empirical application

Our goal is to estimate the intensity of adoption and adoption lags for the different technology-country pairs in our data set. We extend the results above by allowing multiple sectors, each adopting a new technology.¹¹ We do that with a nested CES aggregator, where $\frac{\theta}{\theta-1}$ reflects the between sectors elasticity of demand and $\frac{\mu}{\mu-1}$ is, just as in the one-sector model, the within sector elasticity of demand. We allow $\frac{\theta}{\theta-1}$ to vary across sectors. Further, we allow the growth rate of embodied technological change, γ_τ , and the invention date, \underline{v}_τ , to vary across technologies. We denote the technology measures for which we derive reduced form equations by $m_\tau \in \{y_\tau, k_\tau\}$. Small letters denote logarithms.

These modifications yield the following demand for technology τ output

$$y_\tau = y - \frac{\theta}{\theta-1} p_\tau. \quad (3.25)$$

Combining that with the intermediate goods price (3.24)

$$p_\tau = -\alpha \ln \alpha - a_\tau + (1-\alpha)(y-l) + \alpha r + \alpha \ln \phi_R, \quad (3.26)$$

we obtain the reduced form equation (3.27) for y_τ .

$$y_\tau = y + \frac{\theta}{\theta-1} [a_\tau - (1-\alpha)(y-l) - \alpha r + \alpha \ln \alpha - \alpha \ln \phi_R] \quad (3.27)$$

Similarly, we obtain the reduced form equation for k_τ by combining the log-linear capital demand equation with (3.25) and (3.26),

$$k_\tau = y + \frac{1}{\theta-1} [a_\tau - (1-\alpha)(y-l) - \alpha r + \alpha \ln \alpha - \alpha \ln \phi_R] + \ln \alpha - r - \ln \phi_R. \quad (3.28)$$

These expressions depend on the intensive margin and adoption lag D_τ through their effect on the productivity term, a_τ . Comin and Hobijn (2010) show that, to a first order approxima-

¹¹Comin and Hobijn (2008) derive the multi-sector version of a similar model in detail.

tion,

$$a_\tau \approx (\mu' - 1)n_\tau + z_{v_\tau} + (\mu - 1) \ln(t - T_\tau) + \frac{\gamma_\tau}{2} (t - T_\tau), \quad (3.29)$$

where $T_\tau = \underline{v}_\tau + D_\tau$ is the time when the technology is adopted. In this approximation, the growth rate of embodied technological change, γ_τ , only affects the linear trend in a_τ .¹² Substituting this into (3.25) and (3.28) yields the following reduced form equation

$$m_\tau = \beta_1 + y + \beta_2 t + \beta_3 ((\mu - 1) \ln(t - T_\tau) - (1 - \alpha)(y - l)) + \varepsilon_\tau, \quad (3.30)$$

where ε_τ is the error term. The reduced form parameters are given by the β 's. Note that, the homothetic nature of the production function implies that the coefficient of aggregate demand, y , is equal to one.

According to our theoretical model, the intercept β_1 is given by the following expression which depends on both the intensive and extensive margins of adoption,

$$\beta_1^y = \frac{\theta}{\theta - 1} \left[((\mu' - 1)n_\tau + z_{v_\tau}) - \frac{\gamma}{2} T - \alpha(r + \ln \phi_R - \ln(\alpha)) \right], \quad (3.31)$$

$$\beta_1^k = \frac{1}{\theta - 1} \left[((\mu' - 1)n_\tau + z_{v_\tau}) - \frac{\gamma}{2} T - \alpha(r + \ln \phi_R - \ln(\alpha)) \right] - r - \ln \phi_R. \quad (3.32)$$

We define the intensive margin of adoption of technology τ in country j as

$$\Delta_{j\tau} = ((\mu' - 1)n_\tau + z_{v_\tau}) - \alpha(r + \ln \phi_R - \ln(\alpha)). \quad (3.33)$$

Three different factors affect the intensive margin of adoption: the number of adopters of the technology, n_τ , the normalized productivity level of the technology, z_{v_τ} , and the distortions in the price of capital, $r + \ln \phi_R$.

3.3.2 Identification and estimation procedure

We use the reduced form equation in (3.30) to identify the adoption lags and the intensity of adoption. To this end, we assume that preference parameters (ρ) and technology parameters other than adoption costs (i.e., θ , μ , μ' , γ , α , and z_{v_τ}) are the same across countries. This implies that the equilibrium interest rate (r) is also the same across countries.¹³ The distortions that affect the efficiency of the economy (ϕ_R) and the adoption cost parameters (b_e and b_i) can vary across countries. As a result, the adoption lags (D_τ) and the number of producers that adopt a technology (n) can also differ across countries.

¹²Intuitively, when there are very few vintages in V_τ the growth rate of the number of vintages, i.e. the growth rate of $t - T_\tau$, is very large and it is this growth rate that drives growth in a_τ through the variety effect. Only in the long-run, when the growth rate of the number of varieties tapers off, the growth rate of embodied productivity, γ_τ , becomes the predominant driving force over the variety effect.

¹³Our identification strategy is unaffected if interest rates, r , and the normalized productivity level for the new technologies, z_{v_τ} , varied across countries.

These assumptions impose some cross-country parameter restrictions. Since the intercept term, β_1 , depends on n , D_τ and ϕ_R , it can vary across countries. The trend-parameter, β_2 , just depends on α and γ_τ , so it is assumed to be constant across countries.¹⁴ β_3 only depends on the technology parameter, θ , and is therefore assumed to be constant across countries. We do not estimate μ and α . Instead, we calibrate $\mu = 1.3$, based on the estimates of the markup in manufacturing from Basu and Fernald (1997), and $\alpha = 0.3$ consistent with the post-war U.S. labor share.¹⁵

The parameter β_1 is a technology-country specific constant. Therefore, it can be identified by a technology-country fixed effect. Once we have an estimate of β_1 , we still need an estimate of the adoption lags to obtain an estimate of the intensive margin of adoption. We follow Comin and Hobijn (2010) and identify the adoption lags through the non-linear trend component in equation (3.30), which reflects the variety effect. Intuitively, after controlling for the observables such as GDP or labor productivity, only the adoption lag affects the curvature of m_τ . That implies that, *ceteris paribus*, if we see two countries one with a steeper diffusion curve than the other at a given point in time, this means that the former started adopting the technology later.

For each technology, we report the intensive margin measures relative to the U.S. Note that our estimates of β_1 in (3.31) and (3.32) are not directly comparable for technologies measured with capital and output variables. To construct comparable measures of intensive margin of adoption, we need to eliminate the differential effect of ϕ_R that appears on the technologies measured using capital. To this end, we regress, for each country, the intercepts in (3.31) and (3.32) on a dummy variable that takes value of 1 if the technology is measured in capital units. The coefficient on the dummy captures the differential effect of ϕ_R . We then subtract the dummy coefficient from β_1^k , and correct for the factors $\theta/(\theta - 1)$ and $1/(\theta - 1)$ to obtain the measure of the intensive margin in (3.33).^{16,17}

This identification strategy assumes that the underlying curvature of the diffusion curves is the same across countries. This assumption would be violated, for example, if the efficiency of an economy increased over time non-linearly inducing a similar pattern in the

¹⁴The output elasticity of capital is one minus the labor share in our model. Gollin (2002) provides evidence that the labor share is approximately constant across countries.

¹⁵As argued by Comin and Hobijn (2008), the estimates of the adoption parameters are very robust to alternative calibrations of these parameters.

¹⁶The parameter θ is computed as the average (across technologies) implied estimate from the estimates of β_3 and it is equal to 1.31 for the homothetic case and 1.14 for the non-homothetic case.

¹⁷An alternative approach could be to take advantage of the fact that we have two measures of railways that are measured as output (passengers and freight) and one that is measured as capital (rail lines). Under the plausible assumption that the average intercept of the output measures (passenger and freight measures) corresponds to the rail line measure used in the country, we can back out the additional effect of ϕ_R on the capital measures of technology. Then, we can subtract this additional effect from the other capital measures and can construct comparable measures of the intensive margin. The results we obtain using this procedure are similar to the ones reported in the main text.

technology measure. Of course, a priori, there is no reason why the distortions in the economy evolve to induce such a specific pattern of adoption rather than affecting the trend or adding noise to the evolution of technology. Nevertheless, Comin and Hobijn (2010) take seriously this hypothesis and test formally the identification assumption by allowing β_3 to vary across countries. Then, they see how often they can reject the null that the unrestricted and the restricted estimates are the same (i.e. $\hat{\beta}_3^u = \hat{\beta}_3$). They find that they cannot reject the null in two thirds of the technology-country pairs considered. Further, the estimated adoption lags in the restricted and unrestricted specifications are very highly correlated suggesting that, effectively, the deviations from a constant curvature pattern are not quantitatively important.

Because the adoption lag is a parameter that enters non-linearly in (3.30) for each country, estimating the system of equations for all countries together is practically not feasible. Instead, we take a two-step approach. We first estimate equation (3.30) using only data for the U.S. This provides us with estimates of the values of β_1 and D_τ for the U.S. as well as estimates of β_2 and β_3 that should hold for all countries. In the second step, we separately estimate β_1 and D_τ , using (3.30) conditional on the estimates of β_2 and β_3 based on the U.S. data, for all the countries in the sample besides the U.S.

Besides practicalities, this two-step estimation method is preferable to a system estimation method for two other reasons. First, in a system estimation method, data problems for one country affect the estimates for all countries. Since we judge the U.S. data to be most reliable, we use them for the inference on the parameters that are constant across countries. Second, our model is based on a set of stark neoclassical assumptions. These assumptions are more applicable to the low frictional U.S. economic environment than to that of countries in which capital and product markets are substantially distorted. Thus, if we think that our reduced form equation is more likely to be mis-specified for some countries other than the U.S., including them in the estimation of the joint parameters would affect the results for all countries.

We estimate all the equations using non-linear least squares. Since we estimate β_3 for the U.S., this means that the identifying assumption that we make is that the logarithm of per capita GDP in the U.S. is uncorrelated with the technology-specific error, ε_τ . However, because of the cross-country restrictions we impose on β_3 , the risk of simultaneity bias is not a concern for all the other countries in our sample.

3.3.3 Non-homotheticities

One general concern in structural estimation exercises is model mis-specification. In the context of our model, the place where this concern probably is more relevant is in the elasticity of technology with respect to income. For our model to have a balanced growth path,

the production functions need to be homothetic. This implies that the income elasticity of technology measures is equal to one. In this subsection, we explore the implications for the estimated equations (3.30) if we replace the original production function (3.3) for a more general specification that allows for non-homotheticities.

Consider the following non-homothetic version of (3.3)

$$Y = \frac{1}{\bar{\theta}} \left(\sum_{\tau} \theta_{\tau} Y_{\tau}^{\frac{1}{\theta_{\tau}}} \right)^{\bar{\theta}},$$

where $\bar{\theta}$ is the long-run average of θ_{τ} over τ so that constant returns to scale are guaranteed (in the long-run),

$$Y_{\tau} = Y^{\frac{(\bar{\theta}-1)\theta_{\tau}}{\bar{\theta}(\theta_{\tau}-1)}} P_{\tau}^{\frac{-\theta_{\tau}}{\theta_{\tau}-1}}.$$

This yields the following reduced form equation

$$m_{\tau} = \beta_1 + \beta_y y + \beta_2 t + \beta_3 ((\mu - 1) \ln(t - T_{\tau}) - (1 - \alpha)(y - l)) + \varepsilon_{\tau}, \quad (3.34)$$

which differs from (3.30) in that β_y is not restricted to be equal to 1. It will be greater than one if $\bar{\theta} > \theta_{\tau}$ and smaller otherwise.

As in the homothetic reduced form equation (3.30), we estimate the coefficients in (3.34) that depend on technological parameters (β_2 , β_3 and β_y) for the U.S. and then impose these estimates in the other countries. The only difference is that now we estimate an additional parameter, β_y .

Using the U.S. estimate of β_y has several advantages. First, reliable data on U.S. real GDP is available since 1820. This long time-span facilitates the identification. It also ensures that the estimate is based on the various stages of development that the U.S. has gone through over the last two centuries. As a result, it should capture reasonably well the effect of aggregate demand on technology diffusion for countries at different stages of development. Most importantly, since for each technology-country pair, the intensive margin in our model is constant, by exploiting the time series dimension to identify β_y we avoid the bias due to the potential cross-country correlation between the intensive margin of adoption and income.

Identifying β_y using only the time series dimension is not trivial for two reasons. First, since γ_{τ} may differ across technologies, when estimating β_2 we must include a time trend for each technology. Since in the long run log GDP is approximately linear, GDP and time are co-linear over very low frequencies. Second, most technologies in our sample are embodied in capital goods. The high cyclical and volatility of investment shall induce a high estimate of β_y . This estimate however, would capture the cyclical properties of capital rather than the effect of aggregate demand on our technology measures at lower frequencies.

To overcome these two difficulties, we use a Hodrick-Prescott (HP) filter to decompose

log of real GDP into a high frequency component and a ‘trend’. As is well known,¹⁸ trends that result from HP filters have significant fluctuations at medium term frequencies.¹⁹ By exploiting this variation, we could in principle identify β_y . Thus, we introduce separately the high frequency component and the ‘trend’ in log real GDP and estimate a different elasticity for each.^{20,21} To identify β_y in our data, we estimate simultaneously the system of equations (3.34) for all the U.S. technologies.

Though in our model the parameters that determine the intensive margin are fixed, it is instructive to consider how will our estimate of β_y be affected if they vary in the data. The first thing to realize is that institutions, human capital and other factors that may affect the costs of adoption at the intensive margin in the U.S. have changed very slowly. As a result, this may have effects on our technology measures and on GDP only at very low frequencies.²² These frequencies are so low that most of these effects will be captured by the time trends and will have little effect on the estimate of β_y .

Note further that, since naturally changes in institutions or variables that affect adoption costs would induce a positive co-movement between GDP and technology, the small bias induced on β_y would be upwards. As a result, when identifying the intensive margin we would filter too much aggregate demand inducing a lower cross-country dispersion in the intensive margin. That is, the bias in β_y would bias downwards the importance of the intensive margin for development.

3.4 Results

We consider data for 166 countries and 15 major technologies, that span the period from 1820 through 2003. The technologies can be classified into 6 categories; (i) transportation technologies, consisting of steam and motor-ships, passenger and freight railways, cars, trucks, and passenger and freight aircraft; (ii) telecommunication, consisting of telegraphs, telephones, and cellphones; (iii) information technology, consisting of PCs and Internet users; (iv) medical technology, namely MRI scanners; (v) steel produced using blast oxygen furnaces; (vi) electricity.

The technology measures are taken from the CHAT data set.²³ Real GDP and population data are from Maddison (2007). Appendix 3.6 contains a brief description of each of the 15 technology variables used. For our estimation, we only consider country-technology combi-

¹⁸E.g. Comin and Gertler (2006).

¹⁹In the context of this exercise, medium term frequencies means cycles with periods between 8 years and infinity.

²⁰We have obtained very similar results when including only the ‘trend’ component of GDP.

²¹As conjectured, the short-run elasticity is significantly higher, capturing adjustment over the business cycle.

²²One example would be expansion of high school enrollment in the U.S. between 1910 and 1940.

²³See Comin, Hobijn and Rovito (2009).

nations for which we have more than 10 annual observations. There are 1298 such pairs in our data. The third column of Table 3.1 lists, for each technology, the number of countries for which we have enough data.

We follow Comin and Hobijn (2010) and analyze only the technology-country pairs for which we have plausible and precise estimates of the adoption lags. These are estimates with an adoption date later than the invention year plus 10, and with small standard errors.²⁴ We have plausible and precise estimates for 837 technology-country pairs, which represent approximately two thirds of the total.

3.4.1 Estimated Intensive Margin

3.4.1.1 Dispersion

Table 3.1 presents the descriptive statistics of our estimates of the intensive margin of adoption relative to the United States. The fifth column reports the cross-country average. This statistic is negative for all the technologies but ships and freight railways. This means that for all the technologies but these two, the U.S. intensive margin is higher than the average in our sample.

Column 6 reports the cross-country standard deviation of the intensive margin of adoption by technology while column 10 reports the inter-quartile range.²⁵ The conclusions are robust to using any of these two measures of dispersion so, for brevity, we base our discussion on the cross-country standard deviation. The dispersion in the intensive adoption margin varies significantly by technology ranging from 0.33 (rail passenger) to 0.89 (blast oxygen steel). To have a benchmark, we report, in column 12, $(1 - \alpha)$ times the cross-country standard deviation in log per capita income in 2000 for the same sample of countries for which we have plausible and precise estimates of the adoption margins for each technology.²⁶ The ratio of the standard deviation in the intensive margin to the standard deviation of per capita income is on average 1.2, but it ranges significantly across technologies: from 0.4 (for railways and electricity)²⁷ to 1.72 for (blast oxygen steel). This suggests that technology-specific factors are important drivers of the cross-country variation in the intensive margin.

²⁴Comin and Hobijn (2010) discuss several reasons for obtaining implausible estimates. The 10 year cut off point for plausible estimates is to allow for inference error. The cutoff we use in the standard error of the estimate of T_τ for it to be precise is $\sqrt{2003 - \underline{v}_\tau}$. This allows for longer confidence intervals for older technologies with potentially more imprecise data. Including imprecise estimates in our analysis does not affect the conclusions.

²⁵The inter-quartile range is defined as the difference between the adoption intensities in countries in the 75 and 25 percentiles.

²⁶We scale down log income by the factor $(1 - \alpha)$ because, in addition to the effect through TFP, the extensive margin also induces a higher capital-labor ratio which also affects the level of per capita output.

²⁷In particular, the ratio is 0.4 for passenger-Km moved by railways.

3.4.1.2 Evolution

The evolution of the intensive margin may help us understand the dynamics of growth over the last two centuries. At first sight, the cross-country dispersion of the intensive margin seems to be uncorrelated with the invention date. We test this observation in the first column of Table 3.2, which reports the estimates from a regression of the average intensive margin, the cross-country dispersion and inter-quartile range on the year of invention. The regressions confirm an insignificant relationship between the dispersion of the intensive margin of adoption and the invention date of the technology.

One possible reason for the stationary nature of the dispersion of the intensive margin could be that our estimates of this dispersion for early adopters of early technologies are smaller than they should be because of the effect of the replacement of dominated technologies. The argument would be as follows. Some of the early technologies were dominated by superior technologies a long time ago. For early adopters such as the United States, the level of our measures for their intensive margin has been declining. As a result, the estimated intercept is lower than in late adopters, where these technologies have been dominated more recently. Under this hypothesis, cross-country dispersion in the intensive margin with which early technologies were adopted might have been larger than it now appears in the statistics. The alternative, of course, is that the stationarity of the cross-country dispersion in the intensive margin of adoption correctly represents the empirical facts.

To disentangle these two hypotheses, we re-estimate our baseline regression for the old technologies using only data up to 1939, when presumably none of the early technologies was yet obsolete. Table 3.3 compares the estimates of the average, standard deviation and inter-quartile range of the intensive margin of adoption for the countries for which we can precisely estimate the diffusion equation using data up to 1939.²⁸ The results vary a little by technology but, overall, there are no significant increases in the dispersion of the intensive margin of adoption when we restrict the sample to the pre-1939 period. This implies that the dispersion in the intensive margin of adoption for early technologies is not driven by the fact that early technologies have been dominated sooner in countries that adopted them first.

The lack of convergence in the intensive margin contrasts with the evolution of the dispersion in adoption lags. Column 11 of Table 3.1 reports the cross-country standard deviation of the adoption lags for each technology. It is evident at first sight that dispersion in adoption lags has decreased monotonically over the last two centuries. That is, the difference in adoption lags across countries has been much smaller for technologies invented in the recent past than for those invented in the more distant past. Column 3 in Table 3.2 tests this observation formally. The negative relationship between cross-country dispersion in adoption lags and invention date is statistically significant. In particular, technologies invented

²⁸For obvious reasons, we only report the estimates for technologies invented before 1900.

ten years later have a dispersion that is two years smaller.

3.4.1.3 Variance decomposition

Understanding the sources of variation in the intensive margin of adoption is beyond the scope of this paper. However, we can explore whether this variation is mostly driven by country effects or by technology effects. Specifically, let $\Delta_{j\tau}$ measure the intensive margin of country j in technology τ . We can decompose $\Delta_{j\tau}$ as follows

$$\Delta_{j\tau} = \Delta_j + \Delta_\tau + u_{i\tau}, \quad (3.35)$$

where Δ_j is a country fixed effect, Δ_τ is a technology fixed effect and $u_{i\tau}$ is an error term. The first line of Table 3.4 examines the contribution of the country fixed effects alone. That is, the R^2 when estimating (3.35) with only country fixed effects. Country-specific factors explain approximately 44% of the variation in the intensive margin of adoption. In the second row, we calculate the contribution of technology-specific fixed effects in an analogous manner and find that they explain 34% of the variation. The last row of Table 3.4 shows that country and technology specific factors jointly account for approximately 77% of the variation in the intensive margin. Of this total variation, 43% can be directly attributed to country effects, 33% can be directly attributed to technology effects, and the remaining 1% is due to the covariance between these effects which differs from zero because our panel is unbalanced.

The drivers of the variance in the intensive margin differ quite a bit from the drivers of the variance for the extensive. As shown in Comin and Hobijn (2010), the technology fixed effects account for 65% of the variance in the adoption lags. In contrast, country fixed effects are the main factor when accounting for the variance of the intensive margin of adoption.

3.4.1.4 Non-homotheticities

As discussed in the identification section, we want to explore an alternative approach to dealing with the endogeneity of income in estimating the intensive margin. So far we have used restrictions imposed by the assumption of balanced growth on the elasticity of our technology measures with respect to GDP (equation 3.30). We would like to explore how robust the estimates of the intensive margin are to using other identification schemes. In particular, the alternative scheme consists in identifying the effect of aggregate demand on technology by using the time series variation in GDP in the United States and then imposing the estimate of the U.S. income elasticity when estimating the reduced form equation (3.34) for the other countries.

As discussed above, when estimating the income elasticity for the United States, we want to distinguish between the short and long run income elasticities, since the former is likely

to capture cyclical variation in the demand for investment goods. This presumption is confirmed by our estimates. We find that the long-run income elasticity is 2.2, while the short-run is 6.6. Both estimates are very precise.

The additional flexibility allowed in the model comes at the cost of a lower precision in the estimates of the adoption lag for two U.S. technologies: ships and electricity. This creates the minor problem of having a less precise estimate for the United States in the intensive margin. Since we do not want to have as baseline intensity for the technology an imprecise estimate, for these two technologies, we take France as a reference rather than the United States. Which country is taken as baseline is irrelevant for computing the cross-country dispersion measures. However, the mean intensive margin of adoption may be affected; therefore, the average intensive margin is not directly comparable with the homothetic case.

We obtain plausible and precise estimates for 738 country-technology pairs. This represents 57% of our sample. The statistics for the estimated adoption lags and the intensive margin are reported in Table 3.5. On average, adoption lags are slightly smaller (35 years versus 45 years) when we allow for non-homotheticities. The cross-country dispersion in adoption lags is also slightly smaller (33 years versus 39 years) under non-homotheticities. However, the estimates of the adoption lags under both identification strategies are very highly correlated (see column 3 in Table 3.6). The average correlation across technologies is 0.91, and ranges from 0.79 for electricity to 0.98 for cellphones, MRIs and the Internet.

The cross-country dispersion in both measures of the intensive margin quite similar (0.72 versus 0.68). Column 4 in Table 3.6 shows the correlation between the estimates of the intensive margin of adoption in the homothetic and non-homothetic cases. On average the correlations are high, approximately 0.87. By technology, they range from 0.6 for freight railways to 0.97 for passenger aviation and cellphones.

3.4.1.5 Correlation with per capita income

Before using our model to conduct a development accounting exercise, it is revealing to explore the correlation between per capita income and the intensive margin of adoption for each technology. Table 3.7 reports these statistics for both the homothetic and non-homothetic estimates. The correlations are sizable. In the homothetic case the average correlation across technologies is 67% and in the non-homothetic case it is 64%. We find some variation across technologies. The correlation of the intensive margin with per capita income seems to be lower for the earlier technologies, especially for ships and railways. Contrary to the perception that information technologies may be closing the technological divide between rich and poor countries, we find that the intensive margin of these technologies (i.e. PCs, cellphones, Internet) present quite high correlations with per capita income.

Table 3.7 also reports the correlation between the adoption lags and per capita income.

As shown by Comin and Hobijn (2010) for the homothetic case, the correlation is also fairly high, approximately -46%. In the last column, we show that there is also a significant correlation between the adoption lags in the non-homothetic case and per capita income though slightly lower than in the homothetic case (-30%).

3.4.2 Development accounting

Next, we investigate how the estimated differences in the intensive margin of adoption translate into cross-country differences in per capita income. To answer this question, we have to (i) aggregate the technology-level estimates of the intensive margin to an economy-wide measure of the intensive margin, and (ii) compute the effect of the aggregate intensive margin on per capita income. Both of these computations require the use of a model that maps individual technologies into aggregate productivity. We naturally draw from the equilibrium relationships of the model presented above.

Aggregate production, Y , can be expressed as

$$Y = AK^\alpha L^{1-\alpha}, \quad \text{where } K \equiv \int_{-\infty}^t K_v dv, \quad L \equiv \int_{-\infty}^t L_v dv. \quad (3.36)$$

Aggregate TFP, A , is given by

$$A = \left[\sum_{\tau \in \{0, n\}} (A_\tau)^{\frac{1}{\mu-1}} \right]^{\mu-1} = \left(\frac{\mu-1}{\gamma} \right)^{\mu-1} N^{\mu-1} Z_{\underline{v}} e^{\gamma(t-D-\underline{v})}. \quad (3.37)$$

Aggregate TFP depends on the adoption lag, D , the number of producers that adopter each technology vintage, N , and the normalized level of productivity, $Z_{\underline{v}}$. The adoption lag affects aggregate TFP because a higher D reduces the productivity embodied in the best technology vintage available for production. The number of adopters per vintage, N , affects TFP because their outputs are imperfect substitutes and there are efficiency gains from a greater variety of outputs.

Substituting (3.37) into (3.36) and noting that $\phi_R K R = \alpha Y$ yields the following expression for labor productivity:

$$\frac{Y}{L} = A^{\frac{1}{1-\alpha}} \left(\frac{K}{Y} \right)^{\frac{\alpha}{1-\alpha}} = \left[\left(\frac{\mu-1}{\gamma} \right)^{\mu-1} N^{\mu-1} Z_{\underline{v}} e^{\gamma(t-D-\underline{v})} \right]^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{\phi_R R} \right)^{\frac{\alpha}{1-\alpha}}. \quad (3.38)$$

Taking logs, we obtain:

$$y - l = \bar{c} + \frac{1}{1-\alpha} [(\mu-1)n + z_{\underline{v}} - \alpha(r + \ln \phi_R - \ln(\alpha))] + \frac{\gamma(t-D-\underline{v})}{1-\alpha}$$

where \bar{c} is a constant. The term in squared brackets is equal to the economy-wide intensive margin of adoption. Subtracting the same expression for the U.S. yields

$$(y_j - l_j) - (y_{US} - l_{US}) = \frac{1}{1 - \alpha} \Delta_j - \frac{\gamma^* (D_{US} - D_j)}{1 - \alpha} \quad (3.39)$$

where $(\widehat{y_j - l_j})$, \hat{D}_j and Δ_j denote, respectively, log-labor productivity, the adoption lag and the intensive margin in country j relative to the U.S.²⁹

Using expression (3.39), we can decompose the cross-country variance of labor productivity as

$$1 = \overbrace{\frac{1}{1 - \alpha} \frac{cov[\Delta_j, (\widehat{y_j - l_j})]}{var[(\widehat{y_j - l_j})]}}^{\text{Intensive Mg. Contribution}} - \overbrace{\frac{\gamma}{1 - \alpha} \frac{cov[\hat{D}_j, (\widehat{y_j - l_j})]}{var[(\widehat{y_j - l_j})]}}^{\text{Extensive Mg. Contribution}} \quad (3.40)$$

where $var[X]$ denotes the variance of X , and $cov[X, Y]$ denotes the covariance between X and Y . The first term is the share of the cross-country variance in per-capita income accounted for the intensive margin, while the second term is the share of income differences accounted for the extensive margin.³⁰ Under our model, together these two terms should account for 100% of the cross-country differences in productivity.

To implement the decomposition, we need to compute the aggregate intensive margin, Δ_j . To this end, we assume that the intensive margins we have estimated using our sample of technologies are representative of the average intensive margin of adoption across all the technologies used in production. Under this assumption, the aggregate intensive margin in country j is equal to the average intensive margin in the country across the technologies in our sample.³¹

Figure 3-2 plots the first term in equation (3.39) against log per capita income in 2000 from the Penn World Tables 6.2.³² The thicker dashed line corresponds to the regression line, while the light grey line is the 45°-line. The slope of the regression line is equal to the contribution of the intensive margin in (3.40). We find that the slope of the regression line is 0.44.³³ This implies that the intensive margin of technology adoption accounts for 44% of

²⁹Formally, $(\widehat{y_j - l_j}) \equiv (y_j - l_j) - (y_{US} - l_{US})$ and $\hat{D}_j \equiv (D_j - D_{US})$.

³⁰This variance decomposition follows the decomposition in Klenow and Rodriguez-Clare (1997) and splits the co-variance between the intensive and extensive margins evenly between the two terms.

³¹Formally,

$$\Delta_j \simeq \frac{\sum_{s=1}^{S_j} \Delta_{js}}{S_j}$$

where S_j is the total number of technologies for which we have precise and plausible estimates in country j .

³²Similar results obtain with data from Maddison.

³³The correlation between the two sides is 0.51.

the log per capita GDP differentials observed in the data.

Note that there is an heteroskedastic pattern, as poor countries have more variance in the measure of the intensive margin. This may be explained by the fact that we have fewer observations for poor countries than rich countries, and as a result, our measure of the average intensive margin is more noisy in poor countries.

Finally, we perform a development accounting exercise analogous to the one discussed above for the estimates obtained using a non-homothetic production function. The presence of non-homotheticities prevents the existence of a simple aggregate production function such as we have in our baseline model. To conduct this exercise, we assume, as a first pass, that the aggregate production function does not differ much from the one obtained in the homothetic case. Since our goal is to evaluate the robustness of our findings to alternative identification schemes in the estimation of the intensive margin of technology, it is desirable to use the same production function to draw the aggregate implications of the technology-specific estimates. It turns out that the development accounting exercise for the non-homothetic estimates of the intensive margin yields very similar results to the baseline exercise, as it can be seen in Figure 3-3 The correlation between the intensive margin and income per capita is 0.67, and the coefficient on the regression is 0.52.

To sum up, this section suggests that differences in the intensive margin of adoption account for over 40% of cross-country per capita income differences. The observation that a substantial share of income differences are either caused or amplified by the intensive margin of adoption is robust to the two identification strategies we have followed to deal with the endogeneity of aggregate demand to technology.

3.5 Conclusion

In this paper we have built and estimated a model of technology diffusion and growth. Our model predicts that the diffusion path of individual technologies is determined by the lag with which their different vintages are introduced in the country, the level of aggregate demand and the intensive margin of adoption. Using these predictions and exploiting the panel structure of our data set, we have identified the intensive and extensive margins of adoption for over 800 technology-country pairs that correspond to the diffusion of up to 15 major technologies in 166 countries. The estimates are robust to different strategies used to estimate the elasticity of technology with respect to aggregate demand.

An analysis of the estimates yields significant findings. There is a large cross-country dispersion in the intensive margin, though the dispersion varies significantly across technologies. The cross-country dispersion in adoption lags has declined very significantly over the last two centuries, while we find no such convergence pattern in the intensive margin.

In addition to describing accurately the diffusion patterns of individual technologies, our

model yields a representation of aggregate productivity that allows us to relate income levels to the intensive and extensive margins of technology adoption. We find that approximately 45% of the cross-country variation in per capita income can be attributed to differences in the intensive margin. Comin and Hobijn (2010) reported that differences in the extensive margin account for at least 25% percent of the cross country variation in productivity. Taken together, these results imply that the role of technology is crucial to understanding per capita income differences. In particular, the empirical estimates suggest that 70% of the differences in cross-country income per capita can be explained by differences in technology adoption.

We anticipate that the findings of this paper will stimulate three lines of research. First, we would like to understand better the relative importance of the different drivers of the intensive margin. In particular, we would like to assess what is the role played by adoption costs, institutional constraints that affect the overall efficiency of the economy and distortions that affect the price of capital. Second, we plan to study the underpinnings of the variation in adoption costs which this paper abstracts from. Specifically, our findings beg the question of why adoption costs at both the extensive and intensive margin are so large in developing countries. Finally, the differences we have observed in both the intensive and extensive margins across technologies suggest that maybe the dynamics of technology adoption are important not only to understanding cross-country differences in productivity but to explaining the dynamics of growth over the last two centuries. In particular, we intend to explore whether the dynamics of adoption uncovered in this paper can explain the Great Divergence and the lack of absolute convergence over the post-war period.

Bibliography

- Basu, Susanto, and John G. Fernald. 1997. Returns to Scale in U.S. Production: Estimates and Implications. *Journal of Political Economy*, 105(2): 249-283.
- Basu, Susanto, and David N. Weil. 1998. "Appropriate Technology and Growth." *Quarterly Journal of Economics*, 113(4): 1025-1054.
- Caselli, Francesco, and W. John Coleman. 2001. "Cross-Country Technology Diffusion: The Case of Computers." *American Economic Review - AEA Papers and Proceedings*, 91(2): 328-335.
- Clark, Gregory. 1987. "Why Isn't the Whole World Developed? Lessons from the Cotton Mills." *Journal of Economic History*, 47(1): 141-173.
- Comin, Diego, and Mark Gertler. 2006. "Medium Term Business Cycles." *The American Economic Review*, (96)3: 523-551.
- Comin, Diego, and Bart Hobijn. 2004. "Cross-Country Technology Adoption: Making the Theories Face the Facts." *Journal of Monetary Economics*, 51(1): 39-83.
- Comin, Diego, and Bart Hobijn. 2008. "An Exploration of Technology Diffusion." HBS working paper 08-093.
- Comin, Diego, and Bart Hobijn. 2010. "An Exploration of Technology Diffusion." *The American Economic Review*, forthcoming.
- Comin, Diego, Bart Hobijn, and Emilie Rovito. 2006. "Five Facts You Need to Know About Technology Diffusion." NBER Working Paper #11928.
- Comin, Diego, Bart Hobijn, and Emilie Rovito. 2008. "A New Approach to Measuring Technology with an Application to the Shape of the Diffusion Curves." *Journal of Technology Transfer*, 33(2): 187-207.
- Gollin, Douglas. 2002. "Getting Income Shares Right." *Journal of Political Economy*, 110(2): 458-474.

- Gort, Michael, and Steven Klepper. 1982. "Time Paths in the Diffusion of Product Innovations." *Economic Journal*, 92: 630-653.
- Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, 25(4): 501-522.
- Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics*, 114(1): 83-116.
- Hsieh, Chang-Tai. 2002. "What Explains the Industrial Revolution in East Asia? Evidence from the Factor Markets." *American Economic Review*, 92(3): 502-526.
- Jerzmanowski, Michael. 2007. "Total Factor Productivity Differences: Appropriate Technology vs. Efficiency." *European Economic Review*, 51(8): 2080-2110.
- Klenow, Peter J., and Andrés Rodríguez-Clare. 1997. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *NBER Macroeconomics Annual*, 12: 73-103.
- Maddison, Angus. 2007. *Contours of the World Economy 1-2030 AD: Essays in Macro-Economic History*. Oxford: Oxford University Press.
- Mansfield, Edwin. 1961. "Technical Change and the Rate of Imitation." *Econometrica*, 29(4): 741-766.
- Manuelli, Rodolfo, and Ananth Sheshadri. 2003. "Frictionless Technology Diffusion: The Case of Tractors." NBER Working Paper # 9604.
- Parente, Stephen L., and Edward C. Prescott. 1994. "Barriers to Technology Adoption and Development." *Journal of Political Economy*, 102(2): 298-321.
- Saxonhouse, Gary R., and Gavin Wright. 2004. "Technological Evolution in Cotton Spinning, 1878-1933." in *The Fibre that Changed the World*, eds. Douglas A. Farnie and David J. Jeremy. Oxford: Oxford University Press.
- Young, Alwyn. 1992. "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore." *NBER Macroeconomics Annual*, 7: 13-54.
- Young, Alwyn. 1995. "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." *Quarterly Journal of Economics*, 110(3): 641-680.

3.6 Appendix: Data

The data that we use are taken from two sources. Real GDP and population data are taken from Maddison (2007). The data on the technology measure are from the Cross-Country Historical Adoption of Technology (CHAT) data set, first described in Comin, Hobijn, and Rovito (2006). The fifteen particular technology measures, organized by broad category, that we consider are:

1. **Steam and motor ships:** Gross tonnage (above a minimum weight) of steam and motor ships in use at midyear. *Invention year:* 1788; the year the first (U.S.) patent was issued for a steam boat design.
2. **Railways - Passengers:** Passenger journeys by railway in passenger-KM. *Invention year:* 1825; the year of the first regularly schedule railroad service to carry both goods and passengers.
3. **Railways - Freight:** Metric tons of freight carried on railways (excluding livestock and passenger baggage). *Invention year:* 1825; same as passenger railways.
4. **Telegraph:** Number of telegrams sent. *Invention year:* 1835; year of invention of telegraph by Samuel Morse at New York University.
5. **Telephone:** Number of mainline telephone lines connecting a customer's equipment to the public switched telephone network as of year end. *Invention year:* 1876; year of invention of telephone by Alexander Graham Bell.
6. **Electricity:** Gross output of electric energy (inclusive of electricity consumed in power stations) in KwHr. *Invention year:* 1882; first commercial power-station on Pearl Street in New York City.
7. **Cars:** Number of passenger cars (excluding tractors and similar vehicles) in use. *Invention year:* 1885; the year Gottlieb Daimler built the first vehicle powered by an internal combustion engine.
8. **Trucks:** Number of commercial vehicles, typically including buses and taxis (excluding tractors and similar vehicles), in use. *Invention year:* 1885; same as cars.
9. **Aviation - Passengers:** Civil aviation passenger-KM traveled on scheduled services by companies registered in the country concerned. *Invention year:* 1903; The year the Wright brothers managed the first successful flight.

10. **Aviation - Freight:** Civil aviation ton-KM of cargo carried on scheduled services by companies registered in the country concerned. *Invention year:* 1903; same as aviation - passengers.
11. **Blast Oxygen Steel:** Crude steel production (in metric tons) in blast oxygen furnaces (a process that replaced Bessemer and OHF processes). *Invention year:* 1950; invention of Blast Oxygen Furnace.
12. **Cellphones:** Number of users of portable cell phones. *Invention year:* 1973; first call from a portable cellphone.
13. **Personal computers:** Number of self-contained computers designed for use by one person. *Invention year:* 1973; first computer based on a microprocessor.
14. **MRIs:** Number of magnetic resonance imaging (MRI) units in place. *Invention year:* 1977; first MRI-scanner built.
15. **Internet users:** Number of people with access to the worldwide network. *Invention year:* 1983; introduction of TCP/IP protocol.

3.7 Appendix: Equilibrium and diffusion of the new technology

Let Γ_t denote the total adoption costs at instant t . Then

$$\begin{aligned} \Gamma_t = & \bar{\Psi} (1 + b_e) \left(\frac{\gamma}{\mu - 1} \right) e^{-\frac{\theta}{\mu-1} \gamma D} \left(\frac{Z_0 e^{\gamma \bar{v}}}{A_t} \right)^{\frac{1}{\mu-1}} Y_t (1 - \dot{D}) \\ & + \bar{\Psi} (1 + b_i) (Z_0 A_t e^{\gamma t})^{-\frac{1}{\mu-1}} Y_t \int_{-\infty}^{v_t} Z_v^{\frac{1}{\mu-1}} \dot{N}_v(t) dv. \end{aligned} \quad (3.41)$$

where \dot{D} denotes the time derivative of the adoption lags. Note that along the Balance Growth Path, the distribution over the vintages for which the measure of varieties adopted becomes degenerate around \bar{v}_t and the aggregate costs become $\Gamma_{v_t}^i N_v$.

The equilibrium path of the aggregate resource allocation in this economy can be defined in terms of the following nine equilibrium variables $\{C, K, I, \Gamma, Y, A, N, D, V\}$. Just like in the standard neoclassical growth model, the capital stock, K , is the only state variable. The eight equations that determine the equilibrium dynamics of this economy are given by

(i) The consumption Euler equation.

(ii) The aggregate resource constraint³⁴

$$Y = C + I + \Gamma. \quad (3.42)$$

(iii) The capital accumulation equation

$$\dot{K} = -\delta K + I. \quad (3.43)$$

(iv) The production function, (3.11), taking into account that in equilibrium $L = 1$.

(v) The adoption cost functions (3.6) and (3.7).

(vi) The technology adoption equations, which determine the adoption lag (3.20) and the intensive margin of adoption (3.16).

(vi) The stock market to GDP ratio, (3.14).³⁵

(vii) The aggregate TFP level, 3.12.

3.8 Tables and Figures

³⁴We assume that adoption costs are measured as part of final demand, such that Y can be interpreted as GDP.

³⁵The dynamics of Ψ_t^e and Ψ_t^i are what are considered in the system of equilibrium equations. For example, the law of motion of for Ψ_t^e is (omitting superscripts and subscripts) $\dot{\Psi} = \left\{ \alpha \frac{\epsilon-1}{\epsilon} \frac{Y}{K} - \delta + \frac{1}{\mu-1} \frac{\dot{A}}{A} - \frac{\dot{Y}}{Y} \right\} - \frac{\alpha}{\epsilon} \frac{1}{\Psi}$.

Table 3.1: Estimated Log-Differences in Intensive Margins

Technology	Invention year	Number of Countries	Plausible and Precise	Log Intensive Margin (Δ_j)						sd lag	sd \bar{y}	$\frac{sd \text{ int}}{sd \bar{y}}$
				Mean	sd	p10	p50	p90	IQR			
Ships	1788	61	50	0.45	0.85	-0.50	0.35	1.65	0.95	51.55	0.69	1.23
Rail Passengers	1825	88	62	0.19	0.33	-0.28	0.19	0.65	0.38	32.79	0.83	0.40
Rail Freight	1825	83	42	-0.47	0.35	-0.83	-0.53	0.03	0.49	26.30	0.70	0.50
Telegraph	1835	69	46	-0.47	0.50	-1.18	-0.42	0.14	0.53	32.16	0.61	0.82
Telephone	1876	143	66	-0.78	0.69	-1.71	-0.67	0.05	0.90	32.09	0.78	0.89
Electricity	1882	138	97	-0.40	0.35	-0.92	-0.38	0.01	0.48	19.62	0.87	0.40
Trucks	1885	111	57	-1.16	0.69	-2.18	-1.05	-0.28	0.88	19.48	0.64	1.07
Cars	1885	127	73	-1.22	0.84	-2.31	-1.13	-0.21	1.15	19.09	0.78	1.07
Aviation Freight	1903	96	30	-0.57	0.83	-1.86	-0.37	0.22	0.72	13.99	0.57	1.46
Aviation Passengers	1903	99	51	-0.84	0.63	-1.60	-0.77	-0.08	0.72	12.30	0.60	1.05
Blast Oxygen Furnaces	1950	50	39	-0.79	0.89	-2.25	-0.36	0.02	1.13	7.70	0.52	1.72
PCs	1973	71	68	-0.63	0.79	-1.81	-0.41	0.24	1.02	2.70	0.67	1.18
Cellphones	1973	87	85	-1.50	1.11	-3.18	-1.30	-0.23	1.92	3.95	0.70	1.59
MRI	1977	12	12	-0.44	0.35	-0.87	-0.40	-0.01	0.56	2.30	0.26	1.36
Internet	1983	59	59	-1.00	0.76	-1.99	-0.89	-0.18	1.10	2.19	0.58	1.31
Total		1294	837	-0.69	0.88	-1.90	-0.52	0.19	1.00	38.41	0.73	1.20

Note: Plausible and Precise estimates are defined as having an adoption year greater than 10 years before our invention date (this allows for some inference error) and a standard error for the adoption year smaller than $\sqrt{2003 - \underline{y}_t}$. Sd stands for Standard Deviation, pX denotes the percentile X (e.g., p50 is the median), IQR is defined as the difference between 75th percentile and 25th percentile and sd \bar{y} is $(1 - \alpha)$ times the standard deviation of log of income per capita in 2000. The last line (Total) reports the sum of observations for columns 3 and 4 and the unweighted mean for the technology measures with columns 5 to 13.

Table 3.2: Evolution of the Distribution of the Intensive and Extensive Margin

	100*Log-Intensive Margin		Adoption Lag	
	Homothetic	Non-Homothetic	Homothetic	Non-Homothetic
Mean	-0.50 (0.19)	-0.21 (0.10)	-0.50 (.06)	-0.40 (.06)
Std.Dev.	0.14 (0.12)	-0.04 (0.07)	-0.22 (.02)	-0.23 (.04)
IQR	0.32 (0.18)	0.01 (0.11)	-0.39 (.09)	-0.40 (.09)

Note: Robust standard errors are shown in parenthesis. Mean refers to the average of the intensive or extensive margin. The Homothetic column refers to the baseline model, while the Non-homothetic refers to the estimation with a non-homothetic production function. Std. Dev. refers to the Standard Deviation and IQR, to the Interquartile Range (difference between the 75th and 25th percentiles). All these technology measures are regressed on the year of invention of the technology.

Table 3.3: Comparison of Intensive Margin Estimates up to 1939 versus Whole Sample

Technology	Adoption Year	Number of Countries	Mean 1939	Mean	Std.Dev. 1939	Std.Dev.	IQR 1939	IQR
Ships	1788	12	0.06	-0.11	0.38	0.63	0.39	0.67
Rail Freight	1825	12	-0.49	-0.57	0.25	0.31	0.32	0.45
Rail Passengers	1825	17	0.14	0.24	0.16	0.19	0.17	0.23
Telegraph	1835	20	-0.13	-0.54	0.29	0.54	0.37	0.62
Telephone	1876	9	-0.38	-0.35	0.28	0.28	0.31	0.35
Electricity	1882	18	-0.24	-0.20	0.23	0.18	0.38	0.26
Cars	1885	11	-0.99	-0.75	0.56	0.48	0.63	0.60
Trucks	1885	10	-0.96	-1.01	0.63	0.73	1.00	1.02
Total			-0.31	-0.37	0.51	0.56	0.59	0.69

Note: Number of countries is the number of countries for which we have plausible and precise estimates using data up to 1939. Mean 1939 refers to the mean intensive margin of a technology obtained using data up to 1939. Mean refers to the mean intensive margin of a technology obtained using data for the whole sample (up to 2003). An analogous notation is used for the standard deviation (Std. Dev.) and the interquartile range (IQR).

Table 3.4: Analysis of variance

	Model SS	Country effect	Technology effect	Residual SS	Total SS
Country effect alone	44%	44%		56%	100%
Technology effect	34%		34%	66%	100%
Joint effect	77%	43%	33%	23%	100%

Note: Decomposition of the intensive margin estimates. The percentages are obtained from the ratio of the sum of squares of the country or technology dummy over the total.

Table 3.5: Estimates of Adoption Lags and Intensive Margin allowing for non-homotheticities

Technology	Invention year	Number of Countries	Plausible and Precise	Adoption Lags		Log-Int. Margin	
				Mean	sd	Mean	sd
Ships	1788	61	31	101.61	59.72	0.5	0.82
Railways – freight	1825	88	58	82.28	26.30	0.38	0.41
Rail – passengers	1825	83	40	61.78	31.51	0.09	0.43
Telegraph	1835	69	37	38.39	32.03	0.02	0.59
Telephone	1876	143	64	44.73	31.37	0.15	0.62
Electricity	1882	138	50	34.42	25.60	-0.03	0.63
Trucks	1885	111	53	32.97	19.53	-0.23	0.57
Cars	1885	127	70	36.05	22.05	-0.25	0.7
Aviation – freight	1903	96	36	37.58	14.89	-0.03	0.72
Aviation – passengers	1903	99	53	27.98	13.95	-0.42	0.76
Blast Oxygen Furnaces	1950	50	41	15.63	6.68	0.03	0.5
PCs	1973	71	62	13.96	2.86	0.17	0.54
Cellphones	1973	87	75	13.44	3.85	-0.44	0.76
MRIs	1977	12	12	3.02	2.38	-0.01	0.34
Internet	1983	59	56	7.39	2.09	-0.06	0.47
Total		1294	738	35.87	33.53	-0.04	0.66

Note: Plausible and Precise estimates are defined as having an adoption year greater than 10 years before our invention date (this allows for some inference error) and a standard error for the adoption year smaller than $\sqrt{2003 - \bar{v}}$. Sd stands for Standard Deviation. The last line (Total) reports the sum of observations for columns 3 and 4 and the unweighted mean of technology measures for columns 5 to 8.

Table 3.6: Correlation by Technology of Homothetic and Non-Homothetic Margins

Technology	Observations	Correlation Intensive	Correlation Extensive
Ships	30	0.88	0.96
Rail Passengers	55	0.82	0.93
Rail Freight	38	0.60	0.85
Telegraph	34	0.92	0.91
Telephone	59	0.89	0.96
Electricity	45	0.84	0.79
Trucks	51	0.86	0.88
Cars	63	0.89	0.86
Aviation Freight	29	0.95	0.96
Aviation Passengers	51	0.97	0.90
Blast Oxygen Furnaces	38	0.93	0.96
PCs	60	0.84	0.82
Cellphones	75	0.97	0.98
MRI	12	0.80	0.98
Internet	56	0.90	0.98
Total	696	0.87	0.91

Note: The last line (Total) reports the total sum of observations in the second column and the unweighted average of technology correlations for columns 3 and 4.

Table 3.7: Correlation of Intensive and Extensive Margin with log income per capita in 2000

Technology	Invention year	Homothetic			Non Homothetic		
		Obs.	Intensive Margin	Adoption Lag	Obs.	Intensive Margin	Adoption Lag
Ships	1788	50	0.38	-0.48	31	0.52	-0.47
Rail Passengers	1825	62	0.38	-0.62	58	0.33	-0.49
Rail Freight	1825	42	0.16	-0.67	40	0.46	-0.30
Telegraph	1835	46	0.46	-0.49	37	0.49	-0.28
Telephone	1876	66	0.79	-0.49	64	0.53	-0.37
Electricity	1882	97	0.76	-0.57	50	0.68	-0.27
Trucks	1885	57	0.53	-0.35	53	0.49	-0.20
Cars	1885	73	0.51	-0.31	70	0.63	0.01
Aviation Freight	1903	30	0.85	-0.09	36	0.78	0.20
Aviation Passengers	1903	51	0.80	-0.13	53	0.79	0.23
Blast Oxygen Furnaces	1950	39	0.90	-0.33	41	0.80	-0.35
PCs	1973	68	0.69	-0.31	62	0.66	-0.40
Cellphones	1973	85	0.92	-0.57	75	0.88	-0.51
MRI	1977	12	0.79	-0.39	12	0.61	-0.44
Internet	1983	59	0.95	-0.75	56	0.84	-0.82
Total		837	0.67	-0.46	738	0.64	-0.30

Note: The Homothetic columns refer to the baseline model, while the Non-homothetic refer to the estimation using a non-homothetic production function. The last line (Total) reports the total sum of observations in the third and sixth columns and the unweighted average of the correlations for columns 4, 5, 7 and 8.

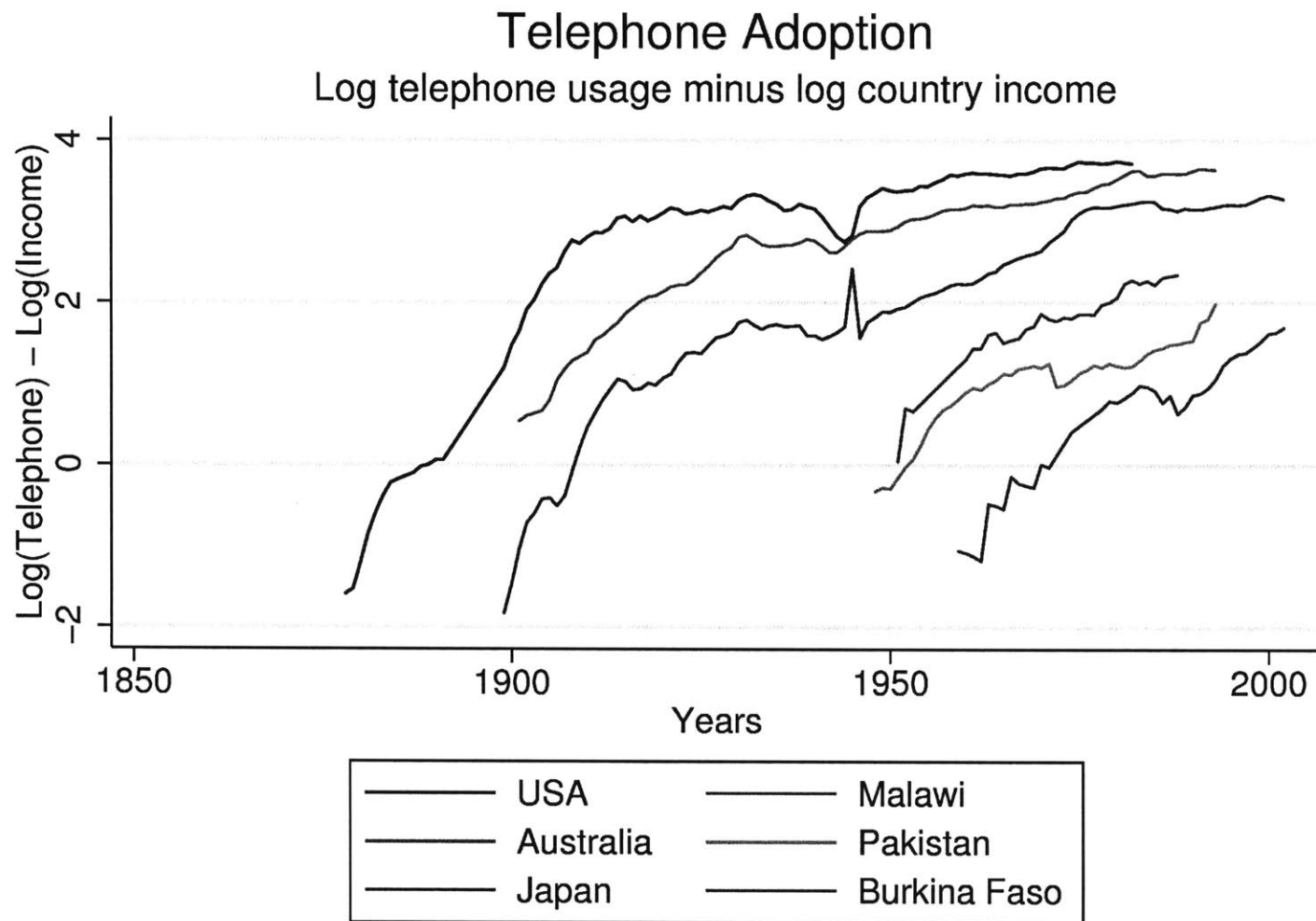


Figure 3-1: Differences in telephone adoption subtracting own country income for four different countries.

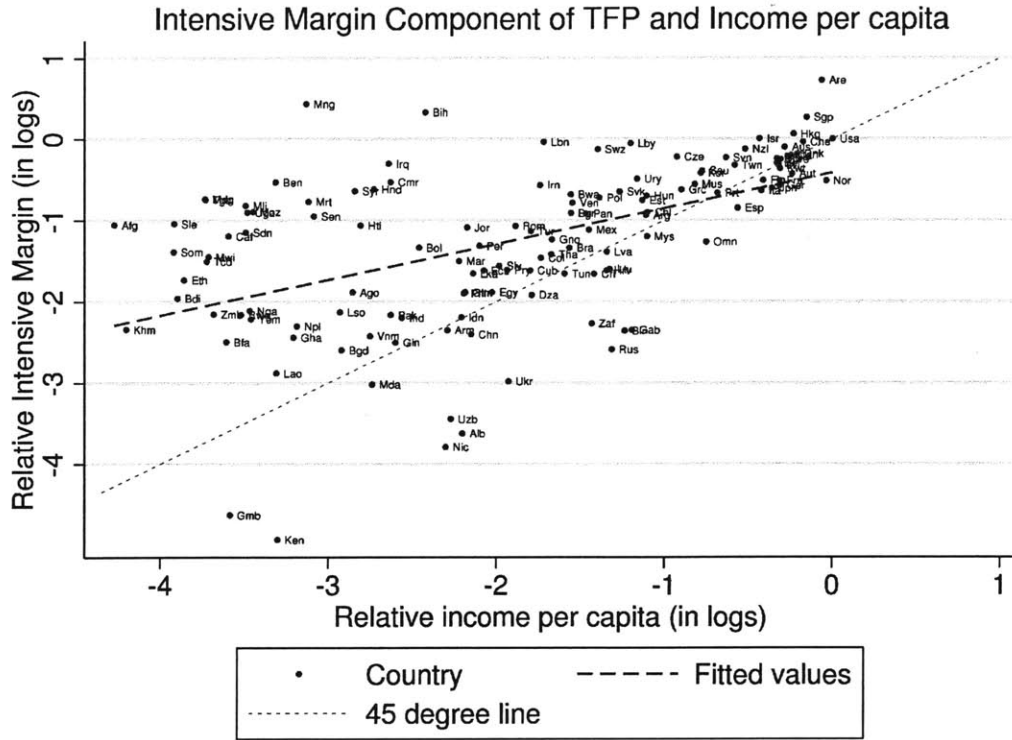


Figure 3-2: Intensive Margin component of TFP and differences in income per capita. The slope of the dashed line is .44.

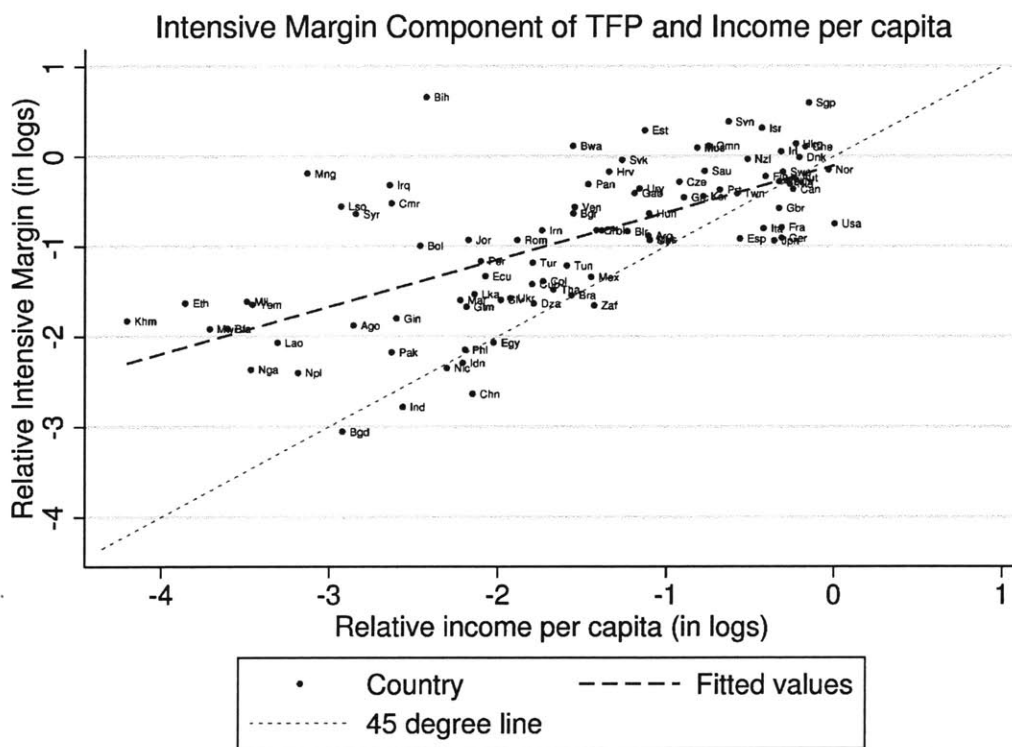


Figure 3-3: Intensive Margin component of TFP and differences in income per capita with non homotheticities. The slope of the dashed line is .52.