Human Capital in Development Accounting and Other Essays in Economics

Hannes Malmberg

Abstract

Human Capital and Development Accounting Revisited. I quantify the effects on development accounting of allowing for imperfectly substitutable labor services. To estimate the degree of substitutability between skilled and unskilled labor services in a cross-country setting, it is sufficient to estimate the relative price of skilled labor services, and I develop a novel method for estimating this relative price using international trade data. My method exploits the negative relationship between relative prices of skilled labor services and relative export values in skill-intensive industries. I find an approximately constant elasticity of substitution with a value of about 1.3. When integrating my results into a development accounting exercise, I find that efficiency differences in skilled labor are more important than uniform efficiency differences in explaining world income differences. Under the traditional development accounting assumption of neutral technology differences, the skilled labor efficiency differences reflect human capital quality differences, and human capital differences can explain a majority of world income differences. Relaxing the assumption of neutral technology differences, an alternative explanation is that there are large skill-biased technology differences between rich and poor countries.

Price Level Determination When Tax Payments Are Required in Money. We formalize the idea that the price level can be determined by a requirement that taxes be paid in money. We show that if households have to pay a money tax of a fixed real value and the money supply is constant, there is a unique stationary price level, and a continuum of non-stationary deflationary equilibria. The non-stationary equilibria can be excluded if we introduce an arbitrarily lax borrowing constraint. Thus, in the basic model, tax requirements can uniquely determine the price level. When money has liquidity value, tax requirements can exclude self-fulfilling hyperinflations.

Swedish Unemployment Dynamics. We decompose the sources of unemployment variations into contributions from variations in different labor market flows. We develop a decomposition method that allows for a distinction between permanent and temporary employment and slow convergence to the steady state, and we apply the method to the Swedish labor market for the period 1987-2012. Variations in unemployment are driven to an approximately equal degree by variations in (i) flows from unemployment to employment, (ii) flows from employment to unemployment, and (iii) flows in and out of the labor force. Flows involving temporary contracts account for 44% of unemployment variation, even though temporary workers only constitute 13% of the working-age population. Neglecting out-of-steady-state dynamics leads to an overestimation of the importance of flows involving permanent contracts.

Supply Chain Risk and the Pattern of Trade. This paper analyzes the interaction of supply chain risk and trade patterns. We construct a model where an industry's risk sensitivity is determined by the number of customized components that it uses, and countries with a low supply chain risk specialize in risk-sensitive goods. Based on our theory, we construct an empirical measure of risk sensitivity from input-output tables and customization measures. Using industry-level trade data and a variety of risk proxies, we show that countries with a low supply chain risk disproportionately export risk-sensitive goods.

Department of Economics

Stockholm University, 106 91 Stockholm

Human Capital in Development Accounting and Other Essays in Economics

Hannes Malmberg

ii

Cover Picture: Graduation photo, Kalmar Teaching College, 1941. Private photo. Author's grandmother second to the right in front – the first in her family to obtain an education beyond primary level.

Doctoral Dissertation
Department of Economics
Stockholm University

Abstracts

Human Capital and Development Accounting Revisited. I quantify the effects on development accounting of allowing for imperfectly substitutable labor services. To estimate the degree of substitutability between skilled and unskilled labor services in a cross-country setting, it is sufficient to estimate the relative price of skilled labor services, and I develop a novel method for estimating this relative price using international trade data. My method exploits the negative relationship between relative prices of skilled labor services and relative export values in skill-intensive industries. I find an approximately constant elasticity of substitution with a value of about 1.3. When integrating my results into a development accounting exercise, I find that efficiency differences in skilled labor are more important than uniform efficiency differences in explaining world income differences. Under the traditional development accounting assumption of neutral technology differences, the skilled labor efficiency differences reflect human capital quality differences, and human capital differences can explain a majority of world income differences. Relaxing the assumption of neutral technology differences, an alternative explanation is that there are large skill-biased technology differences between rich and poor countries.

Price Level Determination When Tax Payments Are Required in Money. We formalize the idea that the price level can be determined by a requirement that taxes be paid in money. We show that if households have to pay a money tax of a fixed real value and the money supply is constant, there is a unique stationary price level, and a continuum of non-stationary deflationary equilibria. The non-stationary equilibria can be excluded if we introduce an arbitrarily lax borrowing constraint. Thus, in the basic model, tax requirements can uniquely

determine the price level. When money has liquidity value, tax requirements can exclude self-fulfilling hyperinflations.

Swedish Unemployment Dynamics. We decompose the sources of unemployment variations into contributions from variations in different labor market flows. We develop a decomposition method that allows for a distinction between permanent and temporary employment and a slow convergence to the steady state, and we apply the method to the Swedish labor market for the period 1987-2012. Variations in unemployment are driven to an approximately equal degree by variations in (i) flows from unemployment to employment, (ii) flows from employment to unemployment, and (iii) flows in and out of the labor force. Flows involving temporary contracts account for 44% of the variation in unemployment, even though temporary workers only constitute 13% of the working-age population. Neglecting out-of-steady-state dynamics leads to an overestimation of the importance of flows involving permanent contracts.

Supply Chain Risk and the Pattern of Trade. This paper analyzes the interaction of supply chain risk and trade patterns. We construct a model where an industry's risk sensitivity is determined by the number of customized components that it uses, and countries with a low supply chain risk specialize in risk-sensitive goods. Based on our theory, we construct an empirical measure of risk sensitivity from input-output tables and customization measures. Using industry-level trade data and a variety of risk proxies, we show that countries with a low supply chain risk disproportionately export risk-sensitive goods.

Till Henrietta och min familj.

The men of experiment are like the ant, they only collect and use; the reasoners resemble spiders, who make cobwebs out of their own substance. But the bee takes a middle course: it gathers its material from the flowers of the garden and of the field, but transforms and digests it by a power of its own. Not unlike this is the true business of philosophy; for it neither relies solely or chiefly on the powers of the mind, nor does it take the matter which it gathers from natural history and mechanical experiments and lay it up in the memory whole, as it finds it, but lays it up in the understanding altered and digested. Therefore from a closer and purer league between these two faculties, the experimental and the rational...much may be hoped.

Francis Bacon, Novum Organum, 1620

Acknowledgments

It feels special to stand at the end of this long journey. Despite hard work and occasional despair, doing a PhD in economics has been the most intellectually exciting undertaking of my life, and it has changed my outlook on the world beyond recognition. Here, I would like to thank some of the many people who have been part of this journey.

I first want to thank my main advisor, Per Krusell, for all his support during these years. Throughout my thesis work, Per has given me the freedom to explore my own more or less realistic research ideas, always being liberal with content and methodology, but dogmatic on quality and relevance. Moreover, at critical moments, he has provided small, but decisive, nudges that have moved my projects in the right direction. For example, after I had presented early-stage work on what later became the main chapter of my thesis, Per noted in passing that, maybe, it could be worth thinking about what my results on human capital meant for development accounting. The title and the main chapter of my thesis testify to the result of exploring this 10 second comment, and this experience was not unique during my thesis writing. I am also extremely grateful for Per's consistent support during the stressful job market experience.

Another piece of helpful advice from Per was that I should talk more to Timo Boppart. This piece of advice led to many exciting discussions between me and Timo on economics, and Timo eventually became my co-advisor. I am grateful for his careful reading and insightful comments that have considerably improved my research, and I am constantly impressed by his open-mindedness, passion for economics, and sound scientific judgment.

I also want to thank the two other members of my job market committee: Torsten Persson and John Hassler. Torsten has been a fantastic source of support. He has always taken the time to patiently deal with the anxieties of a novice researcher, gently

steering me away from patently unproductive research paths, and, more than once, he has helped me situate a collection of unstructured thoughts in the broader framework of our discipline. John has commented extensively on my research, especially on my earlier projects, and with his experience from practical policy-making, he has often helped me think about the relevance of my research.

I started my economics career at the Department of Economics. In the fall of 2010, I did a number of intermediate-level courses, after which Anna Seim and Yves Zenou made me a fantastic service by encouraging me and helping me to apply to the PhD program.

However, over the last five years, the IIES has been my academic home. Its intense intellectual environment with seminars, kitchen discussions, research meetings, and dinners has been truly inspiring. Its equally intense social environment has made me feel at home, and has been the foundation of my intellectual work. I am grateful to all faculty, graduate students, and administrative staff who sustain this great institution.

In addition to my job market committee, I particularly want to thank Ingvild Almås, Timo Boppart, and Konrad Burchardi for inviting me to join their Söderberg project, Philippe Aghion for extensive comments on my work, Assar Lindbeck for consistently providing a historical perspective on our discipline, Kurt Mitman for a brief, but extremely useful, how-to-guide on doing research in quantitative macroeconomics, and David Strömberg for many interesting discussions on methodology, as well as for providing support on the econometric aspects of my job market paper.

I have also enjoyed discussions in the macro group with Tobias Broer, Sebastian Koehne, Alexander Kohlhas, and Kathrin Schlafmann, as well as discussions with Tessa Bold, Tom Cunningham, Arash Nekoei, Peter Nilsson, Robert Östling, Jon de Quidt, Anna Sandberg, and Jakob Svensson. The administrative staff at the IIES has been the source of exceptional support and many fun kitchen discussions over the years. I

warmth, his sharp eye and mind, his strong ethical views, and his deep knowledge of economic theory. I also want to thank David Bloom at Harvard School of Public Health for receiving me on an exchange visit, and to Maximilian Eber and Kinley Salmon at Harvard University for fruitful collaboration and for many exciting conversations. I have also benefited from discussions with Xavier Jaravel at Harvard University.

Economics is but one of many academic disciplines, and academic knowledge is but one of many ways of understanding the world. Throughout my PhD studies, I have drawn intellectual nourishment from extensive work and discussions with people outside of economics, and outside of academia altogether.

The Department of Mathematics at Stockholm University has enabled me to pursue a part-time PhD in mathematical statistics in parallel with my studies in economics, and my advisors Ola Hössjer and Dmitri Silvestrov have been very generous with their time in supporting my mathematical research.

I would also like to thank Zihan Hans Liu, who was my roommate when I was studying at Cambridge, and who transformed my understanding of mathematics. In addition, I've benefited greatly from long discussions with Arvid Ågren on evolutionary biology, Andreas Björdal on modern management, Hanna Björklund on religious history, Paul Igor Costea on microbiology, Johannes Danielsson on Swedish domestic and foreign policy, Emma Holten on literature and modern social theory, Adam Hurst on biotechnology, Jonathan Illicki on medicine, Christian Johansson on mathematics, Oskar Lilja on data sciences, Karin Öberg on astronomy and theology, Philip Pärnamets on cognitive psychology and philosophy, Caroline Walerud on technological innovation, Oliver Vikbladh on neuroscience, and Henrik Widmark on urban history and sociology. Without the willingness of these friends to discuss their perspectives on the world, my outlook on the world would have been much poorer.

Throughout life, my family has been a source of support, inspiration,

and love. My parents, Bo Malmberg and Lena Sommestad, are both social scientists, and, by osmosis, I came to share their interests. Hour-long discussions around the kitchen table taught me the pleasures of academic conversation. For all support, knowledge, and love that I have received from them, I am extremely grateful.

My brother Olof and my sister Anna mean the world to me. Olof has always been my close companion and intellectual sparring partner: often brilliant, sometimes frustrating, but always creative. Anna brings so much joy to my life through her genuine curiosity and openness; I enjoy few things more than discussing with Anna what she has learnt from the world over the last few weeks.

Finally, I want to thank Henrietta for sharing your life with me. We walk through this life together, through big things and small. Your love, excitement, humor, warmth, and humanity provide the backdrop against which things make sense and are valuable. You constantly expand how I see the world, by directing my gaze outwards, away from the sometimes numbing complexities of the internal world, and towards the richness of the lives, feelings, and thoughts of others. For this, and for everything else, I am eternally grateful. You are my greatest support and joy, and I love you.

<div align="right">Hannes Malmberg, Stockholm, May 2017</div>

Contents

Introduction

This thesis consists of four independent chapters on issues in macro-economics, development economics, and trade. While the chapters treat different topics, they jointly reflect my interest in exploring different ways of approaching economic questions. The message of the thesis is one of methodological pluralism: economics is well-served when we use multiple ways of engaging with its subject matter.

In this introduction, I discuss the thesis with a focus on methodology. After briefly summarizing the contents of each chapter, I discuss the methodological approaches that they embody, and the main lessons learnt from each chapter. I conclude with a discussion of how the thesis work has shaped my methodological views. The first chapter of the thesis is its main chapter, but my discussion will reflect the chronology of the writing of the chapters, and will start with Chapters II-IV, and conclude with Chapter I.

Chapter summaries

Chapter II, **Price Level Determination When Tax Payments Are Required in Money**, is joint work with Erik Öberg, and explores the role of tax requirements in determining the value of money.

The value of money has proved difficult to model in economics. The challenge to theory development is that money is intrinsically useless, which means that standard consumption theory predicts that it should have no value. One of the most widely held theoretical views is that the

value of money depends on its conventional use as a medium of exchange: I accept your money as payment, because I expect that others will accept my money in turn. The value of money is sustained by expectations. While such liquidity-based theories of money are plausible, they do not explain why expectations should converge on using government money, and why government money rarely, if ever, has suffered a collapse of value merely due to a collapse of expectations.

These facts suggest that the value of money is somehow connected to government power, which raises a question about the nature of this connection. This chapter shows that one way in which the government can determine the value of money is by requiring that households pay their taxes with government-issued money. The idea that the value of money is connected to tax requirements is old, going back to Adam Smith. In our paper, we model this idea, and show conditions under which the government can enforce a unique equilibrium of the price level by regulating the money supply and requiring that taxes should be paid with money. In cases where money is also valued for its liquidity properties, tax requirements do not necessarily determine the exact value of money, but exclude the possibility that money loses its value through a collapse of expectations.

Chapter III, **Swedish Unemployment Dynamics: New Methods and Results**, is joint work with Niels-Jakob Harbo Hansen and studies the statistical properties of unemployment dynamics. This chapter belongs to a literature which is motivated by the basic observation that an increase in unemployment can either be driven by workers losing their jobs at a higher rate or by unemployed people taking more time to find new jobs. Thus, to understand the nature of any given change in unemployment, we need to analyze the underlying flows of the labor market .

Chapter III develops a new method for decomposing fluctuations in unemployment in the context of a dual labor market featuring a slow convergence to the steady state. The standard assumption in the literature has been that the current level of unemployment can be well

approximated by the steady-state unemployment rate associated with the current flow rates observed in the data. However, it is also known that this approximation works well in a U.S. context, where labor market gross flows are high, but is problematic in a European context, where the flows are much smaller.

At the same time, many European labor markets are dual, with a role for both temporary and permanent contracts. Since the flow properties are likely to differ markedly across these two forms of employment, it can be important to have a method that allows for a distinction between permanent and temporary jobs. To address these issues, we develop a decomposition method which allows for both (i) a slow convergence to the steady state and (ii) an arbitrary number of labor market states. We apply our method to a new Swedish data set on labor market flows covering the period 1987-2012.[1]

Our method is well suited for this application. Indeed, the Swedish labor market is dual, with temporary contracts accounting for 13% of the working-age population, and characterized by relatively low gross flows and a low rate of convergence to the steady state. We find that variations in unemployment are driven to an approximately equal degree by variations in (i) flows from unemployment to employment, (ii) flows from employment to unemployment, and (iii) flows in and out of the labor force.

We find that flows involving temporary contracts account for 44% of the variation in unemployment, while flows involving permanent contracts account for approximately 33% of the variation. The former number is large given the relatively few workers who are employed on temporary contracts. We also show that it is important to account for out-of-steady state dynamics. If we use the decomposition method which relies on approximating the actual state with the steady state, the share of variation attributed to flows involving permanent contracts rises from 33% to 44%. These results are of broader interest in the study of Eu-

---

[1]The most similar method to ours is that of Elsby et al. (2015), which is discussed in the main text.

ropean labor markets, since they suggest that decompositions based on the steady-state assumption are unlikely to be suitable in dual labor markets.

Chapter IV, **Supply Chain Risk and the Pattern of Trade**, is joint work with Maximilian Eber. We explore whether countries that offer reliable supply chains specialize in goods that are sensitive to supply chain risk. We formalize this hypothesis by constructing a trade model with multiple sectors and risky supply chains. Each sector produces a good using a range of intermediate inputs, and the delivery of each input is subject to a failure risk.

The effect of a delivery failure depends on the nature of the input. Some inputs are standardized, which means that delivery risks are pooled through a law of large numbers, and downstream buyers are protected from disruptions. Other inputs are customized, which means that the failure of a supplier disrupts downstream production. As each customized input represents a potential source of failure, an industry's sensitivity to risk depends on its number of customized inputs. We incorporate our sector production model into a trade model, and show that countries with reliable supply chains will specialize in products with many customized inputs.

We test the prediction by regressing industry export flows on fixed effects and an interaction term between proxies of country supply chain reliability and industry risk-sensitivity. This is a method proposed in Romalis (2004) and Nunn (2007). We show that the interaction term is positive and that the result is robust to a wide range of robustness checks.

As previously mentioned, Chapter I, **Human Capital and Development Accounting Revisited**, is the main chapter of the thesis. The chapter analyzes the importance of differences in human capital in accounting for world income differences.

It is an old and prominent idea that differences in human capital are important for understanding world income differences. However, an influential view in economics is that human capital differences play a lim-

ited role in explaining income differences, since microeconomic returns to human capital are relatively small. Rich countries have approximately 8 years more of schooling than poor countries, and the returns to education are approximately 10% per year of schooling. Aggregating microeconomic returns yields a doubling of incomes, which is small as compared to the 30-40 times differences in GDP per worker levels across rich and poor countries.

This argument has been formalized in the literature on development accounting (Hall and C Jones 1999), which shows that human capital differences can only account for a small share of world income differences. Instead, most income differences are accounted for by differences in technology. Here, technology is broadly construed, and refers to anything other than factor inputs that shifts productivity. This result has been influential in development economics, and has led to a focus in the literature on explaining the sources of technology differences. Human capital differences have been presumed to be of second-order importance.

The heart of the argument in traditional development accounting is that given observed skilled wage premia, skilled workers seem to be insufficiently productive for human capital differences to be important. However, an alternative view is that skilled workers are very productive in rich countries, but that the high supply of skilled workers in rich countries pushes down their relative wage. This argument is based on imperfect substitutability between labor services and was proposed by Jones (2014).

Even if it is recognized that imperfect substitutability can lead to an underestimation of the role of human capital, there is less agreement on the quantitative importance of this mechanism. To estimate its quantitative importance, we need to measure how much lower efficiency-adjusted prices of skilled labor services are in rich countries. In this chapter, I bring in new quantitative evidence from international trade to answer this question. I use the fact that the skill intensity of a country's export composition contains implicit information about the relative price of skilled to unskilled labor services.

My results suggest that the relative price of skilled to unskilled labor services is much lower in rich countries as compared to poor countries. As this low relative price of skilled labor services is not reflected in low skilled wage premia, this suggests that skilled workers in rich countries are more efficient than skilled workers in poor countries. Here, efficiency refers to the number of labor services delivered per worker, and a low efficiency of skilled workers increases the price of skilled labor services for a given skilled wage premium.

When I follow traditional development accounting and posit neutral technology differences, these efficiency differences suggest large differences in the quality of skilled labor human capital, and that human capital can explain a majority of the world income differences. If I relax the assumption of neutral technology differences, I cannot exclude technology-based explanations of world income differences. However, in contrast to traditional development accounting, these technology differences will specifically augment the efficiency of skilled labor, and technology-based explanations of world income differences would have to account for the interaction between the economic environment in rich countries, and the efficiency of skilled workers.

Methodological discussion of chapters

In this section, each chapter is discussed from a methodological point of view, with a focus on the main lessons learnt from writing each chapter.

Chapter II on money and taxes is the most theoretical chapter of my thesis. It is only connected to empirics via stylized facts, for example that money usually has a positive stable value and is connected to government power. The lessons from this paper relate to the nature of economic theorizing.

The first lesson was the value of simplification when developing economic theory. Our initial model was complicated, featuring multiple goods, labor in both the government and the private sector, and the price level being set via an arbitrage condition in the labor market. A

monetary economist asked us critically whether all aspects were actually vital to our point. This led us to considerably simplify the model, and the final version consisted of only one good, a money supply rule, and a tax rule. This process of simplification taught me how to better interpret simple models throughout economics, as I could see how to reverse the simplification process to create realistic versions of the models.

The second lesson was the discovery of what I came to dub the "Frankenstein moment" of modeling. This is the moment when you have continually fed assumptions into a model, and the model suddenly starts to return results that you did not expect, but that nevertheless contain insights. These moments taught me in what ways purely theoretical models contribute to our intuitive understanding of the world.

Chapter III is primarily an empirical chapter. Its aim is to decompose the sources of Swedish unemployment variations. Economic theory is a motivation for studying the topic, but the core of the chapter concerns the application of probability theory to economic data.

The main lesson from this chapter was the value of clearly connecting empirical objects to mathematical objects when conducting measurement. Even if the paper used little economic theory, we used stochastic process theory to express labor market transitions as a matrix exponential of an integral of instantaneous flow matrices. With this formulation, it was easy to formulate a consistent treatment of all issues involved in the decomposition exercise.

The connection between theory and measurement is, of course, not unique to our setting; many standard economic measures such as price indices have strong theoretical underpinnings. However, using stochastic process theory for a pure measurement problem in labor economics made the theoretical dimension of measurement salient, which has made me more attentive to the theoretical dimension of measurement in other settings.

Chapter IV is the first chapter in which I more explicitly connected economic theory with data. We used a theoretical model to derive the empirical predictions, and the theory guided measurement as it suggested

that the supply chain sensitivity of an industry should be measured by the number of customized inputs.

However, even though the chapter connected theory and empirics, it suffered from not using an empirical method that allowed for a quantitative assessment of the underlying production model. The trade patterns in Chapter IV were driven by variations in relative unit costs across industries, but lacking a trade elasticity, it was not possible to gauge the size of these relative unit cost deviations. This meant that it was not possible to say anything about the size of productivity damages stemming from supply chain unreliability. In Chapter I, I returned to the question of trade patterns and economic development, with a model that allowed for quantification.[2]

In developing Chapter IV, I also learnt how to build a research idea more from the bottom up, starting with anecdotes, moving to more structured qualitative data, and finishing with a formal model tested on quantitative data. The background to Chapter IV was an interest in understanding why Africa had not become a manufacturing hub like East Asia. Anecdotal evidence suggested that one reason was that the Sub-Saharan African business environment was unreliable, which meant that one could not trust that one would get one's inputs on time.

To explore this hypothesis, I decided to first develop my contextual knowledge. I had long been interested in the opportunities afforded to economics by interviews, case studies, professional knowledge, and cross-fertilizations with other disciplines. Thus, in preparing for the paper, I read books on supply chain management from operations research and management science. Through a research grant from PEDL, I also got the opportunity to travel to Ethiopia together with Kinley Salmon. Kinley Salmon had previous experience from working as a consultant in Addis Ababa, and we conducted an interview study where we talked to NGOs,

---

[2]A second shortcoming of the model in Chapter IV is that the measurement of risk sensitivity was derived from an I/O-table. This meant that the measure captured how many customized industries an industry used for inputs, rather than how many customized inputs any given plant used.

officials, manufacturers, and business facilitators about the role of supply chain risk in nascent manufacturing.

I enjoyed thinking through a familiar topic such as comparative advantage through the unfamiliar lens of supply chain theory and narratives from Ethiopian practitioners. The process of reading and interviewing yielded many different perspectives on supply chain risk. In the end, we managed to integrate some, but not all, of these aspects in our model. Some aspects that were not integrated are potential directions of future work. I also work on a follow-up paper with Kinley Salmon where we seek to interpret the interview results from Ethiopia in light of economic theory.

Chapter I is the main chapter of the thesis, and the one with the closest interaction between data and theory. I analyzed the data through the lens of a gravity model which allowed me to back out quantitatively relevant parameters from my data analysis. In contrast to Chapter IV, this allowed me to make quantitative statements about the importance of economic mechanisms.

It was both an exciting and a humbling experience to use theory for quantitative purposes. The exciting part was that it forced me to think more carefully about the mapping between theory and data, and I learnt a lot of economics by thinking through how different measurement choices would affect my conclusions.

The humbling part consisted of the difficulty in performing quantification based on theory. The world offers us limited sources of variations, and doing quantitative analysis through the lens of theory forces us to place theoretical structure on the data to identify parameters. A large part of Chapter I consisted of checking the robustness to various theoretical modifications, but it is challenging in any setting to state with any generality how dependent any particular conclusion is on the theoretical structure used in the estimation.

Looking ahead, despite the challenges of doing quantitative work, I think that it is important to perform a quantification of theoretical models. Whenever we claim that a model is useful in explaining a phe-

	Data	*Interpretation*
Quantitative measurements	Mathematical models	

Mathematical

Non-mathematical

| Everyday data | Informal theory | |
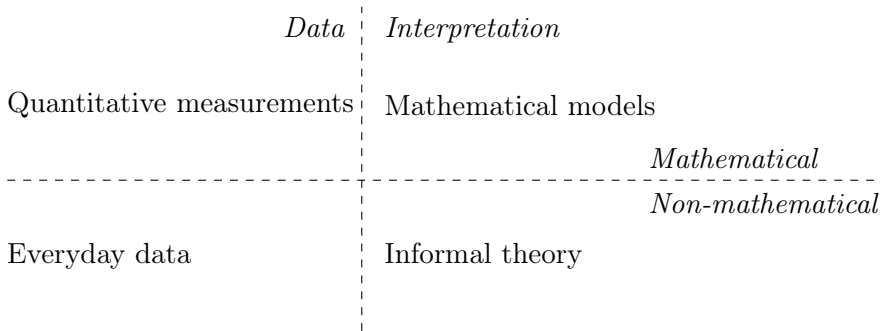
Figure 1: Stylized representation of data vs interpretation, and mathmetical vs non-mathematical knowledge

nomenon, we are committing to a particular quantification of its parameters, explicitly or implicitly. The identification problem does not disappear because it is less explicit. This does not necessarily mean that toy models should be fully calibrated - but whenever we use a model to explain a quantitative phenomenon, we should think about how observable phenomena could inform us about the values of its parameters.

Taking stock

In this section, I take stock of how my methodological thinking has evolved throughout the work on this thesis. The chapters of the thesis tackle different questions using different approaches, and to explain the evolution of my views, I have found it helpful to use a stylized representation of two dimensions of economic knowledge.

The result is in Figure 1 and the figure focuses on two conceptual relationships that I have wrestled with during the writing of this thesis: the relationship between data and interpretation of data, and the relationship between non-mathematical and mathematical approaches to economics. Below, I explain the terms in the figure, before discussing how my views on these relationships have evolved.

The left-hand column in Figure 1 represents different forms of data,

and aims at capturing the distinction between mathematical and non-mathematical forms of data. I call non-mathematical data "everyday data" to indicate that this term describes all the standard ways in which we obtain information about the world: e.g. through newspapers, conversations, interviews, books, and direct observation. Quantitative measurements represent the idealization of everyday data into a mathematical form, and include measurements of GDP, price indices, and income distribution measures.

On the right-hand side, I try to capture a similar distinction between mathematical and non-mathematical approaches to interpretation. Most interpretive knowledge of the world is informal, and consists of the myriad of theories that people use to navigate their environments. This is "informal theory" in the lower right-hand corner, and I include economists' informal intuitions and ideas in this category. Formal mathematical models represent an alternative way of interpreting the world, and mathematical models often represent a formalization and idealization of the informal knowledge base.[3]

With these concepts, my evolving views on economic methodology can be summarized in Figure 2, where the information in black represents my initial views, and the information in red represents the views that I have developed during my graduate studies.

When I started graduate school, my view was, somewhat simplified, that a research project started with some informal ideas about potential economic mechanisms. These informal ideas were formalized into a mathematical model, which could be used to generate testable predictions, which were then tested using some form of quantitative data. I was aware of the potential importance of exploiting qualitative data to build intuition, and of the challenges of quantitative measurement, but

---

[3]Thomas Sargent has, for example, described economics as organized common sense. The idea of treating the mathematical representation of data and theory as an idealization of everyday phenomena bears some resemblance to Edmund Husserl's characterization of the development of modern natural science in his book *The Crisis of European Sciences* (Husserl, 1970).

Data *Interpretation*

Confirmation/Falsification

Quant. meas.   **Math. models**

Estimation
Calibration

(Measurement)

Data collection
Theory-based measurement
Dimension-reduction methods

Refining intuition

Formalization

Math.

Non-math.

Case studies
Interview data
Historical knowledge
Domain-specific expert knowledge

Everyday data   **Informal theory**

(Casual observation)

Figure 2: Evolution of views on research process

I viewed the theory-testing cycle as the core of science.

During the course of my graduate studies, my views have evolved to
include the information in red, with more arrows in the figure, and more
information on each arrow, both reflecting an increased recognition of
different types of research methodologies.

The relationship between informal and mathematical models is now
bidirectional, with models serving to refine our intuition and our informal
knowledge. This reflects my experience of working with model construc-
tion, especially from Chapters II and IV, where the model construction
yielded new insights that aided the intuition.

There is an added line from everyday data to informal theory. This
reflects a view that insofar that qualitative evidence from case studies,
interviews, and outside expertise knowledge can be formalized and stud-
ied rigorously, it constitutes a valid form of scientific evidence. During my
thesis work, I engaged with qualitative data in preparation for Chapter
IV, when I was doing interviews and reading through the supply chain

literature.

Furthermore, the measurement arrow is augmented to capture a richer view of measurement, reflecting that measurement involves an active construction of facts based on data collection, economic theory, and dimension reduction methods. During my thesis work, the measurement exercise from Chapter III illustrated the ways in which data were not just given, but constructed in an interaction between observations and theory.

Finally, I have added estimation and calibration to the ways in which quantitative measurements and mathematical theories interact, reflecting the importance and challenge of using data to quantify economic theory. In my thesis work, this primarily reflects work done on Chapter I, where I used data not just to confirm or falsify a hypothesis, but to estimate the parameters of a model.[4]

Looking back, I can also see that my personal views did not evolve in a vacuum, but were connected to a number of broader developments within the discipline. Over the last 20 years, some of the more exciting research programs in economics have been related to the integration of qualitative evidence, to new facts from careful quantitative measurements, and to a closer interaction between theory and data.

The field experiment movement has highlighted that qualitative and contextual evidence is not just anecdotal, but an important component in research design and in the interpretation of results; the institutional and historical turn in macroeconomic development has further increased the importance of qualitative evidence. Newly available register data sets and analysis tools have generated influential facts regarding inequality, wage dynamics, firm dynamics, productivity dispersion,

---

[4]There is an argument to be made for more lines in Figure 2. For example, one could draw a line from quantitative measurements to informal theory to reflect the use of statistics in the building of intuition. The choice of having a limited set of lines reflects that the diagram does not aim at being a complete description of the process of scientific work, but merely a vehicle for explaining some specific methodological developments.

management quality, and international learning outcomes, to name but a few examples. The heterogeneous agent revolution in macroeconomics has made it possible to connect general equilibrium theory to rich micro data sets, allowing for a closer interaction between data and theory.

In writing my thesis, I have come to appreciate in what ways different approaches to economics – theoretical and empirical, formal and informal – are all distinct, and are all relevant to economic research. This has made me more aware of the values of methodological pluralism, and the advantages and disadvantages of different ways of doing economic research. I hope that writing this thesis has improved my ability to find the right strategy for the right question, and brought me at least a bit closer to the ideal of approaching economic issues with an open, yet critical, mind.

References

Elsby, M. W., Hobijn, B., and Şahin, A. (2015). On the importance of the participation margin for labor market fluctuations. *Journal of Monetary Economics*, 72:64–82.

Hall, R. E. and Jones, C. I. (1999). Why do Some Countries Produce So Much More Output Per Worker than Others? *The Quarterly Journal of Economics*, 114(1):83–116.

Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy.* Northwestern University Press.

Jones, B. F. (2014). The Human Capital Stock: A Generalized Approach. *American Economic Review*, 104(11):3752–3777.

Nunn, N. (2007). Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade. *The Quarterly Journal of Economics*, 122(2):569–600.

Romalis, J. (2004). Factor Proportions and the Structure of Commodity Trade. *American Economic Review*, 94(1):67–97.

Chapter 1

Human Capital and Development Accounting Revisited[*]

1.1 Introduction

Development accounting uses neo-classical production theory in conjunction with price and quantity data to decompose world income differences into contributions from capital-output ratios, human capital stocks and uniform labor efficiency (TFP) differences (Klenow and Rodriguez-Clare, 1997; Hall and C Jones 1999). A key component of development accounting has been to measure the human capital stock by aggregating microeconomic returns to human capital. To aggregate human capital, it has traditionally been assumed that different labor services are perfectly substitutable and that technology differences are neutral across countries, which means that skilled wage premia can be used to con-

vert skilled workers into unskilled equivalent labor units. By assuming that unskilled labor has a similar level of human capital across countries (or by separately modeling/measuring the quality of unskilled labor), human capital stock differences can be estimated. The key finding has been that variations in human capital stocks are much smaller than variations in labor efficiency, which suggests that large uniform TFP differences are needed to explain world income differences. This view has had a considerable influence on the macroeconomic growth and development literature.[2]

However, it is known that the results of traditional development accounting might be sensitive to relaxing the assumption of perfectly substitutable labor services (Caselli and Coleman, 2006; B Jones 2014a; Caselli 2015). In particular, if labor services are not perfect substitutes, traditional development accounting might underestimate the efficiency of skilled workers in rich countries, and overstate the importance of uniform labor efficiency differences. This is due to traditional development accounting equating the skilled wage premium to the relative efficiency of skilled workers. With imperfect substitutability, this is not true, as the relative abundance of skilled services in rich countries pushes down the relative price of skilled labor services. This means that the relative efficiency of skilled labor in rich countries is higher than what is suggested by skilled wage premia.

If skilled labor efficiency differences are more important than uniform TFP differences, this suggests a different set of interpretations of world income differences. If we retain the assumption of skill-neutral technolo-

---

[2]Early contributions to development accounting are Klenow and Rodriguez-Clare (1997) and Hall and C Jones (1999). There has been an ongoing debate about the robustness of development accounting. See, for example, Acemoglu and Zilibotti (2001), Erosa et al. (2010), Schoellman (2011), B Jones (2014a), B Jones (2014b), Manuelli and Seshadri (2014), and Schoellman and Hendricks (2016). There is also a large literature seeking to explain TFP differences. E.g. Parente and Prescott (1999) and Acemoglu et al. (2007) discuss the role of technology diffusion barriers in explaining TFP differences, and Restuccia and Rogerson (2008), Hsieh and Klenow (2009), and Midrigan and Xu (2014) are a few contributions to the large literature that seeks to explain TFP differences by misallocation.

gies from the traditional development accounting literature, a high efficiency of skilled workers in rich countries suggests that skilled workers have a high human capital in rich countries, and that human capital explains a much larger share of world income differences than in traditional development accounting. If we relax the assumption of neutral technology differences, a high efficiency of skilled workers in rich countries can also indicate large skill-augmenting technology differences. This interpretation suggests that technology-based theories of income differences should focus on why differences in the economic environment of rich and poor countries disproportionately affect the efficiency of skilled workers. Thus, imperfect substitutability opens up for a larger role for human capital, and always suggests that the efficiency of skilled labor is relatively more important as compared to uniform technology differences.[3]

Even if it is recognized that imperfect substitutability might be important for development accounting, there is less agreement on its quantitative importance, and how it should be modeled in a cross-country setting. Early contributions in the development accounting literature assumed that labor services were perfect substitutes (Klenow and Rodriguez-Clare, 1997; Hall and C Jones, 1999). Recent contributions have made different assumptions. Caselli and Ciccone (2013) take a non-parametric approach where human capital is aggregated by an arbitrary CRS, concave aggregator. Caselli and Coleman (2006), Jones (2014a), and Caselli (2015) assume that skilled and unskilled labor services are aggregated using a CES-aggregator, where the elasticity of substitution is assumed to

---

[3]The literature on development accounting under imperfect substitutability has made different interpretations of the source of skilled labor efficiency differences. Jones (2014a) interprets skilled labor efficiency differences as reflecting human capital quality differences of skilled workers, whereas Caselli and Coleman (2006) and Caselli (2015) interpret skilled labor efficiency differences as reflecting skill-biased technology differences. The two interpretations are isomorphic in macroeconomic price and quantity data, and this ambiguity has led to conflicting interpretations of the effects of imperfect substitutability on development accounting, with Jones (2014a) finding that imperfect substitutability makes human capital more important, and Caselli and Coleman (2006) and Caselli (2015) finding that imperfect substitutability makes human capital less important.

be in line with US time series and panel estimates.[4] A number of recent contributions in the development accounting literature have also retained the assumption of perfectly substitutable labor services (Erosa et al., 2010; Manuelli and Seshadri, 2014; Schoellman and Hendricks, 2016), as has the recent handbook chapter by C Jones (2015).

In this paper, I bring in new quantitative evidence from international trade data to measure the degree of substitutability between skilled and unskilled workers in a cross-country setting. I use the fact that the skill intensity of a country's export composition contains implicit information about the relative price of skilled and unskilled labor services, and that variations in the relative price of skilled and unskilled labor services can be used to estimate a human capital aggregator. My results suggest that skilled and unskilled labor services are imperfect substitutes, that the cross-country relevant human capital aggregator features an approximately constant elasticity of substitution, and that this elasticity is in the same range as that found in US time series panel studies. Thus, my results provide support for the modeling choices made in Caselli and Coleman (2006), Jones (2014a), and Caselli (2015). When I integrate my findings into development accounting, I find a much smaller role for uniform TFP differences than does traditional development accounting. Instead, I find that income differences are primarily driven by skilled labor efficiency differences.

More formally, my quantitative exercise assumes that there is a human capital aggregator of the form

$$h = G(Q_u u, Q_s s). \tag{1.1}$$

Here, u and s denote the share of unskilled and skilled workers, and Q_u and Q_s denote their respective efficiencies. G is an arbitrary constant returns to scale aggregator, and the setup allows for a nested structure

---

[4]Jones (2014a) further allows for a nested CES-structure where the aggregate flow of skilled labor services is an arbitrary aggregator of heterogeneous types of skilled labor services.

where the aggregate supply of skilled services $Q_s s$ reflects an aggregator of underlying heterogeneous skilled labor services. Thus, the setup nests the case of perfect substitutability, the case with a CES aggregator of skilled and unskilled labor services (Caselli, 2005, 2015), and a nested CES structure (Jones, 2014a). In my quantitative exercise, I estimate Q_u and Q_s, and I identify an appropriate functional form and parametrization of G.

To measure the human capital aggregator G, I assume that labor markets are competitive, which implies that the relative price of skilled and unskilled labor services can be equated to the marginal rate of transformation of the human capital aggregator: $r_s/r_u \dot{=} G_s/G_u$. Then, I note that to measure G, it is sufficient to measure this relative price of skilled and unskilled labor services across countries. Using r_s/r_u, it is possible to back out relative labor efficiencies Q_s/Q_u using that the skilled wage premium equals the relative efficiency of skilled and unskilled workers, times the relative price of skilled and unskilled labor services:

$$\frac{w_s}{w_u} = \frac{Q_s}{Q_u}\frac{r_s}{r_u}.$$

Furthermore, the relationship between r_s/r_u and relative labor service supply $\frac{Q_s s}{Q_u u}$ defines the isoquants of G, which uniquely identify G.

To estimate r_s/r_u across countries, I develop a new method using information contained in international trade flows. My method is based on two premises. First, that low relative prices of skilled services imply low relative unit costs in skill-intensive industries, and, second, that low relative unit costs in skill-intensive industries imply high relative export values in these industries. The first premise means that relative unit cost data are informative about the relative price of skilled services; the second premise means that export value data are informative about relative unit costs. The latter point is important given the lack of detailed industry unit cost data sets covering both rich and poor countries. By applying the connection between relative skilled labor services prices and export composition in reverse, we see that trade data contain information

about relative skilled labor service prices. For quantification, I use a gravity trade model, and I derive a regression specification that combines a trade elasticity estimate with data on export values and industry factor shares.

My trade data analysis suggests that there is a strong negative relationship between country income levels and the relative prices of skilled services. Given that there is also a strong positive relationship between country income levels and relative supplies of skilled labor, this suggests that skilled and unskilled labor services are imperfectly substitutable. Furthermore, I find that the human capital aggregator G can be well approximated by a CES function, and in my baseline specification, the estimated elasticity of substitution is 1.27.

In the next step, I incorporate this finding into the development accounting setting of Hall and C Jones (1999). I posit an aggregate production function

$$\frac{Y}{L} = \left(\frac{K}{Y}\right)^{\frac{\alpha}{1-\alpha}} Ah,$$

where Y is output, L is the size of the workforce, K is physical capital, A is a uniform TFP-shifter, and h is a human capital aggregator of the form in equation (1.1).[5] When I constrain the human capital aggregator G to be additive, I find that variations in the value of G can only explain 12% of world income differences. This is in line with the role attributed to human capital in traditional development accounting. When I allow for imperfect substitutability, the share of income differences explained by differences in the value of G increases to 65%. The importance of TFP-differences decreases correspondingly, and the estimated difference in log TFP between rich and poor countries falls by 66%.

---

[5]The equation expresses output per worker as a function of the capital-output ratio rather than the capital-labor ratio. This follows Hall and C Jones 1999 and takes into account the indirect effect of TFP and human capital on capital accumulation. To separate between TFP and unskilled labor quality Q_u, my baseline analysis follows Hall and C Jones 1999 and assumes that unschooled workers are similar across countries. If Q_u is higher in rich countries, this would further reduce the differences in A across rich and poor countries.

The results are driven by large rich-poor differences in the efficiency of skilled workers. In my estimation, these efficiency differences are due to combining moderate rich-poor differences in the skilled wage premium with large rich-poor differences in the relative price of skilled services. The trade data estimates suggest that rich countries have 4-5 lower log relative prices of skilled services. This difference can either be explained by rich countries having low skilled wage premia, or by rich countries having relatively efficient skilled labor. Wage data suggest that skilled wage premia are indeed lower in rich countries, but not more than approximately one log point lower. Thus, I impute a 3-4 log-point difference in the relative efficiency of skilled labor between rich and poor countries. Intuitively, the German skill premium is not sufficiently low to rationalize high exports of engineering products from Germany, and thus I impute a relatively high efficiency of German skilled labor.

Finally, I discuss the interpretation of the results. I show that the interpretation depends on the source of skilled labor efficiency differences. If I retain the traditional development accounting assumption of skill-neutral technology differences, I find large differences in the human capital of skilled workers, and that human capital explains a majority of world income differences. If I relax the assumption of skill-neutral technology differences, an alternative explanation is that skilled labor efficiency differences reflect skill-biased technology differences. With this interpretation, theories of income differences should still be theories of technology differences, but they should focus more on the interaction between the economic environment and the efficiency of skilled workers. In Section 1.4, I discuss these different interpretations, and potential roads forward in discriminating between different sources of skilled labor efficiency differences.

The outline of the paper is as follows. Section 1.2 develops the estimation strategy for the relative price of skilled labor services r_s/r_u. Section 1.3 presents the development accounting results. Section 1.4 discusses the alternative economic interpretations of my results,

focusing on the interpretation of skilled labor efficiency differences as depending on human capital or skill-augmenting technology differences. Section 1.5 discusses in greater detail the relationship between my paper and B Jones (2014a) who also studies development accounting with imperfectly substitutable labor services. Section 1.6 performs a large number of robustness checks on the baseline results, and Section 1.7 concludes the paper.

Related literature. My paper is part of the development accounting literature, going back to Klenow and Rodriguez-Clare (1997) and Hall and C Jones (1999). This literature is surveyed in Caselli (2005), Hsieh and Klenow (2010a), and C Jones (2015). There has been a number of papers revisiting the contribution of human capital in development accounting, most often in a framework featuring perfect substitutability between different types of labor services. These papers include Hendricks (2002), Erosa et al. (2010), Schoellman (2011), Manuelli and Seshadri (2014), and Hendricks and Schoellman (2016).

A few papers have analyzed development accounting with imperfectly substitutable labor services. These papers include Caselli and Coleman (2006), Caselli and Ciccone (2013), Jones (2014a), and Caselli (2015).

Beyond development accounting, my paper builds on the gravity trade literature to estimate the relative prices of skilled services (Tinbergen, 1962; Anderson et al., 1979; Eaton and Kortum, 2002; Anderson and van Wincoop, 2003; Redding and Venables, 2004; Costinot et al., 2011; Head and Mayer, 2014). A number of papers have used trade data to obtain information about productivities, including Trefler (1993) and Levchenko and Zhang (2016). Morrow and Trefler (2017) is a more recent contribution that integrates trade into development accounting. My paper also relates to the literature that uses industry data to obtain information about economic development, which includes Rajan and Zingales (1998) and Ciccone and Papaioannou (2009). In the context of trade, papers that analyze the relationship between country variables and the industrial structure of trade include Romalis (2004), Nunn (2007), Chor

(2010), Cuñat and Melitz (2012), and Manova (2013). This literature is reviewed in Nunn and Trefler (2015).

1.2 Estimating the relative price of skilled services

The aim of this section is to estimate how the relative price of effective skilled and unskilled labor services r_s/r_u varies across countries. For this purpose, I construct a method for estimating relative factor service prices in general.

My estimation strategy is based on two premises. The first premise is that relative factor service prices influence relative unit production costs. To illustrate this, we can consider a case with two industries. Consider Table 1.1, which shows the factor shares for "Cut and Sew Apparel" (NAICS code 3152) and "Communications Equipment" (NAICS code 3342). Production of Communications Equipment is more skill intensive than production of Cut and Sew Apparel. If the relative price of skilled services rises, we can expect a rise in the relative unit production cost of Communications Equipment as compared to that of Cut and Sew Apparel.[6]

The second premise is that relative unit production costs affect relative export flows, which is a version of the principle of comparative advantage. For example, consider Table 1.2, which presents a number of US and Indonesian export values to Japan. Relative Indonesian-US exports are much higher in Cut and Sew Apparel as compared to Communications Equipment. Applying the principle of comparative advantage, this evidence suggests that Indonesia has a high relative unit production cost of Communications Equipment.

In combination, my two premises suggest that trade data contain information about relative factor service prices. For example, the trade data in Table 1.2 suggest that Indonesia has a high relative unit production cost of Communications Equipment. Furthermore, factor shares in

---

[6]The cost shares are defined as shares of gross output. In Appendix 1.C.3, I describe an alternative method where I decompose the non-tradable component of the intermediate input cost share into cost shares of other inputs using an input-output table. The final results are not affected by whether I use the basic cost shares or perform such a decomposition.

Table 1.1 suggest that Communications Equipment production is more skill intensive than Cut and Sew Apparel production. These two facts together suggest that Indonesia has a high relative price of skilled services.

My estimation strategy formalizes and generalizes this method of obtaining information about relative factor service prices using relative export values conditional on trade destination. For this purpose, I rely on a gravity trade model. My main result is that using a version of a gravity trade model, it is possible to identify relative factor service prices using:

1. Industry factor shares

2. Bilateral industry trade data

3. The price elasticity of export flows

One particular feature of my estimation strategy is that relative unit costs are estimated from trade data. This estimation choice reflects the lack of a data set that provides detailed cross-country comparable industry unit cost data, which cover both rich and poor countries. The best available data set comes from the Groningen Growth and Development Center, which has done important work in constructing a data set of industry unit costs for cross-country comparisons (Inklaar and Timmer, 2008). However, their data set only covers 35 industries in 42 countries, with a limited coverage of poor countries. In contrast, trade data are recorded at a highly detailed industry level in both rich and poor countries. This makes trade data an attractive source of information for development accounting. In Section 1.6.3, I show that for countries where we have both unit cost data and trade data, analyses using unit cost data and trade data yield similar results.

1.2.1 Setup

This section describes the setup of my estimation exercise. The notation is summarized in Table 1.3.

Table 1.1: Factor shares for Cut and Sew Apparel and Communication Equipment

	Cut and Sew Apparel	Communications Equip.
Factor services (f)	US factor shares	US factor shares
Unskilled labor	0.08	0.03
Skilled labor	0.05	0.11
Capital	0.32	0.39
Intermediate inputs	0.54	0.46
Energy	0.01	0.01
Sum	1.00	1.00

Table 1.2: Selected export values from Indonesia and USA to Japan (thousands of US dollars)

Origin	Destination	Industry	Export value
Indonesia	Japan	Cut and Sew Apparel	565,993
USA	Japan	Cut and Sew Apparel	197,100
Indonesia	Japan	Communications Equip.	16,503
USA	Japan	Communications Equip.	236,103

There are $I = 103$ countries, and each country has $K = 84$ indus-tries.[7] The industries correspond to NAICS four-digit manufacturing in-dustries. I observe the value of trade flows $x_{i,j}^k$ from country i to country j in industry k. Each industry produces a good using $F = 5$ factor services. In my baseline analysis, these are services from unskilled labor, skilled labor, capital, intermediate inputs, and energy. $r_{i,f}$ denotes the price of factor service f in country i. The unit production cost c_i^k of industry k in country i is a function of factor service prices. The relationship is given by

$$c_i^k = \frac{C^k(r_{i,1}, \ldots, r_{i,F})}{Z_i}.$$

This assumption implies that there is an industry cost function C^k that is common across countries. In an individual country, the unit cost func-tion c_i^k is derived by deflating the common industry cost function C^k with a country-specific productivity term Z_i, which is common across industries. This particular setup implies that cross-country differences in relative unit costs only stem from cross-country differences in relative factor service prices. However, my development accounting results are not affected if the setup is modified to allow for cross-country differences in factor augmenting technologies.[8]

1.2.2 Key equations

My estimation builds on the following two equations:

$$\log(x_{i,j}^k) = \delta_{i,j} + \mu_j^k - (\sigma - 1)\log(c_i^k) \tag{1.2}$$

$$\log(c_i^k) = \log(c_{US}^k) + z_i + \sum_{f=2}^{F} \alpha_{US,f}^k \log\left(\frac{r_{i,f}/r_{i,1}}{r_{US,f}/r_{US,1}}\right), \tag{1.3}$$

---

[7]The countries correspond to the countries with available data on export val-ues, output levels, capital stocks, schooling levels, and shares of workers in skilled occupations.

[8]In Section 1.6.1, I discuss regression specifications that address other potential confounders in the specification of unit costs.

Table 1.3: Notation

Variable	Description
i	Origin country
j	Destination country
k	Industry
f	Factor service ($f = 1$ unskilled labor services)
$x_{i,j}^k$	Export value of industry k from country i to country j
$r_{i,f}$	Factor service price of factor f in country i
$\alpha_{i,f}^k$	Cost share of factor f in industry k in country i
c_i^k	Unit cost of industry k in country i
σ	Price elasticity of trade

where $z_i = \log\left(\frac{r_{i,1}}{r_{US,1}}\right) - \log\left(\frac{Z_i}{Z_{US}}\right)$ is the log deviation in unskilled labor service prices, adjusted for absolute productivity differences. The first equation (1.2) is a gravity trade equation. The log export value from country i to country j in industry k depends on three terms. The first term is a bilateral fixed effect $\delta_{i,j}$. It captures determinants of overall bilateral trade flows such as the size of the two countries, their bilateral distance, common legal origins, shared language, etc. The second term is a destination-industry fixed effect μ_j^k, which captures the demand for good k in destination j, as well as how good access country j has to industry k, given its other trading partners. The third term captures that conditional on the two fixed effects, exports depend negatively on origin unit production costs, with a price elasticity $\sigma - 1$. In Appendix 1.C.1, I show how equation (1.2) can be derived from both a trade model in the style of Eaton and Kortum (2002), where trade is driven by country-variety specific productivity shocks, and from an Armington model where each country produces a unique variety of each good k.

The second equation (1.3) is a log-linear approximation of industry unit costs around the US cost structure, where $f = 1$ indexes unskilled

labor services. I obtain the approximation in two steps. I first note that

$$
\begin{aligned}
\log(C^k) \quad &\approx \quad \log(C^k) + \sum_{f=1}^{F} \frac{\partial C^k}{\partial r_f} \frac{r_{US,f}}{C^k} \log\left(\frac{r_{i,f}}{r_{US,f}}\right) \\
&= \quad \log(C^k) + \sum_{f=1}^{F} \alpha_{US,f}^{k} \log\left(\frac{r_{i,f}}{r_{US,f}}\right)
\end{aligned}
$$

where C^k is the common cost function of industry k, and $\alpha_{US,f}^{k}$ denotes the US factor share of factor f in industry k. The second line uses Shepherd's lemma applied to the cost function to conclude that $\alpha_{US,f}^{k} = \frac{\partial C^k}{\partial r_f} \frac{r_{US,f}}{C^k}$ when firms are price-takers.

Combining this expression with $c_i^k = \frac{C^k(r_{i,1},...,r_{i,F})}{Z_i}$ gives me

$$
\log(c_i^k) = \log(c_{US}^k) + \log\left(\frac{Z_i}{Z_{US}}\right) + \sum_{f=1}^{F} \alpha_{US,f}^{k} \log\left(\frac{r_{i,f}}{r_{US,f}}\right). \tag{1.4}
$$

I re-arrange this equation to equation (1.3), as my aim is to find the *relative* price of factor services compared to unskilled labor services, $\log\left(\frac{r_{i,f}/r_{1,f}}{r_{US,f}/r_{US,1}}\right)$. This makes it useful to normalize equation (1.4) with the price of unskilled labor services. I use the fact that factor shares sum to 1 to express the unskilled cost share $\alpha_{US,1}^{k}$ in terms of the other cost shares: $\alpha_{US,1}^{k} = 1 - \sum_{f=2}^{F} \alpha_{US,f}^{k}$. Substituting this expression into (1.4) gives me equation (1.3).

Equation (1.3) decomposes log unit cost differences from the US into one term capturing absolute productivity differences, one term capturing differences in the cost of unskilled labor, and a linear combination of relative factor service price differences times US factor shares. Equation (1.3) shows that countries with a relatively high factor service price in factor f (high $\log\left(\frac{r_{i,f}/r_{i,1}}{r_{US,f}/r_{US,1}}\right)$) will have relatively high unit costs in sectors intensive in factor f (relatively high $\alpha_{US,f}^{k}$).

As explained, equation (1.3) comes from a log linear approximation around the US cost structure. If industry production functions are

Cobb-Douglas, this approximation is exact. If industry production functions are not Cobb-Douglas, there is a second-order bias. In Section 1.6.1, I analyze the effect of relaxing the Cobb-Douglas assumption.

1.2.3   Regression specification

To derive my regression specification, I combine the gravity equation (1.2) and the unit cost equation (1.3). I obtain

$$\log(x_{i,j}^k) = \tilde{\delta}_{i,j} + \tilde{\mu}_j^k - (\sigma - 1) \sum_{f=2}^{F} \alpha_{US,f}^k \log\left(\frac{r_{i,f}/r_{i,1}}{r_{US,f}/r_{US,1}}\right).$$

Here, $\tilde{\delta}_{i,j} = \delta_{i,j} - (\sigma - 1)\left(\log\left(\frac{r_{i,1}}{r_{US,1}}\right) - \log\left(\frac{Z_i}{Z_{US}}\right)\right)$ denotes a modified fixed effect that includes the trade bilateral fixed effect, the origin absolute advantage, and the origin unskilled factor service prices. The term $\tilde{\mu}_j^k = \mu_j^k - (\sigma - 1)\log(c_{US}^k)$ denotes a modified fixed effect that includes the trade destination-industry fixed effect μ_j^k and US industry unit costs.

I can use this equation to derive a regression specification. For this purpose, I note that I can measure $x_{i,j}^k$ from international trade data, that I can measure $\alpha_{US,f}^k$ from American industry data, and that I can use the trade literature to obtain estimates of σ.[9] Thus, $\log(x_{i,j}^k)$ is my left-hand variable, and $(\sigma - 1)\alpha_{US,f}^k$ for $f = 2, \ldots, F$ are my explanatory variables. My aim is to estimate the relative factor service price differences $\log\left(\frac{r_{i,f}/r_{i,1}}{r_{US,f}/r_{US,1}}\right)$. This quantity varies on a country-factor basis. Therefore, I want to estimate one parameter for each factor-country combination, and I write $\beta_{i,f}$ for this set of parameters. Given the interpretation of $\beta_{i,f}$ as differences in relative factor service prices compared to those in the US, I normalize $\beta_{i,f}$ by setting $\beta_{US,f} = 0$ for all f.

---

[9]Some papers estimate σ directly from trade data (Broda et al., 2006; Soderbery, 2015), exploiting short-run variations in trade prices and quantities. As I am interested in the long-run elasticity of trade, I choose a calibration approach to select σ.

I obtain the following specification:

$$\log(x_{i,j}^k) = \tilde{\delta}_{i,j} + \mu_j^k - \sum_{f=2}^{F} \left[(\sigma - 1)\alpha_{US,f}^k \right] \times \beta_{i,f} + \varepsilon_{i,j}^k, \qquad (1.5)$$

with the normalization $\beta_{US,f} = 0$ for $f = 2, \ldots, F$. I regress log bilateral trade flows on a bilateral fixed effect, a destination-industry fixed effect, and $-(\sigma - 1)\alpha_{US,f}^k$ for $f = 2, \ldots, F$, allowing for country-factor specific parameters $\beta_{i,f}$. In total, I estimate $(5-1) \times 103 = 412$ parameters: one for each country-factor combination, excluding unskilled labor services. With this regression specification, $\beta_{i,f} = \log\left(\frac{r_{i,f}/r_{i,1}}{r_{US,f}/r_{US,1}}\right)$ identifies the difference between country i and the US in the log relative price of factor service f as compared to unskilled labor services. The difference to the US in the log relative price of skilled labor services is identified by $\beta_{i,skill}$.

1.2.4 Data in trade regression

The regression equation (1.5) requires data on bilateral trade flows $x_{i,j}^k$, US factor shares $\alpha_{US,f}^k$, and a parameter estimate for the trade elasticity σ.

For trade flows, I use the BACI data set which is compiled by CEPII and based on COMTRADE (Gaulier and Zignago, 2010). For each country-destination pair, it reports export values at the HS 2007 six-digit industry level. I use data for 2010.

I measure factor shares by combining data from the NBER-CES Manufacturing Industry Database (Bartelsman and Gray, 1996) with data from the Occupational Employment Statistics (OES) survey. I use the NBER-CES database to obtain the cost shares of capital, labor, materials, and energy. I define the shares of labor, materials, and energy as factor outlays divided by industry gross output, and I define the capital share as 1 minus the other factor shares. To find the shares of skilled and unskilled services, I use the OES to calculate the share of payroll in each industry that goes to workers in occupations with skill levels 3 and 4 in the ISCO-08 classification. This corresponds to the major

occupational groups "Managers", "Professionals", and "Technicians and Associate Professionals". I calculate the skill share as the labor share from the NBER CES times the share of payroll going to skilled workers, and the unskilled share as the labor share times the share of payroll going to unskilled workers. Note that in my regression, I include the materials and energy shares in the regression. Appendix 1.C.3 provides a more detailed discussion of different choices of intermediate input measurement and their effects.

The regression is performed using NAICS four-digit coding, which is the coding scheme of the OES industry data. The trade data are recorded using HS6 codes and the NBER-CES data are recorded using NAICS six-digit codes. The OES occupational data are recorded according to SOC, and they are converted to ISCO-08 to calculate the share of payroll going to skilled workers.

I take my value of the trade elasticity σ from the literature. I look for an estimate of the *long-run* elasticity *between different foreign varieties in the same industry*. This choice reflects the nature of my regression. The regression is run between countries in different parts of the world-income distribution, and aims at capturing persistent cross-country differences. Furthermore, the regression explains a source country's exports conditioned on the total industry imports of a destination country. Thus, the relevant elasticity is the long-run elasticity between different foreign varieties.

I select $\sigma = 10$ as my baseline elasticity. This is a reasonably high estimate of trade elasticity and reflects a conservative choice for estimating the importance of human capital. A higher σ shrinks the importance of human capital since it reduces the estimated differences between countries: differences in relative trade flows translate into smaller unit cost differences. Even though Eaton and Kortum (2002) open up for estimates as high as $\sigma = 14$, my estimate is higher than $\sigma = 5$ found in Simonovska and Waugh (2014), $\sigma = 7.2$ found in Costinot et al. (2011), and the baseline $\sigma = 9.2$ found in Eaton and Kortum (2002).[10]

---

[10]Note that the trade elasticity θ in Eaton and Kortum-style models represents

My baseline estimate $\sigma = 10$ corresponds to the higher range estimates found in Romalis (2007) when he estimates the trade effects of NAFTA. He calculates a pooled trade elasticity by investigating how differential reductions of tariffs due to NAFTA affected trade in the quadrangle USA, Canada, Mexico, and the EU. I select this high estimate to be conservative and due to the fact that the long-run effects of NAFTA studied by Romalis (2007) reflect the type of long-run, foreign-to-foreign substitution that my regression specification seeks to capture. In Section 1.6.1, I discuss the effects of making different assumptions about σ.

1.2.5   Results from trade regression

My main results are displayed in abridged form in Table 1.4. The table presents log relative factor service price estimates for different factors, and for six randomly selected countries in each World Bank Income group. Standard errors are calculated by clustering at the industry-country level.

The table shows that poor countries in general have higher relative factor service prices for skilled services, capital services, and intermediate input services. The pattern is especially pronounced for skilled services. There is some tendency for relative energy service prices to be higher in poor countries, but this pattern is less clear. Relative energy service prices vary more between similar countries and are less precisely estimated.

My primary interest is in the relative prices of skilled services, since these are used in my development accounting exercise. In Figure 1.1, I provide a graphical illustration of the relationship between estimated relative skilled service prices and log GDP per worker. There is a strong negative relationship, and poor countries have approximately 4-5 log points higher relative prices of skilled services. If I take standard errors into account, the results are consistent with a stable, almost linear,

---

the elasticities of export value with respect to price changes, whereas σ represents the elasticity of quantity with respect to price changes. Hence, $\sigma = \theta + 1$ when we convert between the two types of parameters.
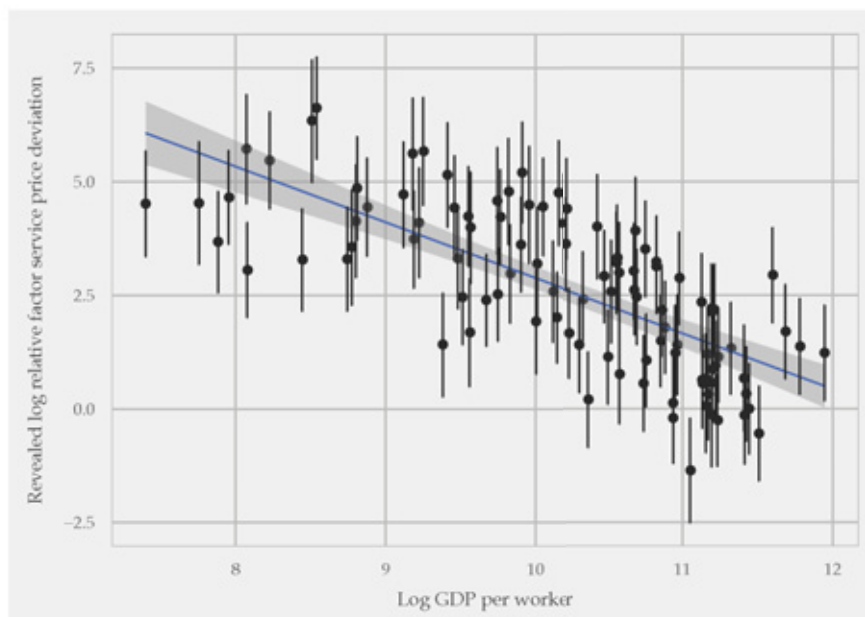
Figure 1.1: Log GDP per worker and $\log\left(\frac{r_{i,skill}/r_{i,unskill}}{r_{US,skill}/r_{US,unskill}}\right)$

relationship between log GDP per worker and the log relative price of skilled services. There is a less clear relationship between log GDP per worker and skilled service prices for the very poorest countries, which could reflect that manufacturing exports are relatively unimportant in these countries. In Section 1.6.3, I analyze the effect of excluding the poorest countries from the analysis, and I show that this increases the estimated importance of human capital.

Even though I do not use the other factor service price estimates in my development accounting exercise, they can be used to evaluate the estimation method. Appendix 1.C.2 discusses the results for other factors in greater detail.

Table 1.4: Regression estimates of log relative factor service price parameters (US = 0)

	Factor services			
	Skilled labor	Capital	Intermediate inputs	Energy
Low income				
Gambia	5.64 (1.01)	2.53 (0.68)	2.56 (0.57)	−1.17 (1.01)
Liberia	4.56 (1.09)	2.73 (0.77)	2.28 (0.64)	3.03 (1.43)
Nepal	5.89 (1.15)	2.83 (0.77)	2.85 (0.74)	5.35 (1.78)
Rwanda	4.54 (1.31)	1.40 (0.75)	1.55 (0.72)	3.04 (1.78)
Tanzania	3.72 (1.03)	1.45 (0.69)	1.54 (0.55)	0.21 (1.25)
Uganda	3.09 (0.96)	0.82 (0.64)	0.93 (0.52)	1.44 (1.32)
Lower middle income				
Indonesia	3.78 (0.98)	1.41 (0.64)	1.65 (0.57)	0.47 (1.20)
Pakistan	4.92 (1.08)	2.14 (0.76)	2.35 (0.68)	2.12 (1.61)
Philippines	1.52 (1.06)	0.57 (0.73)	1.02 (0.58)	1.91 (1.20)
Tunisia	2.62 (1.03)	1.61 (0.66)	1.54 (0.55)	0.81 (1.29)
Ukraine	2.45 (0.92)	0.92 (0.63)	0.96 (0.52)	−2.30 (1.49)
Vietnam	3.60 (1.21)	2.15 (0.78)	2.39 (0.67)	3.10 (1.35)
Upper middle income				
Colombia	3.74 (0.96)	1.01 (0.59)	1.37 (0.50)	−0.87 (1.11)
Dominican Republic	3.27 (1.17)	0.65 (0.73)	1.40 (0.64)	0.75 (1.38)
Paraguay	5.67 (1.18)	1.21 (0.71)	1.27 (0.67)	1.59 (1.82)
Russia	1.12 (0.95)	0.001 (0.65)	−0.10 (0.55)	−4.57 (1.19)
South Africa	1.67 (0.90)	0.46 (0.58)	0.43 (0.47)	−1.61 (1.08)
Turkey	3.59 (0.97)	1.95 (0.60)	2.09 (0.49)	0.40 (1.18)
High income				
Chile	4.13 (1.09)	0.54 (0.65)	0.65 (0.56)	−1.20 (1.46)
Ireland	−0.10 (0.99)	−1.05 (0.68)	−0.53 (0.61)	0.21 (1.66)
Netherlands	0.59 (0.88)	−0.45 (0.56)	−0.17 (0.45)	−0.24 (1.00)
New Zealand	1.54 (0.91)	0.62 (0.62)	0.32 (0.58)	1.15 (1.19)
Taiwan	−0.10 (1.07)	1.37 (0.67)	1.27 (0.58)	0.58 (1.31)
United States	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Observations	453,147			
R^2	0.69			

Note: Standard errors are clustered on origin-industry level

1.3   Development accounting

In this section, I want to use the estimates from Section 1.2 to perform a development accounting exercise. The aim is to decompose the variance in GDP per worker into contributions from differences in capital-output ratios, uniform labor productivity shifters and differences in human capital aggregators.

1.3.1   Aggregate production function

Performing development accounting requires me to put more theoretical structure on the aggregate economy. In Section 1.2, the key assumption was that trade flows followed a gravity relationship. This assumption is consistent with a range of models for the aggregate economy. In contrast, for development accounting, I need an aggregate production function that summarizes both substitution possibilities within and between industries and substitution possibilities between domestic and foreign production.

It can be shown that standard trade models admit a constant returns to scale aggregate production function under relatively mild assumptions (see Appendix 1.A.2), but there is no general result on the functional form of such an aggregate production function. In general, the aggregation of capital, skilled labor, and unskilled labor depends on trade elasticities, the input-output structure, the relative factor shares of traded and non-traded goods, and how traded and non-traded goods are aggregated.

In my baseline specification, I follow the development accounting literature by making the assumption that the aggregate production function can be well approximated by a Cobb-Douglas composite of labor and capital. Moreover, I assume that average human capital is an arbitrary constant returns to scale function of the number and efficiencies of skilled and unskilled workers. This assumption allows me to focus on how labor input aggregation changes the development accounting exercise.

Thus, I posit an aggregate production function:

$$Y = K^\alpha (ALh)^{1-\alpha}, \tag{1.6}$$

where Y is total output, K is the physical capital stock, A is a labor-augmenting technology term, L is the size of the labor force, and h captures the average human capital of the labor force. I allow for a flexible specification of h

$$h = G(Q_u u, Q_s s).$$

Here, G is a constant returns to scale aggregator, u and s are the shares of unskilled and skilled workers, and Q_u and Q_s are the amount of unskilled/skilled services delivered by each unskilled/skilled worker. I will refer to Q_u and Q_s efficiencies of unskilled and skilled labor.

The human capital term h has two potential interpretations. One interpretation of the aggregator is that there are two homogenous skill types, with the respective efficiencies Q_u and Q_s. A second interpretation is that there are two aggregators $H_u = Q_u u$ and $H_s = Q_s s$, which combine heterogeneous types of services into an aggregate flow of unskilled and skilled services. With this interpretation, Q_u and Q_s represent the average flow of unskilled/skilled labor services per unit of unskilled/skilled labor, and w_s and w_u are the average wages of skilled and unskilled workers.

The interpretation with two types of labor services is easier to discuss, whereas the aggregator interpretation is more realistic. I will derive my results in the language of the interpretation with two labor types. I will refer to u and s as the share of unskilled and skilled workers, and to Q_u and Q_s as the (average) efficiencies of unskilled and skilled labor. When I analyze economic mechanisms, I will leverage the mathematical equivalence to interpret my results in light of the aggregator interpretation.

1.3.2 Estimating terms of aggregate production function

To perform development accounting, we need to measure Y, L, α and h.

Data on real output Y, labor force size L, and physical capital stock K are from the Penn World Table Version 8.1. I use data from 2010, and I set the capital share α to 1/3.

To measure $h = G(Q_u u, Q_s s)$, I use data from ILO to measure the share of skilled workers s. I define the share of skilled workers as the share of workers having an occupation requiring skill level 3 or 4. According to ILO, occupations require skill level 3 or 4 when they "typically involve the performance of [...] tasks that require an extensive body of [...] knowledge in a specialized field". In the International Standard Classification of Occupations 2008 (ISCO-08), these are "Managers", "Professionals", and "Technicians and Associate Professionals". Figure 1.2 shows the relationship between the share of skilled workers and log GDP per worker. There is a strong positive relationship, and a linear regression of the skill share on log GDP per worker has an R^2-value of 0.75. My skill definition differs from the literature in being occupation-based instead of schooling-based. I discuss this choice in Appendix 1.B.1.

I calibrate the quality of unskilled labor Q_u using data on schooling levels and Mincerian returns. I define

$$Q_u = e^{\phi(S_u)}, \tag{1.7}$$

where S_u is the average schooling years of unskilled workers, and ϕ is a Mincerian return function capturing the relationship between schooling and wages. I measure S_u using the Barro-Lee schooling data for 2010. I assume that there is perfect positive sorting between years of schooling and working in a skilled profession, which means that unskilled workers correspond to the $1 - s$ share of the workforce with the least schooling. I assume that S_u is the average number of school years in this group.[11] I take the Mincerian return function $\phi(S)$ from Caselli (2005) and define

---

[11]See Appendix 1.B.2 for details on how I calculate the average schooling of unskilled labor.

it as a piecewise linear function with slope 0.13 for $S < 4$, slope 0.1 for $S \in [4, 8)$, and slope 0.08 for $S \geq 8$. This specification was introduced in the literature as a reduced form way of capturing that poor countries have higher Mincerian returns.

To measure the relative efficiency of skilled workers Q_s/Q_u, I assume that labor markets are competitive. This means that the relative wage of skilled and unskilled workers is

$$\frac{w_s}{w_u} = \frac{Q_s}{Q_u} \frac{r_s}{r_u},$$

where

$$\frac{r_s}{r_u} = \frac{f_s}{f_u}.$$

I measure the relative efficiency of skilled and unskilled labor using the equation

$$\frac{w_s}{w_u} = \frac{Q_s}{Q_u} \frac{r_s}{r_u} \iff \frac{Q_s}{Q_u} = \frac{w_s/w_u}{r_s/r_u}. \tag{1.8}$$

The skilled wage premium w_s/w_u is observable, and the relative price of skilled services $\frac{r_s}{r_u}$ was estimated in Section 1.2.[12] This equation states that the skill premium equals the relative *amount* of services provided by skilled and unskilled workers, times the relative *price* of those services. Re-arranging the equation shows that there will be a high estimate of relative skilled labor efficiency if either a) the skill premium is high given the price of skilled services, since this reflects a large amount of services being delivered, or b) if observed skilled service prices are low given the skill premium, since this reflects a high efficiency of skilled labor, thus bringing down the efficiency adjusted price.

I use ILO data to measure the skilled wage premium w_s/w_u. ILO summarizes wage data from multiple sources, and I restrict attention to

---

[12]In Section 1.2, I estimated $\frac{r_s/r_u}{r_{US,s}/r_{US,u}}$. To find r_s/r_u, I need $r_{US,s}/r_{US,u}$. I find this by normalizing US skilled labor efficiency to $Q_{US,s} = 1$. This implies:

$$\frac{r_{US,s}}{r_{US,u}} = \left(\frac{1}{Q_{US,u}}\right)^{-1} \frac{w_{US,s}}{w_{US,u}}.$$

countries where data are available from administrative records, a labor-focused establishment survey, and/or a labor force survey. I use the measure of mean nominal monthly earnings of employees. I combine data on wages and employment across occupations, and I calculate the relative average wage between workers with skill levels 3 or 4 and workers with skill levels 1 or 2. Figure 1.3 shows the relationship between log skilled wage premia and log GDP per worker. Apart from two outliers (Vietnam and Qatar), there is a strong negative relationship. The ILO data only cover a limited set of countries, and there are large variations between countries with similar levels of log GDP per worker. In my development accounting exercise, I want to use a large set of countries, and I am interested in systematic differences between rich and poor countries. Thus, to assign values of the skilled wage premium, I regress the log skilled premium on log GDP per worker (excluding outliers). I assign each country a skilled premium using the fitted value of this regression. This allows me to extend the country coverage beyond the limited set of countries covered in the ILO data, while capturing the systematic changes of the skilled wage premium across the GDP per worker distribution. In Section 1.6.3, I consider how changes in the measurement of the skilled wage premium change my results.

With estimates of r_s/r_u, Q_s/Q_u, and s/u, it is possible to estimate the functional form and parameters of the human capital aggregator G. As G has constant returns to scale, it is characterized up to a constant by the relationship between the marginal rate of transformation f_s/f_u and the relative effective labor supplies $\frac{Q_s s}{Q_u u}$. In particular, a linear relationship between $\log\left(\frac{f_s}{f_u}\right)$ and $\log\left(\frac{Q_s s}{Q_u u}\right)$ suggests that G has a constant elasticity of substitution.

Using the estimates of r_s/r_u from Section 1.2, we can plot the relationship between the (log) marginal rate of transformation and (log) relative effective labor supplies. However, we need to be cautious in interpreting this relationship when relative qualities Q_s/Q_u are derived using equation (1.8). The reason is that relative supplies are calculated
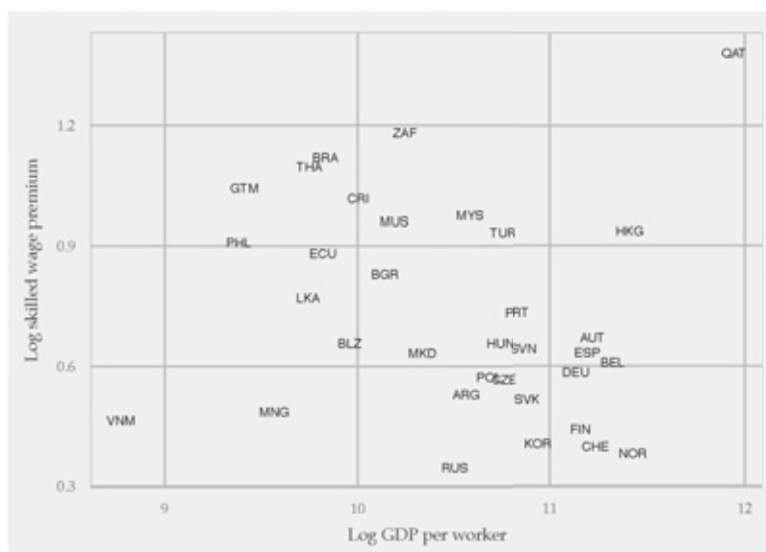
Figure 1.2: Log skilled wage premia versus log GDP per worker

Figure 1.3: Log relative share of workers in skill level 3+4 vs log GDP per worker
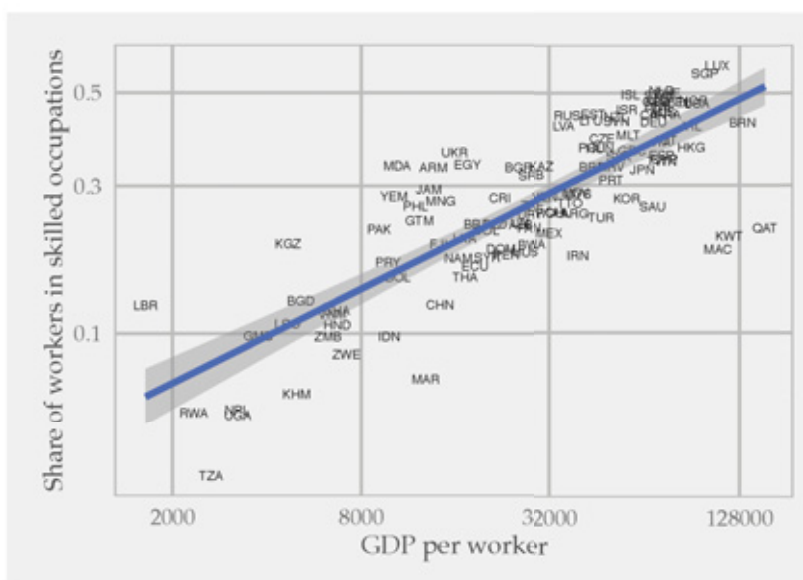
by dividing skilled wage premia by r_s/r_u. This means that if there are measurement errors in r_s/r_u, this could cause a bias of the relationship between $\log\left(\frac{r_s}{r_u}\right)$ and $\log\left(\frac{Q_s s}{Q_u u}\right)$ to be negatively linear with slope -1 due to division bias. One way of addressing this concern is to instrument log relative factor service supplies $\log\left(\frac{Q_s s}{Q_u u}\right)$ with log GDP per capita. Figure 1.5 shows the result both with and without instrumentation. The left-hand panel shows $\log\left(\frac{r_s}{r_u}\right)$ plotted against $\log\left(\frac{Q_s s}{Q_u u}\right)$. The right-hand panel replaces $\log\left(\frac{Q_s s}{Q_u u}\right)$ with the fitted values from a first-stage regression of $\log\left(\frac{Q_s s}{Q_u u}\right)$ on $\log(y)$ and $\log^2(y)$. Both plots suggest an approximately linear relationship between $\log\left(\frac{r_s}{r_u}\right)$ and $\log\left(\frac{Q_s s}{Q_u u}\right)$. Thus, I posit that human capital is aggregated using a constant elasticity of substitution aggregator[13]

$$G(Q_u u, Q_s s) = \left((Q_u u)^{\frac{\eta-1}{\eta}} + a_S(Q_s s)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}.$$

To obtain estimates of the parameters a_s and η, I note that

$$\frac{w_s}{w_u} = \frac{a_S Q_s^{1-1/\eta}}{Q_u^{1-1/\eta}}\left(\frac{s}{u}\right)^{-1/\eta} \Leftrightarrow \log\left(\frac{r_s}{r_u}\right) = \log(a_s) - \frac{1}{\eta}\log\left(\frac{Q_s s}{Q_u u}\right). \tag{1.9}$$

I recover $\log(a_s)$ and $-1/\eta$ as the intercept and slope from a cross-country regression of log relative service prices $\log\left(\frac{r_s}{r_u}\right)$, on log relative service supplies, $\log\left(\frac{Q_s s}{Q_u u}\right)$. This specification is a close cross-country analogue of the regression specification introduced in Katz and Murphy (1992). I estimate a skill share $a_s = 2.06$, and an elasticity of substitution $\eta = 1.27$.[14]

---

[13]As previously noted, G is only defined up to a multiplicative constant. This means that I cannot separate the levels of A and h, but it is still possible to estimate the relative sizes of A and h across countries.

[14]The difference between my estimation and Katz and Murphy is that I use the trade data to obtain an independent estimate of the labor-augmenting terms Q_s/Q_u, whereas Katz and Murphy (1992) identify η by assuming that there is a log-linear time trend in $\frac{Q_s}{Q_u}$ and they estimate the elasticity by deviations around this trend. To test the sensitivity of the estimates to division bias, I have also analyzed the relationship between relative skilled and unskilled factor shares and the relative price of skilled
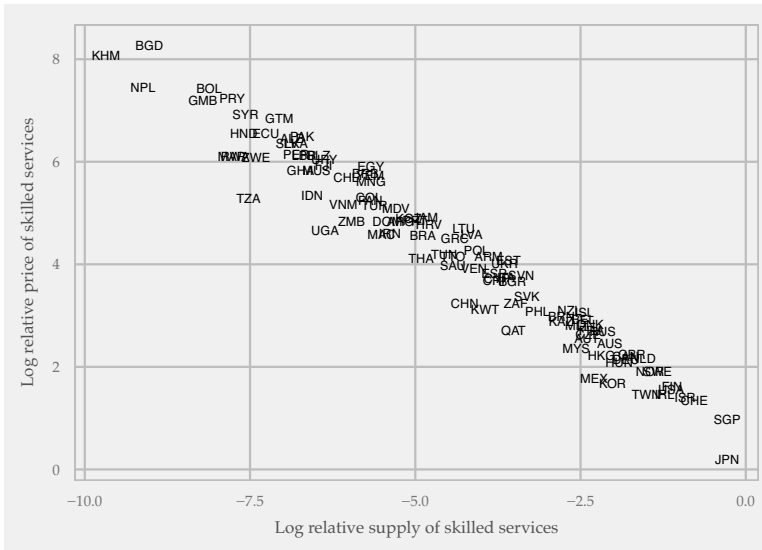
Figure 1.4: Log relative price of skilled services vs log relative effective skilled labor supply
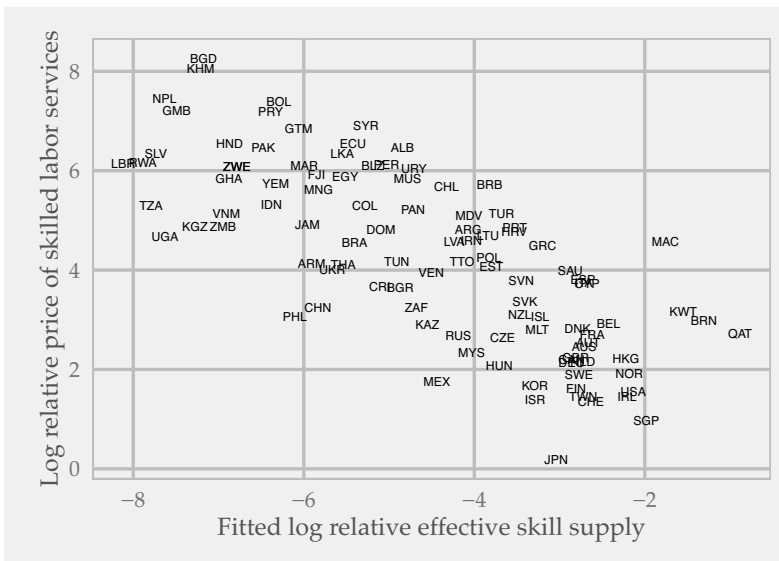
Figure 1.5: Log relative price of skilled services vs fitted log relative effective skilled labor supply

1.3.3 Results

In this section, I perform development accounting. My main outcome variables are the shares of world income differences accounted for by physical capital, the human capital aggregator G, and TFP differences. To evaluate how my human capital measurement method affects development accounting, I compare my results to those obtained when the human capital aggregator is additive in line with traditional development accounting methods.

To decompose income differences into contributions from factors and technology, I re-arrange the aggregate production function (1.6) into

$$y \doteq \frac{Y}{L} = \left(\frac{K}{Y}\right)^{\frac{\alpha}{1-\alpha}} Ah.$$

With this re-arrangement, aggregate output is expressed as a function of the capital-output ratio. This approach follows Hall and Jones (1999) and Hsieh and Klenow (2010b), and takes into account the steady-state effects of human capital and technology differences on capital accumulation.

The aggregate production function admits a linear decomposition of log output per worker:

$$\log(y) \equiv \log\left(\frac{Y}{L}\right) = \frac{\alpha}{1-\alpha} \log\left(\frac{K}{Y}\right) + \log(h) + \log(A).$$

Using this decomposition, I define the shares of income differences at-

---

and unskilled services. As the relative factor shares can be measured independently of r_s/r_u, this specification does not feature any division bias. I similarly find an approximate linear relationship, and similar results for the share of income differences explained by different values of the human capital aggregator G.

tributable to different factors:

$$
\begin{aligned}
\rho^K &= \frac{Cov\left(\frac{\alpha}{1-\alpha}\log\left(\frac{K}{Y}\right), \log(y)\right)}{Var(\log(y))} \\
\rho^h &= \frac{Cov\left(\log(h), \log(y)\right)}{Var(\log(y_i))} \\
\rho^A &= 1 - \rho^K - \rho^h.
\end{aligned}
$$

In addition to share parameters, I define a summary measure of TFP-differences between rich and poor countries. To define this measure, I regress log TFP on log GDP per worker which gives me predicted log TFP as a function of log GDP per worker. My definition of the rich-poor log TFP difference is the change in this predicted value between the 10th and the 90th percentile of the GDP per worker distribution. I write $\Delta\log(A)$ for this difference.

I also calculate the share parameters and the TFP differences using an alternative measure of the human capital aggregator h_{trad}, which is constructed in line with traditional development accounting methods. It is measured by converting skilled workers to unskilled equivalents using the skilled wage premium.[15] I define h_{trad} as

$$
h_{trad} = Q_u\left(u + s\frac{w_s}{w_u}\right), \tag{1.10}
$$

where unskilled labor quality Q_u is defined in equation (1.7).

To compare my measure h_{new} with the traditional development accounting measure h_{trad}, I compare how the share of world income differences explained by the human capital aggregator – ρ^h – changes when

---

[15]My calculation method is analogous to traditional development accounting as it calculates human capital using unskilled equivalents estimated using relative wages. The standard references in development accounting, Hall and C Jones (1999) and Caselli (2005), use a slightly different implementation as they use years of schooling as their skill measure instead of occupation, and they use Mincerian returns instead of occupation-based skilled wage premia to calculate wage differences. They define human capital as $h_i = \exp(\phi(S_i))$ where ϕ is a Mincerian return function and S_i is the average years of schooling in country i. In my setting, their method yields very similar results to using equation (1.10).

I change the human capital measure from h_{trad} to h_{new}. Furthermore, I estimate the reduction in log TFP differences between rich and poor countries when I change the human capital measure from h_{trad} to h_{new}. To measure this reduction, I define the share of TFP differences explained as

$$TFP_{share} = 1 - \frac{\Delta \log(A_{new})}{\Delta \log(A_{trad})}.$$

To interpret this measure, recall that $\Delta \log(A)$ refers to the difference in log TFP between rich and poor countries. If there are no remaining TFP differences between rich and poor countries with my method of measuring human capital, $TFP_{share} = 1$. If the TFP differences between rich and poor countries are the same with my method of measuring the human capital aggregator as with the traditional development accounting method, $TFP_{share} = 0$.

Table 1.5 presents the baseline results of my development accounting exercise. Capital-output variations explain 8% of world income differences. This share does not depend on the method of measuring human capital. The traditional development accounting method attributes 12% of world income differences to human capital, and 79% to TFP.[16] My method attributes 65% of world income differences to human capital, and only 26% to TFP. Estimated log TFP differences between rich and poor countries shrink by 67% when I change the human capital measurement method.

1.3.4   Intuition from country example: Tanzania

To make the development accounting results more concrete, I focus on what they mean for one poor country: Tanzania. In 2010, Tanzania had a GDP per worker of \$2650, which made it the 17th poorest country among the 165 countries in the Penn World Table. I ask the following

---

[16]This estimate is slightly above the $50\% - 70\%$ interval discussed in the review article by Hsieh and Klenow (2010a) and the 70% in the latest handbook chapter written by C Jones (2015). Four percentage points of the difference can be explained by the Mincerian method attributing 14% to human capital. I also use a later version of the Penn World Table and updated data.

Table 1.5: Contribution of factors and TFP to income differences: baseline parametrization

	Baseline
Capital	0.08
Human capital – trad.	0.12
Human capital – new	0.65
TFP – trad.	0.79
TFP – new	0.26
Log TFP diff. – trad.	2.54
Log TFP diff. – new	0.85
TFP-diff. reduction	67%
Elasticity of subst. η	1.27

question: how do different human capital measurement methods predict that Tanzanian GDP per worker would change if the skill levels of the Tanzanian workforce were increased to the levels of the US workforce, keeping the Tanzanian capital-output ratio and TFP constant?

I answer this question using both the traditional development accounting method of aggregating microeconomic returns to schooling as in Hall and C Jones (1999), and by using my way of measuring human capital (here, I assume that differences in the human capital aggregator reflect differences in human capital; in the next section, I discuss the alternative interpretation that it reflects skill-augmenting technology shifters).[17] Granted, it is a complex counterfactual to ceteris paribus increase the skill levels of Tanzanian workers to those of US workers – including specialized computer engineers, world-class researchers, the whole range of the US medical profession, financial experts, corporate lawyers, and so forth. However, the exercise illustrates the effect of varying the method of measuring human capital.

I start with the traditional development accounting approach. For

---

[17]I use the method of Hall and C Jones (1999) instead of equation (1.10). In this setting, they yield very similar results, but it is easier to explain the method of Hall and C Jones (1999) in this context.

2010, the Barro-Lee data estimates Tanzanian average schooling levels to be 5.81 years, and US average schooling levels to be 13.18 years, a difference of approximately 7.5 years. Using the Mincerian return function from Hall and C Jones (1999) and Caselli (2005), these schooling differences translate into an approximately 0.6 log point difference in human capital. Using the aggregate production function (1.6), log Tanzanian GDP per worker increases by the same amount.

This example illustrates that traditional development accounting does not attribute a dominant role to human capital in explaining world income differences. Even if Tanzania increases the skill levels of its workforce all the way to US skill levels, GDP per worker only increases by 0.6 log points, or to $4675. After this change in skill levels, Tanzanian income levels would not move higher than somewhere between Senegal and Bangladesh.

In contrast, my method estimates that there is an approximate 2.6 log point difference in human capital between the US and Tanzania. After increasing the skill levels of the Tanzanian workforce, Tanzania would have a GDP per worker of approximately $36,000, making it approximately as rich as Russia. The lower TFP of Tanzania would still make it substantially poorer than the US (with a GDP per worker of $93,000), but the change in its skill levels would make it an upper middle income country.

1.4   Interpretation of mechanism: High efficiency of skilled labor

1.4.1   Mechanism

Section 1.3.3 showed that my method of aggregating human capital attributes a much smaller share of world income differences to TFP-differences than traditional development accounting did. The key mechanism driving this result is that my method estimates a high efficiency of skilled labor in rich countries. Figure 1.6 shows the relationship be-

tween log GDP per worker and the efficiency of skilled labor according
to the traditional development accounting method which equates rela-
tive skilled labor efficiency with the skilled wage premium and the same
relationship according to my method, which also allows for differences
in the relative price of skilled services (in both cases, I normalize log US
skilled labor efficiency to 0). The figure shows that traditional develop-
ment accounting actually estimates that poor countries have a somewhat
higher efficiency of skilled labor than rich countries. This reflects higher
skilled wage premia in poor countries. My method paints a different pic-
ture. With my method, the efficiency of skilled labor is about four and a
half log points lower in poor countries as compared to rich countries. My
large estimated efficiency differences reflect large estimated differences
in relative skilled service prices. The relative price of skilled services and
the relative efficiency of skilled labor are related through

$$\frac{w_s}{w_u} = \frac{Q_s}{Q_u}\frac{r_s}{r_u},$$

where $\frac{w_s}{w_u}$ is the skilled wage premium. My trade data estimates suggest
that the relative price of skilled services $\frac{r_s}{r_u}$ is 4-5 log points lower in
rich countries. Skilled wage premia are also lower in rich countries, but
only approximately one log point lower. This means that the relative
efficiency of skilled labor is 3-4 log points higher in rich countries. My
results follow from combining this finding with the 0.5 rich-poor log
difference in the quality of unskilled labor. Intuitively, large efficiency
differences are needed to reconcile moderate differences in skilled wage
premia with large differences in trade patterns.

Large skilled labor efficiency differences lead me to attribute more im-
portance to differences in the value of the human capital aggregator G
than does traditional development accounting, and correspondingly less
importance to uniform TFP-differences. Indeed, traditional development
accounting will in general overestimate the importance of uniform effi-
ciency differences when rich countries have a higher efficiency of skilled
labor. The reason is that traditional development accounting relies on
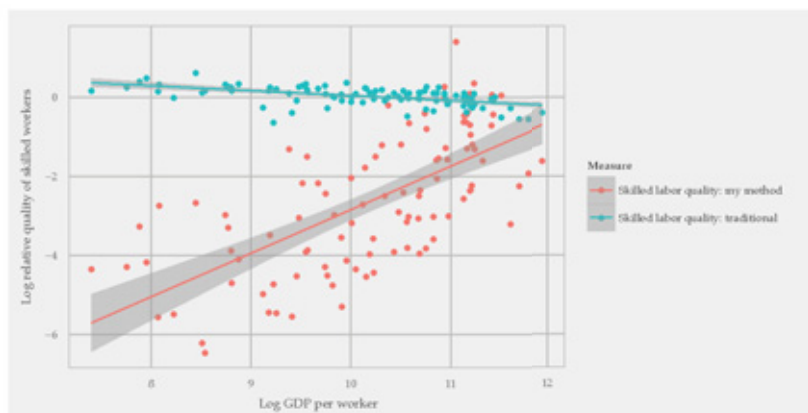
Figure 1.6: Log GDP per worker and log efficiency of skilled workers

the skilled wage premium to capture the output effect of improved efficiency of skilled labor. However, when skilled and unskilled labor services are imperfect substitutes, improved efficiency of skilled labor will not increase the skilled wage premium one-for-one. Instead, skilled labor efficiency improvements lead to two counteracting effects. One effect is a standard productivity effect which increases the skilled wage premium. A second effect is a relative price effect, whereby improvements in the efficiency of skilled labor increase the supply of skilled services and push down the relative price of skilled services. This second effect ensures that the skilled wage premium increases less than one-for-one when skilled labor efficiency improves.[18]

1.4.2   Interpretation of skilled-labor quality differences

The previous section showed that the log efficiency of skilled labor is approximately four times higher in rich countries than in poor countries. This corresponds to rich countries having approximately 50 times higher efficiency of skilled labor. In this section, I discuss my interpretation of these efficiency differences.

---

[18] For further discussions of the role of skilled labor efficiency differences and human capital accounting, see B Jones (2014a) and B Jones (2014b).

If I retain the assumption from traditional development accounting that technology differences across countries are skill neutral, then the large efficiency differences reflect that skilled workers in rich countries have higher human capital. Under this interpretation, a majority of world income differences are explained by human capital differences. This is the interpretation made in Jones (2014a), and it is further elaborated in an unpublished manuscript (Jones, 2014b), where it is discussed that human capital differences among skilled workers can arise as a result of more extensive specialization in rich countries.

If we relax the assumption of neutral technology differences, an alternative explanation is that skilled labor efficiency differences reflect skill-augmenting technology differences. This is the interpretation made in Caselli and Coleman (2006) and Caselli (2015). Under this interpretation, technology differences are still more important than human capital differences, but it is a different form of technology differences than the TFP differences found in traditional development accounting. As the technology shifters are skill-specific, the results suggest that under this interpretation, theories of technology differences should explain how economic environments interact with the efficiency of skilled workers.

With a flexible specification of variation in technology and skilled labor human capital across countries, it is not possible to distinguish the technology and human capital explanations using only price and quantity data. Indeed, human capital quality and factor augmenting technology terms appear in the same way in production functions. Thus, they have the same implications for quantity and price data. Intuitively, price and quantity data alone cannot tell whether a worker is good at hammering, or has a good hammer.

To discriminate between the interpretations, more theoretical structure or other sources of evidence are needed. In the Appendix, I discuss some attempts to distinguish between the two interpretations. In Appendix 1.D.1, I discuss how migration results can be used to shed some light on the respective roles of technology and human capital. The conclusion is that imperfect substitutability makes it more complicated

to interpret what wage changes at migration implies for human capital measurement, as there is no longer a simple mapping between human capital and pre- and post-migration wages with imperfect substitutability. In Appendix 1.D.2, I also analyze human capital versus technology in the case when technology bias is constrained to be endogenous to factor prices (as in, for example, Caselli and Coleman (2006) or Acemoglu (2007)). In this case, I show that large differences in human capital quality are still needed to reconcile differences in Q_s across countries.

Looking ahead, an important task is to analyze the sources of Q_s differences. Promising approaches are likely to involve more detailed data on skilled wages in poor countries, and theoretical specifications that allow us to use migration data and overcome the problems outlined in Appendix 1.D.1.

1.5   Relationship to B Jones (2014)

The paper most closely related to mine is B Jones (2014a), which constructs a theory of development accounting under imperfect substitutability. His key claim is that with a general human capital aggregator, you have to scale traditional development accounting results with the marginal product of unskilled labor to obtain the full value of the human capital aggregator. A general aggregator satisfies

$$
\begin{aligned}
G(H_1, \ldots, H_N) &= G_1 \times \left(H_1 + \sum_{i=2}^{N} \frac{G_i}{G_1} H_i \right) \\
&= G_1 \times \left(H_1 + \sum_{i=2}^{N} \frac{w_i}{w_1} H_i \right)
\end{aligned}
$$

where $G_i = \frac{\partial G}{\partial H_i}$, and where the second line uses a competitive market assumption. The terms in brackets on the second line represent the traditional development accounting aggregator, which has to be scaled up by the marginal product of unskilled labor G_1.

Although using a different formulation than in my paper, Jones also

highlights that traditional development accounting misses quality improvements in skilled labor. In my formulation, traditional development accounting underestimates improvements in the quality of skilled labor as an increased abundance of skilled services depresses the relative price of skilled services. In Jones' formulation, an improvement in the quality of skilled labor increases the marginal product of unskilled labor, which increases the appropriate scaling on the results of traditional development accounting.

Furthermore, Jones recognizes that the quality of skilled labor can be interpreted as resulting from an aggregation of heterogeneous skilled services, which opens up for large quality differences. He emphasizes specialization which is more fully developed in an unpublished paper (Jones, 2014b). As discussed in Section 1.4, other potential mechanisms that can lead to large quality differences include strong complementarities between different skill types (C Jones, 2011), in particular O-ring effects (Kremer, 1993).

However, Jones' positive argument for large quality differences is less strong than his conceptual points. His quantitative argument relies on applying rich country time-series and panel estimates of the elasticity of substitution η to cross-country data. If this elasticity is globally valid, the low supply of skilled labor in poor countries must imply a very high price of skilled services. As these high skilled service prices are not observed in skilled wage premia, the quality of skilled labor must be very low in poor countries.

The challenge to this argument is that most estimates of the elasticity of substitution are medium-run estimates from time series data from rich countries.[19] We do not know a priori whether existing esti-

---

[19] One of few papers that take a long-run perspective is Ciccone and Peri (2005), which estimates long-run elasticities using compulsory schooling reforms and US cross-state data on a decadal level. The estimation method is closer to my desired parameter as it is a long-run estimate, and it uses an instrument to deal with the endogeneity of state-level supply of skilled labor. Their preferred estimate is 1.5 with a standard error of 0.44. The study is unfortunately somewhat limited by weak instruments (the first-stage using the most credibly exogenous instrument has an F-value of 2.56 with 6 instruments). Furthermore, there are five observations for every state and the

mates are the relevant long-run cross-country elasticity estimates to be used in development accounting, or whether it is appropriate to assume a constant elasticity of substitution when analyzing cross-country data. Furthermore, the estimated importance of human capital is sensitive to this elasticity parameter. Using my definition of the share of skilled workers, an elasticity of substitution $\eta = 2$ using Jones' method would bring down the share of world income differences accounted for by human capital to approximately 25% of the world income differences, whereas an elasticity of substitution of approximately $\eta = 1.2$ would mean that all world income differences would be explained by human capital.

Thus, Jones' quantitative argument is difficult to evaluate if we do not have independent estimates of relative skilled service prices in poor countries. My trade data method provides such estimates, and I find that the relative prices of skilled services are indeed very high in poor countries. My estimated elasticity of substitution η is 1.27. In Appendix 1.A.3, I also provide suggestive evidence that a constant elasticity of substitution is appropriate to model cross-country data. My paper thus provides quantitative backing to Jones' conceptual points.

1.6 Robustness and consistency checks

Here, I present various robustness and consistency checks of my results. In Section 1.6.1, I analyze how sensitive my estimates of relative skilled service prices are to varying underlying assumptions and parameters. In Section 1.6.2, I test whether my estimates of relative skilled service prices are consistent with estimates based on unit production cost data when such data are available. In Section 1.6.3, I analyze how my development accounting exercise is affected when I change the measurement of skilled wage premia, and how it is affected when I exclude very poor countries and oil producing countries from the analysis. The discussion of each robustness check is brief, and Appendix 1.E provides more detailed de-

---

standard errors are not clustered at the state level, which opens up for larger standard errors.

scriptions and discussions of the robustness checks.

Across a wide range of specifications and parameter values, the conclusion holds that the role of human capital is considerably expanded as compared to findings based on traditional development accounting methods. Furthermore, for countries where both trade data and unit cost data are available, the two types of analyses give similar results. Excluding the poorest countries and oil producing countries increases the estimated importance of human capital.

1.6.1 Sensitivity of relative skilled service price estimates

I estimate the relative price of skilled services using the regression specification (1.5). In this section, I test the sensitivity of my relative price estimates to variations in the price elasticity of trade, the set of control variables, the functional form of the underlying industry production functions, and the presence of zero trade flows.

Table 1.6 shows how my development accounting results change when I change the elasticity of trade σ. Variations in σ are quantitatively important, and a larger σ means a lower importance of human capital. The intuition is that a larger σ means that less relative unit cost differences are needed to explain the trade data. This reduces the estimated differences in relative skilled service prices which, in turn, imply a reduction in the estimated quality differences of skilled labor. Even though a larger σ implies a smaller role for human capital, the estimated importance of human capital for $\sigma = 15$ is still 4.5 times as large as that found using traditional development accounting methods. When the trade elasticity is $\sigma = 5$, human capital explains more than 100% of world income differences. In Appendix 1.E.1, I discuss the effect of allowing trade elasticities to be different across industries.

A second potential problem in regression (1.5) is omitted variables in the specification of unit costs. The regression specification assumes that variations in relative unit costs are only driven by variations in relative factor service prices. If there are other determinants of unit costs corre-

lated with relative factor service prices, there will be an omitted variable bias. I test for the importance of an omitted variable bias by controlling for potential determinants of unit costs apart from relative factor service prices. In particular, I allow there to be a country-specific penalty on external financing and/or contracting. These penalties increase the log unit cost of an industry in proportion to the financial dependence and/or contracting dependence of the industry. To measure financial dependence and contracting dependence at an industry level, I use measures similar to those developed by Rajan and Zingales (1998) and Nunn (2007), respectively. The results are presented in Table 1.7. Including a term for contracting sensitivity does not affect the importance of human capital, and including a term for financial sensitivity decreases the importance of human capital from 65% to 51%. In Appendix 1.E.2, I describe the definition of industry financial and contracting sensitivities, and how they are included in my regression.

A third potential problem in regression (1.5) is a second-order bias in the log-linearization of unit costs. The regression specification is based on log-linearizing unit costs around the US cost structure. This log-linearization is exact if the industry production functions are Cobb-Douglas. If the industry production functions are not Cobb-Douglas, there will be a second-order bias as industry factor shares vary with relative factor service prices. I analyze how my results change if industry production functions are CES with a common elasticity of substitution $\xi \neq 1$. I test for this bias by creating model generated unit costs from a model where industry production functions are CES. I run my regression specification (1.5) on the model generated data and look for the price differences in the model such that my regressions yield similar results on actual and model generated data. This procedure allows me to gauge the bias in my baseline estimates. Table 1.8 shows the development accounting results for different assumed values of ξ. Appendix 1.D.3 explains the environment, the estimation method, the results, and the economic intuition in greater detail.

A fourth potential problem in regression (1.5) is zero trade flows.

Approximately 62% of the bilateral trade flows on the NAICS four-digit level are zero. Given that regression (1.5) is defined for log trade flows, export flows of value zero are dropped, which risks biasing my estimates. One way of gauging the effects of excluding zeros is to run the regression on a higher level of aggregation, which reduces the numbers of zeros. Figure 1.7 shows the estimated relative skilled service prices when I run the regression on four-digit and three-digit manufacturing industries. The three-digit estimates are less precisely estimated as there are only 21 industries instead of 84. However, there is a very similar relationship between log income per worker and log estimated relative skilled service prices.

Table 1.6: Contribution of factors and TFP to income differences: different σ

	Baseline ($\sigma = 10$)	$\sigma = 5$	$\sigma = 15$
Capital	0.08	0.08	0.08
Human capital – trad.	0.12	0.12	0.12
Human capital – new	0.65	1.33	0.45
TFP – trad.	0.79	0.79	0.79
TFP – new	0.26	-0.4	0.46
Log TFP diff. – trad.	2.54	2.54	2.54
Log TFP diff. – new	0.85	-1.3	1.48
TFP-diff. reduction	67%	153%	42%
Elasticity of subst.	1.27	1.10	1.46

1.6.2 Consistency between trade data and unit cost data

In Section 1.2, I used trade data to substitute for missing unit cost data. However, the Groningen Growth and Development Center has constructed a unit cost measure for 34 industries across 42 countries. A natural consistency check is whether my trade data method yields similar
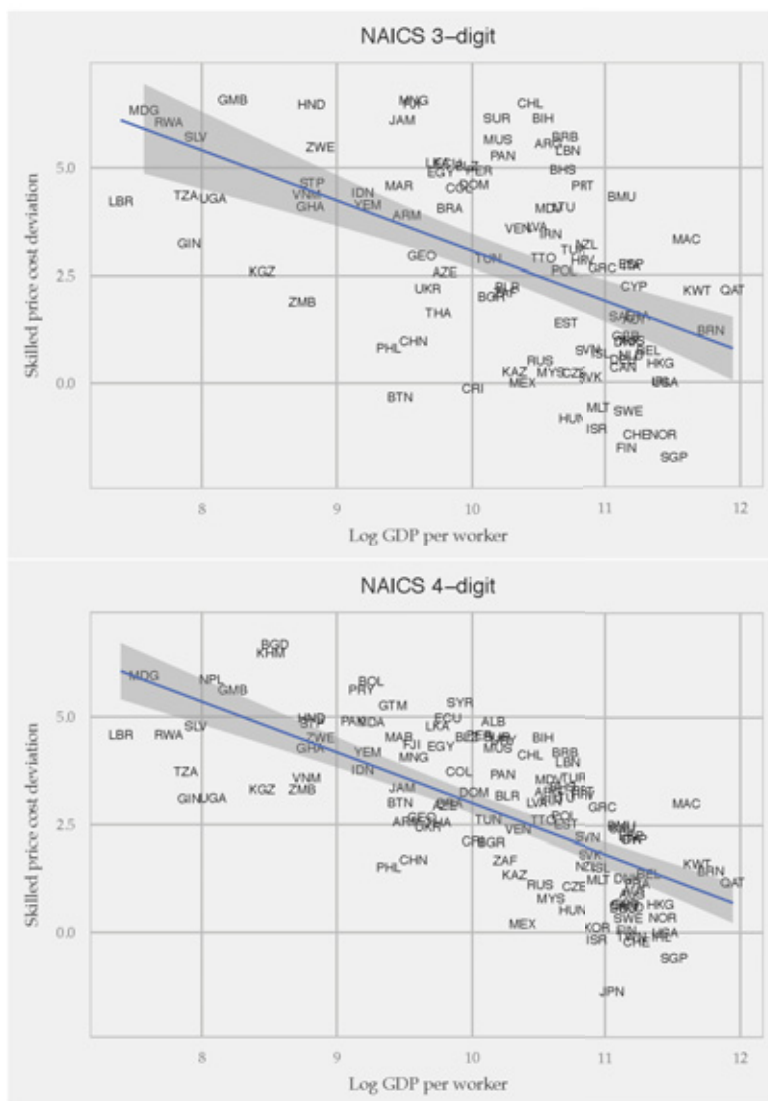
Figure 1.7: Estimated skill price differences with $\sigma = 10$. NAICS 3-digit and 4-digit.

Table 1.7: Contribution of factors and TFP to income differences: different control variables

	Baseline	Contracting	Financing	Both
Capital	0.08	0.08	0.08	0.08
Human capital – trad.	0.12	0.12	0.12	0.12
Human capital – new	0.65	0.63	0.51	0.51
TFP – trad.	0.79	0.79	0.79	0.79
TFP – new	0.26	0.27	0.40	0.40
Log TFP diff – trad.	2.54	2.54	2.54	2.54
Log TFP diff – new	0.85	0.89	1.28	1.28
TFP-diff. reduction	67%	66%	50%	50%
Elasticity of subst.	1.27	1.28	1.35	1.35

conclusions as a unit cost based method on this set of countries.

The GGDC index covers both tradable and non-tradable industries, and manufacturing as well as services. Using the GGDC data set, I can run a unit cost regression to estimate relative factor service prices.[20]

$$\log(c_i^k) = \delta_i + \mu_k + \sum_{f=2}^{F} \alpha_{US,f}^k \tilde{\beta}_{i,f}.$$

Here, δ_i is a country-fixed effect, μ_k is an industry-fixed effect, and $\tilde{\beta}_{i,f}$ identifies the country-factor relative factor service price differences. In Figures 1.8 and 1.9, I plot the relationship between estimated log relative skilled service prices and log GDP per worker, both with country names and with error bars. The results have larger standard errors than the trade based estimates. This reflects the lower number of industries. However, just like the trade based estimates, they exhibit a strong negative correlation with log GDP per worker. The slope parameter of log relative skilled service prices on log GDP per worker is −1.19 using the unit cost data, and −1.53 using the trade data method for the same set

---

[20]In Appendix 1.E.3, I derive this regression specification, and provide more details on all measurements.

Table 1.8: Contribution of factors and TFP to income differences: different ξ.

	$\xi = 0.6$	$\xi = 0.8$	$\xi = 1$	$\xi = 1.2$	$\xi = 1.4$
Capital	0.07	0.07	0.07	0.07	0.07
Human capital – trad.	0.12	0.12	0.12	0.12	0.12
Human capital – new	0.65	0.62	0.63	0.68	0.84
TFP – trad.	0.80	0.80	0.80	0.80	0.80
TFP – new	0.26	0.29	0.28	0.23	0.07
Log TFP diff – trad.	2.56	2.56	2.56	2.56	2.56
Log TFP diff – new	0.85	0.95	0.92	0.75	0.24
TFP-diff. reduction	67%	63%	65%	71%	91%
Elasticity of subst. η	1.36	1.39	1.38	1.34	1.26

of countries. These estimates are similar, and I cannot reject that the two slopes are equal, even when I do not take into account the large standard errors on the unit cost based estimates. Thus, when both types of data exist, the trade data method and the unit cost method paint a similar picture of the relationship between relative skilled service prices and income per worker.

1.6.3   Further robustness tests of development accounting

In this section, I consider further robustness tests of my development accounting exercise. I analyze how my results change when I exclude the poorest countries and when I exclude oil countries, and I analyze how my results change when I change the measurement of skilled wage premia.

My baseline analysis includes all countries with available trade data, ILO data, and PWT data. Hence, my analysis includes very poor countries and countries with significant oil revenues. Including these countries can be problematic as I use manufacturing trade data to estimate the relative price of skilled services prices. Very poor countries have limited manufacturing trade, and the trade patterns of oil countries are primarily determined by their oil endowment. In Table 1.9, I show the results when

Figure 1.8: Skilled price deviation estimates vs log GDP per worker using unit cost data
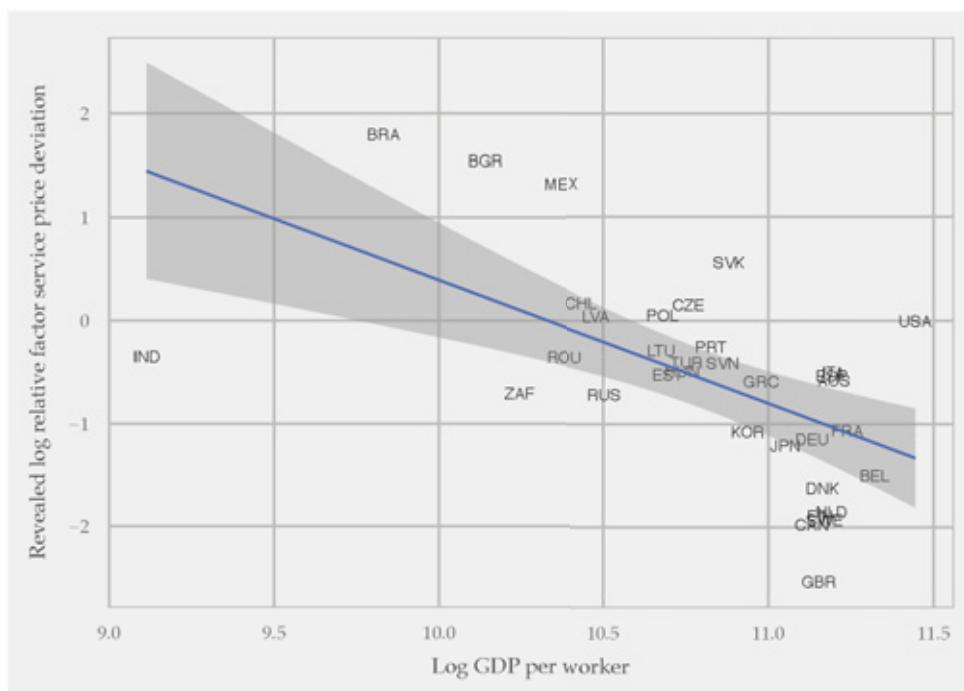
Figure 1.9: Skilled price deviation estimates vs log GDP per worker using unit cost data
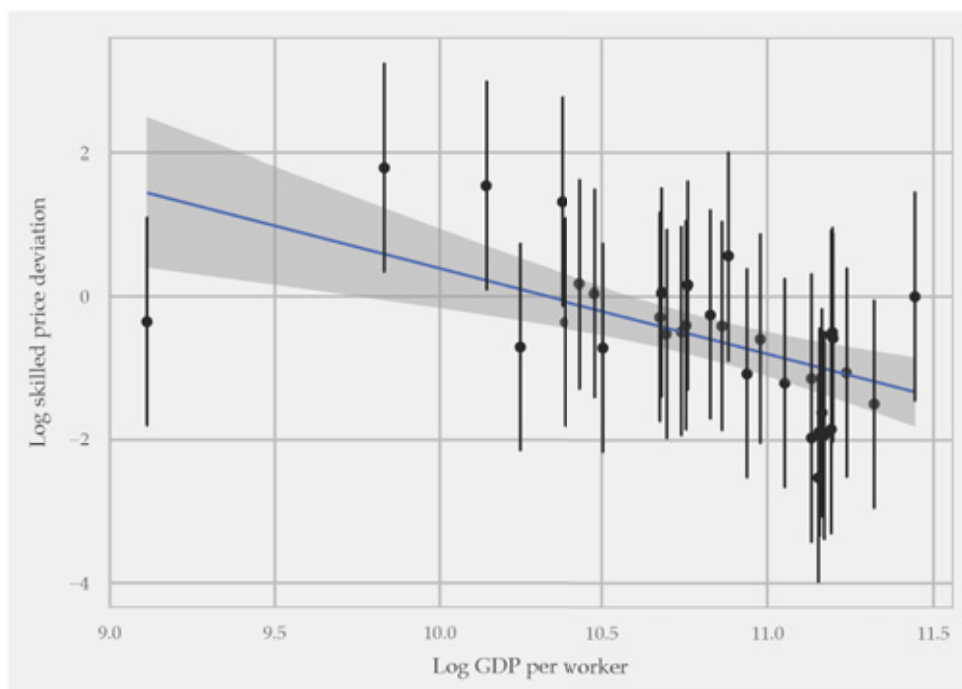
I exclude oil countries and countries with a log GDP per worker of less than 9 in 2010 (corresponding approximately to Ghanaian income levels). Excluding these countries considerably expands the role of human capital, and when both sets of countries are excluded, no TFP differences are needed to explain the income differences among the remaining countries.

I also analyze the robustness of my results to different measurements of skilled wage premia. My skilled wage premia measures are based on limited ILO data, and I want my estimates to be robust to systematic errors in the data on skilled wage premia in poor countries. I am particularly concerned that my measures understate skilled wage premia in poor countries due to the difficulty in measuring the wages of self-employed workers and subsistence farmers. My skilled wage premia measures are based on using a linear relation between log GDP per worker and log skilled wage premia. To test how my results depend on skill premia, I consider how my results change if I allow for a steeper relation between country income and log skilled wage premia keeping rich country skilled wage premia constant. I redo my analysis for different values of the income-skill premia slope $\gamma \leq 0$. The results are presented in Table 1.10.

Variations in the posited slope between skilled wage premia and country income have little effect on the estimated importance of human capital. The reason is that two effects counteract each other. Higher skilled wage premia in poor countries reduce the estimated skill-biased quality differences, but they simultaneously reduce the estimated elasticity of substitution between skilled and unskilled workers. These two effects have opposite consequences for the importance of human capital, and they approximately offset each other. Intuitively, there are two cases. If skilled wage premia are very high in poor countries, it suggests that skilled services are difficult to replace, and poor countries are poor because they have few skilled services. If skilled premia are very low in poor countries, large quality differences in human capital are needed to fit the trade data. Once more, the conclusion is that human capital is

important to account for world income differences.

Table 1.9: Contribution of factors and TFP to income differences: different excluded countries

	Baseline	No v. poor	No oil	Neither
Capital	0.08	0.05	0.08	0.06
Human capital – trad.	0.12	0.10	0.13	0.12
Human capital – new	0.65	0.85	0.68	0.94
TFP – trad.	0.79	0.83	0.77	0.80
TFP – new	0.26	0.08	0.22	-0.0
Log TFP diff – trad.	2.54	2.67	2.49	2.58
Log TFP diff – new	0.85	0.27	0.72	-0.0
TFP-diff. reduction	67%	90%	72%	102%
Elasticity of subst.	1.27	1.20	1.27	1.20

1.7   Concluding remarks

What share of world income differences can be explained by differences in human capital? The development accounting literature has studied this question by aggregating microeconomic returns to schooling. The overall assessment of the importance of human capital has been negative. Even though there are large human capital differences between countries, they cannot explain more than a small fraction of world income differences.

I have revisited the role of human capital in development accounting, using a framework that allows for imperfect substitutability between skilled and unskilled labor services. I have shown that development accounting is possible in this framework if one can estimate the relative price of skilled and unskilled services, and I have developed a method for estimating this relative price using international trade data. My question has been: does development accounting give us sufficient ground to reject a dominant role for human capital in explaining world income differences?

Table 1.10: Contribution of factors and TFP to income differences: different wage premia

Slope coefficients (baseline = -0.12)	Baseline	0.0	-0.2	-0.4
Capital	0.08	0.08	0.08	0.08
Human capital – trad.	0.12	0.12	0.12	0.12
Human capital – new	0.65	0.62	0.65	0.66
TFP – trad.	0.79	0.79	0.79	0.79
TFP – new	0.26	0.29	0.26	0.25
Log TFP diff – trad.	2.54	2.54	2.54	2.54
Log TFP diff – new	0.85	0.95	0.85	0.81
TFP-diff. reduction	67%	63%	67%	68%
Elasticity of subst.	1.27	1.37	1.27	1.18

My results suggest that the answer is no. Using trade data, I find that rich countries have substantially lower relative prices of skilled services. Combining these estimates with data on skilled wage premia suggests that the data are consistent with a substantially higher efficiency of skilled labor in rich countries compared to poor countries. When I include these efficiency differences in my development accounting exercise and interpret them as reflecting a higher quality of human capital of skilled workers, my estimates imply that human capital differences explain 65% of world income differences.

Moving beyond the role of human capital, there is also a broader takeaway from my results: trade data suggest that there are large efficiency differences in skilled labor across rich and poor countries, and that these efficiency differences are large enough to explain a dominant share of world income differences. This conclusion holds regardless of whether these efficiency differences are due to skill-biased technology differences or skill-biased quality differences.

Thus, my paper supports the conclusions of Caselli and Coleman (2006) and B Jones (2014a), who have argued for large cross-country

differences in skilled labor efficiency. Their results build on a different method than my results. They note that even though skill premia are somewhat higher in poor countries, skill premia are not as high as they should be, given the low relative supply of skilled labor in poor countries, at least not if the elasticity of substitution between skilled and unskilled labor is in line with rich country estimates. Both their papers explain this observation by positing that skilled labor efficiency is relatively low in poor countries. Even though they differ in their interpretation of these efficiency differences – Caselli and Coleman argue that they reflect skill-biased technology differences and B Jones argues that they reflect skill-biased human capital differences – they agree on the importance of efficiency differences in skilled labor.

My findings suggest that their results are not just an artifact of assuming that rich country estimates of substitution elasticities are globally valid. When I analyze trade data, a similar pattern emerges. Relative skilled service prices diverge more sharply between countries than skilled wage premia, suggesting large differences in the relative efficiency of skilled labor. Furthermore, the estimated efficiencies of skilled workers are strongly and positively correlated with GDP per worker. Quantitatively, efficiency differences among skilled workers account for a substantial share of the variation in per capita output. By combining my results with the observations made by Caselli and Coleman (2006) and B Jones (2014), we see how skilled labor efficiency differences can provide a unified perspective of the relationship between country income levels, trade patterns, skilled labor supply, and skilled wage premia.

If output differences are primarily driven by efficiency differences in skilled labor, this can influence the research agenda of growth and development economics. First, it means that skilled labor human capital differences can drive a large share of output difference which, in turn, warrants a greater focus on theories of skill acquisition. Potentially interesting areas include the quality of higher education and the incentives and efficiency of on-the-job learning. Second, if the efficiency of skilled labor is driven by skill-specific technology shifters, our technology ex-

planations should put a larger emphasis on why technology differences selectively make skilled labor more productive. This suggests a shift away from general TFP explanations toward more specific theories of technology differences. For example, when we study misallocation, it might be warranted to focus more on how the efficiency of skilled labor is harmed by misallocation – potentially by looking at intersectoral patterns of misallocation. Similarly, when studying technology diffusion, it is warranted to study whether barriers to technology diffusion specifically prevent the diffusion of technologies that are complementary to skilled workers.

References

Acemoglu, D. (2007). Equilibrium Bias of Technology. *Econometrica*, 75(5):1371–1409.

Acemoglu, D., Antràs, P., and Helpman, E. (2007). Contracts and Technology Adoption. *American Economic Review*, 97(3):916–943.

Acemoglu, D. and Autor, D. (2011). Chapter 12 – Skills, Tasks and Technologies: Implications for Employment and Earnings. In *Handbook of Labor Economics*, volume 4, pages 1043–1171.

Acemoglu, D. and Ventura, J. (2002). The World Income Distribution. *The Quarterly Journal of Economics*, 117(2):659–694.

Acemoglu, D. and Zilibotti, F. (2001). Productivity Differences. *The Quarterly Journal of Economics*, 116(2):563–606.

Anderson, J. E., Anderson, and James (1979). A Theoretical Foundation for the Gravity Equation. *American Economic Review*, 69(1):106–16.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1):170–192.

Barro, R. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198.

Bartelsman, E. J. and Gray, W. (1996). The NBER Manufacturing Productivity Database.

Broda, C., Weinstein, D., Broda, C., and Weinstein, D. (2006). Globalization and the Gains from Variety. *Quarterly Journal of Economics*, 121:541–585.

Caselli, F. (2005). Chapter 9 Accounting for Cross-Country Income Differences. *Handbook of Economic Growth*, 1:679–741.

Caselli, F. (2015). *Technology Differences Over Space and Time.*

Caselli, F. and Ciccone, A. (2013). The contribution of schooling in development accounting: Results from a nonparametric upper bound. *Journal of Development Economics*, 104:199–211.

Caselli, F. and Coleman, W. J. (2006). The World Technology Frontier. *American Economic Review*, 96(3):499–522.

Chor, D. (2010). Unpacking sources of comparative advantage: A quantitative approach. *Journal of International Economics*, 82(2):152–167.

Ciccone, A. and Papaioannou, E. (2009). Human Capital, the Structure of Production, and Growth. *Review of Economics and Statistics*, 91(1):66–82.

Ciccone, A. and Peri, G. (2005). Long-Run Substitutability Between More and Less Educated Workers: Evidence from U.S. States, 1950–1990. *Review of Economics and Statistics*, 87(4):652–663.

Costinot, A., Donaldson, D., and Komunjer, I. (2011). What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas. *The Review of Economic Studies*, 79(2):581–608.

Cuñat, A. and Melitz, M. J. (2012). Volatility, Labor Market Flexibility, and the Pattern of Comparative Advantage. *Journal of the European Economic Association*, 10(2):225–254.

Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, 70(5):1741–1779.

Erosa, A., Koreshkova, T., and Restuccia, D. (2010). How Important Is Human Capital? A Quantitative Theory Assessment of World Income Inequality. *Review of Economic Studies*, 77(4):1421–1449.

Gaulier, G. and Zignago, S. (2010). BACI: International Trade Database at the Product-Level (the 1994-2007 Version). *SSRN Electronic Journal*.

Gollin, D. (2002). Getting Income Shares Right. *Journal of Political Economy*, 110(2):458–474.

Hall, R. E. and Jones, C. I. (1999). Why do Some Countries Produce So Much More Output Per Worker than Others? *The Quarterly Journal of Economics*, 114(1):83–116.

Head, K. and Mayer, T. (2014). Gravity Equations: Workhorse,Toolkit, and Cookbook. In *Handbook of International Economics*, volume 4, pages 131–195. Elsevier B.V.

Hendricks, L. (2002). How Important Is Human Capital for Development? Evidence from Immigrant Earnings. *American Economic Review*, 92(1):198–219.

Hendricks, L. and Schoellman, T. (2016). Human Capital and Development Accounting: New Evidence from Wage Gains at Migration.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India *. *Quarterly Journal of Economics*, 124(4):1403–1448.

Hsieh, C.-T. and Klenow, P. J. (2010a). Development Accounting. *American Economic Journal: Macroeconomics*, 2(1):207–23.

Hsieh, C.-T. and Klenow, P. J. (2010b). Development Accounting. *American Economic Journal: Macroeconomics*, 2(1):207–223.

Inklaar, R. and Timmer, M. P. (2008). GGDC Productivity Level Database: International Comparisons of Output, Inputs and Productivity at the Industry Level. *GGDC Research Memorandum*.

Jones, B. F. (2014a). The Human Capital Stock: A Generalized Approach. *American Economic Review*, 104(11):3752–3777.

Jones, B. F. (2014b). The Knowledge Trap: Human Capital and Development Reconsidered.

Jones, C. (2015). The Facts of Economic Growth. Technical report, National Bureau of Economic Research, Cambridge, MA.

Jones, C. I. (2011). Intermediate Goods and Weak Links in the Theory of Economic Development. *American Economic Journal: Macroeconomics*, 3(2):1–28.

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.

Klenow, P. and Rodriguez-Clare, A. (1997). The Neoclassical Revival in Growth Economics: Has It Gone Too Far? Technical report.

Kremer, M. (1993). The O-Ring Theory of Economic Development. *The Quarterly Journal of Economics*, 108(3):551–75.

Krugman, P. (1988). Differences In Income Elasticities and Trends in Real Exchange Rates. Technical report, National Bureau of Economic Research, Cambridge, MA.

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2016). Life-Cycle Human Capital Accumulation Across Countries: Lessons From U.S. Immigrants.

Levchenko, A. A. and Zhang, J. (2016). The evolution of comparative advantage: Measurement and welfare implications. *Journal of Monetary Economics*, 78:96–111.

Manova, K. (2013). Credit Constraints, Heterogeneous Firms, and International Trade. *The Review of Economic Studies*, 80(2):711–744.

Manuelli, R. E. and Seshadri, A. (2014). Human Capital and the Wealth of Nations †. *American Economic Review*, 104(9):2736–2762.

Mattsson, L.-G., Weibull, J. W., and Lindberg, P. O. (2014). Extreme values, invariance and choice probabilities. *Transportation Research Part B: Methodological*, 59:81–95.

Midrigan, V. and Xu, D. Y. (2014). Finance and Misallocation: Evidence from Plant-Level Data †. *American Economic Review*, 104(2):422–458.

Morrow, P. and Trefler, D. (2017). Endowments, Factor Prices, and Skill-Biased Technology: Importing Development Accounting into HOV.

Nunn, N. (2007). Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade. *The Quarterly Journal of Economics*, 122(2):569–600.

Nunn, N. and Trefler, D. (2015). Chapter 5: Domestic Institutions as a Source of Comparative Advantage. In *Handbook of International Economics*, volume 4, pages 263–315.

Oberfield, E. and Raval, D. (2014). Micro Data and Macro Technology. Technical report, National Bureau of Economic Research, Cambridge, MA.

Parente, S. L. and Prescott, E. C. (1999). Monopoly Rights: A Barrier to Riches. *American Economic Review*, 89(5):1216–1233.

Rajan, R. G. and Zingales, L. (1998). Financial Dependence and Growth. *American Economic Review*, 88(3):559–86.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics*, 48(1):7–35.

Redding, S. and Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, 62(1):53–82.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720.

Romalis, J. (2004). Factor Proportions and the Structure of Commodity Trade. *American Economic Review*, 94(1):67–97.

Romalis, J. (2007). NAFTA's and CUSFTA's Impact on International Trade. *http://dx.doi.org/10.1162/rest.89.3.416*.

Schoellman, T. (2011). Education Quality and Development Accounting. *The Review of Economic Studies*, 79(1):388–417.

Schoellman, T. and Hendricks, L. (2016). Human Capital and Development Accounting: New Evidence from Wage Gains at Migration. *2016 Meeting Papers*.

Simonovska, I. and Waugh, M. E. (2014). The elasticity of trade: Estimates and evidence. *Journal of International Economics*, 92(1):34–50.

Soderbery, A. (2015). Estimating import supply and demand elasticities: Analysis and implications. *Journal of International Economics*, 96(1):1–17.

Tinbergen, J. (1962). Shaping the World Economy; Suggestions for an International Economic Policy. *Books (Jan Tinbergen)*.

Trefler, D. (1993). International Factor Price Differences: Leontief was Right! *Journal of Political Economy*, 101(6):961–987.

Appendix

1.A Environment

1.A.1 Heterogeneous skill type aggregator interpretation of Q_u and Q_s

Here, I show that my estimation of the relative quality Q_s/Q_u is consistent with a nested structure where the quality terms Q_u and Q_s arise from aggregation of heterogeneous unskilled and skilled services.

My human capital aggregator is

$$h = \left((Q_u u)^{\frac{\eta-1}{\eta}} + a_s (Q_s s)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}.$$

Before proving the result, I will provide a formal statement of what equivalence means in this context. Assume that the true human capital aggregator is

$$h = \left((H^u)^{\frac{\eta-1}{\eta}} + a_s (H^s)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}},$$

where H^u and H^s are arbitrary constant returns to scale aggregators of heterogeneous unskilled and skilled services. I say that my relative quality estimation is consistent with an aggregator interpretation if the following holds. Given the definition of quality

$$
\begin{aligned}
Q_u &\equiv \frac{H^s}{s} \\
Q_s &\equiv \frac{H^u}{u},
\end{aligned}
$$

the relative quality of skilled and unskilled labor Q_s/Q_u satisfies the equation

$$\frac{w_s}{w_u} = \frac{Q_s}{Q_u} \frac{r_s}{r_u}, \tag{1.11}$$

where $\frac{w_s}{w_u}$ is the relative average wage of skilled and unskilled workers,

and $\frac{r_s}{r_u}$ satisfies

$$\frac{r_s}{r_u} = a_s \left(\frac{H^s}{H^u} \right)^{-1/\eta} .$$

This quality definition defines the quality of unskilled and skilled labor as the average amount of services provided by each worker in each skill category.

I will now prove the equivalence result. I assume that there are $N_u \geq 1$ types of unskilled labor services and $N_s \geq 1$ types of skilled labor services. A share u_{t_u} of the workforce performs unskilled services of type t_u where $t_u = 1, \ldots, N_u$, and a share s_{t_s} of the workforce performs skilled services of type t_s where $t_s = 1, \ldots, N_s$. The average quality of an unskilled worker of type t_u is Q_{u,t_u} and the average quality of a skilled worker of type t_s is Q_{s,t_s}. The workforce shares sum to the aggregate share of skilled and unskilled workers

$$\sum_{t_u=1}^{N_u} u_{t_u} = u$$

$$\sum_{t_s=1}^{N_s} s_{t_s} = s.$$

With this formulation, the quality of unskilled and skilled labor is defined as

$$Q_u \equiv \frac{H^u(Q_{u,1}u_1, \ldots, Q_{u,N_u}u_{N_u})}{u} = H^u \left(Q_{u,1}\tilde{u}_1, \ldots, Q_{u,N_u}\tilde{u}_{N_u} \right)$$

$$Q_s \equiv \frac{H^s(Q_{s,1}s_1, \ldots, Q_{s,N_u}s_{N_u})}{s} = H^s \left(Q_{s,1}\tilde{s}_1, \ldots, Q_{s,N_s}\tilde{s}_{N_s} \right),$$

where a tilde ($\sim$) denotes that we normalize the unskilled and skilled worker shares u_{t_u} and s_{t_s} with the total supply of unskilled and skilled workers s and u.

Now consider an arbitrary unskilled service type t_u and an arbitrary skilled service type t_s. Assuming that the labor market is competitive,

these two types of workers have a relative wage

$$\frac{w_{s,t_s}}{w_{u,t_u}} = \left(\frac{H^s}{H^u}\right)^{-1/\eta} \frac{H^s_{t_s} Q_{s,t_s}}{H^u_{t_u} Q_{u,t_u}} = \frac{r_s}{r_u} \frac{H^s_{t_s} Q_{s,t_s}}{H^u_{t_u} Q_{u,t_u}},$$

where H^s_t and H^u_t denote the partial derivatives of the human capital aggregator functions with respect to their t^{th} elements. The relative wage is a product of i) the relative marginal product of the two aggregators, and ii) the relative marginal contributions of the two skill types to their respective aggregators.

I can use this equation to prove that (1.11) holds. First, I multiply both sides with $\tilde{s}_{t_s}$ and sum over $t_s = 1, \ldots, N_s$ to obtain

$$\frac{w_s}{w_{u,t_u}} = \frac{r_s}{r_u} \frac{Q^s}{H^u_{t_u} Q_{u,t_u}} \tag{1.12}$$

where I use Euler's theorem to obtain

$$Q^s = \sum_{t_s=1}^{N_s} Q_{s,t_s} \tilde{s}_{t_s} H^s_{t_s},$$

and use that average skilled wages are defined by

$$w_s = \sum_{t_s=1}^{N_s} \tilde{s}_{t_s} w_s.$$

I obtain equation (1.11) by applying the same procedure to unskilled labor. I start with equation (1.12), invert the equation, multiply both sides with $\tilde{u}_{t_u}$, sum over $t_u = 1, \ldots, N_u$, and finally, I re-invert the equation.

This proves that an aggregator interpretation of the quality terms is equivalent to a two labor type interpretation when estimating the relative quality of skilled labor Q_s/Q_u. When doing development accounting, I make one further restriction in assuming that the unskilled aggregator is a linear aggregator. This allows me to estimate Q_u from Mincerian return data, and together with my estimation of Q_s/Q_u, I can complete the development accounting exercise.

1.A.2 Supply-side aggregation with multiple industries and trade

I express output with an aggregate production function

$$Y = K^\alpha (ALh)^{1-\alpha}.$$

When estimating the aggregate production function, I assume that the economy consists of multiple industries and that it trades with the outside world. In light of this, the aggregate production function should be interpreted as reflecting substitution possibilities within and between industries, as well as substitution possibilities between domestic and foreign production. Here, I discuss the assumptions needed to have a constant returns to scale aggregate production function with multiple industries and trade. In Appendix 1.A.3, I motivate my particular choice of functional form.

I show that a CRS aggregate production function exists under fairly general conditions when countries are price takers in the world market. However, there are more stringent conditions for the existence of a CRS aggregator in variety-based trade models such as Eaton and Kortum and Armington models. In these models, being small compared to the rest of the world is not sufficient to make a country a price-taker, as every country is a large producer of its own varieties. This means that the terms of trade move against countries as they expand factor supplies. Given that my estimation exercise relies on variety models, this is a potential problem.

However, I show that a CRS aggregate production function is possible under a reasonable modification of variety models. The modification is to assume that quality in an Armington model (and absolute productivity advantage in an Eaton and Kortum style model) is homogenous of degree one in aggregate or industry factor supplies. I demonstrate how this modification yields a CRS representation in an Armington model with many small countries, and a similar mechanism applies to the Eaton and Kortum framework.

To motivate my modification, I first argue that the terms of trade effect is unlikely to be a long-run phenomenon. In particular, if such a long-run effect existed, terms of trade would be sensitive to subdivisions of countries. For example, if Scotland and UK were formally separated, a long-run terms of trade effect from size would imply that both English and Scottish terms of trade should improve with respect to the rest of the world if they split. This feature is unrealistic, and it suggests that whatever scarce resource makes the global demand curve for a country's goods slope downward – restricted number of varieties in an Armington framework, or restricted idea generation in an Eaton and Kortum framework – this scarce resource should scale with size.[21]

Once I modify the Armington model such that qualities scale with factor supplies, a CRS aggregate production function representation is possible. Furthermore, allowing quality to scale with inputs does not affect the key feature of the model: that relative exports across countries and goods are determined by relative trade costs and relative production costs.

Setup

To study the conditions needed for the existence of a CRS representation, I study a general multi-industry model of a country with K industries and F factor services in an open economy $i \in I$. I use a dual formulation. The production technology in country i for each industry is CRS and represented by the unit cost function $c_i^k(r_{i,1}, \ldots, r_{i,F})$. Factor service supplies are $v_{i,f}$. I write y_i^k for production in industry k and x_i^k for consumption in industry k (these two quantities might differ due to trade). I write p_i^k for the domestic price of good k. There exists a rep-

---

[21]This modification is related to Krugman (1988) who shows that growing countries do not face **deteriorating** terms of trade, and he explains this with a variety model of growth. For a contrasting perspective, see Acemoglu and Ventura (2002) who argue that a country's terms of trade deteriorates when it grows through capital accumulation.

resentative consumer whose preferences are defined by an expenditure function $e(\mathbf{p}_i, u_i)$. I assume that these preferences are homothetic, which means that there exists a utility representation of preferences such that the expenditure function can be written

$$e(\mathbf{p}_i, u_i) = \tilde{e}(\mathbf{p}_i)u_i$$

for some function $\tilde{e}$. Throughout this section, I assume that preferences are homothetic and I will write $\tilde{e}$ without a tilde going forward.

A CRS aggregator representation exists if prices are unchanged and output and consumption scale linearly when we scale factor inputs. Formally, I say that a CRS aggregator representation exists if the following condition holds. Let $x_i^k, y_i^k, u_i, r_{i,f}, p_i^k, c_i^k$ be an arbitrary equilibrium given factor supplies $v_{i,f}$. A CRS representation exists if for each such equilibrium, a factor supply $\lambda v_{i,f}$ implies that $\lambda x_i^k, \lambda y_i^k, \lambda u_i, r_{i,f}, p_i^k, c_i^k$ is an equilibrium.

I first consider a model where each country is a price-taker in the world market. In this case, the equilibrium conditions can be written as:

$$
\begin{aligned}
\sum_{k=1}^{K} \frac{\partial c_i^k}{\partial r_{i,f}} y_i^k &= v_{i,f} \quad f = 1, \dots, F \\
\frac{\partial e}{\partial p_i^k} u_i &= x_i^k \quad k = 1, \dots, K \\
c_i^k &\geq p_i^k \quad = 0 \text{ if } y_i^k > 0 \\
e(\mathbf{p}_i)u_i &= \sum_{f=1}^{F} r_{i,f} v_{i,f}
\end{aligned}
$$

The first equation gives clearing conditions for the factor markets, where the left-hand side uses Shepherd's lemma applied to the unit cost function to derive factor demands for each factor f and for industry k. The second equation expresses consumer demand, applying Shepherd's lemma to the expenditure function. The third equation is a zero-profit condition, where the inequality constraint reflects that I allow for zero

production. The fourth equation is the budget constraint for the representative consumer.

By inspection, this system of equations allows for a CRS aggregator representation. If there exists a set of prices such that $y_i^k, x_i^k, u_i, v_{i,f}$ solve the system, then any scaling $\lambda y_i^k, \lambda x_i^k, \lambda u_i, \lambda v_{i,f}$ for $\lambda > 0$ solves the system for the same set of prices.

To study the Armington case, I retain the assumption that the country is small in the aggregate world economy. However, the country is large in its own varieties. I represent this with an Armington model with a continuum of countries and K goods. I write $i \in [0, 1]$ for the country on which I focus.

There are K final goods. Each final good is assembled domestically using a composite of country-industry specific intermediate varieties that are traded between countries. To produce good k, one needs an input variety from each country in the world. I assume that there are no trade costs so that the unit cost C_i^k of assembling final good k in country i is the same in every country and equal to

$$C_i^k \equiv C^k = \left(\int_0^1 a_j^k (c_j^k)^{1-\sigma} dj \right)^{\frac{1}{1-\sigma}} \quad \sigma > 1.$$

I normalize a_j^k so that the unit production costs are $c_j^k = 1$ for all countries $j \neq i$ (our unit of analysis). This means that

$$C^k = 1 \quad k = 1, \ldots, K.$$

Write $q_{i,j}^k$ for the amount of input to industry k that is produced in country i for use in country j. As there are no trading costs and countries are symmetric, $q_{i,j}^k$ does not depend on destination j. Furthermore, using Shepherd's lemma,

$$q_{i,j}^k = \frac{\partial C^k}{\partial c_i^k} x_j^k,$$

where x_j^k is the country j consumption of final goods in industry k.

I can now write down the equilibrium definition.

$$q_{i,j}^k = a_i^k (c_i^k)^{-\sigma} x_j^k$$

$$p_{i,k} = c_{i,k}$$

$$x_i^k = \frac{\partial e(1,\ldots,1)}{\partial P^k} u_i$$

$$\sum_{f=1}^{F} r_{i,f} v_{i,f} = e(1,\ldots,1) u_i$$

$$\sum_{k=1}^{K} \int_0^1 q_{i,j}^k \frac{\partial c_i^k}{\partial r_{i,f}} = v_{i,f}$$

The first equation gives country j's demand for industry k goods produced in country i. The formulation uses that the price index $P_j^k = C_j^k = 1$ for all j. The second equation is a non-profit condition for production in country i. There is no inequality constraint, reflecting that with a CES specification of production technology from intermediates, production of each variety is always positive. The third equation applies Shepherd's lemma to the consumer's expenditure function. It is evaluated at $(1,\ldots,1)$ as all prices $P^k = 1$. The fourth and fifth equations give the consumer budget constraints and the factor market clearing condition.

By inspection, there does not exist a CRS aggregator representation of this system. In the first equation, we see that scaling output will change prices, violating the assumption that there exist scaled equilibria with the same prices. This reflects a terms of trade effect whereby scaling output depresses the terms of trade.

However, there exists a simple modification of the system to obtain a CRS aggregator. If I define $a_i^k = \Phi_i^k(v_{i,1}^k, \ldots, v_{i,F}^k)$ for some CRS aggregator Φ_i^k, there exists a CRS representation of the equilibrium. Allowing the quality term a_i^k to scale linearly with factor supply captures the intuition that subdivision of observation units should not affect trade patterns with third parties. Even with this modification, *relative* trade patterns across industries are still shaped by *relative* costs, and if we

were to add trade costs, then trade costs would affect the distribution between domestic uses and exports, and trade costs would also affect relative exports to different countries.

1.A.3 Functional form of aggregate production function

My aggregate production function has the form

$$Y = K^\alpha (ALh)^{1-\alpha}.$$

As discussed in Appendix 1.A.2, this represents an aggregation taking into account the existence of multiple industries and opportunities for international trade. In this section, I discuss my choice of functional form.

I choose a Cobb-Douglas aggregator between capital and labor services. This is standard in the development accounting literature, and can be motivated by there being constant labor shares across countries (Gollin, 2002).[22]

For the human capital aggregator, I use a CES aggregator of skilled and unskilled labor services, which is standard in the labor economics literature (Acemoglu and Autor, 2011). Ideally, I should have a skill aggregator that was formally aggregated from production functions on the industry level together with a trade model. Unfortunately, there is no straightforward aggregation to a CES representation from industries with heterogeneous factor shares. Thus, the constant elasticity assumption should be interpreted as an approximation to a more freely specified underlying aggregator.

One way of testing my assumption of constant elasticity of substitution is by plotting the cross-country relationship between log relative factor service prices and log relative factor supplies. Theoretically, these

---

[22]Recent studies cast doubt on the Cobb-Douglas assumption (Oberfield and Raval, 2014), and Caselli (2005) suggests that the elasticity of substitution between capital and labor can be a crucial parameter in development accounting. I do not pursue this line of inquiry further here, but it is an interesting avenue of future research. The Cobb-Douglas specification of labor and capital also precludes capital-labor complementarities.

should be related by

$$\log\left(\frac{r_s}{r_u}\right) = \log(a_s) - \frac{1}{\eta}\log\left(\frac{Q_s s}{Q_u u}\right).$$

If the CES assumption is true, the relationship should be linear. The test is not ideal, as my estimated relative quality $\log\left(\frac{Q_s}{Q_u}\right)$ is implicitly present in the relative price of factor services, and thus it appears on both sides of the equation, which biases the relationship towards being linear. However, if the log relative supply of skilled and unskilled workers $\log\left(\frac{s}{u}\right)$ was not linearly related to the relative quality $\log\left(\frac{Q_s}{Q_u}\right)$, the relationship would not be linear. Thus, testing the linearity of this relationship offers an opportunity to falsify the CES assumption. The results are plotted in Figure 1.A.1, which suggests that the linearity assumption is appropriate.

Looking ahead, potential extensions include modifying the functional form to allow for capital-skill complementarities and non-unitary elasticity of substitution between labor and capital.

1.B   Development accounting

1.B.1   Occupational vs schooling based skill cutoff

I define the share of unskilled and skilled workers u and s as the shares of people working in an unskilled and skilled occupation, respectively. This contrasts to the approach taken in Caselli and Coleman (2006), B Jones (2014a), and Caselli (2015) who define the share of skilled workers as the share of individuals having an educational attainment above a pre-specified threshold (for example, primary education and above, high school and above, or college and above).

The distinction between the share of workers with a skilled occupation and the share of workers with a certain educational level does not matter if all countries have the same mapping between educational attainment and occupational skill level. However, there is no a priori reason to believe that this mapping should be the same across countries.
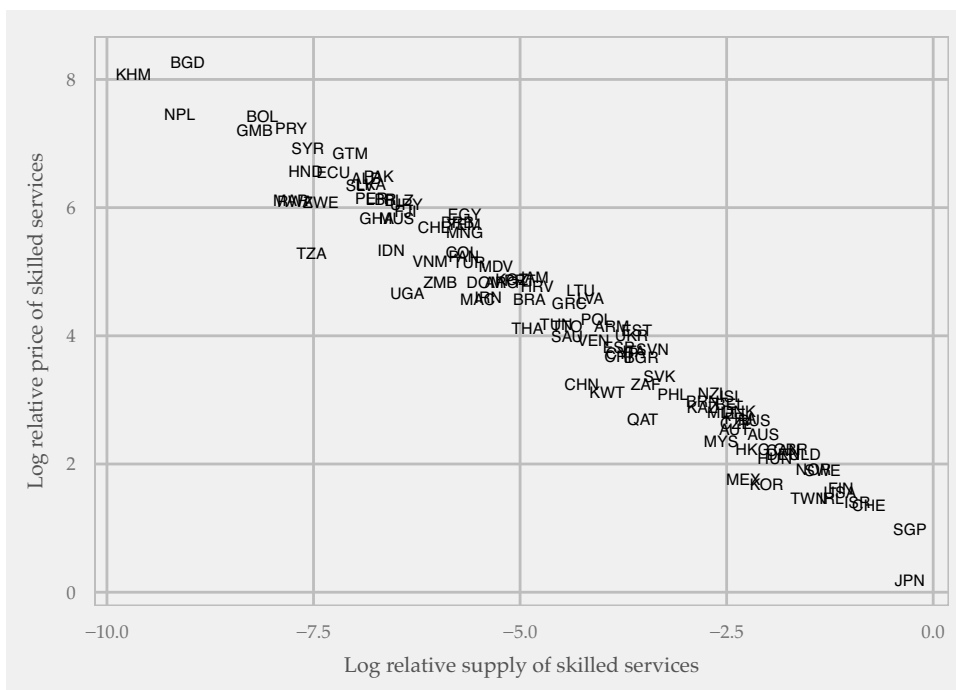
Figure 1.A.1: Testing constant elasticity of substitution

Acemoglu and Autor (2011) have highlighted the importance of distinguishing between educational attainment and tasks when analyzing US time series data as the allocation of skills to tasks is an equilibrium outcome. Their point is more relevant when analyzing differences between countries with very large differences in educational systems. When educational attainment does not map to occupational skill content in the same way across countries, this modeling choice matters.

I choose an occupational definition for two reasons. First, there are multiple ways of acquiring skills, and education is only one of them. Many people learn skilled occupations outside the educational system, and poor quality of schooling increases the risk that schooling does not fully reflect skill acquisition. When skills are not equal to educational attainment, the complexity of the occupation is a proxy for skill. Indeed, as long as there is a positive skilled wage premium, barring compensating differential concerns, people will work in the most complex occupations that they can perform. Second, occupation is closer to the definitions used for skill shares in my trade data exercise, where I define the skill share as the share of gross output that goes to the payroll of workers in certain occupations.

Thus, I measure the share of skilled workers in line with the ILO's ISCO-08 definitions of skill requirements and major occupational groups. The ILO defines 10 major occupational groups and four skill levels. The occupational groups and their respective skill levels are presented in Figure 1.B.1. I use the ILOSTAT database to obtain s as the share of the labor force working as managers, professionals, or technicians and associated technicians, i.e. skill categories 3 and 4 (I define the armed forces as primarily unskilled). I define the unskilled share as $u \equiv 1 - s$.

Figure 1.B.2 compares the results from an education based and occupation based definition of the skill share. Figure 1.B.2 shows that for poor countries, the share of high school educated workers and the share of skilled workers approximately coincide. For rich countries, there are much more high school educated workers than skilled workers. This is evidence that the mapping between educational attainment and skill level

Figure 1.B.1: Mapping of ISCO-08 major groups to skill levels

ISCO-08 major groups	Skill level
1 Managers	3 + 4
2 Professionals	4
3 Technicians and Associate Professionals	3
4 Clerical Support Workers	2
5 Services and Sales Workers	
6 Skilled Agricultural, Forestry and Fishery Workers	
7 Craft and Related Trades Workers	
8 Plant and Machine Operators, and Assemblers	
9 Elementary Occupations	1
0 Armed Forces Occupations	1 + 2 + 4

is different in rich and poor countries, and that the educational cutoff for being in a skilled occupation is lower in poor countries.

These results suggest that education based ratios of skilled and unskilled workers will exaggerate rich-poor differences in the relative supply of skilled and unskilled workers. Overall, my method is therefore more conservative when it comes to finding an important role for human capital. I find that this difference matters when I apply the method in B Jones (2014) using my data definitions. He defines a skilled worker as someone having any education above primary education, and finds that even with an elasticity of substitution of 2, human capital is very important in explaining world income differences. With my definition of skilled labor, an elasticity of substitution of 2 means that human capital is only modestly more important than what is found when using traditional development accounting methods.

Figure 1.B.2: High school and above and share of skilled occupations

1.B.2   Measurement of unskilled labor quality Q_u

I define the quality of unskilled labor Q_u using a Mincerian definition. The quality of unskilled labor is defined as

$$Q_u = \exp(\phi(S_u)).$$

Here, S_u is defined as the average years of schooling of unskilled workers. ϕ is a function capturing the Mincerian returns to education. I use a functional form from Hall and C Jones (1999) and Caselli (2005) where $\phi(S)$ is a piecewise linear function with slope 0.13 for $S < 4$, a slope 0.1 for $S \in [4, 8)$, and a slope 0.08 for $S \geq 8$.

I measure S_u by using the data from Barro and Lee (2013). I assume that there is positive sorting between education and skill levels in occupation, and that S_u represents the average years of schooling of the share u of the population working in unskilled occupations. The Barro-Lee data does unfortunately not record the cumulative distribution of years of schooling, but only total schooling attainment within different levels of schooling. It records the number of schooling years at the primary level, the secondary level, and the higher level.

To calculate S_u, I first note that in the vast majority of countries, the cutoff between skilled and unskilled workers goes below the college level, and I attribute none of the schooling years in higher education to unskilled workers. To calculate the years of schooling in primary and secondary school that should be attributed to low skilled workers, I subtract 7 times the share of skilled workers from both primary and secondary school years, using the approximation that all skilled workers have finished high school and that primary and secondary school both are both 7 years. The results are not sensitive to details in this specification. After this subtraction, I divide the remaining primary and secondary school years with the share of unskilled workers to obtain S_u.

1.C   Estimating the relative price of skilled services

1.C.1   Theoretical derivation of gravity equation

In this section, I show how my gravity specification can be derived from theoretical trade models. I first derive the specification from an Armington style trade model, and then from an Eaton and Kortum style trade model.

Armington model

There are K industries and I countries, indexed i for source countries and j for destination countries. Each country admits a representative household with preferences

$$U_j = \left(\sum_{i=1}^{I} \sum_{k=1}^{K} (a_j^k)^{1/\sigma} (q_{i,j}^k)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \quad j = 1, \dots, I; \; \sigma > 1 \qquad (1.13)$$

where $q_{i,j}^k$ are goods from industry k produced in country i and consumed in country j, σ captures the elasticity of substitution between different varieties, and a_j^k is a country-specific taste term. The taste term is a reduced form way of capturing differences in tastes across countries, including potential non-homotheticities in preferences. The representative consumer maximizes (1.13) subject to a constraint

$$\sum_{i=1}^{I} \sum_{k=1}^{K} P_{i,j}^k q_{i,j}^k \leq Y_j$$

where $P_{i,j}^k$ is the price of good k produced in country i and bought in country j. Y_j is income in country j.

Each variety is produced using a constant returns to scale production

function with the unit cost function

$$c_i^k = C^k \left(r_{i,1}, \ldots, r_{i,F} \right) \tag{1.14}$$

where $r_{i,f}$ is the price of factor service f in country i.

Trade costs take an iceberg form and to consume one unit of a good from country i, a country j consumer has to buy $d_{i,j} \geq 1$ goods from country i. The cost term $d_{i,j}$ satisfies

$$
\begin{aligned}
d_{i,j} &\geq 1 \\
d_{i,i} &= 1 \quad \forall i = 1, \ldots, I \\
d_{i,j} d_{j,l} &\geq d_{i,l}.
\end{aligned}
$$

Output markets are competitive, which implies that prices satisfy

$$P_{i,j}^k = c_i^k d_{i,j}. \tag{1.15}$$

Each country has a supply of factor service flows

$$e_{j,f} \geq 0 \quad i = 1, \ldots, I; \ f = 1, \ldots, F,$$

and country income is given by

$$Y_j = \sum_{f=1}^{F} r_{j,f} e_{j,f} \tag{1.16}$$

An equilibrium is a set of consumption quantities $q_{i,j}^k$, production quantities Q_i^k, factor service prices $r_{i,f}$, unit costs c_i^k, output prices $P_{i,j}^k$, and incomes Y_j such that:

1. $\{q_{i,j}^k\}$ solves the consumer problem given output prices and incomes.

2. Output market clears

$$Q_i^k = \sum_{j=1}^{I} q_{i,j}^k d_{i,j} \forall i, k$$

3. c_i^k and $P_{i,j}^k$ satisfy (1.14) and (1.15) respectively

4. Income is given by (1.16)

5. Factor markets clear

$$e_{i,f} = \sum_k Q_i^k \frac{\partial c_i^k}{\partial r_{i,f}}$$

I will not solve the complete equilibrium, but will only solve for the regression specification relating industry export values to unit costs. In the data, export values between i and j in industry k are presented excluding trade costs (FOB). This corresponds to $P_{i,i}^k q_{i,j}^k$, i.e. the domestic price in i of good k produced in i. Using the competitive output market assumption, this quantity is $c_i^k q_{i,j}^k$.

Consumer optimization implies that for any country-industry pairs (i,k), (i',k')

$$\frac{(a_j^k)^{1/\sigma}(q_{i,j}^k)^{-1/\sigma}}{(a_j^{k'})^{1/\sigma}(q_{i',j}^{k'})^{-1/\sigma}} = \frac{P_{i,j}^k}{P_{i',j}^{k'}}$$

$$\sum_{i=1}^{I}\sum_{k=1}^{K} q_{i,j}^k P_{i,j}^k = Y_j$$

Re-arranging the terms gives us

$$P_{i,i}^k q_{i,j}^k = Y_j \frac{a_j^k (P_{i,j}^k)^{1-\sigma}}{\sum_{j',k'} a_j^{k'} (P_{i,j'}^{k'})^{1-\sigma}} \frac{P_{i,i}^k}{P_{i,j}^k}.$$

Taking logarithms, writing total exports $x_{i,j}^k = P_{i,i}^k q_{i,j}^k$, and substituting

in (1.14) for prices gives me

$$\log(x_{i,j}^k) = \delta_{i,j} + \mu_j^k - (\sigma - 1)\log(c_i^k) \tag{1.17}$$

where

$$\delta_{i,j} = \log(Y_j) - \log\left(\sum_{i',k'} a_j^{k'}(c_{i'}^{k'} d_{i',j})^{1-\sigma}\right) - \log(d_{i,j})$$

$$\mu_j^k = \log(a_j^k).$$

Here, $\delta_{i,j}$ captures all terms that only depend on the bilateral relationship: the income of the buying country, the market access term of the buying country, and all bilateral trading costs between the two countries. μ_j^k captures industry-specific demand effects in the buying country.

Eaton and Kortum model

To derive an industry based gravity equation using an Eaton and Kortum framework, I construct a model close to Chor (2010), who analyzed industry-level trade in an Eaton and Kortum setup. There are I countries where i is an index for a source country and j is an index for a destination country. The model has K goods which are produced domestically, and the production of each good k uses a range of internationally traded intermediate good varieties.

Each country has a representative consumer with preferences

$$U_j = \left(\sum_{k=1}^K a_j^k (Q_j^k)^{\frac{\xi-1}{\xi}}\right)^{\frac{\xi}{\xi-1}} \quad \xi > 1.$$

Each final good k is a composite of internationally traded varieties

$q_i^k(z)$ with $m \in [0,1]$. The price of final good k in country i is

$$P_j^k = \left(\int_0^1 p_j^k(m)^{1-\eta} dm \right)^{\frac{1}{1-\eta}}, \quad \eta > \xi > 1,$$

where $p_j^k(m)$ is the country j price of variety m in industry k. The assumption on the elasticity of substitution means that different varieties are more substitutable than goods from different industries.

As varieties are internationally traded, the price $p_j^k(m)$ paid for a variety will reflect the cheapest available variety for country j. When I specify the cost function for varieties, I am therefore interested in the unit cost of *offered* varieties from country i to country j, which I write $p_{i,j}^k(m)$. The price $p_j^k(m)$ is obtained by minimizing over potential source countries i.

The offered price $p_{i,j}^k(m)$ will depend on a deterministic component of costs in country i and industry k, on trade costs between country i and j, and on a stochastic productivity shock to this particular variety. The deterministic component of costs is

$$c_i^k = C^k(r_{i,1}, \ldots, r_{i,F}) \tag{1.18}$$

where $r_{i,f}$ denotes the factor service price of factor f in country i. Trade costs take an iceberg form and to obtain one unit of an intermediate good from country i, a country j producer has to buy $d_{i,j} \geq 1$ intermediate goods from country i. The cost term $d_{i,j}$ satisfies

$$
\begin{aligned}
d_{i,j} &\geq 1 \\
d_{i,i} &= 1 \quad \forall i = 1, \ldots, I \\
d_{i,j} d_{j,l} &\geq d_{i,l}.
\end{aligned}
$$

The offered price is

$$p_{i,j}^k(m) = \frac{c_i^k d_{i,j}}{z_i^k(m)} \tag{1.19}$$

where $z_i^k(m) \sim Frechet(\theta)$ is a country-industry-variety specific produc-

tivity shock which is Frechét distributed with a parameter θ. A random variable Z is Frechét-distributed with parameter θ if

$$P(Z \leq z) = e^{-z^{-\theta}}.$$

I will not solve a full equilibrium for this model, but only derive the gravity trade equation that results from the model. For each variety m in industry k, country j obtains an offer $p_{i,j}^k(m)$ from each country i given by equation (1.19). The probability distribution of this offer is

$$
\begin{aligned}
P(p_{i,j}^k(m) \leq p) &= P\left(\frac{c_i^k d_{i,j}}{p} \leq z_i^k(m)\right) \\
&= 1 - e^{-\left(\frac{c_i^k d_{i,j}}{p}\right)^{-\theta}} \\
&= 1 - e^{-\left(c_i^k d_{i,j}\right)^{-\theta} p^{\theta}}
\end{aligned}
$$

The best price $p_i^k(m)$ for country i is the minimum of all offers $\min_i p_{i,j}^k(m)$ and has distribution

$$
\begin{aligned}
G(p) &= P\left(\min_i p_{i,j}^k(m) \leq p\right) \\
&= 1 - P(\max_i p_{i,j}^k(m) > p) \\
&= 1 - \prod_i P(p_{i,j}^k(m) > p) \\
&= 1 - \prod_i (1 - P(p_{i,j}^k(m) \leq p) \\
&= 1 - e^{-\sum_i \left(c_i^k d_{i,j}\right)^{-\theta} p^{\theta}}
\end{aligned}
$$

I write

$$\Phi_j^k = \sum_i \left(c_i^k d_{i,j}\right)^{-\theta}. \tag{1.20}$$

This expression summarizes country j's access to industry k. It is decreasing in production costs in industry k and in the bilateral trading costs $d_{i,j}$.

Country j chooses to buy a variety from the country with the lowest price. The probability that country i offers the lowest price is

$$
\begin{aligned}
\pi_{i,j}^k &\equiv P(p_{i,j}^k(z) \leq \min_i p_{i,j}^k(z)) \\
&= \frac{(c_i^k d_{i,j})^{-\theta}}{\Phi_j^k}.
\end{aligned}
$$

If x_j^k is the total amount of intermediate inputs bought by country j in industry k, the trade flow matrix is

$$
x_{i,j}^k = \pi_{i,j}^k x_j^k = \frac{(c_i^k d_{i,j})^{-\theta}}{\Phi_j^k} x_j^k \tag{1.21}
$$

Equation (1.21) requires that the share of import value coming from country i only depends on the share of inputs for which i is the supplier. This property holds as the Frechet distribution has a desirable property called max-stability, which ensures that the best offered price $p_{i,k}(z)$ to country i is independent of the source of the best offer (see Eaton and Kortum (2002) for a derivation in this particular case, and Mattsson et al. (2014) for a more general discussion of this property of random variables) . This means that the total expenditure on imports from one country will be fully determined by the share of varieties $\pi_{n,i}^k$ bought from that country. The reason is that all countries offer identical distributions of variety prices conditioned on them offering the best prices.

Taking the logarithm of both sides of equation (1.21) gives me

$$
\log(x_{i,j}^k) = \delta_{i,j} + \mu_j^k - \theta \log(c_i^k)
$$

where $\delta_{i,j} = -\theta \log(d_{i,j})$ and $\mu_j^k = \log(X_j^k) - \log(\Phi_j^k)$. Thus, the model implies a gravity equation of the right form. Note that when using Eaton and Kortum elasticity estimates θ, there needs to be added a 1 to convert them to the corresponding Armington elasticity estimates σ.

1.C.2   Results for other factors than skilled labor

In Section 1.2, I estimated regression (1.5) to obtain estimates of relative factor service prices across countries. My main interest was in the relative price of skilled services, as this relative price is used directly in development accounting. However, my estimation procedure also yields relative factor service price estimates for capital, intermediate inputs, and energy. Even though I do not use these directly in my development accounting exercising, they are useful to check the plausibility of my factor service price estimation method.

In particular, as capital, intermediate inputs, and energy are partly tradable, we should expect the relative price of these factors compared to unskilled labor to fall with GDP per worker. The reason is that tradable services should have similar prices across countries, whereas we expect the price of unskilled labor services to rise with GDP per worker.

It is possible to quantify how much unskilled service prices should fall with GDP. If we assume that the labor share of output is constant at $1 - \alpha$, the unskilled wage satisfies equation

$$
\begin{aligned}
w_u &= \frac{w_u}{w_u u + w_s s} \times (w_u u + w_s s) \\
&= \frac{1}{u + \frac{w_s}{w_u} s}(1 - \alpha)y
\end{aligned}
$$

where y in the second line denotes output per worker. Using that the price of unskilled labor services is $r_u = w_u/Q_u$ where Q_u is the quality of unskilled workers, I obtain

$$
\log(r_u) = \log(1 - \alpha) + \log(y) - \log(h_{trad})
$$

where $\log(h_{trad}) = \log(Q_u) + \log(u + \frac{w_s}{w_u}s)$ is human capital according to traditional development accounting methods, as defined in equation (1.10). Letting r_t be the price of any tradable input service, its relative

price compared to unskilled labor services will be

$$\log\left(\frac{r_t}{r_u}\right) = \log(r_t) - \log(1-\alpha) - \log(y) + \log(h_{trad}).$$

If $\log(r_t)$ is constant across countries, we can make the following observation: constant $\log(h_{trad})$ across countries implies that relative tradable factor prices decrease one-to-one with GDP per capita. If $\log(h_{trad})$ is positively correlated with GDP, relative tradable factor service prices will fall slower than one-for-one. And even though it is not explicitly modeled in the equation, we can also note that a non-tradable component of t will also make the relative price/GDP-slope less negative.

In my data, $\log(h_{trad})$ increases at approximately $0.15-0.2$ with GDP per capita. Thus, if capital, intermediate inputs, and energy services are fully tradable, they should have a negative slope of between 0.8 and 0.85 with respect to GDP per worker. If they are not fully tradable, the negative relationship should be weaker. The results are presented in Figures 1.C.1-1.C.3. The negative relationship between capital and intermediate input service prices and log GDP per worker is similar at -0.6, which is close to what is predicted by my previous reasoning. The conclusions are less stable for energy prices. Here, there is also a negative relationship, but the data is less precise. This is due to energy having a very small factor share in most industries, and the results for energy are more driven by outliers. Reassuringly, large energy producers such as Saudi Arabia, Kuwait, Russia, and Iran have low revealed energy service prices.

1.C.3   Treatment of intermediate inputs

In my main specification, I include the cost share of intermediate inputs $\alpha^k_{US,int}$. The corresponding estimate $\beta_{i,int}$ identifies $\log\left(\frac{r_{i,int}/r_{i,1}}{r_{US,int}/r_{US,1}}\right)$. This estimate gives the difference between the US and country i in the relative cost of intermediate input and unskilled labor services.

In my interpretation of this parameter, I assume that intermediate

Figure 1.C.1: Log relative capital services prices and log GDP per worker
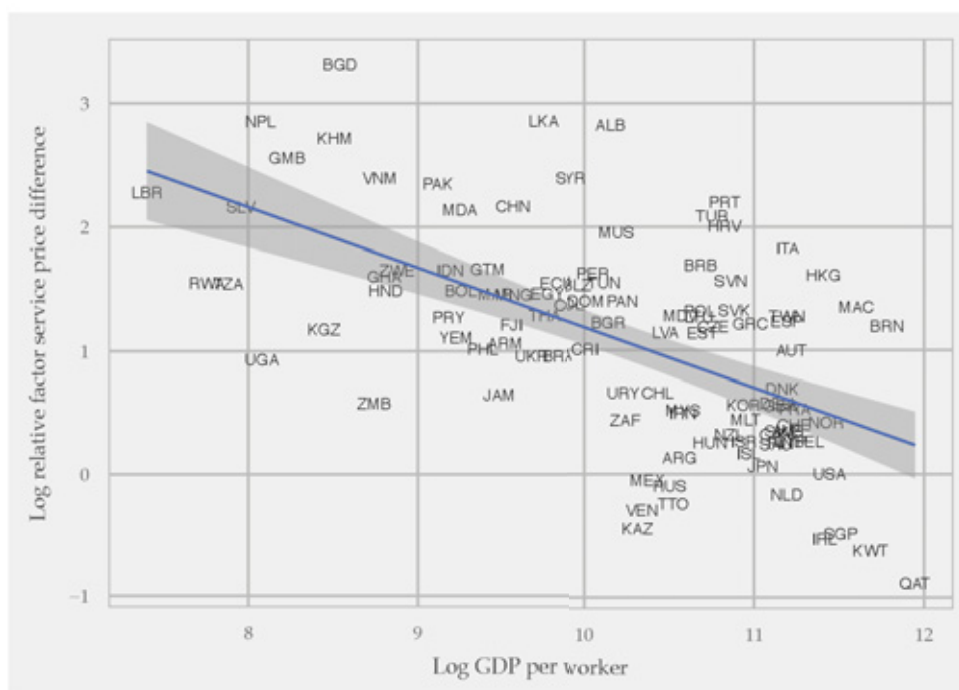
Figure 1.C.2: Log relative intermediate input services prices and log GDP per worker
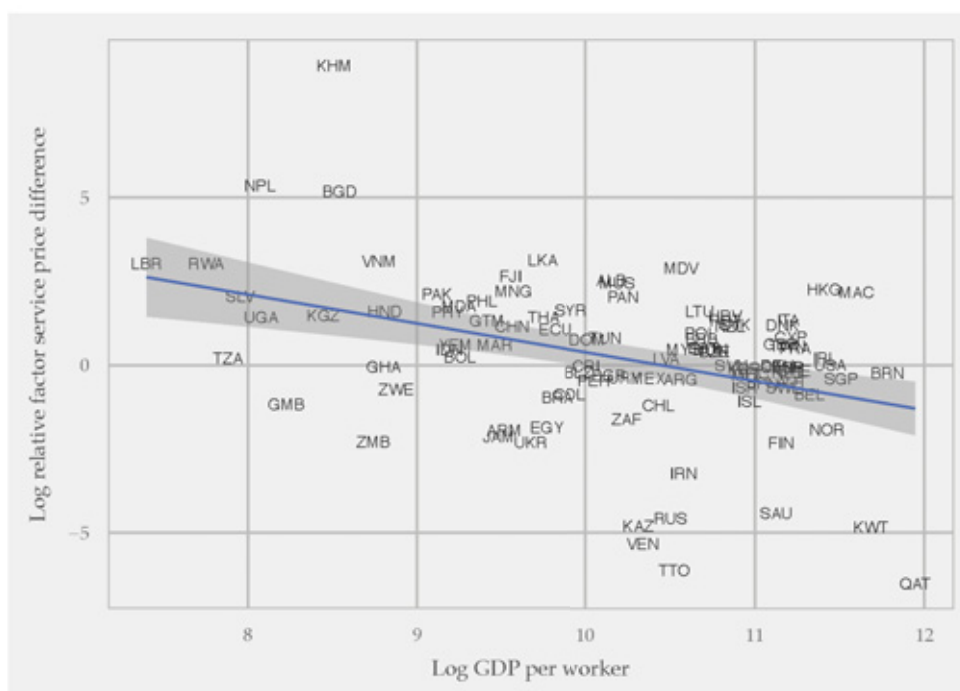
Figure 1.C.3: Log relative energy services prices and log GDP per worker

inputs are traded. I interpret $r_{i,int}$ as a product of an international price of intermediate inputs r_{int}, which is constant across countries, and a country-specific barrier to international intermediate input markets τ_i, which varies across countries.

With this interpretation,

$$\beta_{i,int} = \log(\tau_i/\tau_{US}) - \log\left(\frac{r_{i,1}}{r_{US,1}}\right).$$

$\beta_{i,int}$ varies across countries for two reasons. First, countries differ in their access to international intermediate goods markets τ_i. Bad access to international markets (high τ_i) gives a high revealed price of intermediate input services (high $\beta_{i,int}$). Second, countries differ in their prices of unskilled labor services $\log\left(\frac{r_{i,1}}{r_{US,1}}\right)$. Countries with a low price of unskilled services have a high revealed price of intermediate input services. This has an intuitive interpretation: relatively inexpensive unskilled labor services make internationally traded intermediate inputs relatively expensive.

If intermediate inputs are not traded and the aim is to identify factor service price differences, a different approach is called for. In this case, there is an indirect effect of factor service price differences via input prices. To reflect this, the intermediate input share in an industry k should be resolved into contributions from different factor services, using the input-output structure to determine the factor shares of industry k's intermediate inputs.

To check the robustness of my baseline specification, I develop an approach that allows for both traded and non-traded intermediate inputs. To implement my approach, I use the US input-output table and assume that services are non-traded and that other goods are traded.[23] I use the EU-KLEMS data together with Occupational Employment Survey data

---

[23]There is moderate trade in some services such as entertainment, financial services, and transportation, but the distinction captures the large differences in traded shares between services and other goods in the US input-output table.

to obtain factor shares in service sectors.

I write N_T for the number of traded goods and N_{NT} for the number of non-traded goods. The input-output table L is an $(N_T + N_{NT}) \times (N_T + N_{NT})$ matrix. For each good $k = 1, \ldots, N_T + N_{NT}$, I measure its factor shares including its intermediate input share, and I use these measured factor shares to define the *first-stage* factor shares $\tilde{\alpha}_f^k$. This is the same as normal factor shares with one difference. For intermediate inputs, we define $\tilde{\alpha}_f^k$ as the share of inputs that come from non-tradeable intermediates. In the first stage, I am interested in the cost shares of different factors and of tradable inputs. For each industry, $1 - \sum_{f=1}^{F} \tilde{\alpha}_f^k$ gives the share of costs in industry k going to nontraded factor inputs. These first-stage factor shares are the building blocks of the factor shares α_f^k that will be obtained by resolving the cost share of nontraded intermediate inputs into conventional factors and tradable inputs.

I find the factor shares α_f^k of tradable goods recursively by first finding the factor shares of nontradable goods. I define two matrices L_T and L_{NT} where L_T is an $N_T \times N_{NT}$ matrix giving the input uses of nontraded intermediate inputs in the traded sector, and L_{NT} is an $N_{NT} \times N_{NT}$ matrix giving the cost shares from nontraded inputs in the nontraded sector.

I solve the system recursively. The factor shares of nontraded goods are

$$\alpha_{NT} = \tilde{\alpha}_{NT} + (L_{NT})\alpha_{NT} \Longleftrightarrow \alpha_{NT} = (I - L_{NT})^{-1}\tilde{\alpha}_{NT}$$

where α_{NT} is an $N_{NT} \times F$ matrix, $\tilde{\alpha}_{NT}$ is an $N_{NT} \times F$ matrix, and L_{NT} is an $N_{NT} \times N_{NT}$ matrix. The final matrix α_{NT} gives the factor shares of nontraded services in terms of standard factor shares and traded input shares. All nontraded input shares have been resolved into these constituent parts. Having solved for the factor shares of nontraded goods, the factor shares of traded goods are

$$\alpha_T = \tilde{\alpha}_T + (L_T)\alpha_{NT}.$$

Using this modified definition of factor shares, I can re-estimate my
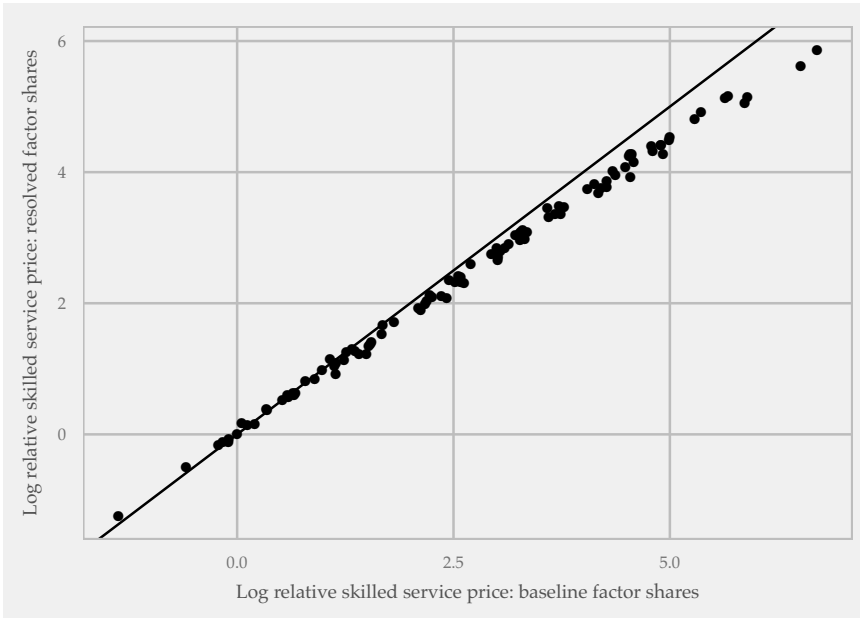
Figure 1.C.4: Comparison of estimated relative skilled service prices with different input measurements

baseline specification. In Figure 1.C.4, I compare the estimates for the estimated skilled service coefficient to my baseline estimation. The new results are very similar to my baseline estimates. The reason is that even though resolving the nontraded factors increases the skilled share in all industries (as I move the skilled component of inputs from the intermediate input share to the skill share), the resolving of nontraded factors does little to alter the *relative* skill shares across industries, which are the bases of my estimation.

1.D   Mechanisms

1.D.1   Interpretation of migration data

In this section, I analyze the relationship between my results and data on migrant wages. Migrant wage data has been an important source of

information in development accounting since Hendricks (2002).[24] Ideally, migration provides a natural experiment to distinguish between human capital based and technology based explanations of world income differences. If selection issues are appropriately addressed, migration data allows us to compare similar workers in two different environments. Human capital is kept constant, and wage differences have to be attributed to some other factor in the environment. Under some conditions, this other factor can be interpreted as technology.

In particular, migrant wage data has been used to argue against a dominant role for human capital in accounting for world income differences. This was the main argument in Hendricks (2002). He showed that migrants from poor countries in the US had dramatically higher wages than workers in their native countries. He argued that this was inconsistent with human capital differences being large enough to explain world income differences. Even though later contributions have tempered this conclusion by using individual data to account for selection (Hendricks and Schoellman, 2016), it remains important for human capital based explanations of world income differences to be consistent with migrant wage data.

Given that I argue that human capital might play a dominant role in explaining world income differences, it is natural to ask how my results relate to migrant wage data. In this section, I show that my results are not inconsistent with existing evidence from migrant wage data. The key difference between my analysis and that of Hendricks (2002) is that I relax the assumption of perfect substitutability and allow for imperfect substitutability between labor services. This leads to a different interpretation of migrant wage data. In particular, imperfectly substitutable labor services imply that human capital is multidimensional and that there is no longer a simple mapping from human capital to pre- and post-migration wages. A worker's wage is the product of the amount of labor services that the worker provides, and the price of those labor ser-

---

[24]In addition to Hendricks (2002), papers that use migrant data include Schoellman (2011), Lagakos et al. (2016), and Hendricks and Schoellman (2016).

vices. Even though wage changes at migration can be due to technology differences, they can also be due to labor service price differences.

I analyze the implications of my results for migrant wage data, and I discuss the implications both for unskilled and skilled migrants. When migrants are unskilled workers, my development accounting results imply that there are limited human capital quality differences between rich and poor countries. I am interested in whether these limited quality differences are consistent with large wage gains for unskilled migrants going from poor to rich countries. When migrants are skilled workers, my development accounting results suggest that there are substantial quality differences between rich and poor countries. In this situation, I am interested in whether these large estimated quality differences necessarily mean that skilled migrants going from rich to poor countries should have much higher wages than local workers, and conversely if skilled migrants going from poor to rich countries necessarily should have much lower wages than local workers.

Starting with unskilled migrants, I note that it is consistent with my results that unskilled workers going from poor to rich countries experience large wage gains. The mechanism is that in rich countries, the high relative supply and quality of skilled workers increase the relative price of unskilled labor services. This relative scarcity of unskilled labor services in rich countries makes unskilled wages higher. Wage gains for unskilled migrants are thus consistent with similar quality of unskilled labor across rich and poor countries, and these wage gains do not rely on large cross-country differences in technology.

For skilled migrants, I begin by analyzing skilled migrants going from rich to poor countries. According to my estimates, these skilled migrants have a much higher quality of human capital than their local counterparts. Does this imply that they will necessarily get much higher wages than local skilled workers? The answer is no. The reason is that high US quality of skilled labor is the result of an aggregation of heterogeneous skilled services. To make sharp predictions of how wages change at migration for a worker with particular skills, we need to know the com-

plementarity and substitutability patterns implicit in the aggregator of skilled services. More concretely, the question is not whether a standard US engineer migrating to Tanzania gets a much higher salary than the average Tanzanian engineer, but whether a US hydraulic engineer specializing in sediment transportation migrating to Tanzania gets a much higher salary than an average Tanzanian engineer. The latter question is not possible to answer without knowing the details of the skilled service aggregator.[25]

Conversely, we can analyze what my results predict about skilled migrants going from poor to rich countries. According to my results, poor countries have a substantially lower quality of skilled labor than rich countries. Do my results predict that a skilled migrant going from a poor country to a rich country necessarily should have a much lower salary than local skilled workers? The answer again is no. To begin with, the argument about complementarity and substitutability patterns that I made concerning skilled workers migrating from rich to poor countries still applies to this situation. Furthermore, even if we neglect potential heterogeneity among skilled and unskilled services, my explanation is still consistent with skilled migrants to rich countries not having dramatically lower wages than local counterparts. The reason is that there is potential for occupational switching at migration. Indeed, note that my results suggest that the relative price of skilled services is lower in rich countries than in poor countries. Thus, a worker that has a comparative advantage in a skilled occupation in a poor country might have a comparative advantage in a low skill

---

[25]Here, I use a low-dimensional representation of labor force heterogeneity to analyze cross-country differences, and a high-dimensional representation of labor force heterogeneity to analyze migration data. This procedure is analogous to the treatment of capital in aggregative growth models. The Solow model and the neoclassical growth models use a one-dimensional representation of capital, and these models are appropriate for capturing broad patterns of growth, output and marginal returns to capital. However, capital aggregation hides an underlying heterogeneity. This means that model predictions from these models are commonly not tested by comparing cross-country differences in rental prices of specialized pieces of equipment. Such comparisons are outside the domain of validity of the aggregate model setup. The same applies to my setup.

occupation in a rich country. For example, a moderately competent computer programmer from a poor country might find it profitable to work in an unskilled profession in the US. If the scarcity of unskilled services in the US has driven up unskilled wages, this is consistent with skilled migrants to rich countries only having moderately lower wages than their local skilled counterparts, compared to the large estimated quality differences in skilled labor. Even though B Jones (2014b) discusses the potential importance of occupational switching for migrant wage data, there has not been any full empirical examination of this mechanism. However, B Jones (2014a) and Hendricks and Schoellman (2016) provide suggestive evidence that occupational downgrading is more common for migrants from poor countries.

In conclusion, it is not directly inconsistent with migrant wage data that human capital might play an important role in explaining world income differences. This consistency is not due to migrant wage data confirming sharp predictions derived from my results. Apart from predicting wage gains for unskilled migrants going to rich countries, my results put weak restrictions on migrant wage data. Given the natural experiment aspect of migration that makes migrant wage data an attractive source of information about human capital differences, an important avenue of future work is to place restrictions on my setup to derive sharper predictions for migrant wage data.[26]

1.D.2   Endogenous skill-biased technology differences and human capital quality

Even though SBTD and quality differences are observationally equivalent with respect to price and quantity data, they are not equivalent in

---

[26]In an unpublished paper, B Jones (2014b) discusses migrant wage data with imperfect substitutability. He also argues for the importance of relative scarcity in accounting for the wages of unskilled migrants, and for occupational switching in accounting for the wages of skilled migrants. The points about complementarity and substitutability patterns are to my knowledge original to this paper.

general. If quality differences in skilled labor explain why countries are rich, development theory needs to explain why countries differ in their quality of skilled labor. If SBTD explain why countries are rich, development theory needs to explain how similar qualities of skilled labor can result in very different levels of skilled labor productivity.

Thus, I try to move beyond price and quantity data to gauge the relative merits of SBTD and quality difference interpretations of my estimates.[27] To this end, I examine whether SBTD reduce estimated quality differences when I put theoretical structure on how technology varies across countries. In particular, I analyze a standard mechanism from the literature where SBTD arise endogenously in response to relative factor service price differences (Caselli and Coleman, 2006; Acemoglu, 2007; Caselli, 2015). In Appendix 1.D.3, I test in general for estimation errors arising from second order errors and endogenous technology differences. There, I provide a detailed description of the environment, my measurement procedure, and my results. Below is a summary.

I set up a simple model of endogenous technological choice in line with Caselli and Coleman (2006) and Acemoglu et al. (2007). Technological bias varies on an industry-country level as a function of factor service prices. For each set of relative factor service prices, I generate unit cost data from the model and run my baseline regression specification on the model generated data. I find the relative factor service prices such that my regression specification gives the same results when applied to model data as when applied to actual data. This gives me the relative service factor prices that are consistent with my regression estimates given that there is endogenous SBTD. By comparing these relative factor service prices with those found under my baseline assumptions, I can test whether my results in Section 1.2 overstates rich-poor quality differences in skilled labor.

The results are mixed but there is no overall tendency for the endogenous SBTD based model to imply lower quality difference than my

---

[27]In Section 1.4.2 I did this by discussing circumstantial evidence for quality differences and economic mechanisms that could make them large.

baseline setup. In many cases, estimated quality differences are actually higher when I allow for SBTD. The exact results depend on parameters and the effects are non-monotone in the size of relative skilled service price differences.

These results can be viewed as somewhat surprising: if there are SBTD, a reasonable expectation is that they would lower the need for price differences in skilled labor services to explain the trade data. One mechanism that helps explain my findings is that there are to two opposing tendencies. SBTD reduce the need for quality differences as they increase the relative productivity of skilled labor. However, when SBTD are endogenous, they also increase the effective elasticity of substitution between skilled and unskilled labor services. In the context of my estimation procedure, this can sometimes mean that there are larger quality differences to explain away. The net effect is ambiguous.

Thus, accounting for SBTD becomes complex when SBTD are endogenous. I have not resolved all issues, and a more thorough investigation of endogenous SBTD in my context is an important avenue for future research. However, in the case when SBTD arise endogenously from relative factor service price differences, they do not unambiguously obviate the need for large quality differences to explain my estimates.

1.D.3   Robustness to industry function specification and endogenous technology bias

My baseline estimates relied on the assumption that it was possible to approximate unit cost differences from the US by log-linearizing around the US cost structure. In terms of assumptions on industry production functions, this assumption amounts to assuming that industry production functions are Cobb-Douglas. Furthermore, to interpret estimates r_s/r_u in terms of human capital, I needed to assume that there were no skill-biased technology differences between countries. This section tests the robustness of my results to deviations from these two assumptions.

The section has three subsections. In Appendix 1.D.3, I describe an

environment featuring CES industry production functions, and endogenous technology bias in response to relative factor service price variations along the lines of Caselli and Coleman (2006), Acemoglu et al. (2007), and Caselli (2015). In Appendix 1.D.3, I show how it is possible to quantify the extent of bias introduced by varying production function assumptions. Appendix 1.D.3 describes the results of the quantification exercise.

Environment

I assume that industry cost functions satisfy

$$
c_i^k \left(\frac{r_{i,1}}{Z_{i,1}^k}, \ldots, \frac{r_{i,F}}{Z_{i,F}^K} \right) = \left(\sum_{f=1}^{F} a_f^k \left(\frac{r_{i,f}}{Z_{i,f}^k} \right)^{1-\xi} \right)^{\frac{1}{1-\xi}} \qquad \xi > 0,
$$

where $r_{i,f}$ is the factor service price of factor f in country i, a_f^k is the factor share of factor f in industry k, $\xi > 0$ is the elasticity of substitution, and $Z_{i,f}^k$ is a factor-augmenting technology term.

The technology terms vary endogenously across countries in response to changes in relative factor prices. In modeling this choice, I follow Acemoglu (2007) and assume that there exists a cost function $G^k(Z_{i,1}^k, \ldots, Z_{i,F}^k)$ capturing the cost of acquiring a technology bundle. I assume that G^k is convex and homogenous of degree $\gamma > 1$. A country's technology bundle in an industry is the solution to

$$
\tilde{c}_i^k = \min_{\{Z_{i,1}^k, \ldots, Z_{i,F}^k\}} \left\{ c \left(\frac{r_{i,1}}{Z_{i,1}^k}, \ldots, \frac{r_{i,F}}{Z_{i,F}^k} \right) + \frac{P_i^k G^k(Z_{i,1}^k, \ldots, Z_{i,F}^k)}{\bar{Z}_i} \right\} \quad (1.22)
$$

where $\tilde{c}_i^k$ is the unit cost of good k in country i taking into account technology acquisition costs, $\frac{1}{\bar{Z}_i}$ is a country specific technology diffusion barrier, and P_i^k is the price of good k in country i. In equilibrium, $P_i^k = \tilde{c}_i^k$.

This specification aims at capturing a mechanism highlighted in the literature: the possibility of endogenous technology bias in response to

variations in relative factor service prices (Caselli and Coleman, 2006; Acemoglu, 2007; Caselli, 2015). Even though other mechanisms might be active, I have chosen a model specification that allows me to focus on this particular mechanism, and exclude other potential mechanisms. By defining technology choice as minimizing a unit cost, I preclude scale effects as my unit cost specification implies that the cost of acquiring technology scales with total industry production. By assuming that technology acquisition costs in an industry are denominated in industry output (which is implicit by including the price P_i^k), I preclude that technology choices are affected by the relative price of output and technology acquisition. Lastly, I assume that technology barriers $\frac{1}{Z_i}$ are common across factors and industries. This precludes that technology choices are affected by industry specific technology diffusion barriers, and it precludes that factor biases in technology arise due to factor specific technology diffusion barriers.

To solve for the technology choice, I take the first-order conditions associated with problem (1.22).

$$(c_i^k)_f \frac{r_{i,f}}{(Z_{i,f}^k)^2} = \frac{P_i^k G_f^k}{\bar{Z}_i} \quad f = 1, \ldots, F \tag{1.23}$$

where the subscripts f on c_k^i and G^k denote partial differentiation with respect to argument number f. Multiplying both sides by $Z_{i,f}^k$, summing over f, and using that c_i^k and G^k are homogenous of degree 1 and $\gamma > 1$ respectively, I obtain

$$c_i^k = \frac{\gamma P_i^k G^k}{\bar{Z}_i}.$$

This means that

$$
\begin{aligned}
P_i^k &= \tilde{c}_i^k \\
&= c_i^k + \frac{P_i^k}{\bar{Z}_i} G^k \\
&= c_i^k \left(1 + \frac{1}{\gamma}\right).
\end{aligned}
$$

Substituting this back into the first-order condition (1.23), I obtain

$$\frac{(c_i^k)_f r_{i,f}}{c_i^k Z_{i,f}^k} = \left(1 + \frac{1}{\gamma}\right) \frac{G_f^k Z_{i,f}^k}{\bar{Z}_i}. \tag{1.24}$$

Noting that the left-hand side is

$$\alpha_{i,f}^k = \frac{(c_i^k)_f r_{i,f}}{c_i^k Z_{i,f}^k},$$

where $\alpha_{i,f}^k$ is the factor share of factor f in industry k for country i, equation (1.24) captures the intuition that a country expands further in a factor-augmenting technology if it has a high share of its costs devoted to that factor.

I can provide a stronger characterization if I put more structure on G^k and assume that it is given by

$$G^k = \sum_{f=1}^{F} \tilde{a}_f^k (Z_{i,f}^k)^\gamma \quad \gamma > 0.$$

In this case, the factor bias can be expressed as

$$\frac{\alpha_{i,f}^k}{\alpha_{i,1}^k} = \frac{\tilde{a}_f^k}{\tilde{a}_1^k} \left(\frac{Z_{i,f}^k}{Z_{i,1}^k}\right)^\gamma$$

I normalize $\tilde{a}_f^k = \alpha_{US,f}^k$ to ensure that the US has no technological bias. In this case, the relative factor bias is

$$\left(\frac{Z_{i,f}^k}{Z_{i,1}^k}\right) = \left(\frac{\alpha_{i,f}^k/\alpha_{US,f}^k}{\alpha_{i,1}^k/\alpha_{US,1}^k}\right)^{\frac{1}{\gamma}} \tag{1.25}$$

The relative factor bias is uniquely determined by the relative factor shares compared to the US. The smaller is γ, the more strongly relative factor technologies react to relative factor service prices.

Quantification

In this section, I quantify how my baseline estimation is affected by the modified assumptions on the industry production functions. In particular, I test how well my baseline method estimates relative factor service prices $\log\left(\frac{r_{i,f}/r_{i,1}}{r_{US,f}/r_{US,1}}\right)$ in this new environment.

For this purpose, I solve for the technology choice $Z_{i,f}^k$ and for unit costs c_i^k given factor prices $r_{i,f}$. I then run a regression

$$\log(c_i^k) = \delta_i + \mu^k + \sum_{f=2}^{F} \tilde{\beta}_{i,f}\alpha_{US,f}^k \quad \tilde{\beta}_{US,f} = 0.$$

I am interested in which factor service price combinations $\tilde{r}_{i,f}$ that generate $\tilde{\beta}_{US,f}$ which are similar to the $\beta_{i,f}$ that I find in my baseline estimation (1.5). By comparing the baseline $\beta_{i,f}$ with $\log\left(\frac{\tilde{r}_{i,f}/\tilde{r}_{i,1}}{\tilde{r}_{US,f}/\tilde{r}_{US,1}}\right)$, I can test how well my baseline estimate $\beta_{i,f}$ captures the relative price of factor services $\log\left(\frac{r_{i,f}/r_{i,1}}{r_{US,f}/r_{US,1}}\right)$ in this new environment. I do not run the full trade regressions, as I only modify how factor prices map to unit costs, and I do not modify how unit costs map to trade flows. I perform the regression on c_i^k excluding technology acquisition costs. As equilibrium technology acquisition costs uniformly scale unit costs, they do not affect the regression.

The detailed implementation of my method is as follows. I assume that there are 84 industries corresponding to the NAICS 4-digit manufacturing industries used in the baseline specification, and that there are two countries: "Poor" and the US. I assume that there are only two countries to reduce the number of parameters to estimate, while still capturing broad differences between rich and poor countries. I normalize US factor prices $r_{US,f} \equiv 1$, and US unskilled technology $Z_{US,1}^k = 1$ for all $k = 1, \ldots, 84$. I set the technology choice parameters to $\tilde{a}_f^k = \alpha_{US,f}^k$ which normalizes US technologies to $Z_{US,f}^k = 1$ for all factors and industries. The normalization of the technology choice function implies a

normalization of the unit cost function, which becomes

$$c_i^k = \left(\sum_{f=1}^{F} \alpha_{US,f}^k \left(\frac{r_{i,f}}{Z_{i,f}^k} \right)^{1-\xi} \right)^{\frac{1}{1-\xi}}. \tag{1.26}$$

Furthermore, as only relative factor service prices are relevant for my estimation exercise, I can without loss of generality normalize $r_{Poor,1} = Z_{Poor,1}^k = \bar{Z}_i = 1$.

My task is to find $\tilde{r}_{Poor,f}$ for $f = 2, \ldots, F$ that replicate my baseline results. First, I use the CES industry production form to derive that

$$\frac{\alpha_{Poor,f}^k}{\alpha_{Poor,1}^k} = \frac{(\tilde{r}_{Poor,f}/Z_{Poor,f}^k)^{1-\xi}}{(\tilde{r}_{Poor,1}/Z_{Poor,1}^k)^{1-\xi}} \frac{\alpha_{US,f}^k}{\alpha_{US,1}^k}.$$

By combining this expression with equation (1.25), I obtain

$$\left(\frac{Z_{Poor,f}^k}{Z_{Poor,1}^k} \right) = \left(\frac{\tilde{r}_{Poor,f}/Z_{Poor,f}^k}{\tilde{r}_{Poor,1}/Z_{Poor,1}^k} \right)^{\frac{1-\xi}{\gamma}} \iff \frac{Z_{Poor,f}^k}{Z_{Poor,1}^k} = \left(\frac{\tilde{r}_{Poor,f}}{\tilde{r}_{Poor,1}} \right)^{\frac{1-\xi}{\gamma+1-\xi}}.$$

Thus, for each set of $\tilde{r}_{Poor,f}$, I can solve for technologies $Z_{Poor,f}^k$ and for unit costs c_{Poor}^k. I run the regression

$$\log(c_i^k) = \delta_i + \mu_k + \sum_{f=2}^{F} \tilde{\beta}_{i,f} \alpha_{US,f}^k \qquad \begin{array}{l} \tilde{\beta}_{US,f} = 0 \\ i = Poor, US \\ k = 1, \ldots, 84 \end{array}.$$

I solve for $\tilde{r}_{Poor,f}$ for $f = 2, \ldots, F$ such that $\tilde{\beta}_{Poor,f}$ matches $\beta_{Poor,f}$ from the baseline specification (I define $\beta_{Poor,f}$ by regressing my estimated $\beta_{i,f}$ on log GDP per worker $\log(y_i)$ for each f, and I define $\beta_{Poor,f}$ as the fitted value for $\log(y) = 9$). By comparing $\log \left(\frac{\tilde{r}_{Poor,s}/\tilde{r}_{Poor,1}}{\tilde{r}_{US,f}/\tilde{r}_{US,1}} \right)$ with $\beta_{Poor,s}$, I can gauge how biased my baseline estimation is in estimating the log relative price of skilled services. I test the effect of this bias on my development accounting exercise by redoing the development accounting

exercise using an estimate of relative skilled service prices

$$\log\left(\frac{r_{i,s}/r_{i,1}}{r_{US,f}/r_{US,1}}\right) = \left(\frac{\log(y_{US}) - \log(y_i)}{\log(y_{US}) - 9}\right)\log\left(\tilde{r}_{Poor,s}\right).$$

Results

Table 1.D.1 shows the share of income differences explained by human capital for different values of the elasticity of substitution ξ and the endogenous technology parameter γ. A large γ means that technology choices are insensitive to variations in relative factor service prices. Unsurprisingly, we see that the endogenous technology choice parameter γ does not matter when $\xi = 1$. In this case, the production function is Cobb-Douglas and all technology differences are neutral. Furthermore, when $\gamma = 5$, the results for different ξ are similar to those found in Table 1.8 when there were no endogenous technology differences. This reflects that with a large γ, technology choices respond weakly to changes in relative factor service prices.

Overall, there is no monotone effect of endogenous technological change on the importance of human capital. For $\xi = 0.6$ and $\xi = 0.8$, making endogenous technology choices more flexible (smaller γ) makes human capital less important. For $\xi = 1.4$, making technology choices more flexible makes human capital dramatically more important. Overall, no specification reduces the importance of human capital below 50%.

Table 1.D.1: Share of income differences explained by human capital for different ξ and γ

	$\gamma = 1.01$	$\gamma = 1.1$	$\gamma = 2$	$\gamma = 5$
$\xi = 0.6$	0.55	0.56	0.62	0.65
$\xi = 0.8$	0.61	0.61	0.62	0.63
$\xi = 1$	0.64	0.64	0.64	0.64
$\xi = 1.2$	0.69	0.69	0.69	0.69
$\xi = 1.4$	6.83	6.48	1.01	0.90

1.E   Robustness

1.E.1   Discussion: Industry-dependent trade elasticities

In my estimates, I assume that the elasticity of trade σ is common across industries. A number of papers in the trade literature has argued for σ varying at an industry level (Broda et al., 2006; Soderbery, 2015). I write σ_k to denote such an industry-varying trade elasticity. Looking ahead, an important extension of my paper is to redo the estimates with a serious treatment of industry-varying σ. However, I have performed a simple robustness check, and tested a number of ways of solving the problem. Here, I also outline which approaches to this that look relatively more promising.

First, I note that it is possible to use residual plots to detect evidence for industry-varying σ_k. If σ_k is higher than average in an industry, a plot of fitted values and residuals will have a positive slope. Indeed, if a country has high fitted trade values in an industry, it suggests that it has low relative costs. If I use an elasticity for that industry which is too low, the fitted value will be low compared to the actual value. The opposite is true when an industry has a low fitted value of trade. If I have underestimated the trade elasticity, actual values will be even lower than fitted values. These effects mean that an underestimated σ_k leads to a positive relationship between fitted values and residuals on an industry level. Conversely, if I have overestimated σ_k, there will be a negative relationship between fitted values and residual values.

By considering industry-by-industry plots of residuals on fitted values, I can obtain information about industry-specific elasticities. I use this method to perform a simple robustness check by excluding all industries with an absolute value of the residual-fitted plot of more than 1 and I find similar results for this restricted set of industries.

I also run the regression specification

$$\log(x_{i,j}^k) = \delta_{i,j} + \mu_j^k - \sum_{f=2}^{F} [(\sigma_k - 1)\alpha_{US,f}^k]\beta_{i,f} + \varepsilon_{i,j}^k \quad \beta_{US,f} \equiv 0$$

and use different estimates of σ_k across industries. I first use the estimates of industry-specific trade elasticities in Broda et al. (2006). To test whether these help resolve the problem with varying trade elasticities, I analyze whether there is less evidence for industry-varying trade elasticities in the fitted-residual plots when I use the industry-specific estimates σ_k from Broda et al. (2006) compared to when I run the regression with a common elasticity of trade corresponding to their median estimate.

I find that using the industry-specific estimates of trade elasticity do not resolve the problem of correlation between fitted values and residuals on the industry level. If anything, using industry-specific elasticity estimates makes the problem worse.

In addition to using the estimates from Broda et al. (2006), I also try an iterative procedure to more directly bring the fitted-residual plots in line. I run the regression with a common $\sigma_k \equiv \sigma$. I iterate and increase the σ_k whenever the fitted-residual slope in industry k is positive, and decrease σ_k whenever the fitted-residual slope in industry k is negative. Unfortunately, the procedure does not converge.

Using estimates from Broda et al. (2006) and the iterative procedure did not solve the problem with varying trade elasticities. One potential reason for this failure is that it is not theoretically correct to modify regression specification (1.5) by just changing σ_k. If trade elasticities vary across industries, they also interact with trade cost terms that are now included in the bilateral fixed effect $\delta_{i,j}$. Thus, this will partly depend on industry k, which means that a standard gravity specification with bilateral fixed effects will not work in this context.

Thus, looking ahead, a proper treatment of varying σ_k will require a way of jointly estimating σ_k across industries and modify the structural trade model to generate a regression specification that fully incorporates varying trade elasticities.

1.E.2 Specification with confounding variables

In Section 1.6.1, I discuss the effects of an omitted variable bias in my specification of unit costs. Here, I explain how I measure and include potential confounders in my regression specification.

I analyze two confounding variables: external financing sensitivity and contracting sensitivity. I assume that there are country-specific contracting and external financing penalties τ_{cont} and τ_{fin} which capture the general quality of a country's judicial and financial systems. Industries are characterized by a contracting intensity $\alpha_{US,con}^k$ and a external financing intensity $\alpha_{US,fin}^k$. Country-level contracting and financial penalties change log unit costs of industries with $\tau_{cont} \times \alpha_{US,con}^k$ and $\tau_{fin} \times \alpha_{US,fin}^k$, respectively. That is, contracting and financing penalties increase the unit costs of industries in proportion to their respective contracting and external financing intensities.

I define an industry's external financing intensity $\alpha_{US,fin}^k$ as the share of investment expenditure not covered by external financing (external financing share of investments) times the share of gross output devoted to investments. I take the external financing share from Rajan and Zingales (1998), and I measure the investment share of total output using NBER CES data. My definition differs from that in Rajan and Zingales (1998) as I multiply the external financing share of investments with the investment share. The reason is that I interpret the country financial penalty τ_{fin} as a markup on external financing needs. A financing penalty increases the unit costs of an industry in proportion to its external financing needs as a share of gross output. To obtain external financing needs as a share of gross output, I multiply the external financing share of investments with the investment share.

I define an industry's contracting intensity by multiplying two terms. The first term is the share of intermediate inputs expenditure that is sensitive to contracting. To measure this term, I follow Nunn (2007) and use the IO table to calculate the share of intermediate good expenditures that are spent on customized inputs, where I define an input as

customized if it is not traded on an exchange nor referenced priced in a trade journal according to the classification of goods in Rauch (1999). The second term in my calculation is the intermediate input cost share, defined as total intermediate good expenditures divided by gross output. I measure this term using NBER CES data. I calculate the contracting intensity $\alpha_{US,con}^k$ as the product of these two terms, i.e. as the product of the share of customized inputs and the intermediate input cost share. My contracting sensitivity method is a slight modification of the measure in Nunn (2007), which only uses the first of my two terms. My modification reflects that I interpret the country contracting penalty τ_{cont} as increasing the cost of contracting sensitive inputs due to the lack of relation-specific investments. The unit cost effect of this is proportional to the total cost of contracting sensitive inputs as a share of gross output. As Nunn's definition only gives the cost of contracting sensitive inputs as a share of intermediate input costs, I multiply his measure with the intermediate input share to obtain my final measure $\alpha_{US,con}^k$.

I include $\alpha_{US,con}^k$ and $\alpha_{US,fin}^k$ in the analysis by adding extra terms to the regression specification (1.5), and run the regression

$$
\begin{aligned}
\log(x_{i,j}^k) \;=\;& \tilde{\delta}_{i,j} + \tilde{\mu}_j^k - \sum_{f=2}^{F}[(\sigma - 1)\alpha_{f,US}^k] \times \beta_{f,i} \\
& -[(\sigma - 1)\alpha_{fin.,US}^k] \times \beta_{fin,i} - [(\sigma - 1)\alpha_{con,US}^k] \times \beta_{con,i} + \varepsilon_{i,j}^k.
\end{aligned}
$$

where $\alpha_{fin.,US}, \alpha_{contr.US}$ give the financial and contracting intensity measured on US data.

1.E.3   Comparison with unit costs

My unit cost analysis uses the Groningen Growth and Development Center's (GGDC) 2005 benchmark producer price index. This data set aims at providing a cross-country comparable producer price index for 34 industries across 42 countries. The index covers both tradable and nontradable industries, and manufacturing as well as services (Inklaar and

Timmer, 2008).

Following recommendations from a creator of the data set, I exclude financial services, business services, real estate, government, health services and education. For these industries, it is difficult to obtain data on output quantities which makes it difficult to make cross-country comparisons in unit costs. I also exclude "private households with employed persons" as this variable is missing for a large number of countries. After my exclusions, I am left with a total of 27 industries and 35 countries with a complete set of observations.

To obtain factor shares, I use the EU KLEMS data set for the US (as my analysis includes non-manufacturing industries, I cannot use the NBER CES database to obtain factor shares). For the US, EU KLEMS provides data on industry level gross output, labor compensation, and intermediate good compensation. I define the labor share as the labor compensation over gross output, and the intermediate share as the intermediate good compensation over gross output. I calculate the skill share by multiplying the labor share with the share of payroll going to skilled workers with an occupational skill level of 3 or 4. I define the capital share as one minus the other factor shares.

I run the regression

$$\log(c_i^k) = \delta_i + \mu_k + \sum_{f=2}^{F} \alpha_{US,f}^k \tilde{\beta}_{i,f} + \varepsilon_i^k$$

where $\tilde{\beta}_{i,f} = \log\left(\frac{r_{i,f}/r_{i,1}}{r_{US,f}/r_{US,1}}\right)$ captures the deviation of relative prices compared to the US.

I compare the results from the unit cost analysis with the trade data analysis by comparing the relationship between GDP per worker and $\tilde{\beta}_{i,f}$ with the relationship between GDP per worker and $\beta_{i,f}$, where $\beta_{i,f}$ comes from the trade data analysis.[28]

---

[28]An alternative way to compare the outcomes would be to regress $\beta_{i,f}$ on $\tilde{\beta}_{i,f}$ and test how close the results are to a 45 degree line. I have chosen my method as I am interested in broad correlations between skilled service prices and GDP per capita, and given the estimation errors in the skill price estimates, regressing them on each

In Figures 1.8 and 1.9, I plot the results from the unit cost data analysis. The slope parameter of log relative skilled service prices on log GDP per worker is -1.19 using the unit cost data, and -1.53 using the trade data method for the same set of countries. I cannot reject that the two coefficients are equal, even without taking into account the large standard errors on the unit costs based parameters $\tilde{\beta}_{i,f}$. Thus, when both types of data exist, the trade data method and the unit cost method paint a similar picture of the relationship between relative skilled service prices and GDP per worker.

1.E.4   Differences in unskilled human capital quality Q_u

In the current setup, I estimate the quality of unskilled labor Q_u by assuming that unschooled labor is of equal quality and that improvements are reflected in Mincerian returns:

$$Q_{U,i} = \exp(\phi(S_{U,i}))$$

where ϕ is a Mincerian return function and $S_{U,i}$ is the average schooling time of unskilled labor.

A number of papers on human capital and development accounting have stressed that there might be uniform quality differences in human capital (Caselli, 2005; Manuelli and Seshadri, 2014). These quality differences might reflect differences in nutrition, health, or the quality of early schooling.

As my paper estimates Q_u and Q_s/Q_u any uniform increase in Q_u will also increase Q_s proportionally.

---

other biases the results down due to measurement error. Regressing $\beta_{i,f}$ on $\tilde{\beta}_{i,f}$ and regressing $\tilde{\beta}_{i,f}$ on $\beta_{i,f}$ both yield a regression coefficient of less than one.

Chapter 2

Price Level Determination When Tax Payments Are Required in Money[*]

2.1 Introduction

> "A prince who should enact that a certain proportion of his taxes should be paid in a paper money of a certain kind might thereby give a certain value to this paper money" – Smith (1776)

> "The modern state can make anything it chooses generally acceptable as money and thus establish its value quite apart from any connection, even of the most formal kind, with gold or with backing of any kind.... if the state is willing to accept the proposed money in payment of taxes and other obligations to itself the trick is done" – Lerner (1947)

"The fiscal theory of the price level recognizes that nominal debt, including the monetary base, is a residual claim to government primary surpluses, just as Microsoft stock is a residual claim to Microsoft's earnings. [...] An equivalent view is that money is valued because the government accepts it for tax payments – the 'public' part of 'This note is legal tender for all debts, public and private.' If the government requires money for tax payments at the end of the day, money will be valued in trade during the day." – Cochrane (2005)

"...paper currencies have value because they're backed by the power of the state, which defines them as legal tender and accepts them as payment for taxes." – Krugman (2013)

From the 18th century to this date, economists have conjectured that the value of government issued money is connected to the requirement that taxes be paid in this money. As indicated by the quotes, the ideas about the nature of this connection differ. One idea is that tax requirements exclude equilibria in which money has no value, another is that tax requirements support the general acceptance of money as a medium of exchange, and another still is that tax requirements provide a mechanism of price level determination that is equivalent to the fiscal theory of the price level.

Recent monetary events highlight the practical importance of understanding the connection between the value of money and tax requirements. The aim of developing bitcoin was explicitly to create a fiat currency with positive value without any government backing. On the other hand, the Greek contingency plan for leaving the euro in the summer of 2015 explicitly relied on tax requirements to sustain the value of a new species of money. The former finance minister Varoufakis (2015) claimed that they planned to make tax credits electronically transferable, and that tax credits would temporarily become the main means of payments in the country.

In this paper, we perform a theoretical analysis of the relationship

between tax requirements in money and the price level. We provide conditions for when tax requirements in money together with a money supply rule uniquely determine the price level.

More specifically, we study an infinite-horizon endowment economy with the added constraint that households have to pay their tax liabilities with their money holdings. We show that when the money supply M and the real value of taxes τ are constant, there is a unique stationary equilibrium. The price level is $P_t = M/\tau$. There is also a continuum of non-stationary equilibria where the price level P_t is less than M/τ and falling over time. Tax requirements in themselves provide an upper bound for the price level, but do not uniquely determine the price level. However, we show that all non-stationary equilibria can be excluded if households face some arbitrarily lax borrowing constraint.

What is the intuition? To find the unique stationary equilibrium, we note that if $P_t > M/\tau$, the money supply M is not sufficient to cover the tax liability τP_t and there is excess demand for money. If $P_t < M/\tau$, the tax constraint is lax, and a strictly positive amount of money $\hat{M}_t = M - \tau P_t > 0$ is carried into the next period. Households only carry money between periods if money earns a return equal to the real interest rate, which means that the price level must fall. Thus, $P_t = M/\tau$ is the unique stationary equilibrium.

To find the non-stationary equilibria, we note that for each initial price level $P_0 < M/\tau$, the economy starts on a deflationary path where the real value of money holdings increases faster than the real interest rate. Even if the value of money holdings grow faster than the real interest rate, the household's transversality condition can be satisfied if the positive money holdings are matched by negative bond holdings of equal size. We show that for each $P_0 < M/\tau$, there exists an equilibrium where this happens. This means that there is a continuum of deflationary equilibria.

This construction also suggests how non-stationary equilibria can be excluded. In all non-stationary equilibria, the real value of bond holdings tends to minus infinity. This means that we can exclude these equilibria

if we impose an arbitrarily lax borrowing constraint, which in our view is a mild assumption.

Motivated by the previously mentioned quote in Cochrane (2005), we also clarify the relationship between our model and the fiscal theory of the price level (FTPL). We introduce initial government debt $B_{-1} = B > 0$ in our baseline model and show that the key difference between our model and the FTPL is how the government can select its surplus sequence. In the FTPL, the government can freely commit to a surplus sequence and the price level adjusts to equate B/P_0 to the present value of the surplus sequence. In our model, the price level $P = M/\tau$ is determined by the money supply and the tax level, and the price level P, in turn, constrains the government's choice of surplus sequence. Indeed, the existence of an equilibrium requires that the government selects a surplus sequence with the present value $B/(M/\tau)$. Thus, our model is different from the FTPL. Our model also introduces novel fiscal-monetary interactions as taxes affect money demand. It is an interesting avenue of future research to fully map out the fiscal-monetary interactions in our model.

We make two extensions to our model to address two issues regarding price level indeterminacy previously discussed in the literature. First, we augment the baseline model with a liquidity motive for holding money in order to analyze the issue of speculative hyperinflations. These hyperinflationary equilibria are a generic feature of models where money has liquidity value; they arise in money-in-the-utility models (Obstfeld and Rogoff, 1983), in the OLG model of money (Woodford, 1984), and in search-and-matching models of money (Williamson and Wright, 2010). Second, we analyze the issue of price level indeterminacy under a balanced budget rule, previously treated in Schmitt-Grohé and Uribe (2000).

We show that tax requirements can exclude hyperinflationary equilibria in a money-in-the-utility model. The reason is that these equilibria depend on a self-fulfilling belief in an ever increasing price level. Since tax requirements place an upper bound on the price level, these equilibria are excluded and the stationary equilibrium is unique.

Price level determinacy under a balanced budget rule is more in-

volved. In this setting, the government has to print money in proportion to the price level if it wants to keep government consumption constant. To avoid equilibrium multiplicity, taxes have to be conditioned on the price level such that taxes increase more than proportionally with the price level. This ensures that there is excess demand (supply) for money when the price level is above (below) some target price level $\bar{P}$, which becomes the unique equilibrium price level. The analysis suggests that a balanced budget rule makes it more challenging, but not impossible, to achieve price level determinacy.

We conclude the paper with a section where we discuss the potential implications of our theory for the conduct of monetary policy. We conjecture that tax requirements will change little if the central bank follows an interest rate rule. However, money supply rules become more attractive with tax requirements, since the government can control money demand.

As seen in the introductory quotes, it is not a new idea to connect the value of money to tax requirements. Apart from verbal expositions, two theoretical contributions are Starr (1974) and Goldberg (2012). Starr analyzes how tax requirements need to be structured in order to give money a positive value in a static environment, but he never extends his analysis to a dynamic setting. Goldberg studies a dynamic search-matching model of money where there is a probability of meeting a tax collector who requires money payments, and provides conditions under which non-monetary equilibria are excluded. To our knowledge, ours is the first paper to study tax requirements in a standard infinite horizon endowment economy.

Throughout the paper, we sketch the proofs to our propositions in the main text, and provide the complete proofs in the Appendix.

2.2   The model

2.2.1   Environment and equilibrium definition

We study a discrete-time infinite horizon endowment economy. There is a representative household which consumes a single good. The household has a per-period utility function u which is concave, increasing and satisfies the Inada conditions. Money holdings do not enter the utility function, and money plays no special role in private transactions.

The household enters each period with money holdings $\hat{M}_{t-1}$, receives bond payouts B_{t-1} in money, and receives an endowment of y goods. The household then trades for consumption goods c_t at price P_t, for money M_t at a normalized price of unity and for nominal government bonds B_t at price Q_t. After trading, the household uses its acquired money holdings M_t to pay nominal taxes T_t. Any remaining money $\hat{M}_t = M_t - T_t \geq 0$ is carried over to the next period. We denote M_t *within-period money* and $\hat{M}_t$ *between-period money*.

With this timing and setup, the household solves:

$$\max_{\{c_t, B_t, M_t, \hat{M}_t\}} \quad \sum_{t=0}^{\infty} \beta^t u(c_t)$$

subject to the following per-period budget constraints

$$c_t + \frac{M_t}{P_t} + \frac{Q_t B_t}{P_t} \leq y_t + \frac{B_{t-1}}{P_t} + \frac{\hat{M}_{t-1}}{P_t} \tag{2.1}$$

and initial conditions

$$\hat{M}_{-1} = B_{-1} = 0. \tag{2.2}$$

The tax requirements are modeled by the two per-period constraints

$$\hat{M}_t + T_t \;\; \leq \;\; M_t \tag{2.3}$$

$$0 \;\; \leq \;\; \hat{M}_t. \tag{2.4}$$

The constraint (2.3) states that money is used to pay taxes and to purchase between-period money $\hat{M}_t$. Constraints (2.3) and (2.4) jointly imply that the household needs to buy enough money to cover its tax liability, i.e. $M_t \geq T_t$. Lastly, the household faces a no-Ponzi constraint

$$\lim_{T \to \infty} \frac{\left(\prod_{t=0}^{T-1} Q_t\right)(\hat{M}_T + B_T)}{P_0} \geq 0. \tag{2.5}$$

The government issues money $M_t - (\hat{M}_{t-1} + B_{t-1})$ and bonds B_t to acquire goods g_t. This formulation of money issuance guarantees that there is always M_t money available for the household to pay its taxes. Trading takes place before taxes are levied, and the per-period government budget constraint is

$$g_t P_t = M_t - (\hat{M}_{t-1} + B_{t-1}) + Q_t B_t. \tag{2.6}$$

After trading, the government levies taxes T_t. A *feasible* government policy is a sequence $\{g_t, M_t, T_t, B_t\}$, such that (2.6) holds for all $\{P_t, Q_t, \hat{M}_{t-1}\}$.

Definition 1. *A competitive equilibrium under a feasible government policy $\{g_t, M_t, T_t, B_t\}$ is a sequence $\{c_t, \hat{M}_t, P_t, Q_t\}$ such that*

- *$\{c_t, B_t, M_t, \hat{M}_t\}$ solves the household problem given $\{T_t, Q_t, P_t\}$.*

- *The markets for goods and money clear:*

$$c_t + g_t = y_t \tag{2.7}$$
$$\hat{M}_t + T_t = M_t. \tag{2.8}$$

2.2.2 Equilibrium characterization

To derive the household's optimality conditions, we use that the per-period budget constraint (2.1) and the tax constraint (2.3) always bind.

We substitute the tax constraint into the budget constraint and obtain:

$$c_t + \frac{T_t}{P_t} + \frac{(1 - Q_t)\hat{M}_t}{P_t} + \frac{Q_t}{P_t}\left(\hat{M}_t + B_t\right) = y_t + \frac{1}{P_t}\left(\hat{M}_{t-1} + B_{t-1}\right).$$
$$(2.9)$$

This expression shows that equilibrium requires a no arbitrage condition on interest rates

$$Q_t \leq 1, \tag{2.10}$$

as the household could otherwise obtain unbounded consumption by increasing $\hat{M}_t$.

The first-order conditions of the household problem with respect to $\hat{M}_t$, B_t, and c_t give us

$$u'(c_t) = \beta \frac{P_t}{Q_t P_{t+1}} u'(c_{t+1}) \tag{2.11}$$

$$(1 - Q_t)\hat{M}_t = 0. \tag{2.12}$$

Equation (2.11) is the Euler equation and equation (2.12) states that money is never held between two periods if the nominal interest rate is positive. Optimality also requires that the solution satisfies a transversality condition, and if we combine this transversality condition with the no-Ponzi constraint (2.5), we obtain an equation for asymptotic wealth holdings:

$$\lim_{T \to \infty} \frac{\left(\prod_{s=0}^{T-1} Q_s\right)\left(\hat{M}_T + B_T\right)}{P_0} = 0. \tag{2.13}$$

The initial condition (2.2), the non-negativity constraint on money (2.4), and equations (2.9)-(2.13) jointly characterize the solution for $\hat{M}_t, B_t$, and c_t given Q_t, P_t, and T_t. Using the fact that the constraint (2.3) binds at the optimum, we can also solve for M_t.

An equilibrium is fully characterized by the market clearing conditions (2.7)-(2.8), the zero lower bound (2.10) and the household optimality conditions (2.2),(2.4), (2.9)-(2.13).

2.2.3 Existence and uniqueness of a stationary equilibrium

We are interested in the existence and uniqueness of a stationary equilibrium. We focus on a parametrization where money supply, the real value of taxes, and government consumption are constant: $M_t = M$, $T_t = \tau P_t$, and $g_t = \tau$. The bond sequence follows from the government per-period budget constraint (2.6), and is given by $B_t = \frac{\tau P_t - M + (\hat{M}_{t-1} + B_{t-1})}{Q_t}$.

In Proposition 1, we show that there exists a unique stationary equilibrium, but also a continuum of non-stationary equilibria with $P_t \to 0$. In Proposition 2 we show that the non-stationary equilibria can be excluded if we impose an arbitrarily lax borrowing constraint on the household.

Proposition 1. *Suppose that the government policy is given by $M_t = M$, $T_t = \tau P_t$, $g_t = \tau$, and $B_t = \frac{\tau P_t - M + (\hat{M}_{t-1} + B_{t-1})}{Q_t}$ for all t. Then, there exists a unique stationary equilibrium, with $P_t = \frac{M}{\tau}$. There also exists an infinite number of non-stationary equilibria with $P_t \le M/\tau$ for all t and $P_t \to 0$.*

Proof. *See Appendix.*

The key to Proposition 1 is that any stationary equilibrium features a positive nominal interest rate. This means that the household carries no money between periods, i.e. $\hat{M}_t = 0$. The constraint (2.3) then implies $\tau P_t = M$, i.e. $P_t = \frac{M}{\tau}$ is the unique stationary equilibrium.

We construct a continuum of non-stationary equilibria by showing that an equilibrium exists for any $P_0 < \bar{P} = M/\tau$ (there can never be an equilibrium with a price level larger than M/τ, as the joint constraints (2.3)-(2.4) would be violated with such a price level). If $P_0 < M/\tau$, the constraint (2.4) is slack, which means that $\hat{M}_0 > 0$ and the household carries between-period money. When the household voluntarily carries between-period money, the nominal interest rate must be zero (i.e. $Q_0 = 1$), as seen from equation (2.12). The Euler equation (2.11) then implies $P_1 = \beta P_0$. As $P_1 < M/\tau$, the tax constraint is also slack for $t = 1$, and

iterating on t we obtain

$$P_t = \beta^t P_0. \tag{2.14}$$

Intuitively, money has to to earn a real return $1/\beta$, and the price level falls over time. Between-period money holdings $\hat{M}_{t-1}$ will approach M as the nominal value of tax payments tends to zero

$$\lim_{t\to\infty} T_t = \lim_{t\to\infty} \tau P_t = 0 \tag{2.15}$$

$$\lim_{t\to\infty} \hat{M}_{t-1} = \lim_{t\to\infty} M - \tau P_{t-1} = M. \tag{2.16}$$

To test whether the transversality condition holds, we are interested in the limiting behavior of household wealth. Here, we note that the real value of money holdings grows faster than the real interest rate. Indeed, if we consider the growth rate of real money holdings

$$\frac{\hat{M}_{t-1}/P_t}{\hat{M}_{t-2}/P_{t-1}} = \frac{M - \tau P_{t-1}}{M - \tau P_{t-2}} \times \frac{P_{t-1}}{P_t},$$

we note that there are two sources of growth. First, the nominal amount of money holdings increases as τP_{t-1} shrinks. Second, the value of nominal money increases with P_{t-1}/P_t, which equals the real interest rate. Therefore, the growth rate of the real value of money holdings is higher than the real interest rate.

How can the fact that money holdings increase faster than the real interest rate be consistent with the household's transversality condition? The reason is that the household simultaneously decreases its real bond wealth $\frac{B_{t-1}}{P_t}$. More precisely, if we substitute the tax policy $T_t = \tau P_t$ and the market clearing condition $c_t + \tau = y$ into the household budget constraint (2.9), we obtain a law of motion for nominal wealth

$$Q_t(B_t + \hat{M}_t) = B_{t-1} + \hat{M}_{t-1}. \tag{2.17}$$

Using the initial condition $B_{-1} = \hat{M}_{-1} = 0$, (2.17) shows that nominal wealth is zero in all periods. The transversality condition trivially holds.

Although the assumptions in Proposition 1 do not give a unique equilibrium, the above argument suggests an easy way of achieving uniqueness. In any non-stationary equilibrium, between-period money holdings $\hat{M}_{t-1}$ converge to M as nominal taxes tend to zero. As total nominal wealth $B_{t-1} + \hat{M}_{t-1}$ is zero, bond wealth B_{t-1} converges to $-M$. As the price level tends to zero, this means that the real value of bond wealth converges to minus infinity. If the household faces a borrowing constraint $-\frac{B_t}{P_t} < \underline{b}$ for any finite $\underline{b} > 0$, unlimited indebtedness is excluded, and so are all deflationary price paths associated with the non-stationary equilibria.

Proposition 2. *Assume that the government policy is the same as in Proposition 1, and that in addition to constraints (2.1)-(2.5), the household faces a borrowing constraint*

$$-\frac{B_t}{P_t} < \underline{b} \tag{2.18}$$

where $\underline{b} > 0$. Then, there exists a unique equilibrium, which has $P_t = \frac{M}{\tau}$ for all t.

Proof. *See Appendix.*

To summarize, we have established that there exists a unique stationary equilibrium if the government policy is stationary. We have further shown that all non-stationary equilibria can be excluded with an arbitrarily lax borrowing constraint.

It is not surprising that there is a stationary equilibrium $P_t = \frac{M}{\tau}$. As soon as the nominal interest rate is positive, the tax constraint binds, and the price level is proportional to the money supply. The mechanism resembles price level determination in a model where there is a money supply rule together with a cash-in-advance (CIA) constraint, which also posits that some expenditure items must be transacted with money.

It might be more surprising that the equilibrium is not unique. This is in contrast with the unique equilibrium obtained in a perfect foresight model with fixed money supply and a standard CIA constraint. The dif-

ference arises because the money supply rules in the two models have different implications for government income. If you keep money supply constant in a CIA environment, there is no seigniorage income. A fall in the price level then leads to an increase in household wealth, which creates excess demand. In our model, a constant money supply requires that the government prints money in every period, since money is withdrawn from the economy via tax payments. In period t, the government needs to issue τP_{t-1} money to compensate for the money withdrawn from the economy in period $t - 1$. The real value of newly issued money is

$$\frac{\tau P_{t-1}}{P_t},$$

and if the price level is falling, this value increases. Given its higher income, the government satisfies its per-period budget constraint by lending to the household. The increase in household debt perfectly matches the increased value of household money holdings. A deflationary path does not create any excess demand and cannot be excluded.

Although this means that a unique equilibrium is not solely obtained by tax requirements, we have also shown that there exists an easy mechanism to obtain uniqueness. It is sufficient to impose an arbitrarily lax borrowing constraint, which in our view is a mild assumption.

2.3   Relationship to the fiscal theory of the price level

Fiscal factors influence the price level in our model, which raises the question of how our model is related to the fiscal theory of the price level (FTPL) as introduced by Leeper (1991), Sims (1994), and Woodford (1995). As seen in the introductory quote of the paper, Cochrane (2005) actually claims that the FTPL is equivalent to the view that money is valued due to tax requirements in money. In this section, we compare the FTPL to our mechanism of price level determination, and show how the two differ.

In the FTPL, the price level is determined by equating the value of nominal government liabilities to the present value of future primary surpluses. More precisely, the theory assumes that there is an initial stock of nominal liabilities B_{-1}, and that the government commits to a surplus sequence $\tau_t - g_t$. Restricting attention to stationary parameterizations $\tau_t = \tau, g_t = g$, the price level at time 0 is determined by the government intertemporal budget constraint:

$$\frac{B_{-1}}{P_0} = \sum_{t \geq 0} \beta^t (\tau - g).$$

To compare the FTPL with our model, we introduce an initial stock of government debt to the model in Section 2.2. Formally, we change $B_{-1} = 0$ to $B_{-1} = B > 0$ in equation (2.2). We prove the following proposition:

Proposition 3. *Suppose that $B_{-1} = B > 0$, $M_t = M$, $\tau_t = \tau$, and $g_t = g$ in the model in Section 2.2. Suppose further that households face a real borrowing constraint $-B_t/P_t < \underline{b}$. Then, an equilibrium exists if and only if τ and g satisfy*

$$\frac{B}{M/\tau} = \sum_{t \geq 0} \beta^t (\tau - g). \tag{2.19}$$

Under these assumptions, the equilibrium is unique and the price level is given by $P = M/\tau$.

Proof. *See Appendix.*

Proposition 3 shows the key difference between our model and the FTPL. In our model, an equilibrium must have $P_t = \bar{P} = \frac{M}{\tau}$. Moreover, this equilibrium exists if and only if the surplus sequence has a present value $B/\bar{P}$. In the FTPL, any surplus sequence with a positive present value can be selected, and the price level adjusts to equate the government intertemporal budget constraint.

To gain more intuition, it is instructive to consider the two cases

$$\frac{\tau - g}{1 - \beta} \lesseqgtr \frac{B}{M/\tau}. \tag{2.20}$$

The left-hand side of (2.20) is the discounted value of the surplus sequence at time 0. The right-hand side is the real value of government debt under the price level $P_0 = \frac{M}{\tau}$. When $\frac{\tau-g}{1-\beta} < \frac{B}{M/\tau}$ and the surplus is too small according to the proposition, the FTPL posits that the price level adjusts to $P_0 > M/\tau$ to equate the value of surpluses to initial government debt. In our model, the intertemporal budget constraint also requires $P_0 > M/\tau$. However, this price level is inconsistent with the tax requirement, and our model has no equilibrium when $\frac{\tau-g}{1-\beta} < \frac{B}{M/\tau}$.

Conversely, if $\frac{\tau-g}{1-\beta} > \frac{B}{M/\tau}$ and the surplus is too large according to the proposition, the FTPL posits that the price level adjusts to $P_0 < M/\tau$ to equate the value of surpluses to initial government debt. In our model, $P_0 < M/\tau$ implies that the tax constraint is slack and the household has to carry between-period money $\hat{M}_0 > 0$. By an identical reasoning to that in Section 2.2, this means that the nominal interest rate is zero, there is deflation at the rate of the real interest rate, and the household accumulates debt of infinite real value. Since these paths are excluded by the borrowing constraint, our model has no equilibrium when $\frac{\tau-g}{1-\beta} > \frac{B}{M/\tau}$.

The mechanism of price level determination in our framework is thus different from the FTPL. Our model also provides novel fiscal-monetary interactions as the tax policy affects money demand. It is an interesting avenue of future research to fully map out the fiscal-monetary interactions in our framework.

2.4   Excluding hyperinflations

Up to this point, we have assumed that the household only holds money to pay taxes. In this section, we extend our analysis to show the effects of tax requirements when the household also holds money for liquidity

reasons.

The purpose of this exercise is to address the issue of speculative hyperinflations. Speculative hyperinflations are a generic feature in models that are designed to capture the liquidity value of money; they arise when money is included in the utility function (Obstfeld and Rogoff, 1983), in OLG models of money (Woodford, 1984), and in search-and-matching models of money (Williamson and Wright, 2010). Under stationary parameterizations, all these models feature one stationary equilibrium where money is not valued, one stationary equilibrium where the value of money is positive, and an infinite number of equilibria where the value of money tends to zero.

We extend the model in Section 2.2 by adding money holdings as an argument in the utility function. The model then becomes similar to that used in the analysis of speculative hyperinflations in Obstfeld and Rogoff (1983). The difference between the models is that our model has tax requirements, and that our model features government bonds but no capital. We first show how speculative hyperinflations arise without tax requirements, and then we show how tax requirements exclude these speculative hyperinflations.

2.4.1 A model with money in the utility function

We start with the baseline model in Section 2.2 and add money as an argument in the utility function. The utility gain from money is separable from consumption and given by the function $v(\cdot)$. v is non-decreasing, concave and satisfies the Inada conditions. Money holdings have a saturation point, so $v(m)$ is constant for large m. Furthermore, we assume that $\lim_{m \to 0} v(m) > -\infty$ and $v''' > 0$.[1] We follow the timing assumption of Obstfeld and Rogoff (1983) and assume that M_t gives a utility flow

---

[1]The assumption $\lim_{m \to 0} v(m) > -\infty$ is introduced in Obstfeld and Rogoff (1983). This assumption is important for the existence of hyperinflationary equilibria, but also plausible. Indeed, there are substitutes for money in trade which means that an infinitely negative utility of zero money holdings seems unreasonable. The assumption of $v''' > 0$ is not made by Obstfeld and Rogoff (1983) and they incorrectly claim that the concavity of v implies convexity of the function $A(\cdot)$ in (2.30). The assumption

in period t. An alternative and arguably more appropriate timing assumption is that households receive utility from ingoing money $\hat{M}_{t-1}$. This alternative timing assumption, however, is mathematically much less tractable. We also include an arbitrarily lax borrowing constraint. Proposition 2 then guarantees that there is a unique equilibrium without money-in-the-utility. We include a borrowing constraint as we are interested in how adding liquidity value to money might give rise to other non-stationary equilibria than those already discussed in Section 2.2. The household solves

$$\max_{\{c_t, B_t, M_t, \hat{M}_t\}} \sum_{t=0}^{\infty} \beta^t \left(u(c_t) + v\left(\frac{M_t}{P_t}\right) \right)$$

$$\text{s.t.} \quad c_t + \frac{M_t}{P_t} + \frac{Q_t B_t}{P_t} \leq y_t + \frac{B_{t-1}}{P_t} + \frac{\hat{M}_{t-1}}{P_t} \quad (2.21)$$

$$\hat{M}_t + T_t \leq M_t \quad (2.22)$$

$$0 \leq \hat{M}_t \quad (2.23)$$

$$-\frac{B_t}{P_t} < \underline{b} \quad (2.24)$$

$$\hat{M}_{-1} = 0 \quad (2.25)$$

$$B_{-1} = B > 0. \quad (2.26)$$

The government per-period budget constraint is the same as in Section 2.2 and is given by (2.6). One difference from Section 2.2 is that we have an initial stock of government debt B. We add this to compensate for seigniorage profits. Indeed, if government consumption equals the real value of taxes in each period, the government takes in too much taxes when the household holds between-period money. To compensate for this effect, we let the household start with a positive holding of government bonds. We say that an allocation and a price vector together form an equilibrium if there exists some initial government debt level B such that the allocation and price vector together with B form an equilibrium according to Definition 1. For the purpose of analyzing speculative hy-

---

$v''' > 0$ ensures the convexity of $A(\cdot)$.

perinflations, we consider a stationary parameterization, and set $y_t = y$, $T_t = \tau P_t$, $g_t = \tau$, and $M_t = M$.

The equilibrium conditions are identical to those in Section 2.2 except that we must replace the optimality condition for money holdings (2.12) with

$$\frac{u'(c_t)}{P_t} = \left[\frac{v'(M_t/P_t)}{P_t}\right] + \beta \frac{u'(c_{t+1})}{P_{t+1}} + \mu_t \qquad (2.27)$$

where μ_t is the Lagrange multiplier on the non-negativity constraint on money holdings (2.23). With money in the utility function, the intertemporal tradeoff in the household choice of money holdings does not only capture forgone consumption and the shadow value of relaxing the non-negativity constraint, but also the flow utility of holding money. To see the equilibrium implications of this condition, we substitute in the market clearing condition for goods (2.7) and money (2.8):

$$\frac{u'(y - \tau)}{P_t} = \left[\frac{v'(M/P_t)}{P_t}\right] + \beta \frac{u'(y - \tau)}{P_{t+1}} + \mu_t. \qquad (2.28)$$

For reasons that will be clear shortly, a useful manipulation of (2.28) is to multiply both sides by the money supply M and write the equation in terms of real money balances $m_t = \frac{M}{P_t}$:

$$m_{t+1} = \frac{1}{\beta} m_t \left(1 - \frac{v'(m_t)}{u'(y - \tau)}\right) - \frac{\mu_t M}{\beta}. \qquad (2.29)$$

Now, define $A(m) = \frac{1}{\beta} m \left(1 - \frac{v'(m)}{u'(y-\tau)}\right)$. Since the tax constraint binds if and only if $m_t = \tau$, an equilibrium to this model is characterized by (2.2), (2.4), (2.7)-(2.11), (2.13) and

$$m_{t+1} \in \Psi(m_t) \qquad (2.30)$$

where $\Psi(\cdot)$ is the correspondence

$$\Psi(m) = \begin{cases} A(m) & \text{if } m > \tau \\ (-\infty, A(m)] & \text{if } m = \tau. \end{cases}$$

We are now ready to analyze the equilibrium outcome. We start by removing tax requirements and show that there are equilibria with speculative hyperinflations. We then show that these equilibria are excluded when we reintroduce the tax requirements.

2.4.2   Speculative hyperinflations

Suppose that the tax requirements are removed by setting $T_t = 0$ in the tax constraint (2.22). Instead, we add a tax $\tau^r = \tau$ to be paid in real goods in the household budget constraint (2.21). This means that $\mu_t = 0$ for all positive money holdings M_t. In this case, the equilibrium price sequences will be identical to those in Obstfeld and Rogoff (1983), and we follow their analysis closely.

Given $\mu_t = 0$, the equilibrium is characterized by the difference equation

$$m_{t+1} = \Psi(m_t) = A(m_t).$$

We analyze the equilibrium properties by plotting $m_{t+1} = \Psi(m_t)$ against the 45-degree line $m_{t+1} = m_t$, as depicted in Figure 2.1.

The shape of the function Ψ is derived from our assumptions on v. First, we can use the assumption that $\lim_{m_t \to 0} v(m_t) > -\infty$ to show that $\lim_{m_t \to 0} m_t v'(m_t) = 0$.[2] This means that $\lim_{m_t \to 0} \Psi(m_t) = 0$. By the Inada conditions, $\Psi(m_t)$ starts below the m_t-axis for small m_t and its slope is $1/\beta$ as $m_t \to \infty$. Thus, $\Psi(m_t)$ intersects the m_t-axis at least once for some $\bar{m}$. As $v'(\cdot)$ is strictly concave, $\Psi(\cdot)$ is strictly convex, and $\Psi(m_t)$ has a unique positive fixed point m^*. As shown in Figure 2.1, this means that there is one stable equilibrium with $m_t = m^*$ for all t, one stable equilibrium where $m_t = 0$ for all t, and a countably infinite number of equilibria where $m_t \to 0$.

The speculative hyperinflations constitute equilibria even though the household foresees that its money holdings will be worthless in finite time. The reason is that in the last period $\bar{t}$ when real money balances

---

[2]See Obstfeld and Rogoff (1983).
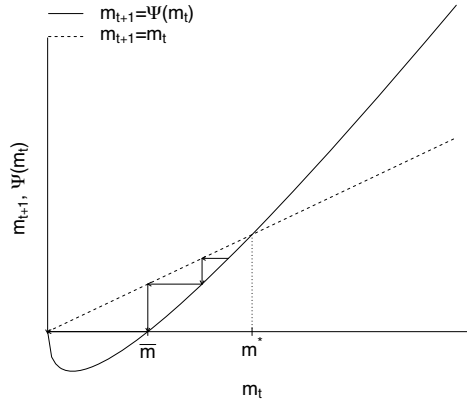
Figure 2.1: Correspondence $\Psi(m)$ without tax requirements
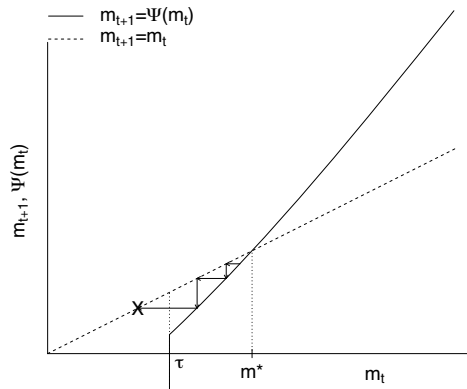
Figure 2.2: Correspondence $\Psi(m)$ with tax requirements, when $0 < \tau < m^*$

are positive, and $m_{\bar{t}} = \bar{m}$, the cost of withholding one unit of extra consumption exactly equals the marginal utility flow from having this real money balance. The household is exactly compensated for the fact that its money holdings will be worthless in the next period. Given that the household foresees this period, it is likewise consistent with optimizing behavior that it holds money willingly in an inflationary path leading to this point.
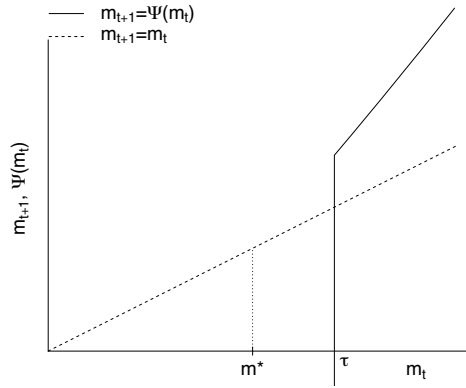
Figure 2.3: Correspondence $\Psi(m)$ with tax requirements, when $\tau \geq m^*$

2.4.3   Reintroducing the tax requirements

Now consider the case where we reintroduce the tax requirements. Under this assumption, there is a lower bound τ on real money balances due to the tax requirements. We split the analysis into two cases depending on whether τ is larger or smaller than m^*.

First, let $0 < \tau \leq m^*$, where m^* is the stationary equilibrium with a positive price level in the model without tax requirements. This case is depicted in Figure 2.2. The unique equilibrium is $m = m^*$. Indeed, any initial value of real money balances $m_0 < m^*$ must lead to inflation and $m_t \to 0$. But then $m_t < \tau$ for some future t. This violates the tax requirements. Equilibrium candidates with $m_0 > m^*$ are deflationary with $m_t \to \infty$. Any such path violates the household's borrowing constraint.

Second, let $\tau > m^*$, which is depicted in Figure 2.3. In this case, there is a unique equilibrium $m = \tau$. Any initial value m_0 starting below this value immediately violates the tax requirements, and any value of m_0 above τ leads to deflation $m_t \to \infty$ which violates the household's borrowing constraint.

We summarize these cases in the following proposition:

Proposition 4. *For any $\tau > 0$, there is a unique equilibrium. If $0 < \tau \leq m^*$, the equilibrium real value of money holdings is $m_t = m^*$, where m^* is given by*

$$A(m^*) = m^*, \quad m^* > 0$$

and if $\tau > m^*$, the equilibrium real value of money holdings is $m_t = \tau$.

Proof. *See Appendix.*

The proposition shows that tax requirements can exclude hyperinflationary equilibria. Tax requirements can therefore provide a potential explanation for the absence of observed speculative hyperinflations. Indeed, even though hyperinflations feature in many models of money, it is questionable whether purely speculative hyperinflations have ever occurred in government backed currencies.[3] If tax requirements are the key mechanism that excludes hyperinflations, it also suggests that we do not need other, more complicated, selection mechanisms – for instance the state-dependent regime switch to a commodity standard as proposed by Obstfeld and Rogoff (1983).

Moreover, if tax requirements are important to prevent speculative hyperinflations, non-government backed fiat money should run the risk of experiencing hyperinflations. Historically, such currencies have been rare, but the development of cryptography based currencies, e.g. bitcoin, have made the case of non-government backed fiat money empirically relevant. The future development of these currencies is an empirical test of the hypothesis that tax requirements are key to avoid speculative hyperinflations. If they keep a stable value over a long time, it will be evidence against tax requirements being central to select a stable equilibrium.

---

[3]All hyperinflations in government issued currencies of which we are aware have been associated with large increases in money supply – such as in Hungary, Germany, and Zimbabwe – or the collapse of government power – like for example the collapse of the Confederate currency at the end of the American civil war. They have not been purely speculative.

2.5   Price level determination under a balanced budget rule

In Section 2.2, we assumed that government consumption g, money supply M, and the real value of taxes τ were constant and that bond issuance B_t varied residually to satisfy the government per-period budget constraint (2.6). Even though $B_t = 0$ in the stationary equilibrium, we did not constrain the government to run a balanced budget out of equilibrium.

This observation raises the question if we can obtain a unique price level under a balanced budget rule. The potential implications of balanced budget rules on price level determinacy has previously been highlighted by Schmitt-Grohé and Uribe (2000). The question is relevant because of the importance of balanced budget rules in practical policy making.[4] This is especially true for countries that might be forced into monetary reform while simultaneously losing access to international financial markets, as exemplified by the possibility of Greece deciding to leave the euro in the summer of 2015.

In this section, we show that when the bond supply is constrained to be zero, the tax policy from Section 2.2 fails to ensure a unique equilibrium price level. This points to a difficulty of price level control under a balanced budget rule. However, there exists a modified tax policy which restores uniqueness. More specifically, for each $\bar{P}$, there exists a tax function $T_t(P_t)$, which respects the constraint $B_t = 0$ and which produces a unique equilibrium $P_t = \bar{P}$.

To construct such a tax policy, we first note that when the bond supply is zero, the money supply rule has to keep real proceeds from

---

[4]A second reason to study balanced budget rules is obtained from the analysis of Bassetto (2002). In a game-theoretic analysis of the fiscal theory of the price level, he shows that equilibrium indeterminacy can arise if the government tries but fails to issue government bonds. This problem is avoided if the government runs a balanced budget.

money sales $\frac{M_t - \hat{M}_{t-1}}{P_t}$ constant and equal to expenditure $g_t = \tau$

$$M_t = \tau P_t + \hat{M}_{t-1}. \tag{2.31}$$

Equation (2.31) differs from the constant money supply function $M_t = M$ from Section 2.2. Importantly, with the money supply function (2.31), the proportional tax function $T_t = \tau P_t$ from Section 2.2 does not give us a unique equilibrium. In fact under $T_t = \tau P_t$, the constant price sequence $P_t = P$ is an equilibrium for any $P > 0$. To see this, note that a constant price level implies a positive nominal interest rate $Q_t < 1$, and zero between-period money holdings $\hat{M}_t = 0$. Then $M_t = \tau P = T_t$ and the tax requirement is satisfied. This is true for all P. Hence, there is a continuum of stationary equilibria when we have a balanced budget rule and the same tax function as in Section 2.2.

Uniqueness fails because price level deviations do not generate excess demand/supply for money. This differs from the equilibrating mechanism in Section 2.2. In Section 2.2, a price level above the stationary equilibrium price level $\bar{P} = M/\tau$ implied that the money supply did not suffice to pay taxes, and a price level below $\bar{P}$ implied a slack tax constraint and that households accumulated money holdings and took on debt, which eventually violated their borrowing constraint.

However, we can modify the tax function to obtain a unique equilibrium in this setting. The tax function needs to generate excess demand for money when the price level is too high, and excess supply of money when the price level is too low. The key observation is that with our new money supply rule (2.31), money supply increases proportionally with the price level. To ensure a unique equilibrium, we need that taxes increase *more than proportionally* with the price level. Therefore, we propose a tax function

$$T_t(P_t) = \tau P_t + \Pi(P_t) \tag{2.32}$$

with $\Pi(\bar{P}) = 0$, $\Pi' > 0$ and where Π' is uniformly bounded away from

zero. Under this policy, we can prove the following proposition.

Proposition 5. *Suppose that the government policy is $B_t = 0, g_t = \tau$, $M_t = \tau P_t + \hat{M}_{t-1}$ and $T_t(P_t) = \tau P_t + \Pi(P_t)$ with $\Pi(\bar{P}) = 0$ for some $\bar{P} > 0$ and $\frac{\partial \Pi}{\partial P_t} > \epsilon$ for some $\epsilon > 0$. Then, there exists an equilibrium with $P_t = \bar{P}$, and this equilibrium is unique.*

Proof. See Appendix.

We sketch the intuition for the proof here. There are three cases:

1. $P_t = \bar{P}$

2. $P_t > \bar{P}$

3. $P_t < \bar{P}$

We can see that Case 1 is an equilibrium. Indeed, if $P_t = \bar{P}$, the tax function gives us nominal taxes $T_t = \tau \bar{P}$ and the money supply is $M_t = M = \tau \bar{P}$. This is the same parameters as in Section 2.2 and by Proposition 1, this is an equilibrium. Case 2 is not an equilibrium since $P_t > \bar{P}$ implies that the level of taxes exceeds the level of money issuance τP_t, and the household will violate the tax requirements. Case 3 is not an equilibrium. $P_t < \bar{P}$ implies that money issuance τP_t exceeds the tax requirements, which means that the price level path decreases with the real interest rate and the value of household money holdings increases faster than the real interest rate. Since $B_t = 0$ for all t, this violates the household's transversality condition.

The tax function (2.32) is not as simple as the tax system in Section 2.2. However, such a tax policy could be constructed in practice by combining a tax requirement $T_t = \tau P_t$, which is fixed in real terms, with a nominal lump-sum transfer/tax. These results suggest that it is possible, but more challenging, to use tax requirements to achieve a determinate price level under a balanced budget rule.

2.6   Discussion

In this paper, we have constructed a sequence of models to analyze how tax requirements affect price level determination. In this section, we discuss potential implications of our theory for the conduct of monetary policy.

We conjecture that tax requirements are unlikely to affect monetary policy when monetary authorities use interest rate rules. We can see this by considering how an interest rate rule would operate in our environment. Under an interest rate rule, given a path of real consumption, the expected inflation rate is determined by the Euler equation. Restricting ourselves to the case where nominal interest rates are positive, the tax constraint binds. The path of money supply is determined by the equation $M_t = \tau P_t$, and money supply responds passively to any changes in the price level. This outcome is essentially identical to that with an interest rate rule in a one-good economy with a cash-in-advance constraint. Apart from sunspot equilibria (Sargent and Wallace, 1975), the equilibrium is unique up to a scaling factor in the initial money supply and price level.

However, the analysis suggests that tax requirements might make money supply rules a more attractive policy instrument. The reason is that tax requirements address a key dilemma with money supply rules: The difficulty of selecting a monetary aggregate for which there is a stable demand, and which is simultaneously possible to control. If the monetary authorities select a narrow monetary aggregate, demand might be unstable as financial innovation can generate close substitutes. The aggregate for which we can expect a reasonably stable demand would be the total amount of all liquid assets. However, the supply of this aggregate is difficult to control for the monetary authorities.

With tax requirements, government tax policy controls the demand for the monetary aggregate required for tax payments. This monetary aggregate can be chosen to be M0, which also gives the government control over its supply. In this situation, the government controls both

the demand and the supply curve of money, and can use this to determine the price level. In light of recent events, where central banks have failed to control the inflation rate using interest rates alone, it is interesting to explore money supply based alternatives. However, more research is needed to exactly clarify how a policy based on tax requirements could be expected to work, and how it could be implemented.

References

Bassetto, M. (2002). A Game-Theoretic View of the Fiscal Theory of the Price Level. *Econometrica*, 70(6):2167–2195.

Cochrane, J. H. (2005). Money as stock. *Journal of Monetary Economics*, 52(3):501–528.

Goldberg, D. (2012). The tax-foundation theory of fiat money. *Economic Theory*, 50(2):489–497.

Krugman, P. (2013). The Antisocial Network. *The New York Times*, available from 'http://www.nytimes.com/', published April 14 [August 6 2015].

Leeper, E. M. (1991). Equilibria under 'active' and 'passive' monetary and fiscal policies. *Journal of Monetary Economics*, 27(1):129–147.

Lerner, A. P. (1947). Money as a Creature of the State. *The American Economic Review*, 37:312–317.

Obstfeld, M. and Rogoff, K. (1983). Speculative Hyperinflations in Maximizing Models: Can We Rule Them Out? *Journal of Political Economy*, 91(4):675.

Sargent, T. J. and Wallace, N. (1975). "Rational" Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule. *Journal of Political Economy*, 83(2):241.

Schmitt-Grohé, S. and Uribe, M. (2000). Price level determinacy and monetary policy under a balanced-budget requirement. *Journal of Monetary Economics*, 45(1):211–246.

Sims, C. A. (1994). A simple model for study of the determination of the price level and the interaction of monetary and fiscal policy. *Economic Theory*, 4(3):381–399.

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations.* George Routledge and Sons, London, Untited Kingdom.

Starr, R. M. (1974). The Price of Money in a Pure Exchange Monetary Economy with Taxation. *Econometrica*, 42(1):45–54.

Varoufakis, Y. (2015). Something is rotten in the eurozone kingdom. *Financial Times*, available from 'http://www.ft.com/', published July 28 [August 10 2015].

Williamson, S. and Wright, R. (2010). New monetarist economics: Models. In *Handbook of Monetary Economics 3*, pages 25–96.

Woodford, M. (1984). Indeterminacy of equilibrium in the overlapping generations model: A survey. *mimeo.*

Woodford, M. (1995). Price-level determinacy without control of a monetary aggregate. *Carnegie-Rochester Conference Series on Public Policy*, 43:1–46.

Appendix

Proofs to Section 2.2

Proof of Proposition 1. To prove the proposition, we proceed in three steps. First, we establish that the proposed stationary equilibrium exists. Second, we show that there are no other stationary equilibria. Finally, we show that there is an infinite number of non-stationary equilibria.

We guess that an equilibrium is formed by

$$c_t \equiv y - \tau \quad \hat{M}_t = 0$$
$$P_t \equiv \frac{M}{\tau} \quad Q_t = \frac{1}{\beta}$$

Recall from the proposition statement that $M_t \equiv M$ and $T_t = \tau P_t$, and that given the values of other variables, we have $B_t = 0$.

The guess is an equilibrium if it satisfies the market clearing conditions (2.7)-(2.8), the consumer optimality conditions (2.2), (2.4), (2.9)-(2.13) and the zero lower bound on interest rates (2.10). We check these conditions one by one.

The non-negativity constraint on money (2.4), the initial conditions on money and bonds (2.2), and the non-negativity constraint on nominal interest rates (2.10) are directly satisfied by assumption. The goods market clears since $c_t = y - \tau$. The money market clearing condition (2.8) holds as

$$T_t + \hat{M}_t = P_t \tau = M = M_t$$

The household per-period budget constraint (2.9) holds since

$$
\begin{aligned}
c_t &+ \frac{T_t}{P_t} + \frac{(1 - Q_t)\hat{M}_t}{P_t} + \frac{Q_t}{P_t}\left(\hat{M}_t + B_t\right) \\
&= (y - \tau) + \tau + (1 - Q_t) \times 0 + Q_t(0 + 0) \\
&= y \\
&= y + \frac{1}{P_t}\left(\hat{M}_{t-1} + B_{t-1}\right).
\end{aligned}
$$

The Euler equation (2.11) holds as the consumption and price level are constant and $Q_t = \frac{1}{\beta}$. The constraint that between-period money holdings are zero if $Q_t < 1$ (2.12) holds as $\hat{M}_t \equiv 0$. The transversality/no-Ponzi condition (2.13) trivially holds since $B_t = \hat{M}_t = 0$. The guess is therefore an equilibrium.

To show that the guess is the unique stationary equilibrium, we do a proof by contradiction and show that there is no equilibrium such that $P_t = \tilde{P}$ for all t and $\tilde{P} \neq \frac{M}{\tau}$.

Assume that there is a stationary equilibrium with $P_t > \frac{M}{\tau}$ for some t. Then $\hat{M}_t < 0$ by the money market clearing condition (2.8). This contradicts the non-negativity constraint on between-period money holdings (2.4).

Assume that there is a stationary equilibrium with $P_t < \frac{M}{\tau}$ for some t. Then $\hat{M}_t > 0$ by the money market clearing condition (2.8). The optimality condition for money holdings (2.12) then requires that $Q_t = 1$. In that case, the Euler equation (2.11) yields $P_{t+1} = \beta P_t < P_t$, contradicting that the price level is constant. Therefore, the proposed stationary equilibrium is the unique stationary equilibrium.

We now want to show that there is a continuum of non-stationary equilibria. This is done by constructing an equilibrium for an arbitrary $P_0 < M/\tau$ (there are no equilibria with $P_0 > M/\tau$ as this would mean that the tax constraint was violated). If $P_0 < M/\tau$, we know that $\hat{M}_0 > 0$ by money market clearing, and optimal choice of money holdings implies

that $Q_t = 1$. The Euler equation, in turn, implies that $P_1 = \beta P_0$. This means that the tax constraint is lax for $t = 1$ as well, and by iterating forward on t, we obtain

$$P_t = \beta^t P_0.$$

We now propose an equilibrium with $B_{-1} = \hat{M}_{-1} = 0$ and for $t \geq 0$

$$c_t \equiv y - \tau \quad \hat{M}_t = M - \tau P_t$$
$$P_t \equiv P_0 \beta^t \quad Q_t = 1$$

We can confirm that under this equilibrium, $B_t = -\hat{M}_t$.

It is clear that the initial conditions (2.2) and the non-negativity constraint on money holdings (2.4) hold. Market clearing for goods and money (2.7)-(2.8) both hold. The household per-period budget constraint holds as

$$c_t + \frac{T_t}{P_t} + \frac{(1 - Q_t)\hat{M}_t}{P_t} + \frac{Q_t}{P_t}\left(\hat{M}_t + B_t\right)$$
$$= (y - \tau) + \tau + 0 \times \hat{M}_t + \frac{1}{P_t} \times (\hat{M}_t - \hat{M}_t)$$
$$= y$$
$$= y + \frac{1}{P_t}\left(\hat{M}_{t-1} + B_{t-1}\right)$$

The Euler equation (2.11) holds as we derived the price level path under the assumption that it did, and the money optimality equation (2.12) holds as $Q_t \equiv 1$. The transversality/no-Ponzi condition (2.13) holds as total nominal wealth is zero in all periods, and the zero lower bound on interest rates (2.10) holds by assumption. $\square$

Proof of Proposition 2. We know that the stationary equilibrium proposed in the proof of Proposition 1 is also an equilibrium under the conditions in the current proposition, since $B_t = 0$ clearly does not violate the new borrowing constraint. To show that the stationary equilibrium

from Proposition 1 is the unique equilibrium, we do a proof by contra-diction and show that there is no equilibrium featuring $P_t \neq \frac{M}{\tau}$.

Assume that there is an equilibrium with $P_t > \frac{M}{\tau}$ for some t. Then $\hat{M}_t < 0$ by the money market clearing condition (2.8). This is inconsistent with the non-negativity constraint on between-period money (2.4).

Assume that there is an equilibrium with $P_t < \frac{M}{\tau}$ for some t. Then, $\hat{M}_t > 0$ by the money market clearing condition (2.8). This means that $Q_t = 1$ from the money optimality condition (2.12). Then the Euler equation (2.11) with constant consumption $y - \tau$ gives us $P_{t+1} = \beta P_t$. Hence, $P_{t+1} < \frac{M}{\tau}$ and the tax constraint is slack and thus $\hat{M}_{t+1} > 0$. By successive iteration, we have that $P_{t+s} = \beta^s P_t$. Thus, $P_{t+s} \to 0$. Money market clearing then implies that

$$\hat{M}_{t+s} = M - P_{t+s}\tau \to M.$$

Furthermore, if we substitute $T_t = \tau P_t$, $c = y - \tau$ and $(1 - Q_t)\hat{M}_t = 0$ (the last from optimality condition (2.12)) into the consumer per-period budget constraint (2.9), we obtain a law of motion for nominal wealth

$$Q_t(B_t + \hat{M}_t) = B_{t-1} + M_{t-1}.$$

Using the initial condition $\hat{M}_{-1} = B_{-1} = 0$, the law of motion implies that $B_t + \hat{M}_t = 0$ for all t. Therefore, if $\hat{M}_t \to M$, this means that $B_t \to -M < 0$, and together with $P_t \to 0$ we get $\frac{B}{P_t} \to -\infty$. This violates the borrowing constraint (2.18). $\square$

Proofs in Section 2.3

Proof of Proposition 3. We first derive the equilibrium relationship be-tween the price level and the government's surplus sequence. Suppose that we have an equilibrium price sequence P_t. In an equilibrium, house-hold consumption is $c = y - \tau$ by goods market clearing. We can iterate

on the consumer's budget constraint to obtain

$$
\begin{aligned}
B_{t-1} + \hat{M}_{t-1} \;=\; & \sum_{s \geq 0} Q_{t,t+s} \left(P_{t+s}(\tau - g) + (1 - Q_{t+s})\hat{M}_{t+s} \right) \\
& + \lim_{T \to \infty} Q_{t,t+T}(\hat{M}_{t+T} + B_{t+T}) \quad\quad (2.33)
\end{aligned}
$$

where $Q_{t,t+s} = \prod_{j=0}^{s-1} Q_{t+j}$ gives the nominal interest rate between time t and $t+s$. Consumer optimality requires that $\lim_{T \to \infty} Q_{t,t+T}(\hat{M}_{t+T} + B_{t+T}) = 0$ by the limit condition (2.13), and $(1 - Q_{t+s})\hat{M}_{t+s} = 0$ as $\hat{M}_{t+s} = 0$ whenever $Q_{t+s} < 1$. Furthermore, $Q_{t,t+s} = \frac{\beta^s P_t}{P_{t+s}}$. (2.33) equation simplifies to

$$
\frac{B_{t-1} + \hat{M}_{t-1}}{P_t} = \frac{\tau - g}{1 - \beta}. \quad\quad (2.34)
$$

Now we can prove our theorem. We consider the three cases

$$
\sum_{t \geq 0} \beta^t (\tau - g) \quad
\begin{matrix}
< \frac{B}{M/\tau} \\
> \frac{B}{M/\tau} \\
= \frac{B}{M/\tau}
\end{matrix}
$$

Case 1. When $\sum_{t \geq 0} \beta^t(\tau - g) < \frac{B}{M/\tau}$, the intertemporal budget constraint (2.34) implies $P_0 > M/\tau$. Money market clearing (2.8) means that $\hat{M}_0 < 0$ which violates the non-negativity constraint on between-period money holdings (2.4).

Case 2. When $\sum_{t \geq 0} \beta^t(\tau - g) > \frac{B}{M/\tau}$, the intertemporal budget constraint (2.34) implies $P_0 < M/\tau$, and together with the money market clearing (2.8) we have that $\hat{M}_0 = M - \tau P_0 > 0$. $\hat{M}_0 > 0$ implies $Q_1 = 1$ by the equation for optimal money holdings (2.12) and thus, $P_1 = \beta P_0$ by the Euler equation (2.11). By induction, $P_t = \beta^t P_0 \to 0$. Thus, $\hat{M}_t \to M$. As for each t, we have

$$
\frac{B_{t-1} + \hat{M}_{t-1}}{P_t} = \frac{\tau - g}{1 - \beta}
$$

we must have $B_{t-1} \to -M$ and $B_{t-1}/P_{t-1} \to -\infty$ which violates the borrowing constraint.

Case 3. When $\sum_{t \geq 0} \beta^t(\tau - g) = \frac{B}{M/\tau}$, there exists a stationary equilibrium $P \equiv M/\tau$, $c_t \equiv y - g$ $\hat{M}_t \equiv 0$, and $Q_t \equiv \beta$. It can be checked that $B_t \equiv B$ in this equilibrium. It is clear that the initial condition for money and bond holdings (2.2) and the non-negativity constraint on money holdings (2.4) are satisfied. The money market and the bond market clear as $c + g = y$ and

$$M_t = M = 0 + \frac{M}{\tau} \times \tau = \hat{M}_t + T_t$$

for all t. The allocation solves the consumer Euler equation (2.11), and as $Q_t < 1$, $\hat{M}_t \equiv 0$ is consistent with optimization of money holdings (2.12). Since nominal wealth and the price level are constant over time, and the nominal interest rate is positive, the allocation is consistent with the limit condition (2.13).

We obtain a unique price level as any other price level would violate the government's intertemporal budget constraint (2.34). The goods market clearing condition and the Euler equation imply $c_t \equiv y - g$ and $Q_t = \beta$. A positive real interest rate implies that $\hat{M}_t$. Hence, the equilibrium is unique. □

Proofs in Section 2.4

Proof of Proposition 4. The proposition has two cases: $0 < \tau \leq m^*$ and $\tau > m^*$. We treat them separately. Recall that m^* is the unique strictly positive solution to the equation $m = A(m)$.

Case 1: $0 < \tau \leq m^*$. We prove that $m_t = m^*$ is the unique equilibrium by first showing that $m_t = m^*$ is an equilibrium, and then do a proof by contradiction to exclude the other candidate equilibria.

To show that $m_t = m^*$, with $P^* = \frac{M}{m^*}$, is an equilibrium, we need to check that there exists an allocation, a price system, and an initial level

of government debt consistent with (2.2),(2.4), (2.7)-(2.11), (2.13) and (2.30). We propose the allocation $c_t = y - \tau$, $\hat{M}_t = M - \tau P^*$, $B_t = 0$ with the initial condition $\hat{M}_{-1} = 0$, $B_{-1} = \hat{M}_t$ and the bond price $Q_t = \beta$.

The initial condition (2.2) is satisfied by assumption. The non-negativity constraint on money holdings (2.4) is satisfied since $\hat{M}_t = M - \tau P^* = M - \tau \frac{M}{m^*} = M(1 - \frac{\tau}{m^*}) \geq 0$. The goods market clearing condition (2.7) and the money market clearing condition (2.8) are satisfied by assumption. The zero lower bound on interest rates (2.10) is satisfied since $Q_t = \beta < 1$. The per-period household budget constraint (2.9) is satisfied in period 0 since

$$
\begin{aligned}
c_t + \frac{M_0}{P_0} + \frac{Q_0 B_0}{P_0} &= (y - \tau) + \tau + \frac{\hat{M}_0}{P_0} + 0 \\
&= y + \frac{\hat{M}_0}{P_0} \\
&= y + \frac{B_{-1} + \hat{M}_{-1}}{P_0}.
\end{aligned}
$$

Note that we use $B_{-1} = \hat{M}_0$. The household budget constraint is satisfied in period $t \geq 1$ since

$$
\begin{aligned}
c_t + \frac{M_t}{P_t} + \frac{Q_t B_t}{P_t} &= (y - \tau) + \tau + \frac{\hat{M}_t}{P_t} + 0 \\
&= y + \frac{\hat{M}_t}{P_t} \\
&= y + \frac{B_{t-1} + \hat{M}_{t-1}}{P_0}.
\end{aligned}
$$

Here we use $\hat{M}_t = \hat{M}_{t-1}$ for $t \geq 1$. The Euler equation (2.11) holds since $Q_t = \beta$ and $P_t = P^*$ for all t. The transversality/no-Ponzi condition (2.13) holds as $B_t = 0$ and $\hat{M}_t$ is constant for all t. Equation (2.30) governing the evolution of m_t holds by assumption as $m^* = M/P^*$ is the fixed point of A.

This shows that there exists an equilibrium with $m_t = m^*$. To show that it is the unique equilibrium, we do a proof by contradiction.

Suppose that $m_t < m^*$ for some t. Either $m_t < \tau$ or $\tau \leq m_t < m^*$. The first case violates the tax constraint (2.23) and cannot be an equilibrium. In the second case, the tax constraint is not binding in period t. From condition (2.30), the evolution of the price level is described by

$$m_{t+1} = A(m_t) \tag{2.35}$$

Since m^* is the unique positive solution to $A(m) = m$, and $A'(m^*) > 1$, we have $m_{t+1} < m_t$. By induction, it follows that $\{m_{t+s}\}_{s=0}^{\infty}$ is a strictly decreasing sequence and therefore it has a limit. Moreover, this limit is a fixed point of A and since m^* is the unique strictly positive fixed point, we know that there exists a limit $\underline{m} \leq 0$ such that $m_t \to \underline{m} \leq 0$. This violates the joint constraints of taxes and non-negative money holdings (2.22) and (2.23). Hence, there is no equilibrium with $m_t < m^*$.

Suppose instead that $m_t > m^*$ for some t. Without loss of generality we can assume that t is the first period in which $m_t > m^*$. As before, the tax constraint is not binding and the price level is again described by (2.35). By successive iteration, we find that

$$m_{t+T} = \frac{1}{\beta^T} m_t \left(\prod_{s=0}^{T} \left(1 - \frac{v'(m_{t+s})}{u'(y-\tau)} \right) \right) \tag{2.36}$$

Since $\left(1 - \frac{v'(m^*)}{u'(y-\tau)} \right) = \beta$ and $v'(\cdot)$ is a decreasing function, we know that there exists an $\epsilon > 0$ such that

$$\left(1 - \frac{v'(m_{t+s})}{u'(y-\tau)} \right) > \beta + \epsilon \tag{2.37}$$

Hence we can deduce:

$$
\begin{aligned}
m_{t+T} &= \frac{1}{\beta^T} m_t \left(\prod_{s=0}^{T} \left(1 - \frac{v'(m_{t+s})}{u'(y-\tau)} \right) \right) \\
&> \frac{1}{\beta^T} m_t (\beta + \epsilon)^T \\
&= \left(1 + \frac{\epsilon}{\beta} \right)^T m_t
\end{aligned}
$$

and hence

$$
\lim_{T \to \infty} m_{t+T} = \infty \tag{2.38}
$$

Since money supply is constant, we thus have that

$$
\lim_{T \to \infty} P_{t+T} = 0. \tag{2.39}
$$

Furthermore, by the saturation assumption v eventually becomes constant. Thus, there exists $\bar{t}$ such that

$$
m_{\bar{t}+T} = \beta^{-T} m_{\bar{t}}
$$

Using this assumption, we can use the Euler equation and substitute $P_t = \frac{\beta^t}{\prod_{s=0}^{t-1} Q_s} P_0$ into the transversality condition (2.13) to obtain

$$
\begin{aligned}
0 &= \lim_{T \to \infty} \frac{\beta^{\bar{t}+T} \hat{M}_{\bar{t}+T} + \beta^{T+\bar{t}} B_{T+\bar{t}}}{P_{T+\bar{t}}} \\
&= \lim_{T \to \infty} \beta^{\bar{t}+T} m_{T+\bar{t}} + \frac{\beta^{T+\bar{t}} B_{T+\bar{t}}}{P_{T+\bar{t}}} \\
&= \beta^{\bar{t}} m_{\bar{t}} + \lim_{T \to \infty} \frac{\beta^{T+\bar{t}} B_{T+\bar{t}}}{P_{T+\bar{t}}}
\end{aligned}
$$

to obtain

$$
\lim_{T \to \infty} \beta^T \frac{B_{T+\bar{t}}}{P_{T+\bar{t}}} = -m_{\bar{t}}
$$

Hence, $\frac{B_{t-1}}{P_t} \to -\infty$ which means that the borrowing constraint (2.24) is violated.

Case 2: $\tau \geq m^*$. We will prove that $m_t = \tau$ is the unique equilibrium by first showing that it is an equilibrium, and then exclude all other cases by contradiction.

To show that $m_t = \tau$, with $\bar{P} = \frac{M}{\tau}$, is an equilibrium, we need to check that there exists an allocation and a price system consistent with (2.2),(2.4), (2.7)-(2.11), (2.13) and (2.30). We propose the allocation $c_t = y - \tau, \hat{M}_t = 0, B_t = 0$ with the initial condition $B_{-1} = \hat{M}_{-1} = 0$ and the bond price $Q_t = \beta$.

(2.2) is satisfied by assumption. (2.4) is satisfied since $\hat{M}_t = 0$. (2.7) is satisfied by assumption. (2.8) is satisfied by assumption. (2.10) is satisfied since $Q_t = \beta < 1$. (2.9) is satisfied since

$$\frac{(1 - Q_t)M - \tau P_t}{P_t} + \frac{Q_t}{P_t}(M - \tau P_t + B_t) = \frac{1}{P_t}(M - \tau P_{t-1} + B_{t-1})$$

$$\Leftrightarrow B_t = \frac{1}{Q_t}\tau(P_t - P_{t-1}) + \frac{1}{Q_t}B_{t-1}$$

$$\Leftrightarrow B_t = \frac{1}{Q_t}B_{t-1}$$

which together with the initial conditions imply that $B_t = 0$ for all t. (2.11) holds since $Q_t = \beta$ and $P_t = \bar{P}$ for all t. (2.13) holds as $B_t = \hat{M} = 0$ for all t. (2.30) holds by assumption.

This shows that there is an equilibrium with $m_t = \tau$. To show that it is the unique equilibrium, we proceed by contradiction.

Suppose $m_t < \tau$ in some period t. This violates the tax constraint (2.23) and cannot be an equilibrium.

Suppose $m_t > \tau$ in some period t. Then we are in the same hyperdeflationary situation in Case 1, and analogous to that case, the solution path will feature a violation of the borrowing constraint as $t \to \infty$. $\square$

Proofs to Section 2.5

Proof of Proposition 5. For existence, we guess that an equilibrium is formed by

$$c_t \equiv y - \tau \qquad B_t = 0 \qquad \hat{M}_t = 0 \tag{2.40}$$

$$P_t \equiv \frac{M}{\tau} \qquad Q_t = \frac{1}{\beta} \tag{2.41}$$

The guess is an equilibrium if it satisfies the market clearing conditions (2.7)-(2.8), the consumer optimality conditions (2.4), (2.2) (2.9)-(2.13) and the zero lower bound on interest rates (2.10). We check these conditions one by one.

The non-negativity constraint on money (2.4), the initial conditions on money and bonds (2.2), and the non-negativity constraint on nominal interest rates (2.10) are directly satisfied by assumption. The goods market clears since $c_t = y - \tau$. The money market clearing condition (2.8) holds as

$$
\begin{aligned}
T_t(\bar{P}) + \hat{M}_t &= \bar{P}\tau + \Pi(\bar{P}) + 0 \\
&= \bar{P}\tau \\
&= M_t(\bar{P})
\end{aligned}
$$

for all t. The Euler equation (2.11) holds as the consumption and price level are constant and $Q_t = \beta$. The constraint (2.12) holds as $\hat{M}_t \equiv 0$. (2.13) trivially holds since $B_t = \hat{M}_t = 0$. The guess is therefore an equilibrium.

For uniqueness, we do a proof by contradiction.

Case 1: $P_t > \bar{P}$. Assume that there is an equilibrium with $P_s > \bar{P}$ in some period s. Then, by the well-ordering principle, there exists a smallest $t = \underline{t}$ such that $P_t > \bar{P}$. Assume that the tax constraint is not violated. From this assumption together with the money market clearing

condition (2.8), we derive

$$T_t(P_t) + \hat{M}_t = M_t(P_t) = P_t \tau_t + \hat{M}_{t-1}$$

As $T_t(P_t) > P_t \tau_t$ whenever $P_t > \bar{P}$ and $\hat{M}_t \geq 0$, we have that $\hat{M}_{t-1} > 0$. That is, as the level of taxes exceeds money issuance, the only way that the money market clearing condition can be satisfied is if the household carried money from a previous period. Now, either $\underline{t} = 0$ which would violate the initial condition $\hat{M}_{t-1} = 0$, or $\underline{t} \geq 1$ in which case the optimality condition for money holdings (2.12) implies $Q_{t-1} = 1$. From the Euler equation (2.11), we then have that $P_t = \beta P_{t-1}$ and so $P_{t-1} > P_t > \bar{P}$, which violates the minimality of $\underline{t}$.

   Case 2: $P_t < \bar{P}$. Assume that there is an equilibrium with $P_t < \bar{P}$ for some t. Then

$$\hat{M}_t = M_t(P_t) - T_t(P_t) > 0$$

$\hat{M}_t > 0$ is consistent with household optimality conditions (2.12) in period t only if $Q_t = 1$. From the Euler equation (2.11), we then have that $\frac{P_{t+1}}{P_t} = \beta$. As such, $P_{t+1} < P_t < \bar{P}$. By induction, $P_{t+k} = \beta^k P_t$, $Q_{t+k} = 1$, $\hat{M}_{t+k} > 0$. However, this means that the transversality condition is violated. To see this, note that

$$\hat{M}_t = \hat{M}_{t-1} + P_t \tau - T_t(P_t) = \hat{M}_{t-1} - \Pi(P_t)$$

and $\Pi(P_t) < 0$ for $P_t < \bar{P}$. Using that $Q_s = 1, B_s = 0$ for all $s \geq t$, we find that

$$
\begin{aligned}
\lim_{T \to \infty} \frac{\left(\prod_{s=0}^{T-1} Q_s \right) (B_T + \hat{M}_T)}{P_0} &= \frac{\left(\prod_{s=0}^{T-1} Q_s \right)}{P_0} \lim_{T \to \infty} \hat{M}_T \\
&= \frac{\left(\prod_{s=0}^{T-1} Q_s \right)}{P_0} \left(\hat{M}_t + \lim_{T \to \infty} \sum_{s=t+1}^{T} -\Pi(P_s) \right) \\
&> 0
\end{aligned}
$$

Therefore, the transversality condition is violated and there does not

exist an equilibrium with $P_t < \bar{P}$ for any t. Thus, $P_t \equiv \bar{P}$ is the unique equilibrium price sequence. $\qquad\square$

Chapter 3

Swedish Unemployment Dynamics[*]

3.1 Introduction

An important question in labor market economics is how important different labor market flows are in driving business cycle fluctuations in unemployment. The standard estimation method in the literature builds on Shimer (2012). He notes that actual U.S. unemployment is close to the steady state implied by the prevailing flow rates and uses this observation to decompose unemployment variations into contributions from unemployment inflows and outflows.

The steady-state assumption can be problematic when working with European data. First, gross flows are much smaller in European countries than in the U.S., which makes the steady-state unemployment rate less likely to be a good approximation of the actual unemployment rate (Elsby et al., 2013). Second, many European labor markets are dual and there are potentially large differences in labor market dynamics between

permanent and temporary jobs. Duality motivates a distinction between permanent and temporary jobs in the decomposition analysis.

We develop a new method for decomposing the sources of labor market fluctuations that jointly allows for (i) a slow convergence to steady state and (ii) an arbitrary number of labor market states. The method uses continuous time Markov chain theory to express the transition probability matrix between different labor market states as the matrix exponential of the integral of instantaneous flow matrices. Our formulation captures the dependence of the current state on the whole history of instantaneous flow rates for an arbitrary number of states, and it also provides a mapping between observed transition rates and underlying instantaneous flow rates. We use the matrix exponential formulation to perform a joint treatment of estimation, time-aggregation, detrending, log-linearization, and variance decomposition of trend-deviations into contributions from different flow rates. The method also allows us to calculate the convergence rate to steady state.

We apply our methodology to a new data set covering Swedish labor market flows through the period 1987-2011. Sweden has a dual labor market with a strong legal distinction between permanent and temporary workers. Temporary workers account for about 13% of the working-age population (and closer to 20% of employment), and experience substantially larger flow rates than permanently employed workers. The labor market is also characterized by a slow convergence to steady state, with a half life of deviations from steady state of approximately two years.

We find that variation in unemployment is driven to an approximately equal degree by variations in (i) flows from unemployment to employment, (ii) flows from employment to unemployment, and (iii) flows in and out of the labor force. Focusing on flows in and out of the labor force, the most important source of variation is changes in the flow from inactivity to unemployment. We find that flows involving temporary contracts account for 44% of the variation in unemployment, while flows involving permanent contracts account for approximately 33% of the variation. The former is sizeable given that workers on temporary

contracts on average only account for 13% of the working-age population during the period.

We also show that it is important to account for out-of-steady-state dynamics. If we use the decomposition method which relies on approximating the actual state with the steady state, the share of variation attributed to flows involving permanent contracts rises from 33% to 44%. The intuition is that small variations in flow rates involving permanent employment have large effects on the implied steady-state employment rate. However, for permanent workers, gross flows are small as compared to population size, which means that the convergence rate to steady state is slow. Hence, variations in the steady state due to changes in flows involving permanent workers are larger than variations in the actual state.

These results are of broader interest in the study of European labor markets. Indeed, they suggest that decompositions based on the steady-state assumption are unlikely to be suitable in dual labor markets. In particular, our results indicate that there is a risk that existing studies on France and Spain (Silva and Vazquez-Grenno, 2013; Hairault et al., 2015) have overestimated the share of variation stemming from flows involving permanently employed workers.

In Section 3.2, we describe our data on the Swedish labor market. Section 3.3 describes our new decomposition method in continuous time and contrasts with a method based on a steady-state approximation. In Section 3.4, we discretize the method and estimate the parameters using the Swedish survey data. In Section 3.5, we use our parametrization to decompose labor market flows using both our new method and the method based on a steady-state assumption. The most closely related paper to ours methodologically is Elsby et al. (2015). We discuss our relationship to this paper in Section 3.6. Section 3.7 concludes the paper.

3.2   Data

3.2.1   Data description

Our data are from the Swedish Labor Force Survey (LFS) and cover the period 1987-2012 (Statistics Sweden, 2011, 2012). During this period, the survey samples 17 000-29 500 individuals each month. The survey began already in 1961, but micro data are only available for 1987 and onwards. The survey samples individuals from a register which aims at containing the entire population in Sweden (RTB). Up to 2001 the sample includes all ages 16-64, but from 2001, the age interval was expanded to 15-74. For consistency, we confine our sample to ages 16-64 throughout the entire period.

The survey sample is rotating (Figure 3.1). Specifically, a participating individual is interviewed about his or her employment status at a given week every third month for two years. Each month, eight groups of individuals are interviewed. Seven of these eight groups have been interviewed previously, while one group is being interviewed for the first time.

The surveyed population is divided into three categories (i) employed, (ii) unemployed, and (iii) outside the labor force (Figure 3.1).

- An individual is defined as employed if she worked at least one hour during the reference week as (a) self-employed (including helping spouse), (b) on a permanent contract (*tillsvidareanställning*), or (c) on a time-limited contract.[1] An individual who has a job but was absent from work due to illness, leave, vacation, military service, a conflict or similar is also counted as employed, as is an individual in a labor market program if she receives some remuneration from her employer.

- An individual is defined as *unemployed* if she is not employed, but

---

[1]An alternative term for a permanent contract is that it is *open-ended*. Time-limited contracts are dominated by temps, object jobs (*objektanställningar*) and "called as needed" (*kallas vid behov*), but also include workers in trial employment, on vacation, and in seasonal employment (Statistics Sweden, 2005)

has been applying for work within the last four weeks and is able to start work within the reference week or in the two following weeks. An individual is also defined as unemployed if she is set to start a job within the next three months, provided that the individual would be ready to start already in the reference week or during the following two weeks.

- An individual is defined as *outside the labor force* if she is not covered by the above definitions. This includes individuals who would and could be able to work, but did not actively seek jobs (*latent unemployment*). This group includes, *inter alia*, pensioners, people engaged in home production, and conscripted soldiers.

The treatment of students deserves a discussion. Up to 2007, full-time students were always defined as belonging to *outside the labor force*. However, in October 2007 the definition was altered such that non-employed students who have been applying for work within the last four weeks and are ready to start work within three weeks, are defined as being unemployed. This change was made in order to comply with guidelines from the International Labour Organization (ILO) and the European Union (EU). For consistency, we use the pre-2007 definition throughout the entire period.

3.2.2 Data selection

We use the definition from the Labor Force Survey to construct four categories. We classify an individual as being (i) employed on a permanent contract, (ii) employed on a temporary contract, (iii) unemployed, or (iv) outside the labor force. We define *self-employed* workers as being employed on permanent contracts.

To compute labor market transition probabilities $P(t, t+1)$ in a quarter t, we restrict the sample to individuals for whom we also have an observation in the next quarter $t+1$. We define 16 indicator variables for each of these observations: I_{t,i,s_1,s_2} for $s_1, s_2 \in [1, 4]$. Here $I_{t,i,s_1=j,s_2=k} =$

1 if the individual i was in state j at time t and in state k at time $t-1$. We compute the flow probability from state j to state k at time t as

$$p_{t,j,k} = \frac{\sum_i I_{i,t+1,s_1=j,s_2=k} w_i}{\sum_i \sum_{s_1} I_{i,t,s_1,s_2=k} w_i} \tag{3.1}$$

The terms w_i are sample-weights provided by Statistics Sweden. These weights reflect the relative over/under sampling of various demographic groups in the survey and are computed separately for both stock and flows on a monthly, quarterly and yearly frequency, respectively. Unfortunately, weights for flows are only available from 2005 and onwards. Thus, we use stock weights from quarter t to calculate transition probabilities from quarter t to $t+1$ for the entire period. We verify the validity of this approach by showing that transition probabilities calculated using flows and stock weights are similar for the period 2005-2012 when both are available.[2]

We clean and seasonally correct data in the transition matrices $P(t, t+1)$ before conducting our analysis. As the entire panel is replaced at the beginning of 2005, the flow probabilities cannot be computed in 2004Q4. In this period, we interpolate all transition probabilities. We correct the diagonal in the interpolated transition matrix to ensure that its rows sum to one.

3.2.3   Labor market stocks and flows

Figure 3.2 shows the stock of employed, unemployed, and inactive in Sweden since 1987. The overall employment rate is falling by approximately 8 percentage points over the period (from 82 to 74 percent), while the inactivity rate is increasing from 16 to 21 percent. However, the change in the overall employment rate masks diversity in the changes for permanent and temporary employment. While the share of the population on permanent employment falls from 72 to 63 percent, the share with temporary employment has risen from 9 to 11 percent.

---

[2]The figures are available upon request.

Figures 3.3-3.4 show the flows moving between the four groups computed as quarterly hazard rates. In Section 3.4, we explain the computation method. Two trends are visible from these figures. First, all flow rates into permanent and temporary employment have decreased. This trend is particularly clear for the flow from unemployment into permanent employment. Second, there has been an increase in the probability of flowing from employment into unemployment. There is no clear trend in the hazard rates for movements into inactivity.

3.2.4 The cyclicality of flow rates

Table 3.2 illustrates the sensitivity of labor market transition rates to the business cycle. We measure the cyclicality by regressing the relevant transition rate (logged) on the unemployment rate and a linear trend. Table 3.2 reports the coefficient on the unemployment rate from this regression. A positive (negative) value in Table 3.2 indicates that the correlation between the relevant transition rate and unemployment is positive (negative). Since unemployment correlates negatively with the business cycle, this means that a negative (positive) value implies a pro-cyclical (counter-cyclical) transition rate. Thus, from Table 3.2 we see that all transition rates into permanent and temporary employment are pro-cyclical, while all rates into unemployment are countercyclical. The transition rate from permanent and temporary employment to inactivity is a-cyclical, while the transition rate from unemployment to inactivity is counter-cyclical.

3.3 Method

In this section, we outline our method for decomposing fluctuations in unemployment into contributions from the underlying labor market flows. Our method simultaneously addresses three issues.

First, it addresses the well-known time aggregation problem. The time aggregation problem arises when individuals can change labor mar-

ket states several times within a measurement interval. In particular, an individual who is coded as being employed for two consecutive quarters might have experienced a spell of unemployment in-between. As pointed out by Shimer (2012), this can lead to an underestimation of the importance of the job finding rate for business cycle dynamics, as a decrease in the job-finding rate mechanically leads to an increase in the measured job loss rate. In a manner similar to that used in Shimer (2012), we address this problem by using a continuous time formulation of the problem.

Our second contribution is to devise a method that can handle a slow convergence rate towards the steady state. A common method in the literature is to assume that steady-state unemployment is a good approximation of actual unemployment and therefore measure how variations in labor market flows contribute to variations in steady-state unemployment. This assumption seems appropriate in the U.S. where a large labor market churn ensures a quick convergence to steady state, but it is more questionable in Europe where both the job-finding and separation rates are lower. Indeed, Elsby et al. (2013) find an average monthly job-finding rate of 13 percent in Europe as compared to 57 percent in the United States and a monthly separation rate of 0.8 percent in Europe as compared to 3.6 percent in the United States.[3] The steady-state assumption is even more questionable when one includes a distinction between permanent and temporary jobs since the flows in and out of permanent employment are low.

Third, our method allows for an arbitrary number of labor market states. By using flow matrices and matrix exponentials to characterize labor market dynamics, we can derive a compact expression for decomposing flows with an arbitrary number of states.

In this section, we first introduce our notation and describe how we use a matrix formulation of continuous time Markov chains to express the current labor market distribution as a matrix exponential of a sequence

---

[3]For Europe, the numbers are simple averages for the countries France, Germany, Ireland, Italy, Norway, Portugal, Spain, and Sweden. The sample start varies from 1968 to 1986.

of flow rates. Second, we express deviations in labor market states from their trend as a function of deviations of labor market flows from their respective trends. Third, we log-linearize the resulting expression around its trend and perform a variance decomposition. Finally, we compare this method to the standard method where current labor market states are approximated by means of the steady-state level associated with current labor market flows.

3.3.1 Notation

We develop our model in continuous time and discretize it later for the purpose of estimation. We write $X(t)$ for variables defined in continuous time and X_t for variables defined in discrete time. We introduce subscripts when necessary to denote indexing, so $X_{i,j}(t)$ denotes the (i, j) element of a matrix defined in continuous time at time t, and $X_{i,j,t}$ denotes the (i, j) element of a matrix defined in discrete time at time t. We use an analogous notation for vectors.

Our first key variable is $\mathbf{x}(t)$ which denotes an $1 \times S$ vector that defines a distribution over labor market states. $x_s(t)$ thus denotes the share of the workforce in state $s \in \{1, \ldots, S\}$, so that

$$\sum_{s=1}^{S} x_s(t) = 1.$$

$Q(t)$ denotes an $S \times S$ instantaneous flow matrix where the $S \times (S-1)$ off-diagonal elements denote the transition rates between labor market states. The rows of Q are defined to sum to zero, which implies that the diagonal elements are $Q_{i,i} = -\sum_j Q_{i,j}$.

$\hat{Q}(t)$ denotes the trend of the instantaneous flow matrix. $\hat{Q}_{i,j}(t)$ is defined as the trend component of Q_{ij} for $i \neq j$, and $\hat{Q}_{ii}(t) = -\sum_{i \neq j} \hat{Q}_{i,j}$. We describe the trend estimation in detail in Section 3.4.

3.3.2   Expressing the labor market distribution using Markov chains in continuous time

Using standard results on continuous time Markov chains (Norris, 1997), the evolution of the labor market distribution from period t to period $t + \tau$ can be expressed as

$$\mathbf{x}(t + \tau) = \mathbf{x}(t) \exp\left(\int_t^{t+\tau} Q(u)du \right).$$ (3.2)

Thus, we can write the distribution at time t as

$$\mathbf{x}(t) = \mathbf{x}(0) \exp\left(\int_{u=0}^{t} Q(u)du \right)$$ (3.3)

This expresses the labor market distribution at time t as a function of an initial state and all subsequent flow rates. Here, exp refers to the matrix exponential defined by

$$\exp(A) = \sum_{n \geq 0} \frac{A^n}{n!}.$$

We are interested in decomposing the variance in labor market states over the business cycle into contributions from changes in different flow rates. For this purpose, we define labor market trend states $\hat{x}(t)$ as the vector generated by starting from an initial state and assuming that the labor market evolves according to the trend instantaneous flow matrix $\hat{Q}(t)$. Formally,

$$\hat{\mathbf{x}}(t) = \mathbf{x}(0) \exp\left(\int_{u=0}^{t} \hat{Q}(u)du \right).$$ (3.4)

Log-linearizing the expression for the actual labor market state (3.3) around the trend labor market state (3.4) yields the following expression:

$$\log\left(\mathbf{x}_s(t)\right) - \log\left(\hat{\mathbf{x}}_s(t)\right) \approx \int_{u=0}^{t} \sum_{i,j \neq i} \left.\frac{\partial \mathbf{x}_s(t)}{\partial Q_{i,j}(u)}\right|_{Q=\hat{Q}} \frac{\hat{Q}_{i,j}(u)}{\hat{\mathbf{x}}_s(t)} \left(\log Q_{i,j}(u) - \log \hat{Q}_{i,j}(u)\right) du,$$

(3.5)

where $\hat{Q}$ denotes the whole history of $\hat{Q}(u)$. Here, the percentage deviation of each labor market state from its trend is a function of the percentage deviation of each labor market flow from its trend, weighted by the elasticity of the relevant labor market state with respect to that particular flow rate. Notice also that the expression is a function of the *entire path* of deviations in flow rates from the trend. An example of an elasticity we are discussing is "the elasticity of unemployment today with respect to a change in the permanent to temporary employment flow rate one year and three months ago, keeping all other flow rates constant". We then integrate over all time periods and sum over all different types of flow rates. This is how the method deviates from the steady-state assumption, which posits that non-contemporaneous flows are irrelevant for understanding the current state.

This expression motivates the definition of a contribution function $\Gamma_{i,j,s}(t)$ which gives the contribution of flow $i \rightarrow j$ to variations in state s at time t. We define $\Gamma_{i,i,s,t} = 0$ for all $i = 1, \ldots, S$ and for $i \neq j$, we define

$$\Gamma_{i,j,s}(t) \equiv \int_{u=0}^{t} \left.\frac{\partial \mathbf{x}_s(t)}{\partial Q_{i,j}(u)}\right|_{Q=\hat{Q}} \frac{\hat{Q}_{i,j}(u)}{\hat{\mathbf{x}}_s(t)} \left(\log Q_{i,j}(u) - \log \hat{Q}_{i,j}(u)\right) du.$$

With this formulation, we obtain

$$\log(\mathbf{x}_s(t)) - \log(\hat{\mathbf{x}}_s(t)) \approx \sum_{i \neq j} \Gamma_{i,j,s}(t).$$

(3.6)

Given (3.6), we wish to decompose the observed business cycle variation in labor market states into contributions from each flow rate. For this purpose, we rely on the linear expression for deviations in the trend

as a function of the flow rates in (3.6). We carry out a statistical variance decomposition of (3.6) in order to find the contribution from each flow rate.

This variance decomposition yields the expression

$$\beta_{i,j,s}^{non.st.st.} = \frac{Cov\left(\log \mathbf{x}(t)_s - \log \hat{\mathbf{x}}(t)_s, \Gamma_{i,j,s}(t)\right)}{Var\left(\log \mathbf{x}(t)_s - \log \hat{\mathbf{x}}(t)_s\right)}. \tag{3.7}$$

Here, $\beta_{i,j,s}^{non.st.st.}$ denotes the contribution of flows $i \to j$ to the variance of state s.

3.3.3   Comparison with the steady-state based method

As mentioned above, the literature often relies on an assumption that the current labor market state can be well approximated by the steady state associated with current labor market flows.

Specifically, we let $\bar{\mathbf{x}}(t)$ be the steady state associated with the observed flow matrix $Q(t)$. This distribution can be found as the (left-) eigenvector to $Q(t)$ with eigenvalue zero. We define the function $\sigma()$, which takes $Q(t)$ as the input and yields $\bar{\mathbf{x}}(t)$ as the output such that:

$$\sigma : Q(t) \in \mathbb{R}^s \times \mathbb{R}^s \longrightarrow \bar{\mathbf{x}}(t) \in \mathbb{R}^s \tag{3.8}$$

$$\text{s.t}$$

$$\bar{\mathbf{x}}(t)Q(t) = 0 \ , \ \sum_s \bar{\mathbf{x}}(t) = 1$$

This way, the steady-state distribution associated with $Q(t)$ is written as:

$$\bar{\mathbf{x}}(t) = \sigma(Q(t)). \tag{3.9}$$

We assume that $Q(t)$ satisfies standard conditions that guarantee the singe-valuedness of σ, in particular that the embedded discrete time Markov chain is irreducible. This is trivially fulfilled in our labor market

since all labor market flow rates are non-zero.[4]

We similarly define the trend steady-state distribution as

$$\tilde{x}(t) = \sigma(\hat{Q}(t)).$$

This expression defines the stationary distribution associated with the trend flow rates. That is, if the trend flow rates were to persist forever, this is what the labor market would converge to. If we live in a labor market where the actual distribution is close to the stationary distribution, the difference between the steady state $\hat{x}(t)$ and the trend steady state $\tilde{x}(t)$ closely mirrors the difference between the trend state of the labor market and the actual state of the labor market.

Log-linearizing (3.9) around the steady state associated with trend flows $\tilde{\mathbf{x}}(t)$ yields

$$\log\left(\bar{\mathbf{x}}(t)\right)_s - \log\left(\tilde{\mathbf{x}}(t)\right)_s \approx \sum_{i,j,i\neq j} \frac{\partial\sigma(Q(t))_s}{\partial Q(t)_{i,j}}\bigg|_{Q=\hat{Q}} \frac{\hat{Q}(t)_{i,j}}{\sigma(\hat{Q}(t))_s}\left(\log Q_{i,j}(t) - \log\hat{Q}_{i,j}(t)\right).$$

(3.10)

Now we can define $\Gamma^{st.st.}_{i,j,s}(t)$ as the contribution of flows $i \to j$

$$\Gamma^{st.st.}_{i,j,s,t} = \frac{\partial\sigma(Q(t))_s}{\partial Q(t)_{i,j}}\bigg|_{Q=\hat{Q}} \frac{\hat{Q}(t)_{i,j}}{\sigma(\hat{Q}(t))_s}\left(\log Q_{i,j}(t) - \log\hat{Q}_{i,j}(t)\right)$$

We can compare this expression with the expression for non steady-state contribution $\Gamma_{i,j,s}(t)$ from (3.6). Just as with $\Gamma_{i,j,s}(t)$, the percentage deviation of each labor market state from the trend is once more a function of the percentage deviation of each flow rate, but they are now multiplied by the elasticity of the *steady-state* labor market state with respect to that flow rate. Here, there is no integral over previous time periods, as the decomposition only depends on contemporaneous flows.

We can once more take advantage of the linearity in (3.10) and con-

---

[4]For standard conditions, see Norris (1997).

duct a variance decomposition to find the contribution from each flow
rate:

$$
\beta_{i,j,s}^{st.st.} = \frac{Cov\left(\log \bar{\mathbf{x}}(t)_s - \log \tilde{\mathbf{x}}(t)_s, \Gamma_{i,j,s}^{st.st.}(t)\right)}{Var\left(\log \bar{\mathbf{x}}(t)_s - \log \tilde{\mathbf{x}}(t)_s\right)}. \tag{3.11}
$$

Here $\beta_{i,j,s}^{st.st.}$ denotes the contribution from the flow rate from state i to
state j to the overall business cycle variation in state s.

3.3.4   Decomposing changes in the *rate* of unemployment

So far, we have decomposed the fluctuations in the population *share*
of people in various labor market states – including the unemployment
share. To be consistent with the literature, we wish to decompose fluc-
tuations in the unemployment *rate*, which is the unemployment share of
the population divided by labor force participation.

To facilitate this calculation, we let unemployment be state U and
assume that the labor force states are from state 1 to state l (in our
case, these states are permanent employment, temporary employment
and unemployment). We can then write the unemployment rate as

$$
\chi_U = \frac{\mathbf{x}_U}{\sum_{s=1}^{l} \mathbf{x}_s} \tag{3.12}
$$

This means that the logarithm of the unemployment rate is written as
the logged unemployment share less the logged participation rate:

$$
\log(\chi_U) = \log(\mathbf{x}_u) - \log(\sum_{s=1}^{l} \mathbf{x}_s) \tag{3.13}
$$

Thus, to determine how much a particular flow contributes to vari-
ations in the unemployment rate, we need to take its contribution to
variations in $\log(x_u)$ derived in the previous section and *subtract* its con-
tribution to variations in the log participation rate.

The first step in doing this is to log-linearize the participation rate

around its trend. For the non steady-state method, we obtain

$$
\log \left(\sum_{s=1}^{l} \mathbf{x}_s(t) \right) - \log \left(\sum_{s=1}^{l} \hat{\mathbf{x}}_s(t) \right) \approx
$$

$$
\sum_{s=1}^{l} \int_{u=0}^{t} \sum_{i,j \neq i} \frac{\partial \mathbf{x}_s(t)}{\partial Q_{i,j}(u)} \Bigg|_{Q=\hat{Q}} \frac{\hat{Q}_{i,j}(u)}{\sum_{s=1}^{l} \hat{\mathbf{x}}_s(t)} \left(\log Q_{i,j}(u) - \log \hat{Q}_{i,j}(u) \right) du.
$$

$$(3.14)$$

For the steady-state based method we obtain

$$
\log \left(\sum_{s=1}^{l} \bar{\mathbf{x}}_s(t) \right) - \log \left(\sum_{s=1}^{l} \tilde{\mathbf{x}}_s(t) \right) \approx
$$

$$
\sum_{s=1}^{l} \sum_{i,j \neq i} \frac{\partial \sigma(Q(t))_s}{\partial Q_{i,j}(t)} \Bigg|_{Q=\hat{Q}} \frac{\hat{Q}_{i,j}(t)}{\sum_{s=1}^{l} \tilde{\mathbf{x}}_s(t)} \left(\log Q_{i,j}(t) - \log \hat{Q}_{i,j}(t) \right) \quad (3.15)
$$

Note that these expressions are almost identical to equations (3.5) and (3.10). The difference is that we sum over all labor force states and divide by the trend participation rate $\sum_{s=1}^{l} \hat{\mathbf{x}}_s(t)$ instead of by $\hat{\mathbf{x}}_s(t)$. Following the definitions from the previous sections, the non steady-state decomposition becomes

$$
\log \left(\sum_{s=1}^{l} \mathbf{x}(t)_s \right) - \log \left(\sum_{s=1}^{l} \hat{\mathbf{x}}_s(t) \right) \approx \sum_{i \neq j} \Gamma_{i,j}^{lfp}(t),
$$

whereas the steady-state decomposition becomes

$$
\log \left(\sum_{s=1}^{l} \mathbf{x}_s(t) \right) - \log \left(\sum_{s=1}^{l} \hat{\mathbf{x}}_s(t) \right) \approx \sum_{i \neq j} \Gamma_{i,j}^{st.st,lfp}(t).
$$

Here

$$
\Gamma_{i,j,t}^{lfp} = \sum_{s=1}^{l} \left[\Gamma_{i,j,s}(t) \frac{\hat{\mathbf{x}}_{\mathbf{s}}(\mathbf{t})}{\sum_{s'=1}^{l} \hat{\mathbf{x}}_{\mathbf{s'}}(\mathbf{t})} \right]
$$

and

$$\Gamma_{i,j,t}^{st.st.,lfp} = \sum_{s=1}^{l} \left[\Gamma_{i,j,s}^{st.st.}(t) \frac{\tilde{\mathbf{x}}_{\mathbf{s}}(\mathbf{t})}{\sum_{s'=1}^{l} \tilde{\mathbf{x}}_{\mathbf{s'}}(\mathbf{t})} \right].$$

Having defined this log-linearization, we can log-linearize the unemployment rate using the non steady-state method as

$$\log\left(\frac{\mathbf{x}_U(t)}{\sum_{s=1}^{l} \mathbf{x}_s(t)}\right) - \log\left(\frac{\hat{\mathbf{x}}_U(t)}{\sum_{s=1}^{l} \hat{\mathbf{x}}_s(t)}\right) \approx \sum_{i \neq j} \Gamma_{i,j,s}(t) - \Gamma_{i,j}^{lfp}(t),$$

while using the steady-state method, we obtain

$$\log\left(\frac{\bar{\mathbf{x}}_U(t)}{\sum_{s=1}^{l} \bar{\mathbf{x}}_s(t)}\right) - \log\left(\frac{\tilde{\mathbf{x}}_U(t)}{\sum_{s=1}^{l} \tilde{\mathbf{x}}_s(t)}\right) \approx \sum_{i \neq j} \Gamma_{i,j,U}^{st.st.}(t) - \Gamma_{i,j}^{st.st.,lfp}(t).$$

Using these expressions, we can define the contribution of each flow $i \to j$ to unemployment by

$$\beta_{i,j}^{nonst.st.,lfp} = \frac{Cov\left(\log\left(\frac{\mathbf{x}_u(t)}{\sum_{s=1}^{l} \mathbf{x}_s(t)}\right) - \log\left(\frac{\hat{\mathbf{x}}_u(t)}{\sum_{s=1}^{l} \hat{\mathbf{x}}_s(t)}\right), \Gamma_{i,j,u}(t) - \Gamma_{i,j}^{lfp}(t)\right)}{Var\left(\log\left(\frac{\mathbf{x}_u(t)}{\sum_{s=1}^{l} \mathbf{x}_s(t)}\right) - \log\left(\frac{\hat{\mathbf{x}}_u(t)}{\sum_{s=1}^{l} \hat{\mathbf{x}}_s(t)}\right)\right)} \quad (3.16)$$

$$\beta_{i,j}^{st.st.,lfp} = \frac{Cov\left(\log\left(\frac{\bar{\mathbf{x}}_u(t)}{\sum_{s=1}^{l} \bar{\mathbf{x}}_s(t)}\right) - \log\left(\frac{\tilde{\mathbf{x}}(t)_u}{\sum_{s=1}^{l} \tilde{\mathbf{x}}(t)_s}\right), \Gamma_{i,j,u}(t) - \Gamma_{i,j}^{lfp}(t)\right)}{Var\left(\log\left(\frac{\bar{\mathbf{x}}_u(t)}{\sum_{s=1}^{l} \bar{\mathbf{x}}_s(t)}\right) - \log\left(\frac{\tilde{\mathbf{x}}_u(t)}{\sum_{s=1}^{l} \tilde{\mathbf{x}}_s(t)}\right)\right)}. \quad (3.17)$$

We use discretized versions of (3.16) and (3.17) together with estimates of Q, $\tilde{Q}$, $\hat{\mathbf{x}}$, $\tilde{\mathbf{x}}$, and Γ to compute the contribution of different flows to the unemployment rate.

3.4   Estimation and discretization

In order to perform the decomposition, we need estimates of the instantaneous flow matrix Q and the corresponding trend matrix $\hat{Q}$. However, as detailed in Section 3.2, the data only record quarterly transition prob-

abilities $P(t, t + 1)$. In this section, we outline how we use $P(t, t + 1)$ to estimate $Q(t)$ and $\tilde{Q}(t)$, and how we discretize formulas (3.16) and (3.17) to allow us to perform the decomposition.

3.4.1 Estimation of the flow matrix $Q(t)$

We start by noting the relationship between the observed transition matrix and the underlying instantaneous flow matrix, where the time unit is one quarter:

$$P(t, t + 1) = \exp\left(\int_t^{t+1} Q(u)du\right). \tag{3.18}$$

That is, the observed transition matrix between period t and $t + 1$ is a function of the continuum of instantaneous flow matrices during the time interval.

To identify a flow matrix from an observed transition matrix, we assume that the transition rates are constant between two measurements. Specifically, we will assume that the flow matrix is constant in the interval $(t, t + 1)$. This allows us to identify $Q(t)$ from (3.18)

$$Q(t) = \log_m P(t, t + 1), \tag{3.19}$$

where $\log_m$ is the matrix logaritm. The resulting flow rates are presented in Figure 3.3, which we described in Section 3.2 above.

3.4.2 Estimation of the trend flow matrix

We then identify the trend flow matrix $\hat{Q}(t)$ element by element using the full time sequence of $Q(t)$.

Specifically, we follow the method from the literature and identify $\hat{Q}_{ij,i\neq j}(t)$ by applying an HP filter to the corresponding time series $\{Q_{i,j}(u)\}_{s=0}^T$. We face a choice of the value of the smoothing parameter λ. Here we follow Gomes (2012) who also works with quarterly data, and choose $\lambda = 10^5$ as our smoothing parameter. In the Appendix, we

vary λ to check the robustness. The diagonal elements, $\hat{Q}_{i,i}(t)$, are computed as residuals such that each row in $\hat{Q}(t)$ sums to 0. The resulting trend flow rates are depicted in Figure 3.5.

3.4.3   Discretization

When we have estimated the discrete versions of $Q(t)$ and $\hat{Q}(t)$, we can define Q_t and $\hat{Q}_t$ as the values they take on the intervals $[t, t+1)$. We can now write the labor market evolution in discrete form as

$$x_t = x_0 \exp\left(\sum_{u=0}^{t-1} Q_u\right).$$

With this expression, we can re-express the contribution function $\Gamma_{i,j,s}(t)$ in a discrete version $\Gamma_{i,j,s,t}$ as

$$\Gamma_{i,j,s,t} = \sum_{u=0}^{t-1} \frac{\partial \mathbf{x}_{s,t}}{\partial Q_{i,j,u}} \frac{\hat{Q}_{i,j,u}}{\hat{\mathbf{x}}_{s,t}} \left(\log(Q_{i,j,u}) - \log(\hat{Q}_{i,j,u})\right)$$

and with the concomitant change to $\Gamma_{i,j,t}^{lfp}$. With this discrete approximation, we define the contribution of a flow $i \rightarrow j$ to variations in unemployment as

$$\hat{\beta}_{i,j}^{nonst.st.} = \frac{\widetilde{Cov}\left(\log\left(\frac{\mathbf{x}_{U,t}}{\sum_{s=1}^l \mathbf{x}_{s,t}}\right) - \log\left(\frac{\hat{\mathbf{x}}_{U,t}}{\sum_{s=1}^l \hat{\mathbf{x}}_{s,t}}\right), \Gamma_{i,j,U,t} - \Gamma_{i,j,t}^{lfp}\right)}{\widetilde{Var}\left(\log\left(\frac{\mathbf{x}_{U,t}}{\sum_{s=1}^l \mathbf{x}_{s,t}}\right) - \log\left(\frac{\hat{x}_{U,t}}{\sum_{s=1}^l \hat{\mathbf{x}}_{s,t}}\right)\right)}$$

where $\widetilde{Cov}$ and $\widetilde{Var}$ denote sample covariances and variances. This is the estimation method we take to the data.

Panel A in Table 3.3 shows the decomposition of the variation in the unemployment rate using our preferred non steady-state method described in Section 3.3. The table suggests that we can divide contributions to unemployment variations into three types of flows of roughly equal sizes. Flows from unemployment to employment explain 35% (=13%+22%), flows from employment to unemployment explain

32% (=15%+17 %), and flows involving non-participation explain 32% of the variation in unemployment.

Panel A in Table 3.3 also tells us how variations in flow rates concerning temporary and permanent contracts contribute to the total variation in unemployment. In particular, we see that variations in flows involving temporary employment account for 44% of the variation in unemployment, while flows concerning permanent employment account for 33%. The former are sizable, as workers on temporary contracts on average (during 1987-2012) only account for 13% of the population aged 16-64, while workers on permanent contracts account for 66%.[5]

The results also suggest that the convergence to steady state is slow. This convergence rate can be calculated using the second largest eigenvalue to the transition matrix, $Q(t)$. The largest eigenvalue is 0 so the second largest eigenvalue is a negative number $\lambda < 0$. Deviations from steady state decay at the rate $-\lambda$. As the second largest eigenvalue is approximately -0.09, the halving time of a deviation from the steady state is approximately $70/9 \approx 8$ quarters ≈ 2 years.[6]

To evaluate the importance of out-of-steady-state dynamics, we calculate the variance decomposition using the steady-state assumption. The results are reported in Panel B of Table 3.3. The main difference compared to the non steady-state method is that flows involving permanent workers are attributed a larger role. The share explained by flows involving permanent employment rises from 33% to 44% (30+14).

Why are the contributions stemming from permanent employment overestimated under the steady-state assumption? A likely reason is that the stock of individuals in permanent employment is sizable. Consequently, small swings in the in- and outflow rates from this state will have a large impact on the steady-state level of unemployment. However, as the flows in and out of permanent employment are low, it takes a long time to reach the new steady state. Consequently, short-lived swings in

---

[5]Recall from Section 3.2 that we have characterized self-employed workers as being on permanent contracts.

[6]For reference, see Norris (1997).

the transition in and out of regular employment will cause large variations in the steady state but small ones in the actual distribution.

Taken together, our results suggest that properly accounting for out-of-steady-state dynamics is important when analyzing a dual labor market. In a Swedish context, failing to do so leads to an overestimation of the variation from permanent employment. We think that this point is potentially also important in a broader European context, where labor markets have a substantial share of temporary contracts. When analyzing labor market duality in France and Spain, existing studies (Hairault et al., 2015; Silva and Vazquez-Grenno, 2013) have not accounted for the out-of-steady-state dynamics.

In Table 3.4 - Table 3.9 we check the robustness of our results in three dimensions. First, we redo the decomposition using non seasonally adjusted data (Table 3.4). Second, we vary the value of the smoothing parameter we utilize to compute the path of trend flow rates, $\tilde{Q}(t)$. This is done in Table 3.5 - Table 3.6. Third, we vary the time period. We do this to (i) exclude the latest recession (Table 3.7), (ii) exclude the recession in the early 1990s (Table 3.8) and (iii) exclude both these recessions (Table 3.9). None of these modifications change our results considerably.

3.5   Results

Panel A in Table 3.3 shows the decomposition of the variation in the unemployment rate using our preferred non steady-state method described in Section 3.3. The table suggests that we can divide contributions to unemployment variations into three types of flows of roughly equal sizes. Flows from unemployment to employment explain 35% (=13%+22%), flows from employment to unemployment explain 32% (=15%+17 %), and flows involving non-participation explain 32% of the variation in unemployment.

Panel A in Table 3.3 also tells us how variations in flow rates concerning temporary and permanent contracts contribute to the total variation in unemployment. In particular, we see that variations in flows involving

temporary employment account for 44% of the variation in unemployment, while flows concerning permanent employment account for 33%. The former are sizable, as workers on temporary contracts on average (during 1987-2012) only account for 13% of the population aged 16-64, while workers on permanent contracts account for 66%.[7]

The results also suggest that the convergence to steady state is slow. This convergence rate can be calculated using the second largest eigenvalue to the transition matrix, $Q(t)$. The largest eigenvalue is 0 so the second largest eigenvalue is a negative number $\lambda < 0$. Deviations from steady state decay at the rate $-\lambda$. As the second largest eigenvalue is approximately -0.09, the halving time of a deviation from the steady state is approximately $70/9 \approx 8$ quarters ≈ 2 years.[8]

To evaluate the importance of out-of-steady-state dynamics, we calculate the variance decomposition using the steady-state assumption. The results are reported in Panel B of Table 3.3. The main difference compared to the non steady-state method is that flows involving permanent workers are attributed a larger role. The share explained by flows involving permanent employment rises from 33% to 44% (30+14).

Why are the contributions stemming from permanent employment overestimated under the steady-state assumption? A likely reason is that the stock of individuals in permanent employment is sizable. Consequently, small swings in the in- and outflow rates from this state will have a large impact on the steady-state level of unemployment. However, as the flows in and out of permanent employment are low, it takes a long time to reach the new steady state. Consequently, short-lived swings in the transition in and out of regular employment will cause large variations in the steady state but small ones in the actual distribution.

Taken together, our results suggest that properly accounting for out-of-steady-state dynamics is important when analyzing a dual labor market. In a Swedish context, failing to do so leads to an overestimation of

---

[7]Recall from Section 3.2 that we have characterized self-employed workers as being on permanent contracts.

[8]For reference, see Norris (1997).

the variation from permanent employment. We think that this point is potentially also important in a broader European context, where labor markets have a substantial share of temporary contracts. When analyzing labor market duality in France and Spain, existing studies (Hairault et al., 2015; Silva and Vazquez-Grenno, 2013) have not accounted for the out-of-steady-state dynamics.

In Table 3.4 - Table 3.9 we check the robustness of our results in three dimensions. First, we redo the decomposition using non seasonally adjusted data (Table 3.4). Second, we vary the value of the smoothing parameter we utilize to compute the path of trend flow rates, $\tilde{Q}(t)$. This is done in Table 3.5 - Table 3.6. Third, we vary the time period. We do this to (i) exclude the latest recession (Table 3.7), (ii) exclude the recession in the early 1990s (Table 3.8) and (iii) exclude both these recessions (Table 3.9). None of these modifications change our results considerably.

3.6   Comparison to Elsby et al. (2015)

Methodologically, the most similar paper to ours is Elsby et al. (2015), which seeks to estimate the importance of the participation margin for labor market fluctuations in the US. To compare their paper to ours, note that the key to our methodology is to formulate labor market changes in terms of the matrix exponential of instantaneous flow rates:

$$x(t+\tau) = x(t) \exp \left(\int_t^{t+\tau} Q(s)ds \right). \qquad (3.20)$$

This formulation allows us to (i) derive the (average) instantaneous flow rate in a period as the matrix logarithm of the observed transition probability matrix, and (ii) log-linearize changes in labor market states around any path of flow matrices.

Elsby et al. (2015) do not use the matrix formulation in equation (3.20), but they develop a method to perform variance decomposition, taking into account the history of flow rates in the case when linearization is performed around lagged flow rates. To derive flow rates from

observed transition probability matrices, they derive our matrix logarithm formulation from first principles. They note that if flow rates Q are constant between t and $t+1$, the labor market flows are given by

$$\dot{x}(t) = x(t)Q, \tag{3.21}$$

which is a linear differential equation with the solution

$$x(t+1) = x(t)V\Gamma V^{-1},$$

where V is a matrix of eigenvectors of Q and Γ is a diagonal matrix with diagonal elements e^{λ_i} where λ_i are the eigenvalues of Q. If P is the transition matrix between t and $t+1$, this implies $P = V\Gamma V^{-1}$ and Q can be derived from P by

$$Q = V\log(V^{-1}PV)V^{-1}, \tag{3.22}$$

where $\log(\cdot)$ denotes an element-wise logarithm. By noting that the general solution of equation (3.21) is $x(t+1) = x(t)\exp\left(\int_t^{t+1} Q(s)ds\right)$, we see that (3.22) estimates the average flow rate Q_t over a given time period. It can also be shown that the right-hand side of (3.22) gives the matrix logarithm of P.

To perform a variance decomposition, their paper linearizes $\Delta x(t)$ around lagged flow rates $\Delta Q(t)$. Under this linearization, one does not need to directly use equation (3.20) since it is possible to express $\Delta x(t)$ as a function of the history of $\Delta\bar{x}(t)$'s, where $\bar{x}(t)$ denotes the implied steady state given $Q(t)$. Thus, given that they are interested in $\Delta x(t)$, they are able to fully express the effects of lagged flow rates as a function of their effects on the implied steady-state distribution. The method differs from the standard steady-state approximation since it takes into account the full history of changes in steady-state distributions.

3.7   Conclusion

This paper decomposes the total variation in unemployment in Sweden, where the labor market is dual and the convergence to steady state is slow. In doing so, we make two methodological contribution to the literature. First, we extend existing decomposition methods in the literature by allowing for an arbitrary number of labor market states. This is helpful in order to analyze the Swedish labor market, where both temporary and permanent contracts exist. Second, in this setting we allow for non steady state dynamics. This is important in a Swedish context, due to low flows between states and consequently a slow rate of convergence towards the steady state.

Using this setup we show that the contribution to unemployment variability from in- and outflow from unemployment is roughly 60/30. We furthermore show that flows involving temporary contrasts account for 44% of the total variation, while flows involving permanent contracts account for 33%. The former is substantial given that on average only 13% of the working-age population is on temporary contracts.

We also show that properly accounting for out of steady state dynamics is important. Indeed, applying the standard decomposition from the literature, which relies on an assumption of fast convergence to steady state, leads to overestimating the contribution from permanent contracts and underestimating the contribution from temporary contracts. We think this point is relevant for existing studies that decompose the variability in unemployment in a European context (Silva and Vazquez-Grenno, 2013; Hairault et al., 2015).

References

Elsby, M. W., Hobijn, B., and Şahin, A. (2015). On the importance of the participation margin for labor market fluctuations. *Journal of Monetary Economics*, 72:64–82.

Elsby, M. W. L., Hobijn, B., and Şahin, A. (2013). Unemployment Dy-

namics in the OECD. *Review of Economics and Statistics*, 95(2):530–548.

Gomes, P. (2012). Labour market flows: Facts from the United Kingdom. *Labour Economics*, 19(2):165–175.

Hairault, J. O., Le Barbanchon, T., and Sopraseuth, T. (2015). The cyclicality of the separation and job finding rates in France. *European Economic Review*, 76:60–84.

Norris, J. (1997). *Markov Chains*. Cambridge University Press, Cambridge, United Kingdom.

Shimer, R. (2012). Reassessing the Ins and Outs of Unemployment. *Review of Economic Dynamics*, 15(2):127–148.

Silva, J. I. and Vazquez-Grenno, J. (2013). The Ins and Outs of Unemployment in a Two-Tier Labor Market. *Labour Economics*, 24:161–169.

Statistics Sweden (2005). *Flödesstatistik från AKU*.

Statistics Sweden (2011). *Basic Principles for Labour Force Surveys (LFS) and Labour Market Statistics*.

Statistics Sweden (2012). *Arbetskraftsundersökningarna*.

Figures
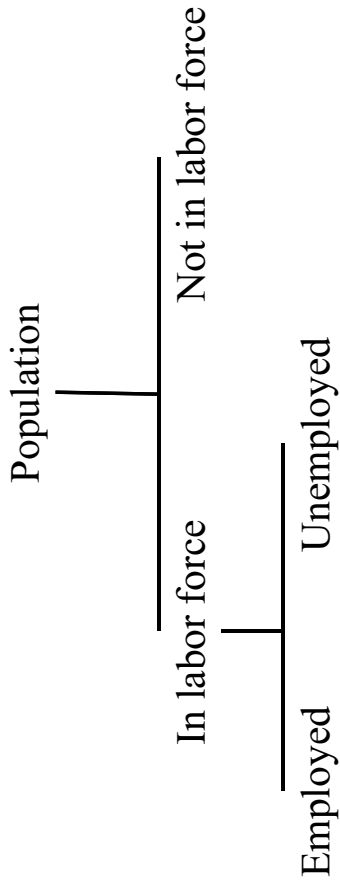
Figure 3.1: Basic classification of population in labor force survey
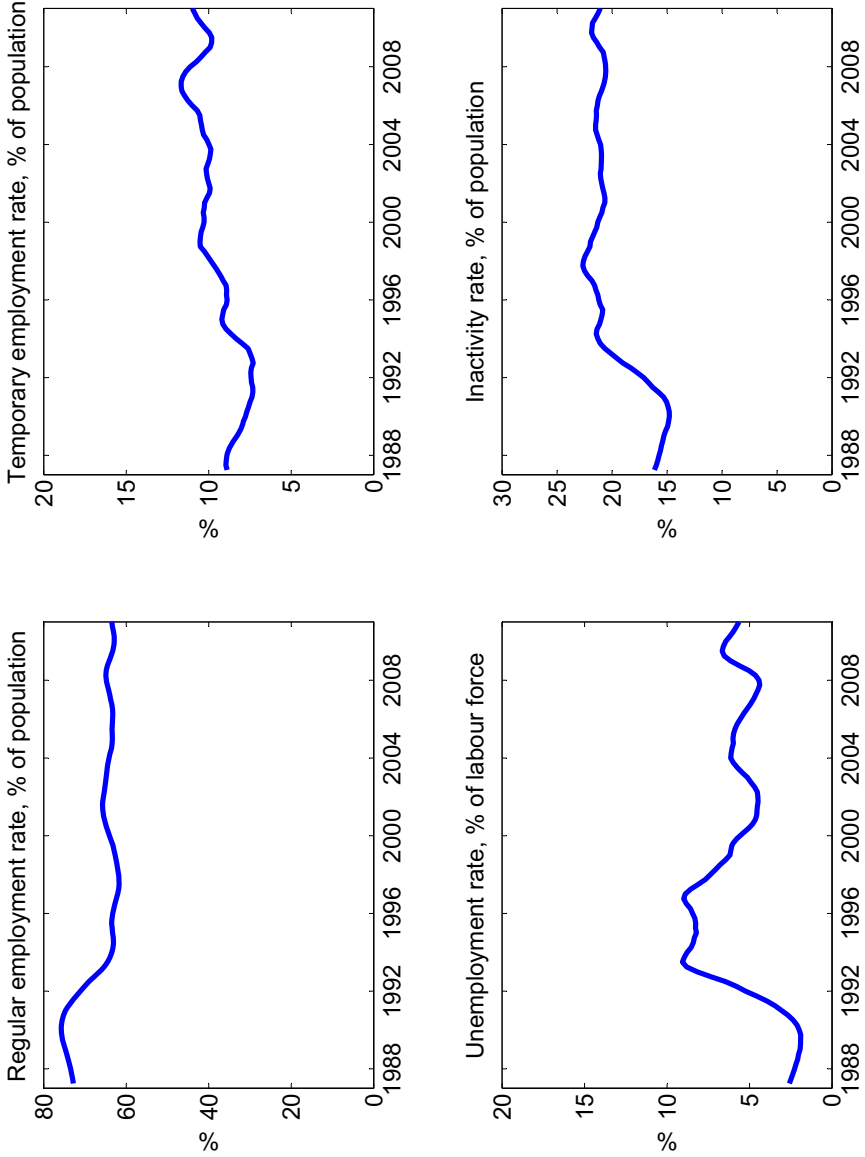
Figure 3.2: Labor market stocks across time
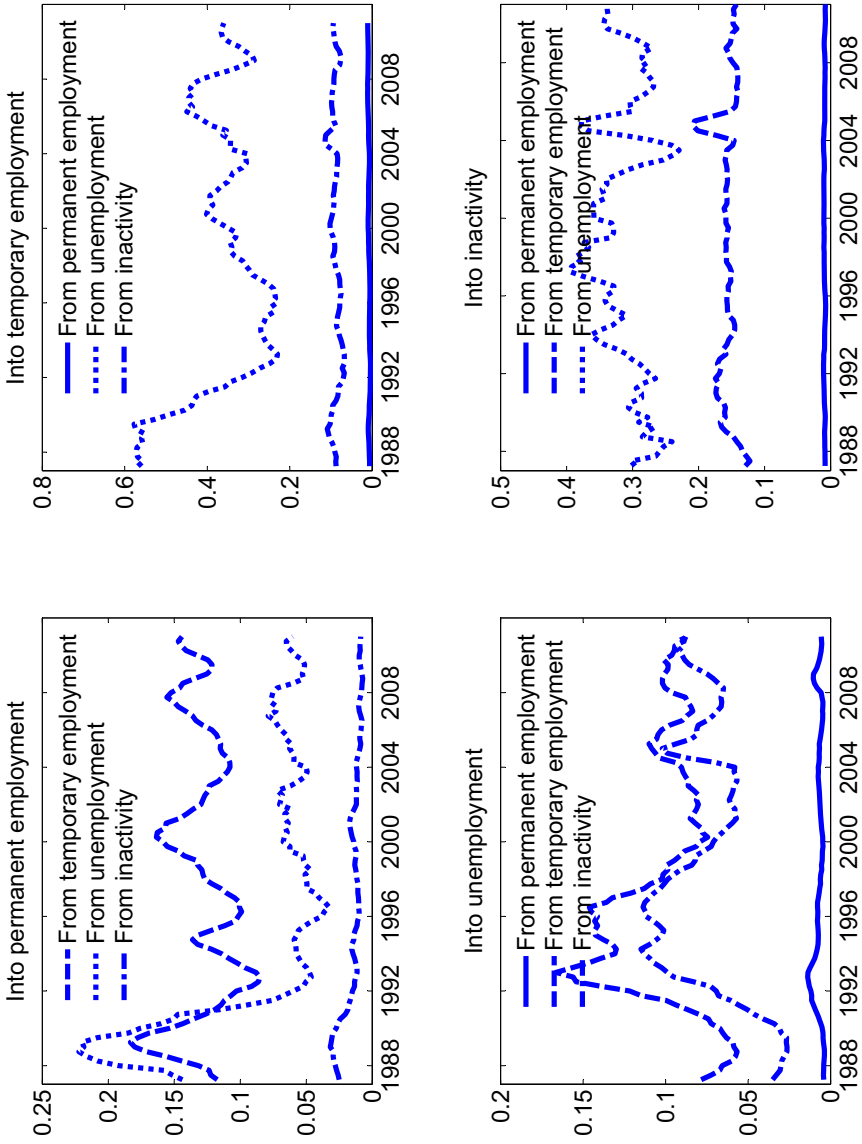
Figure 3.3: Hazard rates for transition across labor market states
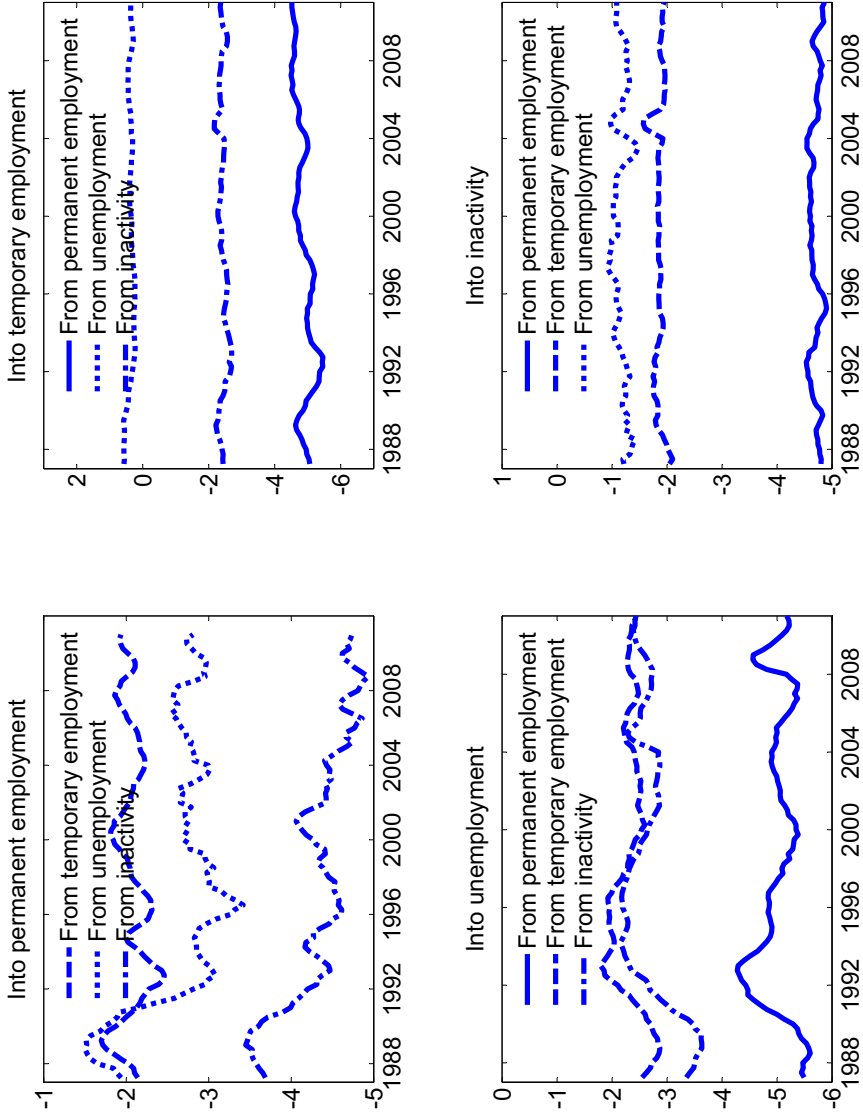
Figure 3.4: Hazard rates for transition across labor market states, logged
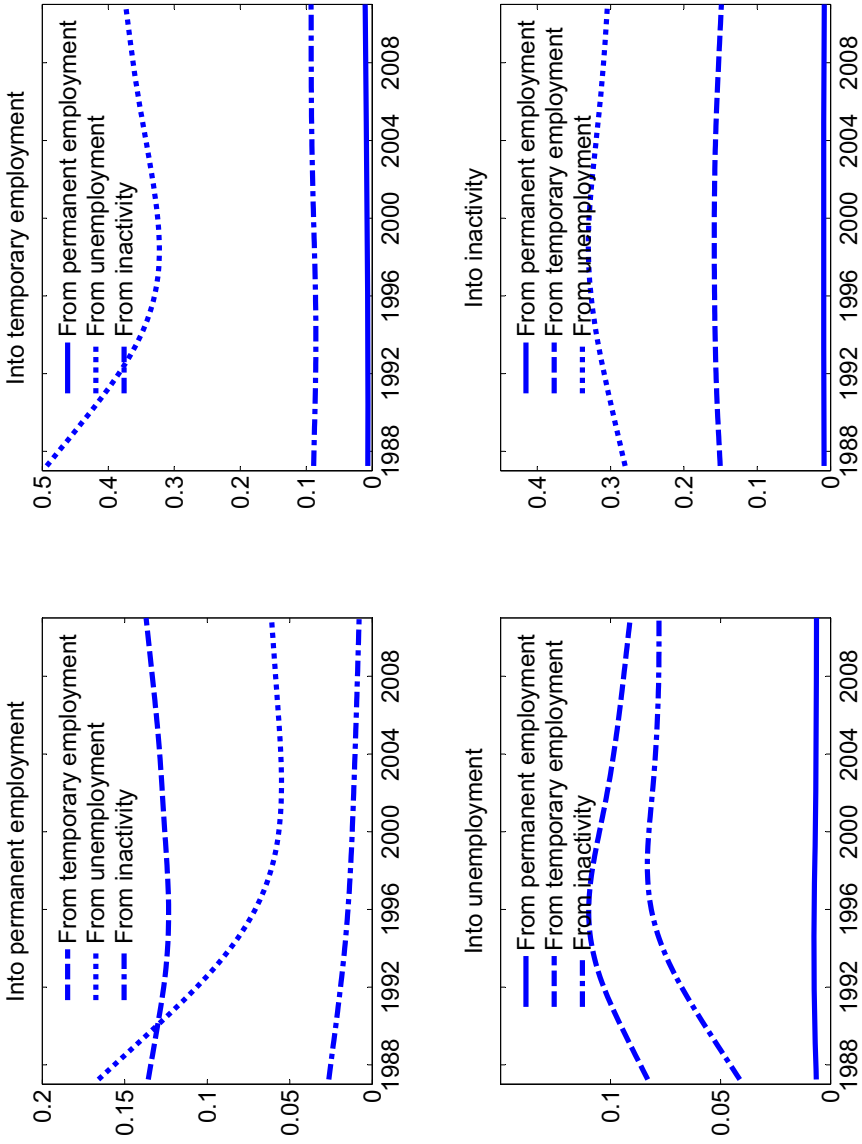
Figure 3.5: Trend flow rates computed using a HP-filter with $\lambda = 10^5$
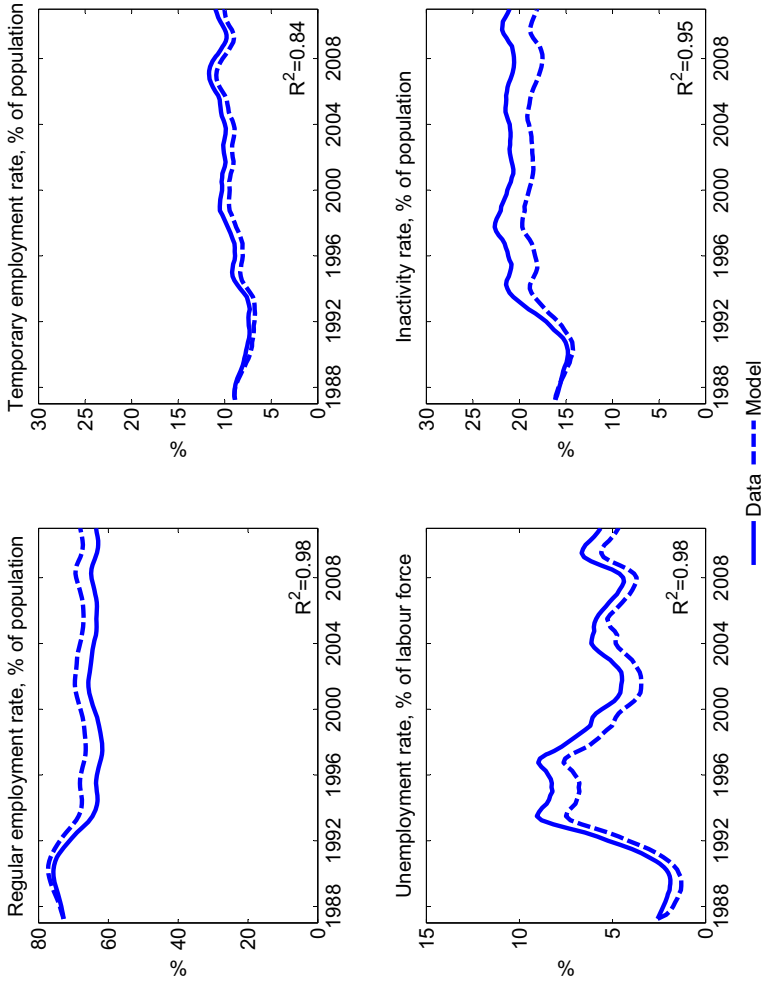
Figure 3.6: Actual and simulated labor market distribution, seasonally adjusted data
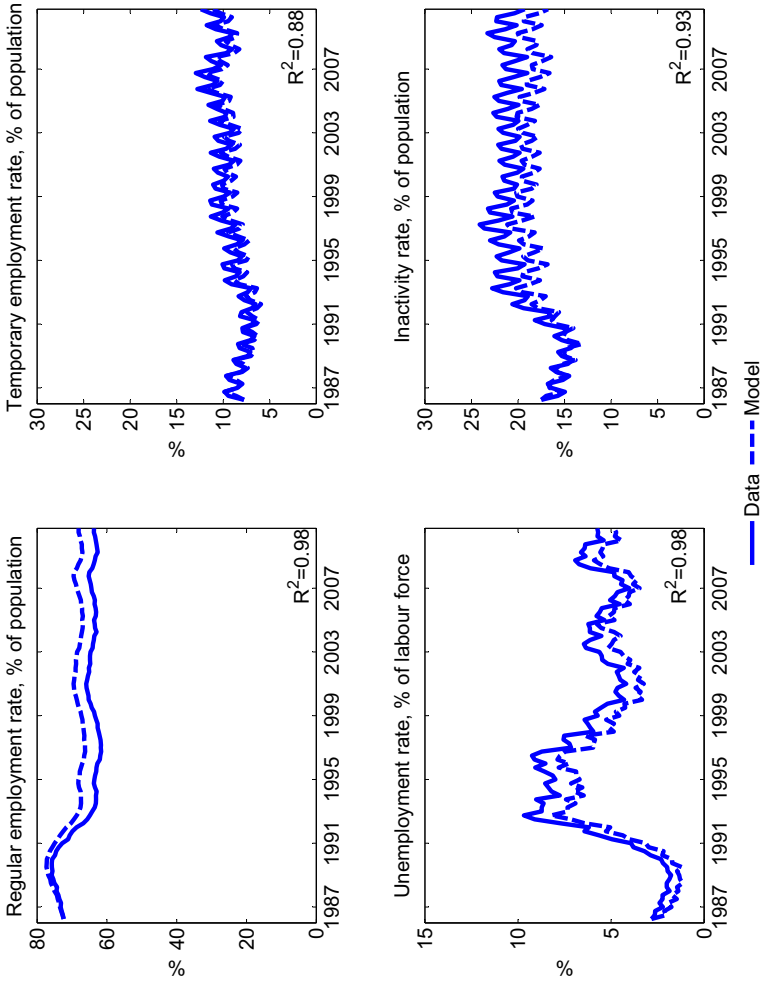
Figure 3.7: Actual and simulated labor market distribution, non seasonally adjusted data

Tables

Table 3.1: Illustration of the rotating panel structure

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
Month 0	$\mathbf{G_0}$	G_{-3}	G_{-6}	G_{-9}	G_{-12}	G_{-15}	G_{-18}	G_{-21}
Month 1	G_1	G_{-2}	G_{-5}	G_{-8}	G_{-11}	G_{-14}	G_{-17}	G_{-20}
Month 2	G_2	G_{-1}	G_{-4}	G_{-7}	G_{-10}	G_{-13}	G_{-15}	G_{-19}
Month 3	G_3	$\mathbf{G_0}$	G_{-3}	G_{-6}	G_{-9}	G_{-12}	G_{-14}	G_{-18}
Month 4	G_4	G_1	G_{-2}	G_{-5}	G_{-8}	G_{-11}	G_{-14}	G_{-17}
Month 5	G_5	G_2	G_{-1}	G_{-4}	G_{-7}	G_{-10}	G_{-13}	G_{-16}
Month 6	G_6	G_2	$\mathbf{G_0}$	G_{-3}	G_{-6}	G_{-9}	G_{-12}	G_{-16}
Month 7	G_7	G_4	G_1	G_{-2}	G_{-5}	G_{-8}	G_{-11}	G_{-14}
Month 8	G_8	G_5	G_2	G_{-1}	G_{-4}	G_{-7}	G_{-10}	G_{-13}
Month 9	G_9	G_6	G_3	$\mathbf{G_0}$	G_{-3}	G_{-6}	G_{-9}	G_{-12}
Month 10	G_{10}	G_7	G_4	G_1	G_{-2}	G_{-5}	G_{-8}	G_{-11}
Month 11	G_{11}	G_8	G_7	G_4	G_1	G_{-2}	G_{-5}	G_{-8}
Month 12	G_{12}	G_9	G_6	G_3	$\mathbf{G_0}$	G_{-3}	G_{-6}	G_{-9}

Notes: G_n is group of individuals who entered the survey in month n. Each group is surveyed with an interval of 3 months. In each month 7/8 of the sample has been surveyed before. 1/8 of the sample is surveyed for the first time.

Table 3.2: Cyclical variation in hazard rates

From \ To	Permanent emp.	Temporary emp.	Unemployment	Inactivity
Permanent emp.		−0.0599***	0.0546***	−0.0032
Temporary emp.	−0.4190***		0.5632***	0.0124
Unemployment	−1.7459***	−3.9956***		0.9163***
Inactivity	−0.1924***	−0.2939***	1.1750***	

Notes: The reported figure is the the coefficient of the unemployment rate in a regression of the relevant hazard rate logged. Time trends are included in the regression. *(**)[***] denotes significance on 10(5)[1] pct. level. Sample period is 1987Q1-2011Q3.

Table 3.3: Decomposition of the variance in Swedish unemployment rate, seasonally adjusted data, $\lambda = 100000$, 1987-2012

A. Non-steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.02	0.15	-0.00	0.13
TE	0.05	0.00	0.17	-0.00	0.21
U	0.13	0.22	0.00	-0.02	0.33
I	0.03	0.03	0.28	0.00	0.34
$\sum$	0.20	0.23	0.60	-0.03	1.00

B. Steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.05	0.18	0.00	0.14
TE	0.10	0.00	0.14	-0.00	0.24
U	0.15	0.19	0.00	-0.00	0.34
I	0.05	0.02	0.21	0.00	0.28
$\sum$	0.30	0.17	0.54	-0.00	1.00

C. Difference (panel B. - panel A.)

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.03	0.03	0.00	0.01
TE	0.06	0.00	-0.02	0.00	0.03
U	0.02	-0.03	0.00	0.02	0.01
I	0.02	-0.00	-0.07	0.00	-0.05
$\sum$	0.09	-0.06	-0.06	0.02	0.00

Notes: Panel A and B are computed using (3.7) and (3.11), respectively. PE: Permanent employment. TE: Temporary employment. U: Unemployment. I: Inactivity. The total variation has been normalized to the contribution from flows from/to these 4 states. $\lambda = 100000$ in HP-filter.

Source: Own calculations on data from Statistics Sweden

Table 3.4: Decomposition of the variance in Swedish unemployment rate, non seasonally adjusted data, $\lambda = 100000$, 1987-2012

A. Non-steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.02	0.15	-0.00	0.13
TE	0.05	0.00	0.17	-0.01	0.21
U	0.13	0.22	0.00	-0.03	0.31
I	0.03	0.03	0.30	0.00	0.35
$\sum$	0.20	0.23	0.61	-0.04	1.00

B. Steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.04	0.18	0.01	0.15
TE	0.09	0.00	0.15	0.01	0.24
U	0.15	0.22	0.00	-0.01	0.35
I	0.04	0.02	0.19	0.00	0.25
$\sum$	0.28	0.20	0.52	0.01	1.00

C. Difference (panel B. - panel A.)

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.02	0.03	0.01	0.03
TE	0.04	0.00	-0.02	0.01	0.04
U	0.02	-0.00	0.00	0.02	0.04
I	0.01	-0.01	-0.10	0.00	-0.10
$\sum$	0.08	-0.03	-0.09	0.05	0.00

Notes: Panel A and B are computed using (3.7) and (3.11), respectively. PE: Permanent employment. TE: Temporary employment. U: Unemployment. I: Inactivity. The total variation has been normalized to the contribution from flows from/to these 4 states. $\lambda = 100000$ in HP-filter.

Source: Own calculations on data from Statistics Sweden

Table 3.5: Decomposition of the variance in Swedish unemployment rate, $\lambda = 500000$, 1987-2012

A. Non-steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.02	0.13	-0.00	0.11
TE	0.04	0.00	0.16	-0.00	0.21
U	0.14	0.23	0.00	-0.03	0.34
I	0.03	0.02	0.29	0.00	0.35
$\sum$	0.21	0.24	0.59	-0.04	1.00

B. Steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.04	0.16	0.00	0.12
TE	0.09	0.00	0.14	0.00	0.24
U	0.16	0.20	0.00	-0.01	0.34
I	0.05	0.02	0.23	0.00	0.30
$\sum$	0.30	0.18	0.54	-0.01	1.00

C. Difference (panel B. - panel A.)

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.02	0.03	0.00	0.01
TE	0.05	0.00	-0.02	0.00	0.03
U	0.02	-0.03	0.00	0.02	0.01
I	0.02	-0.00	-0.06	0.00	-0.05
$\sum$	0.08	-0.06	-0.05	0.03	0.00

Notes: Panel A and B are computed using (3.7) and (3.11), respectively. PE: Permanent employment. TE: Temporary employment. U: Unemployment. I: Inactivity. The total variation has been normalized to the contribution from flows from/to these 4 states. $\lambda = 500000$ in HP-filter.

Source: Own calculations on data from Statistics Sweden

Table 3.6: Decomposition of the variance in Swedish unemployment rate, $\lambda = 1600$, 1987-2012

A. Non-steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.02	0.20	0.01	0.19
TE	0.03	0.00	0.17	0.00	0.20
U	0.11	0.18	0.00	0.01	0.30
I	0.01	0.02	0.27	0.00	0.31
$\sum$	0.15	0.19	0.64	0.03	1.00

B. Steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.06	0.23	0.02	0.18
TE	0.13	0.00	0.14	0.00	0.27
U	0.14	0.16	0.00	0.02	0.32
I	0.04	0.03	0.15	0.00	0.22
$\sum$	0.32	0.12	0.51	0.04	1.00

C. Difference (panel B. - panel A.)

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.04	0.03	0.01	-0.01
TE	0.11	0.00	-0.04	-0.00	0.07
U	0.04	-0.02	0.00	0.01	0.02
I	0.03	0.01	-0.12	0.00	-0.09
$\sum$	0.18	-0.06	-0.13	0.01	0.00

Notes: Panel A and B are computed using (3.7) and (3.11), respectively. PE: Permanent employment. TE: Temporary employment. U: Unemployment. I: Inactivity. The total variation has been normalized to the contribution from flows from/to these 4 states. $\lambda = 1600$ in HP-filter.

Source: Own calculations on data from Statistics Sweden

Table 3.7: Decomposition of the variance in Swedish unemployment rate, $\lambda = 100000$, 1987-2007

A. Non-steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.02	0.15	-0.00	0.13
TE	0.05	0.00	0.17	-0.00	0.21
U	0.13	0.22	0.00	-0.02	0.33
I	0.03	0.02	0.28	0.00	0.33
$\sum$	0.20	0.22	0.60	-0.03	1.00

B. Steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.05	0.18	-0.00	0.13
TE	0.10	0.00	0.15	-0.00	0.25
U	0.15	0.19	0.00	-0.01	0.34
I	0.05	0.02	0.22	0.00	0.29
$\sum$	0.30	0.17	0.54	-0.01	1.00

C. Difference (panel B. - panel A.)

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.03	0.03	-0.00	-0.00
TE	0.06	0.00	-0.02	0.00	0.04
U	0.02	-0.03	0.00	0.02	0.01
I	0.02	-0.00	-0.06	0.00	-0.05
$\sum$	0.10	-0.06	-0.06	0.02	0.00

Notes: Panel A and B are computed using (3.7) and (3.11), respectively. PE: Permanent employment. TE: Temporary employment. U: Unemployment. I: Inactivity. The total variation has been normalized to the contribution from flows from/to these 4 states. $\lambda = 100000$ in HP-filter.

Source: Own calculations on data from Statistics Sweden

Table 3.8: Decomposition of the variance in Swedish unemployment rate, $\lambda = 100000$, 1992-2012

A. Non-steady state method

From \ To	PE	TE	U	I	Σ
PE	0.00	-0.02	0.14	-0.01	0.11
TE	0.05	0.00	0.16	-0.01	0.20
U	0.13	0.26	0.00	-0.03	0.36
I	0.03	0.03	0.27	0.00	0.33
Σ	0.21	0.28	0.56	-0.05	1.00

B. Steady state method

From \ To	PE	TE	U	I	Σ
PE	0.00	-0.05	0.16	-0.00	0.11
TE	0.12	0.00	0.13	-0.00	0.25
U	0.15	0.22	0.00	-0.00	0.37
I	0.05	0.03	0.19	0.00	0.27
Σ	0.32	0.20	0.48	-0.01	1.00

C. Difference (panel B. - panel A.)

From \ To	PE	TE	U	I	Σ
PE	0.00	-0.03	0.02	0.00	0.00
TE	0.06	0.00	-0.02	0.01	0.05
U	0.03	-0.04	0.00	0.03	0.01
I	0.02	-0.00	-0.09	0.00	-0.06
Σ	0.11	-0.07	-0.08	0.04	0.00

Notes: Panel A and B are computed using (3.7) and (3.11), respectively. PE: Permanent employment. TE: Temporary employment. U: Unemployment. I: Inactivity. The total variation has been normalized to the contribution from flows from/to these 4 states. $\lambda = 100000$ in HP-filter.

Source: Own calculations on data from Statistics Sweden

Table 3.9: Decomposition of the variance in Swedish unemployment rate, $\lambda = 100000$, 1992-2007

A. Non-steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.02	0.14	-0.01	0.11
TE	0.05	0.00	0.16	-0.01	0.20
U	0.13	0.26	0.00	-0.03	0.36
I	0.03	0.03	0.27	0.00	0.33
$\sum$	0.20	0.27	0.57	-0.05	1.00

B. Steady state method

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.05	0.15	-0.01	0.10
TE	0.12	0.00	0.14	-0.00	0.25
U	0.16	0.22	0.00	-0.00	0.37
I	0.05	0.03	0.19	0.00	0.28
$\sum$	0.33	0.20	0.49	-0.01	1.00

C. Difference (panel B. - panel A.)

From \ To	PE	TE	U	I	$\sum$
PE	0.00	-0.03	0.02	0.00	-0.01
TE	0.07	0.00	-0.02	0.01	0.05
U	0.03	-0.04	0.00	0.03	0.02
I	0.02	-0.00	-0.08	0.00	-0.05
$\sum$	0.12	-0.07	-0.08	0.04	0.00

Notes: Panel A and B are computed using (3.7) and (3.11), respectively. PE: Permanent employment. TE: Temporary employment. U: Unemployment. I: Inactivity. The total variation has been normalized to the contribution from flows from/to these 4 states. $\lambda = 100000$ in HP-filter.

Source: Own calculations on data from Statistics Sweden

Appendix

Algorithm Description

In this section, we describe the algorithm we use to calculate the contribution terms for the non-steady state $\beta_{i,j,s}^{nonst.st.}$ and $\beta_{i,j}^{nonst.st.,lfp}$. We are interested in calculating

$$\beta_{i,j}^{nonst.st.,lfp} = \frac{Cov\left(\log\left(\frac{\mathbf{x}(t)_3}{\sum_{s=1}^3 \mathbf{x}(t)_s}\right) - \log\left(\frac{\hat{x}(t)_3}{\sum_{s=1}^3 \hat{\mathbf{x}}(t)_s}\right), \Gamma_{i,j,3,t} - \Gamma_{i,j,t}^{lfp}\right)}{Var\left(\log\left(\frac{\mathbf{x}(t)_3}{\sum_{s=1}^3 \mathbf{x}(t)_s}\right) - \log\left(\frac{\hat{x}(t)_3}{\sum_{s=1}^3 \hat{\mathbf{x}}(t)_s}\right)\right)}, \qquad (3.23)$$

where

$$\Gamma_{i,j,s,t} = \int_{u=0}^t \sum_{i,j\neq i} \frac{\partial \mathbf{x}(t)_s}{\partial Q(u)_{ij}} \frac{\hat{Q}(u)_{ij}}{\hat{\mathbf{x}}(t)_s} \left(\log Q_{ij}(u) - \log \hat{Q}_{ij}(u)\right) du$$

and

$$\Gamma_{i,j,t}^{lfp} = \Gamma_{i,j,1,t} \frac{\hat{x}(t)_1}{\sum_{s=1}^3 \hat{x}(t)_s} + \Gamma_{i,j,2,t} \frac{\hat{x}(t)_2}{\sum_{s=1}^3 \hat{x}(t)_s} + \Gamma_{i,j,3,t} \frac{\hat{x}(t)_3}{\sum_{s=1}^3 \hat{x}(t)_s}.$$

Discretization

The first step in implementing the algorithm is to rewrite the expression of the current labor market state in the discretized version as

$$x(t) = x(0) \exp(\sum_{u=0}^{t-1} Q_u),$$

where $Q_u = \int_{z=u}^{u+1} Q(z)dz$ is the average flow matrix over one period. With this expression, we express $\Gamma_{i,j,s,t}$ as

$$\Gamma_{i,j,s,t} = \sum_{u=0}^{t-1} \frac{\partial \mathbf{x}(t)_s}{\partial (Q_u)_{ij}} \frac{(\hat{Q}_u)_{ij}}{\hat{\mathbf{x}}(t)_s} \left(\log(Q_u)_{ij} - \log(\hat{Q}_u)_{ij}\right)$$

and similarly construct $\Gamma^{lfp}_{i,j,t}$. With this discrete approximation, we define the contribution of a flow $i \to j$ to variations in unemployment as

$$\hat{\beta}^{nonst.st.}_{i,j} = \frac{\widetilde{Cov}\left(\log\left(\frac{\mathbf{x}_{3,t}}{\sum_{s=1}^{3}\mathbf{x}_{s,t}}\right) - \log\left(\frac{\hat{\mathbf{x}}_{3,t}}{\sum_{s=1}^{3}\hat{\mathbf{x}}_{s,t}}\right), \Gamma_{i,j,3,t} - \Gamma^{lfp}_{i,j,t}\right)}{\widetilde{Var}\left(\log\left(\frac{\mathbf{x}(t)_3}{\sum_{s=1}^{3}\mathbf{x}(t)_s}\right) - \log\left(\frac{\hat{x}(t)_3}{\sum_{s=1}^{3}\hat{x}(t)_s}\right)\right)},$$

where $\widetilde{Cov}$ and $\widetilde{Var}$ denote sample covariances and variances.

Recursive formulation

To find $\Gamma_{i,j,s,t}$, we define it recursively over t. We write $\Gamma_{i,j,\cdot,t}$ for the row vector with elements $\Gamma_{i,j,s,t}$ and $s = 1, 2, 3, 4$.

Proposition 6. $\Gamma_{i,j,\cdot,t}$ *satisfies*

$$\Gamma_{i,j,\cdot,0} = \mathbf{0}$$
$$\Gamma_{i,i,\cdot,t} = 0 \quad \forall i = 1, 2, 3, 4$$

and

$$\Gamma_{i,j,\cdot,t} = \left.\frac{\partial \mathbf{x}_t}{\partial Q_{i,j,t-1}}\right|_{\hat{Q}} \cdot \left(\frac{1}{\hat{\mathbf{x}}_t}\right)\hat{Q}_{i,j,t-1} + \left[\Gamma_{i,j,\cdot,t-1} \cdot \hat{x}_{t-1} \cdot \left(\frac{1}{\hat{x}_t}\right)\right]\exp\left(\hat{Q}_{t-1}\right)$$

where $\frac{1}{\mathbf{v}}$ denotes a vector with elements $1/v_i$ and the dot $\cdot$ denotes element-by-element multiplication.

Proof. $\Gamma_{i,j,\cdot,0} = \mathbf{0}$ as it is an empty sum, and diagonal elements are zero by how we defined the derivative with respect to a flow matrix (adjusting the diagonal element to keep all row sums zero).

Furthermore, we can use that

$$x_t = x_{t-1}\exp\left(Q_{t-1}\right)$$

to obtain

$$\left.\frac{\partial x_t}{\partial Q_{i,j,u}}\right|_{\hat{Q}} = \left.\frac{\partial x_{t-1}}{\partial Q_{i,j,u}}\right|_{\hat{Q}}\exp(\hat{Q}_{t-1})$$

whenever $u < t - 1$. Using this expression, we obtain that whenever $u < t - 1$, we have

$$
\frac{\partial \mathbf{x}_t}{\partial Q_{i,j,u}} \cdot \left(\frac{1}{\hat{\mathbf{x}}_t}\right) \hat{Q}_{i,j,u} \left(\log(Q_{i,j,u} - \log(\hat{Q}_{i,j,u})\right) =
$$

$$
\frac{\partial \mathbf{x}_{t-1}}{\partial Q_{i,j,u}} \exp(\hat{Q}_t) \cdot \left(\frac{1}{\hat{\mathbf{x}}_t}\right) \hat{Q}_{i,j,u} \left(\log(Q_{i,j,u}) - \log(\hat{Q}_{i,j,u})\right) =
$$

$$
\left[\left(\frac{\partial \mathbf{x}_{t-1}}{\partial Q_{i,j,u}} \cdot \left(\frac{1}{\mathbf{x}_{t-1}}\right) \hat{Q}_{i,j,u} \left(\log(Q_{i,j,u}) - \log(\hat{Q}_{i,j,u})\right)\right) \cdot \mathbf{x}_{t-1}\right]
$$

$$
\cdot \exp(\hat{Q}_{t-1}) \cdot \left(\frac{1}{\hat{\mathbf{x}}_\mathbf{t}}\right).
$$

We can use this expression to calculate

$$
\begin{aligned}
\Gamma_{i,j,\cdot,t} &= \sum_{u=0}^{t-1} \frac{\partial \mathbf{x}_t}{\partial Q_{i,j,u}} \hat{Q}_{i,j,u} \cdot \left(\frac{1}{\hat{\mathbf{x}}_\mathbf{t}}\right) \left(\log(Q_{i,j,u}) - \log(\hat{Q}_{i,j,u})\right) \\
&= \frac{\partial \mathbf{x}_t}{\partial Q_{i,j,t-1}} \hat{Q}_{i,j,t-1} \cdot \left(\frac{1}{\hat{\mathbf{x}}_\mathbf{t}}\right) \left(\log(Q_{i,j,t-1}) - \log(\hat{Q}_{i,j,t-1})\right) \\
&+ \sum_{u=0}^{t-2} \frac{\partial \mathbf{x}_t}{\partial Q_{i,j,u}} \hat{Q}_{i,j,u} \cdot \left(\frac{1}{\hat{\mathbf{x}}_\mathbf{t}}\right) \left(\log(Q_{i,j,u}) - \log(\hat{Q}_{i,j,u})\right) \\
&= \frac{\partial \mathbf{x}_t}{\partial Q_{i,j,t-1}} \hat{Q}_{i,j,t-1} \cdot \left(\frac{1}{\hat{\mathbf{x}}_\mathbf{t}}\right) \left(\log(Q_{i,j,t-1}) - \log(\hat{Q}_{i,j,t-1})\right) \\
&+ \left[\left(\sum_{u=0}^{t-2} \frac{\partial \mathbf{x}_{t-1}}{\partial Q_{i,j,u}} \cdot \left(\frac{1}{\mathbf{x}_{t-1}}\right) \hat{Q}_{i,j,u} \left(\log(Q_{i,j,u}) - \log(\hat{Q}_{i,j,u})\right)\right) \cdot \mathbf{x}_{t-1}\right] \exp(\hat{Q}_{t-1}) \cdot \left(\frac{1}{\hat{\mathbf{x}}_\mathbf{t}}\right) \\
&= \frac{\partial \mathbf{x}_t}{\partial Q_{i,j,t-1}} \hat{Q}_{i,j,t-1} \cdot \left(\frac{1}{\hat{\mathbf{x}}_\mathbf{t}}\right) \left(\log(Q_{i,j,t-1}) - \log(\hat{Q}_{i,j,t-1})\right) \\
&+ (\Gamma_{i,j,\cdot,t-1} \cdot \hat{\mathbf{x}}_{\mathbf{t-1}}) \cdot \exp(\hat{Q}_{t-1}) \cdot \left(\frac{1}{\hat{\mathbf{x}}_\mathbf{t}}\right),
\end{aligned}
$$

which completes the proof.

Implementation in arrays

We use the following for arrays

$$
\begin{aligned}
\mathbf{x} &= S \times (T+1) \\
\hat{\mathbf{x}} &= S \times (T+1) \\
Q &= S \times S \times (T+1) \\
\hat{Q} &= S \times S \times (T+1) \\
\Gamma &= S \times S \times S \times (T+1) \\
\Gamma^{lfs} &= S \times S \times (T+1).
\end{aligned}
$$

We observe/derive $\mathbf{x}, \hat{\mathbf{x}}, Q, \hat{Q}$. We then initialize $\Gamma_{i,j,s,0} = 0$ for all i, j, s. We iterate on t and for each $t = 1, \ldots, T$ we define that diagonal elements are zero:

$$
\Gamma_{i,i,s,t} = 0 \ \forall \quad i = 1, 2, 3, 4.
$$

For each $i \neq j$ we define a row vector $\Gamma_{i,j,\cdot,t}$ by

$$
\begin{aligned}
\Gamma_{i,j,\cdot,t} &= \hat{x}_{\cdot,t} \left. \frac{\partial \exp(Q_{\cdot,\cdot,t-1})}{\partial (Q_{i,j,t-1})} \right|_{Q_{t-1} = \hat{Q}_{t-1}} \cdot \left(\frac{1}{\hat{\mathbf{x}}_{\mathbf{t-1}}} \right) \hat{Q}_{i,j,t-1} \left(\log(Q_{i,j,t-1}) - \log(\hat{Q}_{i,j,t-1}) \right) \\
&+ (\Gamma_{i,j,\cdot,t-1} \cdot \hat{\mathbf{x}}_{\cdot,\mathbf{t-1}}) \exp(\hat{Q}_{t-1}) \cdot \left(\frac{1}{\hat{\mathbf{x}}_{\cdot,\mathbf{t}}} \right)
\end{aligned}
$$

Lastly, we define

$$
\Gamma^{lfp}_{i,j,t} = \Gamma_{i,j,1,t} \frac{\hat{x}(t)_1}{\sum_{s=1}^{3} \hat{x}(t)_s} + \Gamma_{i,j,2,t} \frac{\hat{x}(t)_2}{\sum_{s=1}^{3} \hat{x}(t)_s} + \Gamma_{i,j,3,t} \frac{\hat{x}(t)_3}{\sum_{s=1}^{3} \hat{x}(t)_s}.
$$

and we can calculate the required sample covariances.

Chapter 4

Supply Chain Risk and the Pattern of Trade[*]

4.1 Introduction

This paper is motivated by a growing concern in the policy and business community over supply chain risk. The 2011 tsunami in Japan illustrates how important reliability is for modern production. For example, General Motors had to close a factory in Louisiana due to a lack of Japanese-made parts.[1] The Inter-American Development Bank notes that "firms fragmenting production internationally are likely to look for locations with adequate transport and logistics infrastructure to reduce disruptions in the supply chain" (Blyde, 2014). Similarly, the US Department of Commerce argues that "Expected gains from offshoring can often be

[1]Source: http://www.nytimes.com/2011/03/20/business/20supply.html (last accessed Nov 17th, 2015)

erased by [...] unexpected delays."[2] In this paper, we study the relationship between supply chain risk and observed trade patterns.

Countries vary in the degree of supply chain risk. Some countries offer high-quality infrastructure and predictable, quick bureaucratic services. In other countries, poor logistics systems and low government effectiveness increase supply chain risk: components get stuck in the port; roads rain away; land rights are not transparent; import permits are delayed, and foreign currency availability is uncertain.

Variation in country-level supply chain risk induces comparative advantage when goods vary in their risk sensitivity. Consider two industries, plain t-shirt production and cars. For t-shirts, there are few separate intermediate inputs used in production. In this case, a risky supply chain is not too problematic. On the other hand, a modern car factory based on lean production principles requires a continuous flow of hundreds—if not thousands—of customized components. In car manufacturing, supply chain reliability is crucial. If a country's infrastructure and its institutions create severe supply chain risk, the country can be expected to have a comparative advantage in t-shirt production, and a comparative disadvantage in modern car production.

We formalize this intuition by constructing a model in which each sector produces a final good using intermediate inputs. Intermediate input production is subject to *disruption risk,* which means that production (including delivery) fails with some positive probability. Proximate causes of production failures are infrastructure problems, strikes, political instability, and unpredictable bureaucratic procedures.

Some inputs are pre-committed before uncertainty is resolved, and these inputs are lost if complementary inputs are not delivered. In this way, upstream supply disruptions damage downstream productivity. It is an important assumption that some inputs are pre-committed. Indeed, without precommitment, upstream idiosyncratic supply chain disruptions would not have downstream consequences as inputs could be real-

---

[2]http://www.esa.doc.gov/economic-briefings/assess-costs-everywhere-shipping (last accessed Nov 17th, 2015)

located to production units which had not experienced supply chain disruptions. The combination between supply chain risk and pre-committed inputs damages productivity.

A key feature of the model is that the effect of disruption risk depends on whether intermediate goods are *standardized* or *customized.* Standardized inputs are homogenous and traded on centralized exchanges, whereas customized inputs are delivered directly from input producers to final goods producers. For standardized inputs, the centralized exchange insulates producers from disruption risk through a law of large numbers. There will be a steady supply of goods even if some suppliers fail for idiosyncratic reasons. This is not the case for customized inputs, and upstream disruptions in customized input production are transmitted to dowstream producers.

We derive a novel aggregation result that makes the model highly tractable. We show that aggregate supply and demand of a sector can be characterized by a representative firm with deterministic production, even though the underlying firms experience stochastic shocks. Supply chain risk enters the sectoral production function as a productivity penalty. As all customized intermediate inputs are essential for production, this productivity penalty grows exponentially with the number of customized intermediate inputs used in production.[3] The theory also suggests a measure of industry risk sensitivity. Because each customized intermediate good represents an independent source of error, the risk sensitivity of a product depends on the *number of customized components* used in production.

We embed this sectoral production structure in a simple trade model. In the model, we let goods vary in their number of customized inputs m, and let countries vary in the disruption risk π. We show that productivity is log-submodular in π and m. Thus, we can use the insight from Costinot (2009a) to show that there will be negative sorting between π and m. In other words, risky countries (high π) will produce goods with few

---

[3]We assume that input disruption shocks are independent.

customized intermediate inputs (low m).

In the empirical part of the paper, we test this hypothesis in trade data using the methodology in Romalis (2004). In a first step, we use input-output tables and the definition of customized goods developed by Rauch (1999) to construct a measure of how many customized intermediate inputs each industry uses. To proxy for disruption risk, we use the World Governance Indicator (WGI) for government effectiveness and the World Bank's Logistics Performance Index (LPI). We then test whether countries with more effective government and logistics systems export relatively more goods with a large number of customized components. We show that this effect exists and is statistically and economically meaningful. The effect is present in a wide range of specifications, even when we (over-) control for country income levels.

The effects we find are somewhat smaller but of a similar order of magnitude as other institutional determinants of trade patterns, for example contracting quality (Nunn, 2007). However, our theory builds on a different mechanism. Nunn (2007) emphasizes that a bad contracting environment leads to a higher cost of customized intermediate inputs via lower levels of relation-specific investments. Therefore, he measures the *proportion in value terms* of inputs that comes from customized intermediate inputs. Our paper focuses on disruption risk and therefore considers the *number* of customized intermediate inputs.

Different perspectives on the sources of comparative advantage also imply different policy levers. Many countries are actively trying to attract "sophisticated" industries such as advanced electronics manufacturing. If relationship-specific investments are key—as implied by models such as Antràs (2003) and measured by Nunn (2007)—then the main task for governments is to improve the quality of the contracting environment and the rule of law. If supply chain risk matters—as suggested by our results—then it is also important for governments to improve the reliability of the business environment through, for example, more predictable bureaucracy and infrastructure.

Our theory also has implications for the measurement of the quality

of the business environment. Few existing indicators focus on uncertainty and risk. For example, the World Bank's Doing Business Indicators have been used widely to illustrate the challenges for businesses in poor countries. It measures the de jure time and cost to complete a wide range of functions such as the time to export, import, receive electricity, and open a business. However, for most tasks, the indicator only provides a single estimate per country of the time required for a given task, and it contains little information about the variability of its implementation. We stress the importance of risk and uncertainty in the business environment. Our findings suggest that characterizations of the business environment should include measures of risk. For example, surveys would benefit from reporting not only the average time to obtain a permit, but also the variance associated with the time to obtain such a permit and the risk of not obtaining a permit at all.

We discuss related literature in Section 4.2. Section 4.3 provides the model. We bring the model to the data in Section 4.4. Section 4.5 concludes the paper.

4.2 Literature

The paper connects to a number of different literatures. The production structure in which all components are vital for production relates to Kremer's O-Ring theory of production (Kremer, 1993). Furthermore, we analyze how institutional features interact with risk to shape countries' trade patterns. Therefore, the paper also adds to the literature on institutional sources of comparative advantage. Lastly, it contributes to a new literature analyzing the role of supply chain disruptions in production networks.

O-Ring theory and sequential production Our production process features a number of vital inputs that are necessary for production. Thus, the most closely related model is the O-Ring Theory proposed by Kremer (1993). He analyzes a production process where all tasks have to be

performed succesfully for production to be succesful. In his analysis, low human capital plays a similar role as supply chain unreliability plays in our setting.

A paper that applies an O-Ring like mechanisms to trade is Costinot (2009b). He proposes a model of comparative advantage where firms trade off the value of specialization against the risk of disruption when they select team sizes. Disruption comes from poorly enforced contracts. As the gains from specialization stems from economizing on fixed training costs, his definition of sensitivity is the total training cost for workers in an industry. In contrast, we have a different source of disruption risk, and therefore focus on the number of specialized inputs as the measure of risk sensitivity.

More generally, production processes in which all components are vital are related to sequential production, in which goods have to pass through a number of pre-defined steps. Economists have long noted the potential implications of sequential production, and also analyzed trade patterns in the context of sequential production models. Dixit and Grossman (1982) is an early attempt of analyzing the role of sequential production in shaping trade patterns. More recently, Costinot et al. (2013) and Antràs and Chor (2013) have proposed novel models of sequential production and used them to interpret sorting along global supply chains.

Institutional sources of comparative advantage   We analyze how variations in government effectiveness and logistics systems quality shape trade patterns. This connects the paper to the growing literature on institutional determinants of comparative advantage. Existing work on the institutional determinants of comparative advantage focuses on the role of technological differences Eaton and Kortum (2002), factor endowments (Romalis, 2004), contracting quality (Nunn, 2007; Antràs, 2003; Antras and Helpman, 2004), financial development (Manova, 2013), or labor market institutions (Cuñat and Melitz, 2012). Nunn and Trefler (2014) provide a recent survey of the literature.

In particular, Levchenko (2007) treats institutional quality as a source

of comparative advantage. In the model, he focuses on imperfect contract enforcement in the spirit of Grossman and Hart (1986). A contrast to our model is that we posit that institutions determine the amount of risk that firms face, which in turn shapes comparative advantage. Blyde and Molina (2015) provides evidence that foreign direct investment is related to logistics infrastructure. Similar to our paper, he stresses that production of complex goods is challenging when the environment is risky.

Disruptions in production networks A number of recent papers have analyzed the effect of disruption in production networks. Carvalho et al. (2016) analyze the production network consequences of the 2011 Great East Japanese Earthquake. Barrot and Sauvagnat (2016) provide evidence for a key assumption in our paper, namely that upstream disruption shocks are particularly damaging when upstream producers produce specialized goods.

4.3   Model

We construct a trade model where intermediate input production is risky. Countries vary in their degree of supply chain risk and goods vary in their risk sensitivity. This generates specialization across countries according to comparative advantage. We first develop a parsimonius characterization of production with risky inputs. For each sector, we derive a sector level aggregate production function, which summarizes how supply chain risk and industry characteristics interact to determine sector level productivity. We then use these sectoral production functions in a trade model to characterize how supply chains shape trade patterns.

A sector s consists of a continuum of final goods producers which produce a good using labor and multiple intermediate inputs. The final goods producers combine intermediate inputs using a CES aggregator where inputs are gross complements. Therefore, every input is essential for production. Intermediate inputs are produced using labor, and the

production process in the intermediate goods sector is risky. This means
that for each intermediate good producer, there is a possibility that pro-
duction—or delivery—will fail, and failures are independent across dif-
ferent suppliers.

For intermediate inputs, the model makes a distinction between stan-
dardized and customized intermediate inputs. Standardized intermediate
inputs are traded on a centralized market, and all input producers ship
to this market. Idiosyncratic delivery risks average out through a law
of large numbers and there is a deterministic flow of products to the
centralized market. This means that the final goods producers face no
delivery risk for standardized intermediate inputs despite production and
delivery risk for input producers. The situation is different for customized
intermediate inputs. Here, each final goods producer matches with a spe-
cific customized input producer and pre-commits to use this particular
supplier. If there is a production disruption with this supplier, the final
goods producer will not get anything of that particular input. Figure 4.1
illustrates the market structure.[4]

Figure 4.1: Model structure
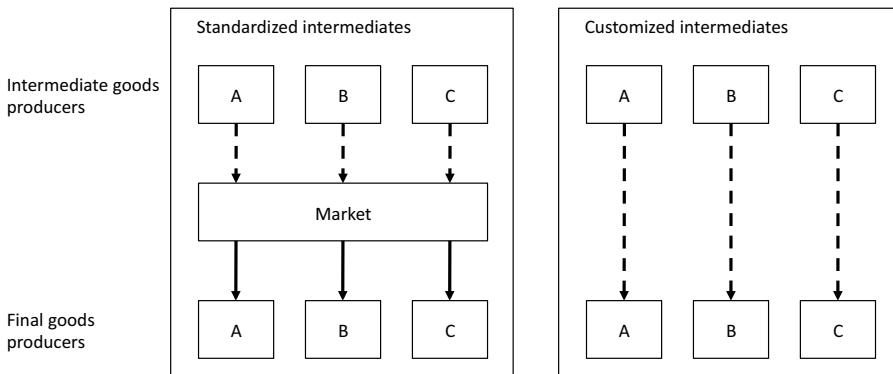
With this production and market structure, final goods production

---

[4]Our distinction between customized and standardized inputs has received empir-
ical support by Barrot and Sauvagnat (2016) who show that downstream producers
are particularly negatively affected by upstream shocks when inputs are customized.

succeeds only when all customized intermediate inputs are succesfully delivered. We define the failure probability π and assume that failures are independent. Then, production succeeds with probability $(1 - \pi)^{m_s}$ where m_s is the number of customized intermediate inputs in sector s. We assume that labor supply and other customized input supplies are pre-committed before the resolution of production risk. Hence when production fails, those inputs are wasted. One insight in the model is that there need to be pre-committed costs to generate large costs of supply chain disruptions. If costs were not pre-committed, production factors would be re-allocated to firms which did not suffer supply chain disruptions, which would limit the negative consequences of supply chain disruptions.

Even though the model features idiosyncratic risk, we show that aggregate output and aggregate labor demand of every sector can be summarized by an optimizing representative firm. This representative firm has a deterministic production function which is linear in labor. The supply chain risk appears as a productivity penalty proportional to $(1 - \pi)^{m_s(1-\gamma)+\gamma}$. where γ is the cost share of standardized intermediate goods. The interpretation of the productivity penalty is that supply chain risk confers a $(1 - \pi)^{m_s}$ penalty on the productivity of labor and customized intermediate inputs, as they are pre-commited but only utilized when production succeeds. There is a $1-\pi$ productivity penalty on standardized intermediate inputs due to production and delivery risk, but the centralized market means that this effect is not amplified by downstream effects in the production chain. Combining these two penalties using the factor shares yields the aggregate productivity penalty.

Once we have characterized each sector using a representative firm, we can build a trade model that incorporates supply chain risk. We create a world economy in which sectors vary in their number of intermediate inputs m_s and countries vary in their degree of supply chain risk π_c. We represent the production technology of each country-sector pair using the previously derived representative firm. This gives us a trade model with country-industry-specific productivity penalties $(1 - \pi_c)^{m_s(1-\gamma)+\gamma}$.

We note that these productivity terms are *log-submodular* in π_c and m_s. It is well-known in trade theory that there is a a close connection between log-submodularity in productivity and negative sorting, and we prove that our model indeed features negative sorting between π_c and m_s. Countries with high supply chain risk will specialize in goods with a low number of customized inputs.

We abstract from risk mitigation in this model. In practice, firms can mitigate supply chain risk by holding inventory, having redundancy in the supply chain, and by making costly investment in speeding up delayed processes. However, even if firms could mitigate supply chain risks, the trade pattern consequences would be similar if there is a fixed cost associated with mitigation per input: for example, if there is a fixed costs of contracting with an extra supplier to have redundancy, or in setting up an inventory for a particular input. The functional form of the relation between productivity, number of customized inputs, and supply chain risk would change if there were mitigation possibilities, but the sorting prediction only depends on a positive interaction between costs of supply chain disruptions and the number of customized inputs.

In Section 4.3.1, we set up the production environment for a sector and derive a representative firm to characterize the sector's aggregate behavior. In Section 4.3.2, we insert these sectors into a trade model and derive the pattern of specialization.

4.3.1   Sector level supply function

A sector s features a unit interval of final goods producers $j \in [0, 1]$ (we will suppress this subscript when we talk about firm behavior). Final goods production requires labor, a composite of customized intermediate inputs X, and a composite of standardized intermediate inputs Z for production.

It will be important to distinguish between variables that are determined before and after the resolution of production and delivery uncertainty. In particular, the realized intermediate input supplies will be

stochastic as they depend on the realization of a collection of production and delivery shocks. We will use the convention to put a tilde ($\sim$) on top of variables to denote stochastic variables that are determined after the resolution of uncertainty. The production function is given by

$$\tilde{y} = \kappa l^{\alpha} \tilde{X}^{\beta} \tilde{Z}^{\gamma}, \quad \alpha + \beta + \gamma = 1$$

We introduce the normalization $\kappa = \alpha^{-\alpha} \beta^{-\beta} \gamma^{-\gamma} (1 - \gamma\pi) m^{\frac{\beta}{1-\eta}} n^{\frac{\gamma}{1-\eta}}$ for notational convenience. The composite intermediate goods are produced according to

$$
\begin{aligned}
\tilde{X} &= \left(\sum_{i=1}^{m} \tilde{x}_i^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \\
\tilde{Z} &= \left(\sum_{i=1}^{n} \tilde{z}_i^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}.
\end{aligned}
$$

Timing matters as firm decisions can take place before or after the resolution of uncertainty. In our model, firms decide on labor use and customized input orders before the resolution of uncertainty. They decide on standardized input purchases after the resolution of uncertainty. Our choice of timing is motivated by considering the possibility of reallocating inputs in case of input delivery failure. We think it is reasonable that labor is difficult to reallocate quickly, and customized goods orders involve pre-commitment as the producer needs to specialize a production batch to a particular buyer. In contrast, for standardized inputs with deep markets, it is reasonable that inputs can be reallocated from firms with disruptions to those without disruptions relatively easily. Hence, the firm first decides on labor input l and customized input orders x_i^f. This particular assumption is not crucial for our conclusions regarding trade patterns, but it is important that some costs are pre-committed as supply chain disruption otherwise would not create costs for downstream buyers.

From the point of view of a firm, labor has a pre-determined wage w

and the firm gives a take it or leave it offer to customized intermediate input producers to pay $p_i^x x_i^f$ in case of successful delivery.[5] After the resolution of uncertainty, the final goods firm decides how much of the standardized intermediate inputs to buy. We denote this quantity $\tilde{z}_i^f$ to emphasize that it is a stochastic choice variable depending on the realization of production disruption shocks. The firm pays p_i^z per unit of standardized goods.[6] We assume that firms behave competitively in the standardized input market and that they can buy an arbitrary amount of goods at the prevailing price p_i^z. There is no delivery uncertainty, and in equilibrium p_i^z will adjust to clear the market. Taken together, the firm solves

$$\max_{l, x_i, \tilde{z}_i} \quad \mathbb{E}\left(P\tilde{y} - wl - \sum_{i=1}^m p_i^x \tilde{x}_i - \sum_{i=1}^m p_i^z \tilde{z}_i\right) \quad s.t. \quad y = \kappa l^\alpha \tilde{X}^\beta \tilde{Z}^\gamma,$$

and subject to the customized input supplier accepting the offer.

We simplify this expression in steps to clarify the optimization problem. We first note that the randomness can be reduced to two cases: either all customized inputs arrive or at least one is missing. When a customized input is missing, production will fail ($\tilde{y} = 0$) regardless of the purchased amount of standardized inputs. Clearly, the firm will then choose not to buy any standardized inputs. Thus, there is only one state of the world in which the firm buys standardized inputs: when all custumized input goods arrive. We write z_i without a tilde ($\sim$) to denote the purchased amount of standardized inputs in this case. As all failures of

---

[5]We place all the bargaining strength on the buyer side and we do not introduce any contracting frictions. These assumptions can be relaxed to analyze the interaction between contracting frictions and production uncertainty. It is a non-consequential assumption that firms only pay when delivery is successful as firms are risk neutral, but if firms were risk averse, the pricing scheme would embody some form of risk-sharing. This notion could be useful to analyze the selection of payment terms in international trade. The choice of writing total payment as $p_i(j)x_i^f(j)$ is only an inconsequential reparametrization of total payments $T_i(j)$.

[6]In a general version we would write $\tilde{p}_i^z$ to denote that the price of standardized inputs is determined after the realization of production shocks, but in this case there are no aggregate production shocks, and the price will be independent of the realized shocks with probability one.

customized goods are independent and happen with probability π, the probability that all deliveries will suceed is $(1-\pi)^m$. We can rewrite the optimization problem as

$$\max_{l^f, x_i, z_i} (1-\pi)^m Py - wl - \sum_{i=1}^{m} p_i^x (1-\pi)x_i - \sum_{i=1}^{n}(1-\pi)^m p_i^z z_i,$$

subject to production being given by $y = \kappa l^\alpha X^\beta Z^\gamma$. Here, $(1-\pi)$ in the x_i-terms stems from our assumption that firms only pay customized goods suppliers upon succesful delivery, and X, Z are the values taken by $\tilde{X}, \tilde{Z}$ when there are no disruptions. With this formulation, we can derive the relative demand for different factors using standard methods. We can note that the presence of supply chain risks bias firms to have relatively more standardized inputs versus customized inputs than implied by their price ratio p_i^z/p_i^x. The intuition is that supply chain risks bias firms away from pre-commited inputs, as these are wasted in case of supply chain disruptions.

$$\frac{x_i}{l} = \frac{\beta}{\alpha} \frac{1}{m(1-\pi)} \frac{w}{p_i^x}$$

$$\frac{z_i}{l} = \frac{\gamma}{\alpha} \frac{1}{n(1-\pi)^m} \frac{w}{p_i^z}$$

The customized intermediate input sector has a linear production function in labor. When they employ labor l_i^x, they produce output l_i^x with probability $1-\pi$ and zero output with probability π. The firms obtains an order x_i for which it is paid $p_i^x x_i$ upon delivery and 0 otherwise. Conditional on producing, it is always optimal for the firm to employ x_i units of labor to fill the order exactly. Firms can also choose not to produce at all. Thus, they choose between accepting or not accepting an order. They solve

$$\max_{x_i' \in \{0, x_i\}} x_i' \left((1-\pi)p_i^x - w\right).$$

Just as customized intermediate input producers, standardized interme-

diate input producers have linear production functions in labor and pro-
duction is risky. Thus, they employ l_i^z workers and produce l_i^z goods with
probability $1 - \pi$ and 0 goods with probability π. When successful, they
sell their output to the centralized market at price p_i^z. Producers choose
$l_i^z \geq 0$ to maximize their expected profit

$$\Pi_i^z = p(1 - \pi)l_i^z - wl_i^z \tag{4.1}$$

We analyze a single sector which will be inserted into a trade model.
Therefore, our primary interest is how the sector's aggregate labor de-
mand and aggregate output vary with prices. That is, we are interested
in

$$
\begin{aligned}
Y &= \int_0^1 \tilde{y}(j)dj \\
L &= \int_0^1 l(j)dj + \sum_{i=1}^m \int_0^1 l_i^x(j)dj + \sum_{i=1}^n \int_0^1 l_i^z(j)dj
\end{aligned}
$$

and how they depend on the final goods price P and wages w. Our
main result is that the sector's aggregate behavior can be described by
a representative firm where supply chain risk enters as a productivity
term. We first define the aggregate net supply of the sector $\mathcal{S}_{sto}(P, w)$
as the set of sector outputs and labor demands that are consistent with
profit maximization for some intermediate good prices. More formally,
a pair of output and labor demand (Y,L) belongs to the net supply
correspondence $\mathcal{S}_{sto}$ if we can find some intermediate input prices, order
quantities, and labor demands such that:

- The quantities and labor demands are optimal for both final goods
  producers and intermediate goods producers given intermediate
  intermediate input prices and aggregate prices P and w

- Total production of final goods is Y

$$Y = \int_0^1 \tilde{y}_j dj = (1 - \pi)^m y \quad a.s.$$

- Total labor demand from final and intermediate good producers is
  L

$$\int_0^1 \left(l(j) + \sum_{i=1}^m l_i^x(j) + \sum_{i=1}^n l_i^z(j) \right) dj = L \quad a.s.$$

- Standardized goods markets clear almost surely

$$\int_0^1 \tilde{z}_i(j)dj = \int_0^1 l_i^z(j) \mathbb{I}\left(success_i(j) = 1\right) dj \quad a.s. \quad i = 1, \dots, n.$$

Here $success_i(j)$ is an indicator variable taking value 1 if there is no disruption for firm j in standardized intermediate input sector i. Exploiting the fact all firms behave symmetrically, we can write the labor demand equation and the market clearing equation for standardized inputs as

$$
\begin{aligned}
l + \sum_{i=1}^m l_i^x + \sum_{i=1}^n l_i^z &= L \quad a.s. \\
(1-\pi)^m z_i &= (1-\pi) l_i^z \quad a.s.
\end{aligned}
\tag{4.2}
$$

The market clearing condition in the standardized input markets (4.2) is non-standard. The left-hand side reflects that only a fraction $(1-\pi)^m$ of firms demands standardized input goods, whereas the right-hand side reflects that only a fraction $(1-\pi)$ of all standardized input producers are successful in their production. In all equilibrium equations, we use the formulation almost surely (a.s.) because the statement is probabilistic. There do exist events where the market does not clear, but due to a law of large numbers, these events have probability zero.

Now we want to show that this sector aggregate supply correspondence $\mathcal{S}_{sto}$ is identical to the aggregate supply correspondence of a representative firm with a linear deterministic production function

$$Y = (1-\pi)^{m(1-\gamma)+\gamma} L \tag{4.3}$$

The intuition behind this representative firm is that there is a $(1-\pi)^m$ probability that a firm will produce. For customized inputs and labor

input, the productivity penalty is $(1 - \pi)^m$ as they are pre-committed. For standardized intermediate inputs, the productivity penalty is just $(1 - \pi)$ as firms do not pre-commit to use them. Given that the shares of labor, customized, and standardized intermediate inputs are α, β, γ, we obtain an aggregate productivity penalty

$$[(1 - \pi)^m]^\alpha [(1 - \pi)^m]^\beta [(1 - \pi)]^\gamma = (1 - \pi)^{m(1-\gamma)+\gamma}$$

using the fact that $\alpha + \beta + \gamma = 1$. Given our proposed representative firm, the profit of the firm is given by

$$P(1 - \pi)^{m(1-\gamma)+\gamma}L - wL$$

and we define the supply correspondence $\mathcal{S}_{rep}$ of the representative firm as all pairs Y and L that are consistent with profit maximization for the representative firm. More formally, (Y, L) belongs to $\mathcal{S}_{rep}$ if L maximizes profit and $Y = (1 - \pi)^{m(1-\gamma)+\gamma}L$. We can now state our representative firm theorem:

Proposition 7. *(Representative Firm) The aggregate behavior of a sector can be described by representative firm, i.e.*

$$S_{sto}(P, w) = S_{rep}(P, w) \quad \forall P, w > 0$$

Moreover, when $w/P = (1 - \pi)^{m(1-\gamma)+\gamma}$, both the sector supply correspondence $S_{sto}(P, w)$ and the representative firm supply correspondence $S_{rep}(P, w)$ are given by

$$Y = (1 - \pi)^{m(1-\gamma)+\gamma}L, \quad L \geq 0.$$

When $w/P < (1 - \pi)^{m(1-\gamma)+\gamma}$, both correspondences are empty as there is no finite labor demand consistent with optimization. When $w/P > (1 - \pi)^{m(1-\gamma)+\gamma}$, both correspondences are $\{(0,0)\}$ as zero production is the only firm choice consistent with optimization.

Proof. See appendix.   □

This result means that we can use the representative firm's production function to analyze the aggregate behavior of a sector. Provided we find a general equilibrium featuring prices P, w, and aggregate sectoral output and labor demand Y and L, we can find intermediate input prices and micro-level firm behavior which is optimal given P, w and produces the aggregate outcome (Y, L). Conversely, there is no micro-behavior that is consistent with optimization and produces other aggregate outcomes than $\mathcal{S}_{rep}$. Therefore, without loss of generality, we can assume that sectoral production is represented by (4.3) when we analyze trade patterns.

4.3.2   Trade model

In this section, we use the representative firm from Section 4.3.1 to derive trade patterns with risky supply chains. We posit a world economy in which industries differ in the number of customized intermediate inputs m and countries differ in terms of risk levels π. Under these conditions, we show that high-π countries will produce low-m goods.

There are k industries $m_1 < m_2 < \cdots < m_k$ indexed by the number of customized intermediate inputs. All goods have a common number n of standardized intermediate inputs and common intermediate input shares for standardized inputs γ. There is a continuum of countries indexed by production risk $\pi \in [\underline{\pi}, \bar{\pi})$ with common labor supplies L. The production function for good m_j in country π is given by

$$Y_{\pi,j} = (1 - \pi)^{m_j(1-\gamma)+\gamma} \, \ell_{\pi,j} \tag{4.4}$$

and in each country, the representative firm in each sector maximizes profits

$$\Pi_{\pi,j} = p_j \, (1 - \pi)^{m_j(1-\gamma)+\gamma} \, \ell_{\pi,j} - w_\pi \ell_{\pi,j}.$$

Consumers in country π maximize

$$U(c_{\pi,1}, \ldots, c_{\pi,k}) \quad s.t. \quad \sum_{i=1}^{k} c_{\pi,j} p_j \leq w_\pi L,$$

where U is strictly concave and satisfies the Inada conditions.

Equilibrium

An equilibrium in the economy consists of prices p_j, wages, w_π, labor allocation $l_{\pi,j}$, production $Y_{\pi,j}$, and consumption $c_{\pi,j}$ such that

- The labor allocation maximizes firm profits

- Output is given by the production function: $Y_{\pi,j,} = (1-\pi)^{m_j + (1-\alpha)} \ell_{\pi,j}$

- Firms make zero profits

$$\Pi_{\pi,j} \leq 0$$
$$\Pi_{\pi,j} = 0 \quad if \ l_{\pi,j} > 0$$

- Goods and labor markets clear

$$\int_{\underline{\pi}}^{\bar{\pi}} Y_{\pi,j} d\pi = \int_{\underline{\pi}}^{\bar{\pi}} c_{\pi,j} d\pi \quad \forall j = 1, \ldots k$$
$$\sum_{j=1}^{k} \ell_{\pi,j} = L \quad \forall \pi \in [\underline{\pi}, \bar{\pi})$$

- If good m_j is produced in country π, there exists δ such that m_j is produced in all countries $\pi' \in [\pi, \pi + \delta)$. This assumption is technical and ensures that the function assigning countries to goods is right-continuous (see Costinot et al. (2013) for the use of a similar assumption).

We are interested in how countries sort according to comparative advantage. The following proposition describes the equilibrium allocation.

Proposition 8. *(Unique Equilibrium) There exists a unique equilibrium. It features k cutoff points*

$$\underline{\pi} = \pi_k < \pi_{k-1} < \cdots < \pi_1 < \pi_0 = \bar{\pi}$$

such that

$$\ell_{\pi,j} > 0 \quad if \ \pi \in [\pi_j, \pi_{j-1})$$
$$l_{\pi,j} = 0 \quad if \ \pi \notin [\pi_j, \pi_{j-1})$$

Proof. (see appendix) □

Proposition (8) states that unreliable (high-π) countries produce goods with few customized intermediate inputs (low-m). This is the prediction that we take to the data.

4.4 Empirical evidence

In this section, we test our model of comparative advantage using country-industry export data. We follow the standard methodology in the empirical comparative advantage literature (Romalis, 2004; Nunn, 2007) and estimate the equation

$$log\,(x_{ig}) = \beta\,(r_i \times n_g) + \mu_i + \theta_g + \epsilon_{ig} \tag{4.5}$$

Here, $x_{i,g}$ denotes country i's exports in industry g, r_i is a measure of risk and n_g is the risk sensitivity of industry g. We include country and industry fixed effects, μ_i and θ_g, respectively. Any country level variable that is common to all industries is subsumed in the country fixed effect. Importantly, this includes the total exports of the country. The industry fixed effects capture cross-industry effects that are common across countries. For example, exports are generally higher for goods that are easy to ship and that have a high expenditure share among consumers. Therefore, the coefficient β measures the *tilt* in countries'

trade pattern towards certain industries depending on country-industry characteristics. The interpretation is the same as in Romalis (2004).

The logic of the specification can be illustrated with an example. Suppose, for the sake of argument, that there are two industries, electronics and cement production. The former is highly sensitive to disruptions while the latter is relatively robust. Assume further that there are two countries, a large safe and a small risky country. First, we might expect the large country to have higher exports in both industries. The country fixed effect takes this into account. Second, we might assume that electronics are generally more traded than cement. The industry fixed effect takes this into account. Any remaining effect then ideally reflects the interaction of industry and country variables. The safe country is expected to export more electronics than cement, since electronics are risk-sensitive. This is the effect that the coefficient β aims to captures.

We adopt the convention that high values of r_i correspond to high reliability (low risk), and our theoretical prediction is therefore $\beta > 0$: countries with high scores on reliability measures specialize in industries that are sensitive to risk.

4.4.1   Data Sources and Concordances

To measure trade flows, we use the BACI dataset which is compiled by CEPII and based on the COMTRADE data (Gaulier and Zignago, 2010). We use total value of exports for each country in each HS 2007 six digit level industry. We use data for 2012. To categorize inputs as specialized vs customized we use Rauch's classification into goods which are traded on exchange, goods which are referenced in a trade journal, and goods which are neither (Rauch, 1999). For measurement of government properties, we use the World Banks' World Governance Indicators (WGI) (Kaufmann et al., 2011). The logistics quality is measured by the World Bank's Logistics Performance Index (Arvis, 2010). GDP and country factor endowments are obtained for 2011 data in the Penn World Table 8.1 (Feenstra et al., 2015). For GDP, we use expenditure-side real

GDP at chained PPPs in million 2005 USD. We measure the capital stock per worker by dividing the total capital stock at current PPPs in millions of 2005 dollars with the number of engaged persons measured in millions. For human capital we use an index of human capital provided by the PWT constructed based on years of schooling (Barro and Lee, 2013) and returns to schooling (Psacharopoulos, 1994). Our measures of number of inputs and their contract sensitivity are taken from the 2007 US Input-Output tables published by the BEA[7]. To measure capital and skill intensity across different industries we use the NBER CES database (Bartelsman and Gray, 1996). Capital intensity is defined as the total value of capital divided by total payroll (dividing by payroll instead of number of workers give an approximation of human capital instead of physical labor input). The skill intensity of an industry is defined as the ratio of non-production payroll to total payroll.

Table 4.1: Data Sources and Industry Classifications

Dataset	Code
NBER CES	NAICS 1997 6-digit
IO-table	IO 2007 6-digit
Rauch	SITC rev.2 4-digit
BACI	HS 2007 6-digit

Table 4.1 provides a list of the industry level codes for the various datasets. The regressions are performed in NAICS 2012 6-digit and we use a set of concordances to map our industry level variables into NAICS 2012 6-digit. We use a concordance between HS 2007 10-digit and and NAICS 2007 6-digit to convert the trade data to NAICS 2007 6-digit. We use a procedure where trade flows coded in HS 2007 6-digit are allocated equally to all 10-digit extensions, and these are then mapped to NAICS 2007 6-digit code. We create chains of concordances from NAICS 2007 to NAICS 2002 and NAICS 1997 to convert the capital and skill intensities to NAICS 2007, and use the trade flows coded in NAICS 2007 6-digit to

---

[7]http://www.bea.gov/industry/io_benchmark.htm (last acessed Nov 24th, 2015)

create the weights used in these concordances. We create a concordance
from NAICS 2007 to NAICS 2012 to convert all data into NAICS 2012.

For the customization variable, we use a concordance between SITC
rev.2. 4-digit and HS 2007 6-digit to convert the variable into HS 2007
6-digit. Again we use the trade data now in HS 2007 6-digit to create
the weighting scheme. We then map to IO 2007 via NAICS 2007 6-
digit as we use the Rauch data in the IO-table to calculate industry
characteristics. The IO-data together with the Rauch variables are then
mapped to NAICS 2012 via NAICS 2007.

In the Appendix, we describe in detail which sources we use for the
concordances, how concordances are weighted, and how the weights are
used in the transformations.

4.4.2   Measuring Products' Sensitivity to Unreliability

Motivated by our theory, we propose a novel measure of industries' risk
sensitivity. In the model, we distinguish between standardized and cus-
tomized components. Standardized components are traded in liquid mar-
kets. As a consequence, final goods producers are not materially affected
by idiosyncratic supply failures. By contrast, customized components
cannot be replaced easily and, therefore, the final goods producers is
exposed to the risk that an component cannot be sourced. This could
be due to outright failure of a supplier or the failure of a port author-
ity, bad infrastructure, and so forth. Furthermore, as all components are
gross complements, a non-zero amount of each components is essential
to production. Hence, the *number of customized components* that an
industry uses is an appropriate measure of risk sensitivity.[8]

We classify components as customized using the methodology devel-
oped by Rauch (1999). For each industry, he records if a good is traded
on an exchange or reference-priced in a trade journal. We define a com-
ponent as customized if it belongs to an industry which Rauch records

---

[8]Our model does not feature mitigation of supply chain risk. A similar result
obtains if there is a fixed cost of mitigation per distinct input.

as neither traded on an exchange nor reference-priced. Using the US input-tables, we look up the list of other industries that a given industry buys from and count the number of those industries that are customized according to Rauch. Tables 4.1 and 4.2 list the top and bottom thirty industries sorted by the number of customized components. Our risk sensitivity measure leads to an intuitive classification of most industries. Motor vehicle components and semiconductor production, for example, are classified as sensitive to risk, whereas farming and cement manufacturing are not. Industry-level variables are visualized in Figure 4.1b.

We infer the number of inputs used from industry level data. To the extent that firms are heterogeneous, this introduces a problem of aggregation. Consider two firms in the same industry that use 50 inputs each. If the firms' business models are not exactly identical, only 30 out of 50 might be the same for the two firms. On aggregate, however, we would observe the industry using 70 inputs, despite each firm using only 50. To protect us against the extreme case when a very small fraction of firms in an industry uses a particular input, we re-estimate the main regression excluding input industries which contributes less than 0.1% and 0.01% of total intermediate input value. The results are robust against this modification. Moreover, as long as this shortcoming is similar across industries it will not affect our results, which is based on the ranking of industries.[9]

Our measure of sensitivity to unreliability can justifiably be called complexity as it denotes how many specialized components a product uses. Thus, we can contrast it with our proposed measures of complexity in the literature. Nunn (2007) develops one such measure. He also uses the Rauch (1999) measure of product differentiation. He measures an industry's sensitivity to contract quality as the share of total input

---

[9]Further problem that we have not addressed is firstly that an industry can use multiple components from a single industry. This would lead to an underestimation of the number of inputs. Lastly, there can be a problem in that the fineness of the IO-table classification is endogenous to the US production structure, which might mean that there is a bias in that US-concentrated industries appear to have more inputs because of how the IO-table is subdivided.

value that comes from customized inputs. The motivation behind his measure is that cost-saving investments in specialized goods production are relation specific, and will be provided less if contract protection is poor. In light of this, it is reasonable to use the proportion of component costs as lack of relation-specific investments can plausibly be expected to increase costs proportionally. While it is theoretically motivated to weigh the customized intermediate good content by value in the context of his study, our model suggests an independent role for the *number* of components that are customized.

Another measure that has been used to capture complexity is one minus the Herfindahl index of input suppliers. The Herfindahl index is a concentration index of an industry's input suppliers. It is high if an industry's intermediate good demand is skewed towards few industries. This measure of complexity is used in Blanchard and Kremer (1997) and Levchenko (2007). Levchenko (2007) explicitly discusses why they choose to use the Herfindahl index instead of the number of intermediate inputs: "If intermediate input use is dominated by one or two inputs (high concentration), and all the other intermediates are used very little, then what really matters to the final good producer is the relationship it has with the largest one or two suppliers." Our theory suggests that all suppliers of customized inputs might matter independent of size. Former Apple executive Tony Fadell illustrated this point well when the Japanese tsunami threatened to disrupt global supply chains: "lacking some part, even if it costs just dimes or a few dollars, can mean shutting down a factory".[10]

In Figure 4.1a, we compare our measure to the measure proposed in Nunn (2007). Generally, the correlation is strong and positive. Both measures classify Automobile Manufacturing as risk-sensitive and contract-intensive. On the other end of the spectrum, Soybean Farming is classified as neither risk-sensitive and nor contract intensive. However, there are some differences as well. Classification differs for the

---

[10]http://www.nytimes.com/2011/03/20/business/20supply.html?pagewanted=2&_r=0 (last accessed: November 19th, 2015)

Figure 4.1: Comparing Complexity Measures

In this figure, we compare our measure of risk sensitivity—the number of non-substitutable inputs—to two measures that have been used in the literature. In the first panel, we compare our measure to contract intensity as defined by Nunn (2007). In the second panel, we compare our measure to the industry Herfindahl as defined by Blanchard and Kremer (1997) and Levchenko (2007). All measures are calculated at 6-digit level and standardized. We calculate trade-weighted averages at 3-digit level (printed in bold). We omit Petroleum and Coal Products Manufacturing (ID 324) to improve visibility but include it when calculating the line of best fit.

(a) Number of Non-Substitutable Inputs vs. Contract Intensity

(b) Number of non-substitutable inputs vs. Herfindahl

textile-related industries . Nunn's 2007 measure classifies textile-related industries (NAICS 313, 314, 315, and 316; see bottom-right area in the graph) as complex, whereas our measure categorizes textile-related industries as non-complex.

In Figure 4.1b, we compare our measure to the Herfindahl measure used in Blanchard and Kremer (1997) and Levchenko (2007). The two measures are strongly correlated and tend to classify broad industries in similar ways. A notable exception is the transportation sector (NAICS 336), which our measure tends to classify as more risk-sensitive than than the measure based on the Herfindahl Index.

4.4.3   Measuring Countries' Reliability

We are interested in measuring disruption risk in different countries. In this context, we need to take a stand on likely causes of production and delivery disruption. For this, we focus on two country characteristics: logistics systems quality and overall government effectiveness. The motivation for including the quality of logistics system is clear: disruption is more likely if third-party logistics providers have low quality, goods clear customs slowly, and transportation infrastructure is subject to frequent failures.[11] We also include government effectiveness which we define as the quality of bureaucratic procedures and government provided services. We include this firstly as red tape is another possible cause of supply chain disruptions. Disruption risks in this area include delays in permits for starting production, or delays in permits for bringing in inputs and foreign worker. It also captures poorly functioning bureaucracy in customs, as well as uncertain land rights. The quality of government

---

[11]In the current model, intermediate input suppliers are all domestic, which means that the final goods supplier does not get through customs to obtain intermediate inputs. However, even in cases where you only source domestically, we believe it is plausible that customs problems affect reliability through its effect on your intermediate input suppliers. Explicit modeling of this channel would involve intermediate good trade and bilateral delivery risks which are not in the current model. Formally showing how customs risk interacts with intermediate goods trade is an interesting area of further research.

provided services is important as failures in electricity, water supply and infrastructure are sources of potential supply chain disruptions.

When it comes to measurement, we proxy logistics systems quality with the World Bank's Domestic Logistics Performance Index (Arvis, 2010). The index is based on surveys with global freight forwarders and express carriers, and combines it with quantitative measures of some components of supply chains. As of 2014 it encompasses 160 countries. For bureaucratic quality, we use the Government Effectiveness-measure from the Worldwide Governance Indicators (Kaufmann et al., 2011). It is an aggregated measure derived from a large number of measures including the quality of bureaucracy, extent of red tape, infrastructure quality, and the quality of various government provided services. Figure 4.1a visualizes the distributions of the main country-level variables in our data.

4.4.4   Results

In Table 4.2, we present the main results for the baseline specification (equation 4.5). We are interested in the interaction of industries' risk sensitivity measured by the number of customized components (Cust. Inp.) and countries' reliability. Our two preferred measures of country reliability are government effectiveness (Gov. Effectiveness) and logistics performance (LPI). We report interactions with three additional World Governance Indicators: regulatory quality (Reg. Quality), political stability and absence of violence (Stability), and control of corruption (Corruption). All these indicators proxy for an environment that is amenable to the production of risk-sensitive products.

The results are consistent with the hypothesis that risk-sensitive industries are disproportionately produced by reliable countries. Consider an industry that is one standard deviation above the mean in terms of risk sensitivity (Cust. Inp.). Increasing a country's government effectiveness (Gov. Effectiveness) by one standard deviation is associated with 10.4% (column 1) more exports in this industry, compared to a coun-

try with an average Logistics Performance Index. Increasing a country's
Logistics Performance Index (LPI) by one standard deviation is associ-
ated with 10.1% (column 6) more exports in this industry, compared to
a country with an average Logistics Performance Index. The coefficients
are of very similar magnitude for the other institutional variables that we
use to proxy for reliability. The main coefficient is statistically significant
at the 1% level for all measures considered.

4.4.5   Relationship to other results in the literature

As previously discussed, Nunn (2007) tests whether contracting quality
affects the pattern of trade. Might we just be capturing the effect that
stable countries also tend to have good contracting environments? In
Table 4.2, we replicate Nunn's main result (column 3). Countries with
high scores on the rule of law index (Rule of Law) tend to export contract-
intensive goods (Contract int.). Given that both our country-measures
(rule of law vs. government effectiveness) and our industry measures
(contract intensity vs. risk sensitivity) are correlated, our main result in
column 1 might be spurious. However, as we show in column 6, the two
estimates remain quantitatively similar and significant when analyzed
jointly (column 6). This result suggests that our mechanism is distinct
from the role of contracting. If we compare our quantitative effect to the
one found in Nunn (2007), ours is somewhat smaller at 10% compared
to Nunn's 28% in his baseline specification.

   Of course, the two explanations are not mutually exclusive. In fact,
when component producers fail to deliver a component on time they
typically also violate a contract. However, in Nunn, poor contracting is
analyzed in terms of relationship-specific investments and we consider
our theory as an additional explanation for observed trade patterns. The
distinction matters since the policy implications differ: Nunn (2007) im-
plies that countries can attract sophisticated industries by improving
contract enforcement. Our story suggests that a complementary policy
lever is the reduction of supply chain risk.

Table 4.2: Baseline Regression

This table presents estimates of the main specification (equation 4.5). The unit of observation country-industry pairs. The outcome variable is the natural logarithm of total trade volume. The variable Cust. Imp. refers to the industry's number of customized inputs, which is the measure of risk sensitivity implied by our model. risk sensitivity is interacted with country-characteristics. Our two preferred measures of country reliability are government effectiveness (Gov. Effectiveness) and logistics performance (LPI). We report interactions with four additional World Governance Indicators for robustness: regulatory quality (Reg. Quality), political stability and absence of violence (Stability), and control of corruption (Corruption). Standard errors are double-clustered at the country and industry level. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

	Dependent Variable: Log Exports				
	(1)	(2)	(3)	(4)	(5)
Cust. Imp. x Gov. Effectiveness	0.108*** (0.029)				
Cust. Imp. x Reg. Quality		0.110*** (0.029)			
Cust. Imp. x Stability			0.098*** (0.028)		
Cust. Imp. x Corruption				0.087*** (0.025)	
Cust. Imp. x LPI					0.105*** (0.029)
Industry FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Observations	35,458	35,458	35,458	35,458	35,458
R^2	0.815	0.815	0.815	0.815	0.815
Adjusted R^2	0.813	0.813	0.813	0.813	0.813
Residual Std. Error	1.653	1.653	1.654	1.654	1.653

Table 4.3: Comparison with Numn (2007)

This table compares our estimates to Numn (2007), who posits that countries with good legal systems (ruleoflaw) specialize in contract-intensive goods (numn). Contract intensity is defined as the share of inputs that are neither traded on exchanges nor reference priced. Our main result is provided in column 1. We replicate the main result of Numn (2007) in column 4. In column 6, we estimate the effect of both Contract int. ×Rule of Law and Cust. Imp ×Gov. Effectiveness at the same time. Standard errors are double-clustered at the country and industry level.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

			Dependent Variable: Log Exports			
	(1)	(2)	(3)	(4)	(5)	(6)
Cust. Imp. x Gov. Effectiveness	0.108***				0.085***	
	(0.029)				(0.028)	
Cust. Imp. x Rule of Law		0.100***				
		(0.026)				
Contract Int. x Gov. Effectiveness			0.103***		0.061**	0.159
			(0.030)		(0.028)	(0.115)
Contract Int. x Rule of Law				0.094***		−0.055
				(0.028)		(0.107)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,458	35,458	35,458	35,458	35,458	35,458
R²	0.815	0.815	0.815	0.815	0.816	0.815
Adjusted R²	0.813	0.813	0.813	0.813	0.813	0.813
Residual Std. Error	1.653	1.653	1.653	1.654	1.652	1.653

Table 4.4: Heckscher-Ohlin Controls and Interaction with GDP

This table conducts additional robustness checks. Column (1) reports our benchmark results. In column (2) we estimate specialization with Heckscher-Ohlin variables similarly toRomalis (2004): skill abundance (ln_hl) interacted with industries' skill intensity as well as capital abundance interacted with capital intensity. In column (3), we interact our measure of complexity (Cust. Inp.) with the logarithm of countries' per capita income. In column (4), we re-estimate our benchmark specification while also controlling for the income interaction. In column (5), we also add Heckscher-Ohlin controls. Standard errors are double-clustered at the country and industry level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

	Dependent Variable: Log Exports					
	(1)	(2)	(3)	(4)	(5)	(6)
Cust. Inp. x Gov. Effectiveness	0.108***	0.100***		0.086***		0.103***
	(0.029)	(0.028)		(0.032)		(0.031)
Skill Intensity x Skill Abundance		0.473***		0.471***		0.475***
		(0.162)		(0.161)		(0.161)
Capital Intensity x Capital Abundance		0.006		0.007		0.006
		(0.022)		(0.022)		(0.023)
Cust. Inp. x log(GDP/Capita)			0.082***	0.015		
			(0.025)	(0.028)		
Cust. Inp. x Skill Abundance					0.370***	−0.022
					(0.130)	(0.131)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,458	35,458	35,458	35,458	35,458	35,458
R^2	0.815	0.816	0.815	0.816	0.815	0.816

Table 4.5: Additional Robustness Checks

This table conducts additional robustness checks. Column (1) reports our benchmark results. In column (2) we estimate specialization allowing there to be country specific effects on the export of capital-intensive goods, skill-intensive goods, and contracting intensive goods as defined by Numn (2007). Standard errors are double-clustered at the country and industry level. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

	Dependent Variable: Log Exports	
	(1)	(2)
num_cust_x_effectiveness	0.108***	
	(0.029)	
num_cust_x_regquality		0.079***
		(0.027)
Country-Capital Interaction	No	No
Country-Skill Interaction	Yes	Yes
Country-Contract Interaction	Yes	Yes
Observations	35,458	35,458
R^2	0.815	0.826
Adjusted R^2	0.813	0.822
Residual Std. Error	1.653	1.611

4.4.6 Robustness Checks

The main result is biased if our regressors are correlated with the error term, and this can happen in many different ways. First, our measures of reliability can be correlated with other country characteristics which give a comparative advantage in goods with many customized components. Second, the number of customized components can be correlated with other industry features, and high reliability can give a comparative advantage due to these industry features as well. There can also be some mixture of these two effects, for example that high government reliability is correlated with high financial development, and that a large number of customized components is correlated with having high external financing needs. We assess the robustness of our results by including other interaction terms between country and good characteristics in our regression specification.

First, we test whether Heckscher-Ohlin effects can explain the results by controlling for the interaction between factor endowments and factor intensity of different industries similar to Romalis (2004). It could be the case that reliable countries are simply countries with a large endowment of skilled labor and risk-sensitive industries tend to be skill-intensive. In column (2) of Table 4.4, we replicate the result that skill-abundant countries specialize in skill-intensive industries (the coefficient on the interaction between country skill abundance and industry skill intensity). Importantly, our main estimate (Cust Inp. $\times$ Gov. Effectiveness) barely changes when we control for factor endowments. In unreported results, we confirm that the same is true for other measures of reliability.

Second, we add an interaction term between industry risk sensitivity and the logarithm of income as a catch-all term for variables that might proxy for being a rich country. It should be noted that this is over-controlling: countries might be rich because they have reliable supply chains, which makes them productive in complex goods. Hence, we control for an outcome. Despite that, as shown in Table 4.4, our main result remains statistically significant (column 5). Quantitatively, the es-

timate becomes only marginally weaker when we (over-) control for log income. This suggests that alternative explanations connected to country income levels do not eliminate the effects from our risk proxies.

Lastly, we make a demanding robustness check by running a regression specification

$$log\,(x_{ig}) = \beta\,(r_i \times n_g) + \mu_i + \mu_i^{skill}h_g + \mu_i^k k_g + \mu_i^{nunn}(contrac\_int_g) + \theta_g + \epsilon_{ig}$$

Here, k_g is a measure of an industry's capital intensity, h_g is a measure of an industry's skill intensity, and $contract\_int_g$ is Nunn's measure of an industry's contracting intensity. The μ_i-terms capture a country-specific term which allows for country-specific tilts of trade patterns in favor of capital intensive, skill intensive, and contracting intensive industries. This specification is more flexible than the Heckscher-Ohlin controls where we assumed that the tilt in favor of capital and skill intensive goods were due to capital and skill abundance on the country level. Similarly, it is more flexible than the regression where we included our variable together with Nunn's, as that specification constrained countries to be tilted towards contracting intensive industries only due to good rule of law. The current specification does not place any such constraints on the pattern of specialization. We present the results in Table 4.5. There is a small change in the point estimate, and the effect is still significant effect at the 1%-level. This suggests that our risk sensitivity measure captures a dimension of trade patterns distinct from that attributable to capital, skill, and contracting intensity.

4.5   Conclusion

This paper provides a tractable model of the effect of supply chain risk on trade patterns. Supply chain risk damages productivity through a combination of input supply disruption risk combined with pre-commited inputs. In our model, we show how the behavior of a sector with idiosyncratic delivery risk can be described by a representative firm. Supply

chain risk enters as a productivity penalty, which grows exponentially with the number of specialized inputs. Therefore, the appropriate measure of supply chain sensitivity on a sectoral level is the number of specialized inputs. In an international setting, the theory implies that low risk countries specialize in risk-sensitive industries, and this prediction is borne out in the data.

Our paper suggests a number of policy relevant conclusions. First, it suggests that reducing risk can attract industries that produce risk-sensitive goods. The paper also implies that measures of the business environment would be more informative if they described the variability in outcomes. The World Bank's Doing Business Survey, for instance, measures the time to start a business. However, it does not contain the risk of severe delays during the process, which might also be important.

Looking ahead, there are several natural extensions to the paper. One extension is to include trade in intermediate inputs. The model now assumes that all inputs are produced domestically, and that there is a country-specific risk of supply chain disruption. In practice, intermediate input trade is important, and variations in supply chain risk is a potentially important factor for understanding patterns in intermediate input trade. A modificaction of the model to include intermediate input trade should also include that some delivery risks are only relevant for cross-border trade. For example, customs procedures might be slow and frictions to international contracting can make deliveries uncertain. We conjecture that such an extension could have rich predictions for the connection between country variations in supply chain risk, the spatial organization of production, and the structure of intermediate input trade.

Given the potential endogeneity concerns in our empirical work, we are also interested in extensions to improve identification. One such extension would be to use the panel dimension of trade data. The World Governance Indicators goes back to 1996 and the BACI trade data goes back to 1995. This would allow us to test whether countries that improve on institutional measures also see a concomitant rise in trade of

risk sensitive goods.

References

Antràs, P. (2003). Firms, contracts, and trade structure. *The Quarterly Journal of Economics*, 118(4):1375–1418.

Antràs, P. and Chor, D. (2013). Organizing the global value chain. *Econometrica*, 81(6):2127–2204.

Antras, P. and Helpman, E. (2004). Global sourcing. *Journal of Political Economy*, 112(3):552–580.

Arvis, J.-F. (2010). *Connecting to compete 2010: trade logistics in the global economy; the logistics performance index and its indicators.* World Bank.

Barro, R. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950-2010. *Journal of Development Economics*, 104:184–198.

Barrot, J.-N. and Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, page qjw018.

Bartelsman, E. J. and Gray, W. (1996). The NBER Manufacturing Productivity Database.

Blanchard, O. and Kremer, M. (1997). Disorganization. *The Quarterly Journal of Economics*, 112(4):1091–1126.

Blyde, J. and Molina, D. (2015). Logistic infrastructure and the international location of fragmented production. *Journal of International Economics*, 95(2):319–332.

Blyde, J. S. (2014). The drivers of global value chain participation: Cross-country analyses. In *Synchronized Factories*, pages 29–73. Springer.

Carvalho, V. M., Nirei, M., Saito, Y. U., and Tahbaz-Salehi, A. (2016). Supply chain disruptions: Evidence from the great east japan earthquake.

Costinot, A. (2009a). An Elementary Theory of Comparative Advantage. *Econometrica*, 77(4):1165–1192.

Costinot, A. (2009b). On the origins of comparative advantage. *Journal of International Economics*, 77(2):255–264.

Costinot, A., Vogel, J., and Wang, S. (2013). An elementary theory of global supply chains. *The Review of Economic Studies*, 80(1):109–144.

Cuñat, A. and Melitz, M. J. (2012). Volatility, Labor Market Flexibility, and the Pattern of Comparative Advantage. *Journal of the European Economic Association*, 10(2):225–254.

Dixit, A. K. and Grossman, G. M. (1982). Trade and protection with multistage production. *The Review of Economic Studies*, 49(4):583–594.

Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, 70(5):1741–1779.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *The American Economic Review*, 105(10):3150–3182.

Gaulier, G. and Zignago, S. (2010). BACI: International Trade Database at the Product-Level (the 1994-2007 Version). *SSRN Electronic Journal*.

Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy*, 94(4):691–719.

Kaufmann, D., Kraay, A., and Mastruzzi, M. (2011). The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law*, 3(2):220–246.

Kremer, M. (1993). The O-Ring Theory of Economic Development. *The Quarterly Journal of Economics*, 108(3):551–75.

Levchenko, A. A. (2007). Institutional quality and international trade. *The Review of Economic Studies*, 74(3):791–819.

Manova, K. (2013). Credit Constraints, Heterogeneous Firms, and International Trade. *The Review of Economic Studies*, 80(2):711–744.

Nunn, N. (2007). Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade. *The Quarterly Journal of Economics*, 122(2):569–600.

Nunn, N. and Trefler, D. (2014). Domestic institutions as a source of comparative advantage. *Handbook of International Economics*, 4:263.

Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World development*, 22(9):1325–1343.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics*, 48(1):7–35.

Romalis, J. (2004). Factor Proportions and the Structure of Commodity Trade. *American Economic Review*, 94(1):67–97.

Additional Results

Table 4.1: Top 10 Industries by Risk Sensitivity

| Code	Name
336390	Other Motor Vehicle Parts Manufacturing
326191	Plastics Plumbing Fixture Manufacturing
326199	All Other Plastics Product Manufacturing
321920	Wood Container and Pallet Manufacturing
321991	Manufactured Home (Mobile Home) Manufacturing
321992	Prefabricated Wood Building Manufacturing
321999	All Other Miscellaneous Wood Product Manufacturing
333618	Other Engine Equipment Manufacturing
336120	Heavy Duty Truck Manufacturing
334413	Semiconductor and Related Device Manufacturing

Table 4.2: Bottom 10 Industries by Risk Sensitivity

| Code	Name
312140	Distilleries
311710	Seafood Product Preparation and Packaging
113310	Logging
113210	Forest Nurseries and Gathering of Forest Products
311221	Wet Corn Milling
311920	Coffee and Tea Manufacturing
311213	Malt Manufacturing
311212	Rice Milling
311211	Flour Milling
331410	Nonferrous Metal (except Aluminum) Smelting and Refining

Figure 4.1: Country and Industry Variables

This figures presents histograms of country and industry characteristics. All variables are standardized. Data sources for country-level variables: From World Governance Indicators (Kaufmann et al., 2011), we collect Government Effectiveness (effectiveness), Regulatory Quality (regquality), Political Stability and Absence of Violence/Terrorism(stability), Voice and Accountability(voice), Control of Corruption(corruption), and Rule of Law (ruleoflaw). We add the World Bank's Logistics Performance Index (lpi; see Arvis, 2010). In robustness checks, we also use Penn World Tables (Feenstra et al., 2015) to account for skilled labor (ln_hl), capital(ln_kl), and the logarithm of per capita GDP (ln_y). Data sources for industry-level variables: We define the number of inputs (num) using the US input-output tables. The number of customized inputs is calculated by counting the number of inputs that are neither reference-priced nor traded on an exchange according to Rauch (1999). Contract intensity (nunn) is calculated as in Nunn (2007). Skill and capital intensity (sk_int, cap_int) are measured in the NBER CES

(a) Country-level
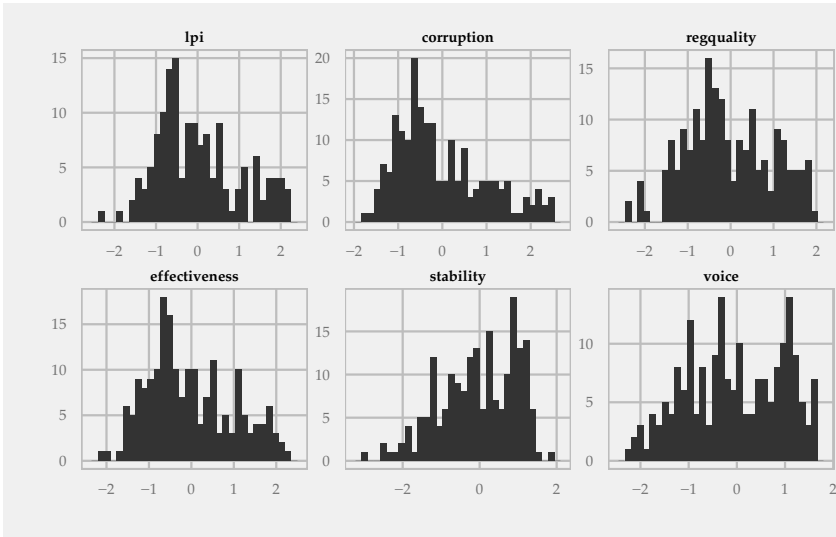
(b) Industry-level
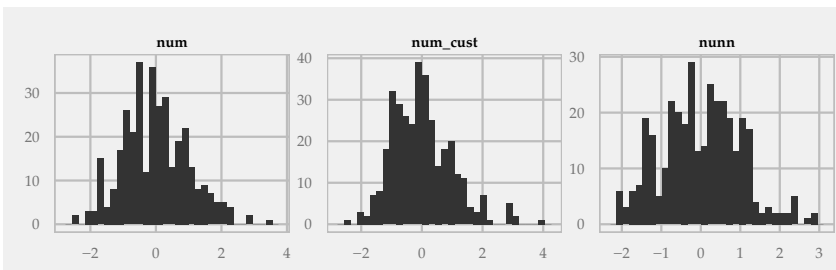
Appendix

Proof of Proposition 7

The aim of the proof is to show that $\Omega_{det}(\Gamma) = \Omega_{sto}(\Gamma)$ for all $\Gamma = (P, w, \pi)$. Furthermore we want to show that

$$\Omega_{sto}(\Gamma) = \begin{cases} \emptyset & \text{if } \frac{w}{P} < (1 - \pi)^{m+(1-\alpha)} \\ \{(L, F(L; \gamma)) : L \geq 0\} & \text{if } \frac{w}{P} = (1 - \pi)^{m+(1-\alpha)} \\ \{(0, 0)\} & \text{if } \frac{w}{P} > (1 - \pi)^{m+(1-\alpha)} \end{cases}.$$

We first note that it is obvious that

$$\Omega(\Gamma) = \begin{cases} \emptyset & \text{if } \frac{w}{P} < (1 - \pi)^{m+(1-\alpha)} \\ \{(L, F(L; \gamma)) : L \geq 0\} & \text{if } \frac{w}{P} = (1 - \pi)^{m+(1-\alpha)} \\ \{(0, 0)\} & \text{if } \frac{w}{P} > (1 - \pi)^{m+(1-\alpha)} \end{cases}.$$

Indeed, if real wage is below unit cost, no finite L solves the firm's problem. If real wage is above unit cost, 0 is the only profit maximizing production level. If real wage equals unit cost, firms are indifferent about production size.

Thus, the interesting thing is to show $\Omega_{det}(\gamma) = \Omega_{sto}(\gamma)$. We go through the three cases of $\frac{w}{P}$ and show that $\Omega_{det}(\Gamma) \subseteq \Omega_{sto}(\Gamma)$ and $\Omega_{sto}(\Gamma) \subseteq \Omega_{det}(\Gamma)$ for each case.

Case 1: $\frac{w}{P} < (1 - \pi)^{m+\gamma}$

It is trivial that $\emptyset \subseteq \Omega_{sto}(\Gamma)$. To prove that $\Omega_{sto}(\Gamma) = \emptyset$, we note that if $(Y, L) \in \Omega_{sto}(\Gamma)$ we need $p_i^x \leq w/(1 - \pi)$ for $i = 1, \ldots, m$ and $p_i^z(j) \leq \frac{w}{1-\pi}$ for $i = 1, \ldots, n$ as there would otherwise be infinite labor demand in the intermediate goods sector. But with this assumption, unit cost in the final goods sector becomes

$$\frac{w^\alpha \left(\sum_{i=1}^{m}((1-\pi)p_i^x(j))^{1-\eta}\right)^{\frac{\beta}{1-\eta}} \left(\sum_{i=1}^{n}(p_i^z)^{1-\eta}\right)^{\frac{\gamma}{1-\eta}}}{m^{\frac{\beta}{\eta-1}}n^{\frac{\gamma}{\eta-1}}(1-\pi)^m} \quad \leq \quad \frac{w}{(1-\pi)^{m+\gamma}}$$

$$< \quad P$$

which means that labor demand is unbounded in the final goods sector. Thus, no finite L is consistent with optimization.

Case 2: $\frac{w}{P} = (1-\pi)^{m+\gamma}$

First, we want to show that $\Omega_{det}(\Gamma) \subseteq \Omega_{sto}(\Gamma)$, that is we want to show that $(L, F(L;\Gamma)) \in \Omega_{sto}(\Gamma)$ for any L. To do this, consider prices $p_i^x(j) = \frac{w}{1-\pi}$ for $i = 1, \ldots, m$ (more precisely that the offered payment is $x_i^F(j)p_i^x(j)$) and $p_i^z = \frac{w}{1-\pi}$ for $i = 1, \ldots, n$, and allocations

$$
\begin{align}
l^F(j) &= \alpha L \tag{4.6} \\
x_i^F(j) &= \frac{\beta L}{m} \quad i = 1, \ldots, m \tag{4.7} \\
l_i^x(j) &= \frac{\beta L}{m} \tag{4.8} \\
z_i^F(j) &= \frac{\gamma L(1-\pi)}{n} \quad i = 1, \ldots, n \tag{4.9} \\
l_i^z(j) &= \frac{\gamma L}{n} \quad i = 1, \ldots, n \tag{4.10}
\end{align}
$$

It is clear that labor demand sums to L. Intermediate goods producers are indifferent between different production levels, so their choices are optimal. The final goods producer's problem is equivalent to solving a deterministic problem with price $P(1-\pi)^m$ and where the price of customized components is modified to $p_i^x(j)(1-\pi)$ to reflect that the final goods producer only pays in case of delivery. Given the symmetry within the classes of standardized and intermediate components, it is clear that the firm chooses the same amount x, z of all of them. So the firm solves the problem

$$\max_{l^f, X, Z} P(1-\pi)^m \kappa l^\alpha x^\beta m^{\frac{\beta\eta}{\eta-1}} z^\gamma m^{\frac{\gamma\eta}{\eta-1}} - lw - mx(1-\pi)p_i^x(j) - nzp_i^z.$$

Standard optimization gives that $\frac{l^f}{x} = \frac{\alpha}{(\beta/m)}$ and $\frac{l^f}{z} = \frac{\alpha}{(\gamma/n)}$, and we can check that profits are zero for all l^f when these two conditions are satisfied. Thus, the proposed allocation solves the final goods producer's problem.

Total production is given by

$$
\begin{aligned}
\tilde{Y} &= \int_0^1 \tilde{y}(j)dj \\
&= \kappa(1-\pi)^m (\alpha L)^\alpha \left(m \left(\frac{\beta L}{m} \right)^{\frac{\eta-1}{\eta}} \right)^{\frac{\beta\eta}{\eta-1}} \left(n \left(\frac{\gamma L(1-\pi)}{n} \right)^{\frac{\eta-1}{\eta}} \right)^{\frac{\gamma\eta}{\eta-1}} \\
&= \Omega_{det}(1-\pi)^{m+\gamma}L \\
&= F(L;\Gamma).
\end{aligned}
$$

Hence, $(L, F(L;\Gamma)) \in \Omega_{sto}(\Gamma)$ and $\Omega_{det}(\Gamma) \subseteq \Omega_{sto}(\Gamma)$.

Second, we want to show that $\Omega_{sto}(\Gamma) \subseteq \Omega_{det}(\Gamma)$. So consider an arbitrary (L, Y). If $L = 0$, then $Y = 0$ trivially and we are done, as $(0,0) \in \Omega_{det}(\Gamma)$. So let us assume that $Y, L > 0$. As $L > 0$, we need that $l^f(j) > 0$ for some j. Let $\mathcal{S}$ be set of j for which this is true and assume without loss of generality that $\mathcal{S} = [0,1]$ (size is indeterminate, but if $\mathcal{S} \neq [0,1]$ we can just divide everything with the measure of $\mathcal{S}$).

If final goods producers optimally choose positive labor component, optimality implies that they also choose positive amounts of all intermediate components. Thus, market clearing implies that for all i, there exists some j, such that $l^z(j) > 0$ which means that $p_i^z = \frac{w}{1-\pi}$ for all i. Similarly, $l_i^x(j) > 0$ for all i, j which mean that offers are given with $p_i^x(j) = \frac{w}{1-\pi}$. The necessary condition for optimality for final goods producers derived above gives us the relative demand for labor and different intermediate goods components. Using the market clearing condition for

intermediate goods products and labor, we get that the labor allocations and intermediate good demands are given by equations (4.6)-(4.10). This means that total production is $Y = F(L; \Gamma)$, and $(L, Y) \in \Omega_{det}(\Gamma)$.

Hence, $\Omega_{sto}(\Gamma) = \Omega_{det}(\Gamma)$.

Case 2: $\frac{w}{P} > (1 - \pi)^{m+\gamma}$

We want to show that $\Omega_{sto}(\Gamma) = \{(0, 0)\}$. We first show that $(0, 0) \in \Omega_{sto}(\Gamma)$. Now, suppose that $p_i^z(j) = p_i^z(j) = \frac{w}{1-\pi}$ for all i, j. Then no production lies in the optimal set for all intermediate good producers. Furthermore, we can check that no production is also optimal for the final good producers by noting that their unit cost exceeds their price. Thus, $(0, 0) \in \Omega_{sto}(\Gamma)$.

Next, we want to show that $\Omega_{sto}(\Gamma) \subseteq \{(0, 0)\}$. We prove this by contradiction. Suppose that $(L, Y) \in \Omega_{sto}(\Gamma)$ with $L > 0$. This means that $l^f(j) > 0$ for some j. This also means that for each $i \in \{1, \ldots, n\}$, there exists a $j' \in [0, 1]$ such that $l_i^z(j') > 0$. Hence, $p_i^z = \frac{w}{1-\pi}$ for all i. Furthermore, optimality together with the restriction that the customized goods producers accept their offers, requires that $l_i^x(j) = \frac{w}{1-\pi}$ for all $i \in \{1, \ldots, m\}$. But now we can check that these prices make it optimal for final goods producers to choose zero production, thus contradicting our assumption that $L > 0$.

Hence, we again have $\Omega_{sto}(\Gamma) = \Omega_{det}(\Gamma)$.

Proof of Proposition 8

We proceed in steps. First we prove that each country has positive production and that each good is produced in equilibrium. Then we characterize the sorting behavior and show that the equilibrium is unique.

Lemma 1. *For each m_j, there exists a π with $\ell(\pi, m_j) > 0$, and for each π, there exists an m_j with $\ell(\pi, m_j) > 0$.*

The Inada condition means that every good is produced in equilibrium, which proves the first part of the proposition. The second part of the proposition follows directly from the labor clearing condition.

Second, we prove a lemma that captures the sorting of m_j and π. It states that if there are a high risk and a low risk country, as well as a complex and a simple good, then if the low risk country produce the simple good in equilibrium, the high risk country will not produce the complex good. This excludes reversals of comparative advantage and is used to prove sorting.

Lemma 2. *Suppose that $\pi' < \pi$ and $m_j < m_{j'}$. Then $\ell(\pi', m_j) > 0$ implies $\ell(\pi, m_{j'}) = 0$.*

Proof. We proceed by contradiction. Suppose $\ell(\pi', m_j), \ell(\pi, m_{j'}) > 0$. Then the no profit conditions give us

$$
\begin{aligned}
p_{m_{j'}}(1 - \pi)^{m_{j'}} &= w_\pi \\
p_{m_j}(1 - \pi)^{m_j} &\leq w_\pi \\
p_{m_{j'}}(1 - \pi')^{m_{j'}} &\leq w_{\pi'} \\
p_{m_j}(1 - \pi')^{m_j} &= w_{\pi'}
\end{aligned}
$$

From which we derive the contradiction

$$
\begin{aligned}
\frac{p_{m_{j'}}}{p_{m_j}} &\geq (1 - \pi)^{m_j - m_{j'}} \\
\frac{p_{m_{j'}}}{p_{m_j}} &\leq (1 - \pi')^{m_j - m_{j'}}
\end{aligned}
$$

This is a contradiction as $\pi' < \pi$ and $m_j < m_{j'}$ implies that $(1 - \pi')^{m_j - m_{j'}} < (1 - \pi)^{m_j - m_{j'}}$ so no price ratios satisfy the two inequalities simultaneously. $\square$

Corollary 1. *$\ell(\pi, m_j) > 0$ implies $\ell(\pi, m_{j'}) = 0$ for all $j' \neq j$. I.e. each country only produces one good.*

Proof. Suppose that $m_j < m_{j'}$ are both produced in country π, i.e. $\ell(\pi, m_j), \ell(\pi, m_{j'}) > 0$. Our assumption of continuity means that

we can find δ such that $\ell(\pi', m_j) > 0$ for all $\pi' \in (\pi, \pi + \delta)$. But then $\ell(\pi, m_{j'}), \ell(\pi', m_j) > 0$ which contradicts the lemma.

With aid of this corollary, we can obtain a full characterization of the sorting behavior. I.e. that there exist

$$\underline{\pi} = \pi_k < \pi_{k-1} < \cdots < \pi_1 < \pi_0 = \bar{\pi}$$

such that

$$
\begin{aligned}
\ell(\pi; m_j) &> 0 \quad if \; \pi \in [\pi_j, \pi_{j-1}) \\
l(\pi; m_j) &= 0 \quad if \; \pi \notin [\pi_j, \pi_{j-1})
\end{aligned}
$$

Define the correspondence

$$\Psi(\pi) = \{j \in \{1, \ldots, k\} : \ell(\pi, m_j) > 0\}$$

From previous results, this correspondence is always non-empty, single valued, and weakly decreasing in π. Define

$$\pi_1 = \inf \{\pi : \Psi(\pi) = 1\}$$

The set $\{\pi : \Psi(\pi) = 1\}$ is non-empty, and the assumption that $\ell(\pi, m_j) > 0$ implies $\ell(\pi', m_j) > 0$ for $\pi' \in [\pi, \pi + \delta)$ implies that Ψ is right-continuous, so $\Psi(\pi_1) = 1$. By weak monotonicity $\Psi(\pi) = 1$ for $\pi \in [\pi_1, \pi_0)$. Define π_2 analogously and continue in the same way. The Proposition is thus proved. □

Concordance construction

To generate concordances and map data across coding system, we create a general mathematical framework to treat the problem. In this Web Appendix, we describe how the general system works, and then we show how we use it to convert our particular data.

The basic building block of our concordance system is a many-to-many concordance between coding systems A and B where we have weights on both A and B. We call such concordances *two-weighted concordances*. An example of such a concordance is provided below:

A	B	A_w	B_w
1	a	10	70
2	b	20	50
2	c	20	100
3	c	15	40
4	d	5	70
5	d	25	70
6	e	30	90

Note that each code in system A can be converted to multiple B codes (in this example, code "2" in System A maps to both code "b" and "c" in System B). The converse is also true: both code "4" and "5" map to code "e". The weights code how important the respective industries are. This could for example be total value of shipments, total trade value, etc. Notice the weights are both on A and B, and that they are constand whenever they stand for the same industry.

We can define this mathematically as there being two sets A, B with measures w_A, w_B giving the mass on each code, and a concordance being a correspondence

$$\phi : A \rightrightarrows B.$$

We will write results in terms of this mathematical definition, but also in terms of examples to show the working of the system.

We will go through three operations relating to two-weighted concordances:

1. How to transform quantity variables such as total industry sales using a two-weighted concordance

2. How to transform property variables such as capital intensity using

a two-weighted concordance

3. How to create a two-weighted concordance using a unweighted concordance and a weighting scheme for one of the variables (e.g. when we want to create a two-weighted concordance between HS and SITC and only have total trade in HS codes).

Transform quantity variables using two-weighted concordances

Starting with quantity variables, suppose that we have total trade flows in industry code A. We then want to allocate it across different codes in coordinate system B. In this case, for each element in A we look at all elements in B that it maps to. It then allocates the quantity in A across the elements in B in proportion to their weights. The quantity attributed to element B is then the sum of the contributions over all elements in A.

A	vship
1	1000
2	3000
3	6000
4	2000
5	3000
6	4000

Consider the table above, where we want to convert from the coordinate system A to the coordinate system B using the previous correspondence. We will explain what value of shipments we will attribute to industry c in system B. The pre-image of "c" is "2" and "3" in system A, so we can look how much of the shipments of these two A-industries that will be attributed to "c". Industry "2" ships 3000 in value, and it corresponds to both industry "b" and "c" in System B. As the relative weights of "b" and "c" are 50 and 100 respectively, 1000 will be attributed to "b" and 2000 will be attributed to "c". However, in the concordance, we see

that 3 only maps to "c", so all 6000 shipments from 3 will be attributed to c. Hence, total attribution to "c" is $2000 + 6000 = 8000$.

We can write this in terms of the mathematical representation Φ as well, together with the weights μ_A and μ_B. If

$$f_A : A \to \mathbb{R}$$

is an arbitrary quantity measure on A we convert it to B by

$$f_B(y) = \sum_{x \in \Phi^{-1}(y)} f_A(x) \times \frac{\mu_B(y)}{\sum_{y' \in \Phi(y)} \mu_B(y')}.$$

The equation is quite difficult to parse, but it says that we take all the values from the pre-image to y. The value of each of those pre-images x attributed to y is equal to the relative weight of $\mu_B(y)$ compared to the total weights of those codes in B that x maps to.

Transform property variables using two-weighted concordances

The situation is different when we have so-called property variables, for example capital intensity, skill intensity or other industry level properties. We can see how these differs by means of an example. Suppose that we have a concordance between HS 2007 six-digit and HS 2007 ten-digit data. If we want have data on trade flows on six-digit level and want to convert these to ten-digit level. Then, the reasonable thing is to split it up across the ten digits according to some weighting scheme.

However, if we instead have measured capital intensity on the six-digit level, the natural thing is to give this capital intensity as a prediction for the capital intensity in all ten-digit descendant categories (if we have no additional information on capital intensity on ten-digit level). Similarly, if we wanted to convert from ten-digit to six-digit, trade flows ought to be summed, whereas for properties it is appropriate to take a weighted average of industry-level properties on the ten digit level.

Thus, we see that property variables translate across coding systems in a fundamentally different way from quantity variables. We define the

transformation scheme for property variables by saying that for each code $y \in B$ in the target system, we define its property as a weighted average of the properties that its pre-images $x \in A$, where we use the weights on A as a weighting scheme. For example, in our example concordance, we would attribute c a property which is the weighted average of $2, 3$ in System A, using the measures $\mu_A(\{2\}) = 20$ and $\mu_A(\{3\}) = 15$ as weights.

More formally, if we have a property measure

$$g_A : A \to \mathbb{R}$$

defined on A, then we translate it to B using ϕ by the equation

$$g_B(y) = \frac{\sum_{x \in \phi^{-1}(y)} g_A(x)\mu_A(x)}{\sum_{x \in \phi^{-1}(y)} \mu_A(x)}.$$

Construct a two-side weighted concordance from a one-sided weighted concordance

Above we defined how you translate between different coordinate systems if you have a two-sided weighted concordance. However, sometimes we only have a one-sided concordance. For example, if we have total trade data in HS 2007 six-digtit and want to create a concordance between HS 2007 6-digit and NAICS 2007 it might be that we do not have data to create a natural weighting scheme for NAICS 2007 data.

For this case, we have a procedure to create a two-sided weighted concordance from a one-sided weighted concordance. It is quite similar to the quantity transformation above. Suppose that we have a concordance ϕ and a measure μ_A on A and want to create a measure μ_B on B. The question is how much weight we should attribute to each element $y \in B$. In this case, we go through each element $x \in A$ and take its weight $\mu_A(x)$ and portion it out equally on all elements y' in B that x maps to. This gives us how much weight element x gives to element y, that is $\frac{\mu_A(x)}{|\phi(x)|}$ where $|\phi(x)|$ gives how many codes x maps to. By summing over all x

we get the total contribution to y. In mathematical terms

$$\mu_B(y) = \sum_{x \in \phi^{-1}(y)} \frac{\mu_A(x)}{|\phi^{-1}(x)|}.$$

Practical implementation

The process above allows us to define three primitive operations: creating a two-sided concordance, using it to convert between property variables, and use it to convert between quantity variables. We can use these three operations to create arbitrary chains of concordances between data. Below we list the actual concordances we create, which weights are used, and how we use these concordances to translate everything into NAICS 2012 six-digit data.

Created concordance sequence:

1. Create concordance between HS 2007 six-digit and HS 2007 ten-digit from one sided concordance with total world trade as weight on HS 2007 six digit.

2. Create concordance from HS 2007 10-digit to NAICS 2007 six digit from a one sided concordance using [...] as a basic concordance and the HS 2007 10-digit weights obtained from previous exercise

3. Create concordance from NAICS 2007 six digits to NAICS 2002 six digits using a one sided concordance with [...] as basic concordance and the NAICS 2007 six digits weights obtained from previous step

4. Create concordance from NAICS 2002 six digit to NAICS 1997 six digit analogously to previous step

5. Create concordance from NAICS 2007 six digit to NAICS 2012 six digit analogously to previous step

6. Create concordance between IO 2007 six-digit and NAICS 2007 six digit directly using [....] as basic concordance, total production as

weight on IO-codes and previously constructed weights from step 2 for NAICS 2007 six digit

7. Create concordance between HS 2007 six digit and SITC rev.2 four digits using a one-sided concordance with [...] as basic concordance and total world trade as weight on HS 2007 six digit.

Once we created these concordances, we can translate all variables to NAICS 2012 six-digit code. We use the following transitions.

Source data set	Code	Path
NBER CES	NAICS 1997 6 digits	NAICS 1997 6 digits
		NAICS 2002 6 digits
		NAICS 2007 6 digits
		NAICS 2012 6 digits
IO-table	IO 2007 6 digits	IO 2007 6 digits
		NAICS 2007 6 digits
		NAICS 2012 6 digits
BACI Trade data	HS 2007 6 digits	HS 2007 6 digits
		HS 2007 10 digits
		NAICS 2007 6 digits
		NACIS 2012 6 digits
Rauch	SITC rev 2 4 digits	SITC rev 2 4d
		HS 2007 6 digits
		HS 2007 10 digits
		NAICS 2007 6 digits
		NAICS 2012 6 digits

Sammanfattning

Denna avhandling består av fyra oberoende kapitel som berör frågor i makroekonomi, utvecklingsekonomi och utrikeshandel.

Kapitel I, **Human Capital and Development Accounting Revisited**, är huvudkapitlet i avhandlingen. Kapitlet undersöker humankapitalets roll i att förklara inkomstskillnaderna mellan världens länder.

Att förstå de stora skillnaderna mellan rika och fattiga länder är en av de största utmaningarna för nationalekonomin, och en framträdande teori har varit att skillnader i humankapital är viktiga för att förstå skillnader i inkomstnivåer (Lucas, 1988).

Teorin tycks fånga ett antal samband väl. Vi vet att högkvalificerade arbetare är mer produktiva än lågkvalificerade arbetare, och en slående skillnad mellan rika och fattiga länder är de stora skillnaderna i utbildningsnivåer. Ett liknande mönster framträder när vi analyserar skillnaderna i yrkesfördelningen mellan rika och fattiga länder. I de fattigaste länderna jobbar färre än 10 procent av arbetskraften inom ett högkvalificerat yrke, medan motsvarande siffra överstiger 50 procent i vissa rika länder.

Inom utvecklingsekonomi har det dock funnits ett inflytelserikt argument mot att humankapital skulle spela en central roll i att förklara inkomstskillnaderna mellan världens länder. Argumentet bygger på att inkomstskillnaderna mellan länder är långt större än den mikroekonomiska avkastningen på humankapital. Med andra ord, skillnaden mellan högkvalificerades och lågkvalificerades löner *inom*

länder är långt mindre än medelinkomstskillnaderna *mellan* länder. Om lönen reflekterar produktiviten hos högkvalificerade arbetare, kan inte ett högre utbud av högkvalificerade arbetare förklara rika länders välstånd.

Argumentet är enkelt att förstå med överslagsräkning. Människor i rika länder utbildar sig i snitt omkring åtta år längre än människor i fattiga länder, och avkastningen på utbildning är ca 10 procent per skolår. Om man använder avkastningen på utbildning som ett riktmärke, tyder detta på att fattiga länder ungefär kan fördubbla sina inkomstnivåer genom att öka sina utbildningsnivåer till rika länders nivåer. Detta är självklart en stor inkomstökning, men den räcker inte på lång väg för att överbrygga inkomstgapet mellan fattiga och rika länder: världens fattiga länder är cirka 30 till 40 gånger fattigare än de rikaste länderna.

Inom utvecklingsekonomi har det här argumentet formaliserats i litteraturen om "utvecklingsbokföring", eller *development accounting* (Klenow and Rodriguez-Clare, 1997; Hall and Jones, 1999). Slutsatserna från den litteraturen är att humankapitalskillnader kan förklara en begränsad – om icke oansenlig – del av inkomstskillnaderna i världen. Huvuddelen av inkomstskillnaderna går inte att förklara med skillnader i vare sig fysiskt kapital eller humankapital, och hänförs därför till en icke-observerad restterm som ofta kallas teknologi. Denna term kan inkludera skillnader i vad som traditionellt sett ses som teknologi, men den innefattar också allt annat som förändrar inkomster utan att mängden insatsvaror förändras. Man kan jämföra teknologi med begreppet mörk energi i fysik: det är en storhet som måste postuleras för att modellen ska gå ihop.

Det har varit ett inflytelserikt resultat i utvecklingsekonomin att humankapital enbart kan förklara en liten del av världens inkomstskillnader. Resultatet har lett till att forskningen har fokuserat på att förklara vad de stora teknologiskillnaderna mellan länder består av, medan skillnader i humankapital har förutsatts spela en mer begränsad roll.

Det centrala argumentet i traditionell utvecklingsbokföring är att

lönepremien reflekterar kvalificerade arbetares produktivitet. Då löne-premien inom länder är långt mindre än inkomstskillnaderna mellan länder antas humankapital spela en begränsad roll. Det finns dock en alternativ syn, nämligen att kvalificerade arbetare är väldigt produktiva i rika länder, men att deras relativlöner begränsas av det höga utbudet av kvalificerade tjänster i rika länder. Detta argument har framförts i Jones (2014) och bygger på en nedåtlutande efterfrågekurva för kvali-ficerade tjänster. Formellt sker detta om okvalificerade och kvalificerade tjänster inte är perfekt utbytbara.

Avsaknad av perfekt utbytbarhet kan leda till att humankapitalets roll underskattas, men det finns mindre samsyn på hur kvantitativt viktig denna mekanism är. För att bedöma denna mekanism behöver man göra en skattning av *hur mycket* lägre det relativa priset på kvalificerade tjänster är i rika länder.

I mitt kapitel använder jag evidens från utrikeshandelsdata för att göra detta. Jag noterar att länder med relativt dyra kvalificerade tjänster kan förväntas exportera relativt få kunskapsintensiva produkter. Jag visar sedan hur den här insikten kan användas för att skatta skillnader i relativpriset mellan kvalificerade och okvalificerade tjänster med hjälp av internationell handelsdata.

Min analys av handelsdatan tyder på att kvalificerade tjänster är relativt sett mycket dyrare i fattiga jämfört med i rika länder. Priset på arbetstjänster beror både på hur mycket en arbetare kostar (lönen) samt hur många tjänster en arbetare levererar (effektiviteten). Jag visar att det höga priset på kvalificerade tjänster i fattiga länder både beror på en hög relativlön, samt på en låg relativ effektivitet hos kvalificerade arbetare.

Mina resultat tyder på att humankapital kan spela en större roll i att förklara världens inkomstskillnader än vad som hävdas i traditionell utvecklingsbokföring. Om jag följer traditionell utvecklingsbokföring och antar att effektivitetsskillnader hos kvalificerad arbetskraft speglar humankapital, kan humankapital förklara en majortiet av världens inkomstskillnader. Det här tyder på att vi inte kan utesluta en central

roll för humankapital när vi försöker modellera ekonomisk utveckling.

Att tolka effektivitet av kvalificerad arbetskraft som humankapital är i linje med hur traditionell utvecklingsbokföring fungerar. Däremot är det svårt att utesluta andra tolkningar, då pris- och kvantitetsdata inte låter oss skilja mellan humankapital och andra källor av produktivitet. Intuitivt sett är det svårt att se utifrån styckpris och lönebetalningar om någon är bra på att hamra eller har en bra hammare. Vad vi däremot kan säga är att alternativa förklaringar till effektivitetsskillnader fortfarande måste förklara den selektivt höga effektiviteten av kvalificerad arbetskraft i rika länder. Därför är slutsatsen att inkomstskillnader antingen beror på rena humankaptitalskillnader, eller på skillnader mellan rika och fattiga länder som selektivt ökar produktiviteten av kvalificerad arbetskraft i rika länder.

Kapitel II, **Price Level Determination When Tax Payments are Required in Money**, är författat tillsammans med Erik Öberg. Det utforskar hur ett krav på att skatter ska betalas i pengar påverkar hur pengar värderas.

Penningvärdet har visat sig vara svårt att modellera inom nationalekonomin. Utmaningen är att pengar är värdelösa i sig, vilket innebär att vanlig konsumtionsteori förutspår att de saknar värde. En klass av modeller som har försökt förklara penningvärdet bygger på att pengar har ett likviditetsvärde, då pengar av sedvänja används som transaktionsmedel i ekonomin. Intuitionen är att jag accepterar dina pengar som betalning under antagandet att andra kommer att acceptera dem i sin tur. Pengars värde upprätthålls genom ett nät av förväntningar.

Även om sådana likviditetsteorier är rimliga, kan de inte förklara varför detta nät av förväntningar nästan alltid leder till att statens pengar blir det allmänna transaktionsmedlet. Ett likviditetsvärde skiljer sig också från andra typer av konsumtionsvärden då det är självrefererande, i den bemärkelsen att en varas likviditetsvärde beror på dess pris: pengar har ett värde som transaktionsmedel just för att pengar är värdefulla. Detta betyder att likviditetsteorier inte kan utesluta att

pengar plötsligt förlorar sitt värde genom att det nät av förväntningar som upprätthåller deras värde bryts.

Men för statliga pengar tycks vi sällan, om ens någonsin, observera att pengar förlorar sitt värde på detta sätt. Det sedvanliga sättet för statliga pengar att bli värdelösa är istället genom en övertryckning av pengar – som i Tyskland under 1920-talet eller mer nyligen i Zimbabwe – eller genom att den understödjande staten kollapsar – som för Sydstaternas valuta efter amerikanska inbördeskriget eller för ett antal japanska ockupationsvalutor efter andra världskriget.

Dessa fakta tyder på att penningvärdet på något plan hänger ihop med statlig makt. Detta väcker en fråga om hur denna koppling ser ut. Kapitel I visar att ett sätt för staten att tillförsäkra pengar ett värde är genom att kräva att skatter ska betalas med pengar.

Idén att penningvärdet kan förklaras med ett krav på skattebetalningar är gammal inom nationalekonomin. En informell version finns redan hos Adam Smith. I kapitlet formaliserar vi denna idé i en allmän jämviktsmodell, och visar under vilka villkor staten kan tillförsäkra pengar ett specifikt värde genom att reglera penningutbudet samt genom att kräva att skatter ska betalas med pengar.

I vår grundmodell saknar pengar ett intrinsikalt värde. Vi utforskar även hur resultaten påverkas om vi antar att pengar dessutom har ett likviditetsvärde. Vi visar att penningvärdet under detta antagande inte bestäms direkt av skattekraven, men att skattekraven utesluter möjligheten att pengar kan förlora sitt värde enbart för att folk gemensamt tappar tron på dem.

Kapitel III, **Swedish Unemployment Dynamics: New Methods and Results**, är författat tillsammans med Niels-Jakob Harbo Hansen och studerar de statistiska egenskaperna hos arbetslöshetsvariationer.

Kapitlet behandlar en känd fråga inom arbetsmarknadsekonomi. Vi vet att små nettoflöden på arbetsmarknaden döljer stora bruttoflöden. Det betyder att en ökning av arbetslösheten antingen kan drivas av att arbetstagare förlorar sina arbeten i en högre takt, eller att det tar längre tid för arbetslösa att hitta ett nytt jobb. För att förstå vad som driver

en viss förändring på arbetsmarknaden måste man studera arbetsmarknadens underliggande flöden.

Kapitlet utvecklar en ny metod för att bryta ned variationerna i arbetslöshet till underliggande flöden. Metoden modifierar ett antal antaganden i de mest använda metoderna inom litteraturen för att anpassa dem till analys av europeiska arbetsmarknader. Det sedvanliga antagandet som har använts i litteraturen är att arbetsmarknaden befinner sig nära sitt *fortvarighetstillstånd* (Shimer, 2007).[1]

Detta antagande fungerar väl för att beskriva den amerikanska arbetsmarknaden som kännetecknas av stora flöden och en snabb konvergens mot sitt fortvarighetstillstånd. Det är dock välkänt att antagandet fungerar sämre för europeiska arbetsmarknader, där storleken på arbetsmarknadens flöden är mindre, vilket gör att arbetsmarknaden långsammare når sitt fortvarighetstillstånd efter en förändring av de underliggande flödena. Vidare kännetecknas europeiska arbetsmarknader av *dualitet*, med en mer eller mindre skarp skiljelinje mellan fasta och tillfälliga anställningar. Då flödesegenskaperna är mycket annorlunda för olika typer av anställningsformer är det värdefullt att använda analysmetoder som tillåter en analys av arbetsmarknader med olika anställningsformer.

För att hantera dessa frågor utvecklar vi en metod som tillåter både i) långsam konvergens mot arbetsmarknadens fortvarighetstillstånd, och ii) ett godtyckligt antal tillstånd som arbetare kan befinna sig i. Vi tillämpar vår metod på en ny svensk datamängd som täcker åren 1987-2012. Metoden passar väl för den svenska tillämpningen. Den svenska arbetsmarknaden kännetecknas av dualitet – under perioden är ca 13 procent av den arbetsföra befolkningen anställda på tillfälliga kontrakt – och en tämligen långsam konvergenstakt mot sitt fortvarighetstillstånd.

Vi finner att arbetslöshetsvariationer i ungefär lika stor grad drivs av förändringar i (i) flöden mellan arbete och arbetslöshet, (ii) flöden från

---

[1]Begreppet är en översättning av engelskans *steady-state*, och används inom arbetsmarknadsekonomi för att beskriva det tillstånd som skulle råda om dagens flödesstorlekar fortsatte för alltid.

arbetlöshet till arbete, samt (iii) flöden in och ut ur arbetskraften. Flöden som innefattar tillfälligt anställda bidrar med 44 procent av variationen i arbetslöshet, vilket är mycket med tanke på den relativt låga andelen av arbetskraften som har tillfälliga kontrakt.

Vi visar också att det är viktigt att ta den långsamma konvergenstakten till fortvarighetstilltståndet i beaktande. Om vi antar att arbetsmarknaden omedelbart når sitt fortvarighetstillstånd, stiger den skattade betydelsen av flöden till och från fasta anställningar. En trolig orsak till detta är att fortvarighetstillståndet nås långsamt när det sker en förändring av flöden till och från fasta anställningar, då det finns många fasta anställda och flödena är små. Antagandet om omedelbar konvergens gör att flödesförändringens betydelse överdrivs. Dessa resultat är av bredare intresse för studier av europeiska arbetsmarknader, då de tyder på att det kan vara problematiskt att använda antagandet om omedelbar konvergens när man studerar duala arbetsmarknader.

Kapitel IV, **Supply Chain Risk and the Pattern of Trade**, är författat tillsammans med Maximilian Eber. Vi utforskar huruvida länder som erbjuder tillförlitliga försörjningskedjor specialiserar sig i att producera varor som är känsliga för risker i försörjningskedjan.

Vi formaliserar den här hypotesen genom att bygga en modell av internationell handel med flera sektorer och riskabla försörjningskedjor. Varje sektor producerar en vara och använder ett antal olika insatsvaror, och det finns en risk att leveransen av en insatsvara inte kommer att fungera.

Effekten av ett leveransmisslyckande beror på insatsvarans karaktär. Vissa insatsvaror är standardiserade, såsom vete, och för dessa insatsvaror påverkas inte det sammantagna utbudet av enskilda leverantörers problem. Enskilda problem tidigt i kedjan stör inte produktionen nedåt i kedjan, då det är lätt att byta leverantör. Andra insatsvaror är specialgjorda, i vilket fall störningar tidigt i leveranskedjan sprids nedåt. Då varje specialgjord insatsvara är en potentiell felkälla, är industrier med många specialgjorda insatsvaror mer känsliga för störningar i försörjningskedjan. Vi bygger in den

här mekanismen i en handelsmodell, och visar att länder med säkra försörjningskedjor kan förväntas specialisera sig i industrier med många specialgjorda insatsvaror.

Vi testar den här förutsägelsen genom att utföra en regressions-analys på internationella handelsflöden. Vi följer en metod som utvecklats i litteraturen för att testa teorier om komparativa fördelar. Metoden går ut på att utföra en regression av handelsflöden på fixeffekter av land och industri, samt på en *interaktionsterm* mellan en landegenskap och en industriegenskap (Romalis, 2004; Nunn, 2007; Nunn and Trefler, 2015). I vårt fall är landegenskapen ett mått på försörjnings-kedjans tillförlitlighet, och industriegenskapen är antalet specialgjorda insatsvaror som en industri använder. Vi visar att interaktionstermen är positiv, och att en standardavvikelses ökning i en försörjningskedjas tillförlitlighet och en industris störningskänslighet leder till en ökning av exporten om 10 procent. Denna effekt är robust med avseende på en rad modifikationer i den empiriska modellspecifikationen.

Litteraturförteckning

Hall, R. E. and Jones, C. I. (1999). Why do Some Countries Produce So Much More Output Per Worker than Others? *The Quarterly Journal of Economics*, 114(1):83–116.

Jones, B. F. (2014). The Human Capital Stock: A Generalized Approach. *American Economic Review*, 104(11):3752–3777.

Klenow, P. and Rodriguez-Clare, A. (1997). The Neoclassical Revival in Growth Economics: Has It Gone Too Far? Technical report.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42.

Nunn, N. (2007). Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade. *The Quarterly Journal of Economics*, 122(2):569–600.

Nunn, N. and Trefler, D. (2015). Domestic Institutions as a Source of Comparative Advantage. *Handbook of International Economics*, 4:263–315.

Romalis, J. (2004). Factor Proportions and the Structure of Commodity Trade. *American Economic Review*, 94(1):67–97.

Shimer, R. (2007). Reassessing the Ins and Outs of Unemployment. *NBER Working Paper*, 13421.

MONOGRAPH SERIES

1. Michaely, Michael *The Theory of Commercial Policy: Trade and Protection*, 1973

2. Söderström, Hans Tson *Studies in the Microdynamics of Production and Productivity Change*, 1974

3. Hamilton, Carl B. *Project Analysis in the Rural Sector with Special Reference to the Evaluation of Labour Cost*, 1974

4. Nyberg, Lars and Staffan Viotti *A Control Systems Approach to Macroeconomic Theory and Policy in the Open Economy*, 1975

5. Myhrman, Johan *Monetary Policy in Open Economies*, 1975

6. Krauss, Melvyn *International Trade and Economic Welfare*, 1975

7. Wihlborg, Clas *Capital Market Integration and Monetary Policy under Different Exchange Rate Regimes*, 1976

8. Svensson, Lars E.O. *On Competitive Markets and Intertemporal Resources Allocation*, 1976

9. Yeats, Alexander J. *Trade Barriers Facing Developing Countries*, 1978

10. Calmfors, Lars *Prices, Wages and Employment in the Open Economy*, 1978

11. Kornai, János *Economics of Shortage*, Vols I and II, 1979

12. Flam, Harry *Growth, Allocation and Trade in Sweden. An Empirical Application of the Heckscher-Ohlin Theory*, 1981

13. Persson, Torsten *Studies of Alternative Exchange Rate Systems. An Intertemporal General Equilibrium Approach*, 1982

14. Erzan, Refik *Turkey's Comparative Advantage, Production and Trade Patterns in Manufactures. An Application of the Factor Proportions Hypothesis with Some Qualifications*, 1983

15. Horn af Rantzien, Henrik *Imperfect Competition in Models of Wage Formation and International Trade*, 1983

16. Nandakumar, Parameswar *Macroeconomic Effects of Supply Side Policies and Disturbances in Open Economies*, 1985

17. Sellin, Peter *Asset Pricing and Portfolio Choice with International Investment Barriers*, 1990

18. Werner, Ingrid *International Capital Markets: Controls, Taxes and Resources Allocation*, 1990

19. Svedberg, Peter *Poverty and Undernutrition in Sub-Saharan Africa: Theory, Evidence, Policy*, 1991

20. Nordström, Håkan *Studies in Trade Policy and Economic Growth*, 1992

21. Hassler, John, Lundvik, Petter, Persson, Torsten and Söderlind, Paul *The Swedish Business Cycle: Stylized facts over 130 years*, 1992

22. Lundvik, Petter *Business Cycles and Growth*, 1992

23. Söderlind, Paul *Essays in Exchange Rates, Business Cycles and Growth*, 1993

24. Hassler, John A.A. *Effects of Variations in Risk on Demand and Measures of Business Cycle Comovements*, 1994

25. Daltung, Sonja *Risk, Efficiency, and Regulation of Banks*, 1994

26. Lindberg, Hans *Exchange Rates: Target Zones, Interventions and Regime Collapses*, 1994

27. Stennek, Johan *Essays on Information-Processing and Competition*, 1994

28. Jonsson, Gunnar *Institutions and Incentives in Monetary and Fiscal Policy*, 1995

29. Dahlquist, Magnus *Essays on the Term Structure of Interest Rates and Monetary Policy*, 1995

30. Svensson, Jakob *Political Economy and Macroeconomics: On Foreign Aid and Development*, 1996

31. Blix, Mårten *Rational Expectations and Regime Shifts in Macroeconometrics*, 1997

32. Lagerlöf, Nils-Petter *Intergenerational Transfers and Altruism*, 1997

33. Klein, Paul *Papers on the Macroeconomics of Fiscal Policy*, 1997

34. Jonsson, Magnus *Studies in Business Cycles*, 1997

35. Persson, Lars *Asset Ownership in Imperfectly Competitive Markets*, 1998

36. Persson, Joakim *Essays on Economic Growth*, 1998

37. Domeij, David *Essays on Optimal Taxation and Indeterminacy*, 1998

38. Flodén, Martin *Essays on Dynamic Macroeconomics*, 1999

39. Tangerås, Thomas *Essays in Economics and Politics: Regulation, Elections and International Conflict*, 2000

40. Lidbom, Per Pettersson *Elections, Party Politics and Economic Policy*, 2000

41. Vestin, David *Essays on Monetary Policy*, 2001

42. Olofsgård, Anders *Essays on Interregional and International Political Economics*, 2001

43. Johansson, Åsa *Essays on Macroeconomic Fluctuations and Nominal Wage Rigidity*, 2002

44. Groth, Charlotta *Topics on Monetary Policy*, 2002

45. Gancia, Gino A. *Essays on Growth, Trade and Inequality*, 2003

46. Harstad, Bård *Organizing Cooperation: Bargaining, Voting and Control*, 2003

47. Kohlscheen, Emanuel *Essays on Debts and Constitutions*, 2004

48. Olovsson, Conny *Essays on Dynamic Macroeconomics*, 2004

49. Stavlöt, Ulrika *Essays on Culture and Trade*, 2005

50. Herzing, Mathias *Essays on Uncertainty and Escape in Trade Agreements*, 2005

51. Bonfiglioli, Alessandra *Essays on Financial Markets and Macroeconomics*, 2005

52. Pienaar, Natalie *Economic Applications of Product Quality Regulation in WTO Trade Agreements*, 2005

53. Song, Zheng *Essays on Dynamic Political Economy*, 2005

54. Eisensee, Thomas *Essays on Public Finance: Retirement Behavior and Disaster Relief*, 2005

55. Favara, Giovanni *Credit and Finance in the Macroeconomy*, 2006

56. Björkman, Martina *Essays on Empirical Development Economics: Education, Health and Gender*, 2006

57. Larsson, Anna *Real Effects of Monetary Regimes*, 2007

58. Prado, Jr., Jose Mauricio *Essays on Public Macroeconomic Policy*, 2007

59. Tonin, Mirco *Essays on Labor Market Structures and Policies*, 2007

60. Queijo von Heideken, *Virgina Essays on Monetary Policy and Asset Markets*, 2007

61. Finocchiaro, Daria *Essays on Macroeconomics*, 2007

62. Waisman, Gisela *Essays on Discrimination and Corruption*, 2008

63. Holte, Martin Bech *Essays on Incentives and Leadership*, 2008

64. Damsgaard, Erika Färnstrand *Essays on Technological Choice and Spill- overs*, 2008

65. Fredriksson, Anders *Bureaucracy, Informality and Taxation: Essays in Development Economics and Public Finance*, 2009

66. Folke, Olle *Parties, Power and Patronage: Papers in Political Economics*, 2010

67. Yanagizawa-Drott, David *Information, Markets and Conflict: Essays on Development and Political Economics*, 2010

68. Meyersson, Erik *Religion, Politics and Development: Essays in Development and Political Economics*, 2010

69. Klingelhöfer, Jan *Models of Electoral Competition: Three Essays on Political Economics*, 2010

70. Perrotta, Maria Carmela *Aid, Education and Development*, 2010

71. Caldara, Dario *Essays on Empirical Macroeconomics*, 2011

72. Mueller, Andreas *Business Cycles, Unemployment and Job Search: Essays in Macroeconomics and Labor Economics*, 2011

73. Von Below, David *Essays in Climate and Labor Economics*, 2011

74. Gars, Johan *Essays on Macroeconomics of Climate Change*, 2012

75. Spiro, Daniel *Some Aspects on Resource and Behavioral Economics*, 2012

76. Ge, Jinfeng *Essays on Macroeconomics and Political Economy*, 2012

77. Li, Yinan *Institutions, Political Cycles and Corruption: Essays on Dynamic Political Economy of Government*, 2013

78. Håkanson, Christina *Changes in Workplaces and Careers*, 2013

79. Qin, Bei *Essays on Empirical Development and Political Economics*, 2013

80. Jia, Ruixue *Essays on the Political Economy of China's Development*, 2013

81. Campa, Pamela *Media Influence on Pollution, and Gender Equality*, 2013

82. Seim, David *Essays on Public, Political and Labor Economics*, 2013

83. Shifa, Abdulaziz *Essays on Growth, Political Economy and Development*, 2013

84. Panetti, Ettore *Essays on the Economics of Banks and Markets*, 2013

85. Schmitt, Alex *Beyond Pigou: Climate Change Mitigation, Policy Making and Distortions*, 2014

86. Rogall, Thorsten *The Economics of Genocide and War*, 2015

87. Baltrunaite, Audinga *Political Economics of Special Interests and Gender*, 2016

88. Harbo Hansen, Niels-Jakob *Jobs, Unemployment, and Macroeconomic Transmission*, 2016

89. Stryjan, Miri *Essays on Development Policy and the Political Economy of Conflict*, 2016

90. Karadja, Mounir *On the Economics and Politics of Mobility*, 2016

91. Kitamura, Shuhei *Land, Power, and Technology: Essays on Political Economy and Historical Development*, 2016