The Return to Capital: New Facts and Interpretation

Julia Faltermeier

TESI DOCTORAL UPF / ANY 2019

DIRECTOR DE LA TESI
Jaume Ventura i Luca Fornaro
Departament d'Economia i Empresa

To my family

Acknowledgements

First of all I would like to thank my advisors Luca Fornaro and Jaume Ventura for their guidance, encouragement, and patience throughout this research project. It has been very rewarding to learn from them how to distinguish the necessary from the superfluous and capture the essence of an idea in a simple model. I would also like to thank Manual Gracia-Santana for all his help, especially throughout the job market.

I was fortunate to learn and receive honest feedback during many seminars and meetings with CREi members, UPF faculty and visiting researchers. I am particularly grateful to Vladimir Asriyan, Fernando Broner, Antonio Ciccone, Alejandro Cuñat, Davide Debortoli, Julian di Giovanni, Jan Eeckhout, Harold Fadinger, Jordi Gali, Alberto Martin, David Nagy, Barbara Rossi, Edouard Schaal, Robert Zymek, and all the participants of the CREI Macroeconomics Breakfast, the CREI International Lunch, the Barcelona GSE Jamboree 2017 and the UPF Student Seminar for their helpful comments and suggestions. Marta Araque and Laura Agustí provided amazing support throughout the PhD, especially throughout the job market - it would not have been possible without you! I am also grateful to Bruno Guallar, Marta Ledesma and Esther Xifre for all the administrative support at the BGSE. Also, I am thankful for the excellent teaching offered at Newnham College, University of Cambridge, the Barcelona Graduate School of Economics and during the MRes at Universitat Pompeu Fabra which has prepared me for the PhD.

I thank the Fundació LaCaixa for the generous financial support from the La Caixa-Severo Ochoa Fellowship.

The PhD would have been a lonely experience without companions to embark on this long journey. Derrick and Christoph, you were the best office mates I could have imagined. Marta, Federica, Yimei, Christian, thanks for all the inspiring conversations. Cristina, thanks for all the research we did together and the many Python tricks I learnt from you.

This thesis is dedicated to my family whose love and support has brought me here today. You have guided me in my decisions and always provided a peaceful retreat with time to recharge and reflect. Miquel, you have made me feel at home in Barcelona. I am excited about all the adventures waiting for us. Moltes gràcies!

Abstract

This thesis presents new facts on the aggregate return to capital and shows their implications for underlying drivers of investment and the efficiency of the allocation of capital.
The first chapter analyzes aggregate returns to capital across countries. Although international capital flows were small, returns have converged since the 1970s. Trade integration appears a key driving force behind this trend because factor income shares of capital move with the aggregate return. As barriers to international trade fall, specialization in capital- or labor-intensive industries can explain a significant share of the convergence.
The second chapter examines why the return to capital in the U.S. has not fallen along with interest rates. Using firm-level data in production function estimation addresses some shortcomings of aggregate data. The results suggest that higher capital frictions dampened investment demand despite low interest rates. Rising markups, on the other hand, appear to have limited explanatory power.

Resum

Aquesta tesi presenta noves aportacions pel que fa al retorn agregat del capital i mostra les seves implicacions amb els factors d'inversió subjacents i l'eficiència de l'assignació de capital.
En el primer capítol s'analitza el retorn agregat del capital dels països. Tot i que els fluxos de capital internacional eren escassos, les rendibilitats han convergit des de la dècada de 1970. La integració del comerç és causant d'aquesta tendència, ja que el factor ingressos del capital es mou amb el retorn agregat. Com que les barreres al comerç internacional cauen, l'especialització en les indústries intensives en capital o en mà d'obra poden explicar una part important d'aquesta convergència.
El segon capítol examina per què el retorn del capital als Estats Units no ha caigut amb els tipus d'interès. L'ús de dades d'empreses en l'estimació de la funció de producció corregeix algunes deficiències de les dades agregades. Els resultats suggereixen que les friccions de capital més elevades han fet disminuir la demanda d'inversió, malgrat els baixos tipus d'interès. L'augment dels marges, però, no pot explicar aquest desenvolupament.

Preface

The dynamics of investment and the allocation of capital are central components to understand the drivers of economic growth. This thesis documents new facts on the return to capital and proposes how they may be reconciled with existing theories of international finance, trade, and investment. In the absence of frictions, arbitrage between investment opportunities will equalize the return across projects and outside uses of funds, such as investing in government bonds. The return to capital, through its close link with the marginal product of capital, thus helps to differentiate if a low investment is the result of a lack of profitable investment opportunities or if there are frictions that prevent the reallocation of funds across countries or firms. Thus studying the dispersion in returns to capital and how it evolves over time can inform the type of frictions economic modeling needs to take into account and can direct policy-making towards effective measures to improve the efficiency of the allocation of capital. Despite its usefulness, the return to capital is nearly not as well explored as its labor counterpart, wages, mainly due to measurement challenges. Whereas wage payments are usually observable, the return to capital on the firm- or country level can only be inferred using at least a minimum of assumptions. Moreover, the importance of financial frictions and the impact of misallocation of resources have received more attention, both with the global financial crisis but also with lack of convergence of many emerging markets to advanced economy levels of income. Drawing on recent revisions in aggregate data across countries and firm-level data in the second chapter, this thesis contributes through careful measurement of the return to capital, the description of a number of stylized facts and an explanation of how they may fit into the existing economic theory.

In the first chapter, I find that aggregate marginal products of capital across countries have converged over time and I propose trade integration as an important driving force behind this trend. Using newly available data, I construct measures of the marginal product of capital across countries and over time following Caselli and Feyrer (2007)'s methodology. I show that countries have converged in marginal products since 1970 which indicates that the productive efficiency of the allocation of capital across countries has improved. However, this is not the result of large-scale reallocation through international capital flows. Instead, I demonstrate how the dramatic increase in world trade over the last decades explains a significant fraction of the convergence of marginal products. I write a tractable multi-country, multi-sector model that features both inter- and intraindustry trade. In my model, trade integration leads to convergence through two channels. The first one is specialization in labor vs. capital-intensive sectors commonly-known from Heckscher-Ohlin trade theory. The second channel acts through the response of aggregate savings to movements in wages and returns which affect capital accumulation and lead to convergence in the relative factor endowments. The predictions of this dynamic model are consistent with the new empirical facts and other well-known characteristics of the cross-country income distribution. After I calibrate the model to match capital-output ratios, capital shares, and bilateral trade flows, I find that trade integration can explain about 30 % of the convergence in marginal products.

In the second chapter, I document that standard measures of the U.S. aggregate return to capital do not match the decline in real interest rates. This paper evaluates the contribution of three economic developments to explain this trend, (i) constraints in firms' ability to invest in additional capital, (ii) over-estimation of returns due to a rise in pure profits, and (iii) improvements in the efficiency of the allocation of capital across firms. Aggregate macro data provide limited information to differentiate between the first two explanations and is silent on capital misallocation across firms. Therefore, this paper exploits firm-level evidence to disentangle the underlying structural forces driving the aggregate trend. The evidence points towards capital frictions as the main reason for low investment. Despite rising markups, profits explain only around a third of the divergence. Finally, the paper compares the U.S. experience with the data for the six largest European economies and documents heterogeneity across countries and industries in Europe.

Contents

List of Figures

List of Tables

Chapter 1

IS THE DISTRIBUTION OF CAPITAL ACROSS COUNTRIES EFFICIENT? - CONVERGENCE IN RETURNS TO CAPITAL AND THE ROLE OF TRADE

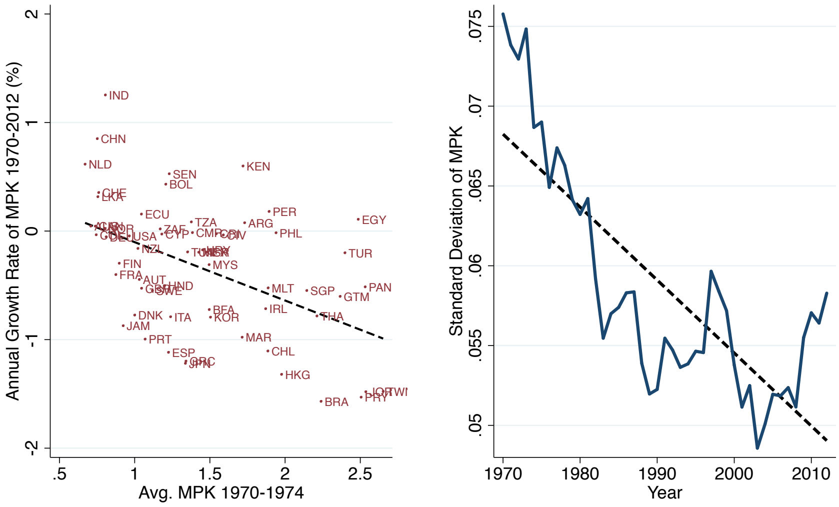
1.1 Introduction

The distribution of aggregate marginal products of capital (MPKs) indicates whether the allocation of capital across countries is efficient. Persistent differences in MPKs suggest unexploited gains from reallocating capital across borders. Following Caselli and Feyrer (2007)'s methodology, Figure 1.1a presents a measure of the aggregate MPK in 1970 against its growth rate between 1970 to 2012. There is clear mean-reversion - economies with a high MPK in the 70s reduced their MPK relative to low-MPK ones. Figure 1.1b plots the standard deviation of MPKs across countries over time. The dispersion decreases over time. Both figures illustrate that MPKs have equalized over time. This convergence suggests that the productive efficiency of the allocation of capital across countries has improved.

What drives the convergence in MPKs across countries? The answer to this question provides us with a better understanding of the economic forces at play in globalizing markets and matter for informed policy decisions. Financial integration appears to be a likely explanation. As capital accounts open, international financial markets reallocate capital towards high-MPK economies which would result in convergence of MPKs. Figure 1.2 shows capital inflows between 1981

Figure 1.1: Convergence in MPK 1970-2012

(a) Mean-reversion (b) Fall in dispersion



Notes: (a) Geometric growth rate and MPK at the beginning of the sample are calculated relative to the world. The OLS coefficient is -.5377, significant at the 1% level. (b) The standard deviation of MPKs reduces on average by .0005 each year.

and 2007 against the change in MPK, but there is no significant correlation. In general, countries which reduced their MPKs did not receive substantial capital inflows with some, in particular among the Asian Tigers, even lending internationally. Moreover, net capital inflows are small in terms of initial output.

Instead, converging economies are characterized by above-average increases in trade over GDP. Can trade integration account for the convergence in MPKs? To answer this question, note that the aggregate MPK equals the share of income paid to capital per dollar of capital stock. A decomposition shows that MPKs adjust through both capital accumulation and movements in the capital share. This paper shows that the global fall in trade cost affects MPKs through both of these channels and can explain the convergence in MPKs. Although all types of trade have increased dramatically over the period, the growth in North-South trade was particularly strong. This rise in interindustry trade has enabled economies to specialize in the production of goods that align best with their factor endowments. This specialization manifest itself in the observed changes in factor income shares and factor prices. Instead of factor reallocation across borders, factors are embodied in goods trade.

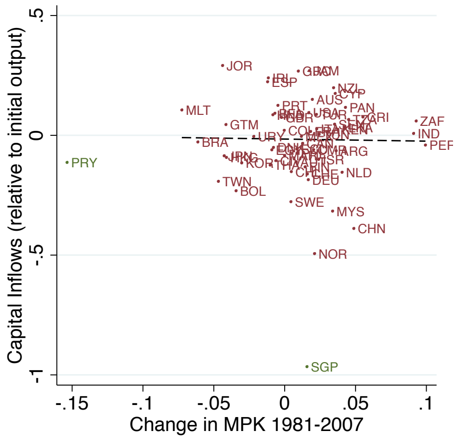
After the first part of the paper discusses empirical evidence in detail, the second section provides a theory to interpret these findings. A tractable dynamic, multi-country Heckscher-Ohlin-Krugman model of costly trade shows how a fall in trade cost shifts the world equilibrium to a new steady state with almost equalized MPKs. In this model, MPKs converge across countries when relative factor endowments become more similar or when countries specialize in production which uses abundant factors intensively. The fall in trade cost affects MPKs through both of these channels. Firstly, the increase in demand for exports raises the demand for relatively abundant factors which fosters specialization along factor-intensities and eliminates differences in factor prices. Secondly, aggregate savings respond to the movements in wages and returns to capital. In particular, the increase in labor income in labor-abundant, developing economies stimulates savings and increases the steady-state level of capital per worker. The resulting convergence in factor endowments further equalizes MPKs across countries.

Finally, the third part of the paper evaluates the importance of trade integration compared to other drivers, namely changes in productivities and preferences. I estimate factor-specific productivities, trade costs and rates of time preference such that the model matches capital-output ratios, capital shares and bilateral trade flows for ten economies. Then I compare the initial steady state in the 1970s to the integrated steady state just before the global financial crisis. For a reasonable choice of the model parameters, this accounting exercise suggests a substantial fall in trade costs, but also significant movements in productivities. A counterfactual analysis quantifies the relative contribution of the decline in trade costs and attributes around 30 % of the convergence to trade integration alone.

1.2 Related Literature

This paper is closely related to two strands of literature. The first one examines the (productive) efficiency of the allocation of capital. The second one examines the effects of international trade and financial integration. Since Lucas (1990), economists have tried to reconcile the large differences in capital-labor ratios across countries with increasing economic integration. Heathcote and Perri (2013) provide an excellent overview of both current theory and empirical evidence to understand international efficiency. This paper builds considerably on the influential paper by Caselli and Feyrer (2007). They propose a simple methodology to measure the aggregate MPK which this paper follows. The measure uses widely available national accounts data and requires relatively mild assumptions. Despite the vast differences in economic development, income and capital endowment per worker, they find that MPKs are mostly equal. Their findings suggest no substantial gain from reallocating capital internationally which seems to indicate that international financial markets work well. Cross-country income differences

Figure 1.2: Capital Inflows and Changes in MPKs



Notes: Using data from Lane and Milesi-Ferretti (2007) net capital flows are computed from the cumulative current account deficit converted into constant international dollars using the price of investment goods provided in the PWT. The sample covers 60 countries from 1981-2007 (excluding financial crisis). Singapore and Paraguay are excluded as outliers from the regression. Capital flows data is not available for some countries before 1981.

appear to be mainly due to productivity differences.

This result is a striking. Considering what the recent financial crisis has taught economists about the importance of financial frictions, one wonders why they would not matter at least equally if not more for the international financial system. Although the empirical results in this paper also suggest similar MPKs towards the end of the sample, the missing correlation with capital flows and the movement in capital shares question the role of financial integration in equalizing MPKs. Instead, this paper proposes trade integration and productivity changes as the main drivers of convergence which suggests that international financial frictions may well be significant and co-exist with equalized MPKs.

This paper is not the first one to be interested in Caselli and Feyrer's results. On the one hand, several papers, some still unpublished, re-examine the measure, confirm that MPKs are not higher in developing countries and propose different explanations. Lowe et al. (2012) attribute the finding to distortions between the private and the public sector in developing countries. Chirinko and Mallick (2008) highlight the role of adjustment costs. Chatterjee and Naknoi (2010) take a business cycle perspective and model predicted capital flows driven by shocks to the efficiency of producing investment goods. They find minimal welfare gains of financial integration.

On the other hand, some papers challenge Caselli and Feyrer's findings. David et al. (2016) also take a time-series perspective and see higher returns to capital in developing countries using a measure that is similar to Caselli and Feyrer (2007). They explain the difference with risk premia using an asset pricing framework. However, their approach focuses on cross-country differences in average returns and does not consider convergence over time. A recent paper by Monge-Naranjo et al. (2015) re-calculates MPKs with recently available data for natural resources. They find substantial misallocation of capital across countries which they attribute to the role of human capital accumulation. Although the present paper builds on similar empirical insights on the MPK, it links them to trade integration and structural change. Instead of quantifying the degree of inefficiency, it focuses on the forces that have reduced misallocation.

Secondly, this paper builds on an extensive literature on the effects of international trade and financial integration. Aizenman et al. (2007) document that most capital is self-financed, suggesting that economies accumulate capital mostly through domestic savings and not through international borrowing. Similarly, an influential paper by Gourinchas and Jeanne (2013) finds that against the predictions of neoclassical theory fast-growing emerging economies did not receive capital inflows, now widely known as *allocation puzzle* or *upstream capital flows*. Under the premise of open international capital markets, they conclude that the main divergence between neoclassical theory and the data lies in the savings response. In accord with Caselli and Feyrer (2007), the absence of noteworthy differences in returns suggests that investment distortions are not the source of the discrepancy. Their puzzle has sparked a broader literature to re-examine international capital flows and develop theories to rationalize the *savings puzzle*. For example, Alfaro et al. (2014) show that upstream flows are driven by public instead of private flows. This paper offers an alternative framework to interpret these findings. If equalized MPKs are in fact not the product of large-scale financial integration, but the result of trade integration, the premise of open capital markets can be misleading. This interpretation suggests that international financial frictions may well be important to understand aggregate savings and international capital flows.

Also, the theory proposed in this paper builds on a vast literature on the effects of trade integration. The model is close to Romalis (2004) which combines the classical Heckscher-Ohlin model with Krugman (1980) to introduce a continuum of sectors. Zymek (2015) and Fadinger (2011) show that the Heckscher-Ohlin model can provide useful insights in understanding the growth in factor-trade and factor-specific productivity. Ventura (1997) uses HO trade in a Ramsey growth model to explain the East Asian Miracle and Acemoglu and Ventura (2002) shows how trade can lead to a stable world income distribution. Cuñat and Maffezzoli (2007) also introduce dynamics to a Heckscher-Ohlin model although in an infinite horizon setting whereas the present model focuses on life-cycle savings. Jin

(2012) shows how factor trade can generate upstream capital flows along the business cycle, but the framework does not explain differences in MPKs. Instead, this paper views factor trade and financial flows as substitutes as already found in Samuelson (1971).

Finally, the findings in this paper are also related to the recent interest in the decline of the labor share as shown in Karabarbounis and Neiman (2014) and Elsby et al. (2013). This literature has revised the common assumption that factor income shares are constant and proposes various explanations for the trends observed in numerous countries. In this paper, I confirm the variation over time and across countries and show that they relate systematically to movements in MPKs. This is compatible with structural change driven by trade integration, the key mechanism, although not necessarily inconsistent with alternative explanations.

1.3 Section I - Empirical Evidence

This section shows how the aggregate marginal product of capital can be measured, the data used and potential limitations. Then it discusses in detail what explains differences in MPKs across countries and which are the margins of adjustment. This part concludes with additional evidence on trade integration, financial flows, and world aggregates.

1.3.1 Measurement

The Return to Capital

The return to capital shows the payoff from buying and renting out one unit of capital. Large differences in returns across countries imply gains from reallocating capital between countries. Let's denote the aggregate produced capital stock at time t as K_t, the rental rate as r_t and the price of capital as P_t^K. The return to capital R_t consists of the rental income $r_t P_t^K K_t$ plus the capital gain $(1-\delta)P_{t+1}^K K_t$ which depends on the future price of capital P_{t+1}^K and the depreciation rate δ. As pointed out by Caselli and Feyrer (2007), under the assumptions of perfect competition and constant returns to scale in production, the rental rate equals the marginal product of capital (MPK) and it can be computed without assuming a specific production function. Aggregate capital income can be written as $r_t P_t^K K_t = \alpha_t P_t^Y Y_t$ where α_t denotes the capital share in nominal GDP $P_t^Y Y_t$. Thus the return to capital in time t is given by:

$$\text{Return}_t = \underbrace{\alpha_t \frac{P_t^Y Y_t}{P_t^K K_t}}_{\text{MPK}} + \underbrace{(1-\delta_t)\frac{P_{t+1}^K}{P_t^K}}_{\text{Capital Gain}} \tag{1.1}$$

Caselli and Feyrer (2007) emphasize that to test the efficiency of the capital allocation correctly one should compare the return from investing one more dollar in capital, and not the return from investing one more unit of capital. This difference matters because as shown by Hsieh and Klenow (2007) the relative price of investment goods tends to be higher in developing countries. Thus failing to correct for the relative price of capital would bias the estimate for developing countries upwards. Moreover, they propose to adjust the capital share for land and natural (or non-produced) capital. Why? Normally, the capital income share is calculated as the residual income not recorded as labor income. For economies rich in natural resources such as oil, this methodology would falsely attribute the depletion of natural resources to capital income whereas the capital stock obtained from the perpetual inventory method only includes produced (reproducible) capital. The measure in this paper takes the relative price of capital as well as natural resources into account. Caselli and Feyrer (2007) did not have access to data on country-specific depreciation rates and thus assumed it to be equal focusing only on the MPK-part of the return, whereas I also analyze the capital gain. However, the MPK-term remains at the center of the analysis because it is closely linked to long-run trends whereas the capital gain is mainly driven by short-run fluctuations in prices. Apart from this extension, I follow the same methodology and all remaining differences stem from data revisions.

Data

The Penn World Tables Version 9.0 is the primary data source for aggregate time series made comparable across countries such as real GDP, the labor share, the capital stock, price levels of output and capital and depreciation rates. The extensive documentation of the Penn World Tables as Feenstra et al. (2013) provides more details. Additional data on the natural resources rents come from the World Bank database. Thus I can subtract the share of non-reproducible capital directly and do not need to impute it from wealth data as proposed by Caselli & Feyrer. Appendix A contains more information on the measurement. I drop outliers and only use countries for which all the required variables are available over the whole period to construct a balanced-panel from 1970-2012. This procedure avoids that outliers or new countries entering the sample drive the results, at the cost of potentially underestimating the variation.

Table 1.1: Which components account for difference between the original vs. revised MPK?

Variable	**% contribution to revision**
Capital share α	74.8
Rel. price $\frac{P^Y}{P^K}$	23.7
Output-capital ratio Y/K	1.6

Notes: These numbers are based on 1996. The decomposition comes from $Var(\overline{MPK}) = Cov(\overline{MPK}, \overline{\alpha}) + Cov(\overline{MPK}, \overline{\frac{P^Y}{P^K}}) + Cov(\overline{MPK}, \overline{\frac{Y}{K}})$ where $\overline{x} = log(x^{Revised}) - log(x^{CF})$

The final dataset contains 60 countries - of which 23 are OECD members. In 2011, the poorest country in the sample is Burkina Faso with a real GDP per capita of 1,344 USD. The richest country is Norway with a real GDP per capita of over 80,000 USD. Computing the world return as a cross-country average weighted by the capital stock, on average the return to capital is 1.12 over the whole sample. It is slightly higher for 1970-1974 with 1.19 and decreases over time to 1.09 in the last five years of the dataset (2008-2012). However, the downwards trend disappears once we account for inflation. Sri Lanka has the lowest return on average over the sample, with just 7 %, Panama has the highest average return with 25 %. In 1970 the standard deviation is .095, and it falls to .073 by 2012.

Comparison with Caselli and Feyrer

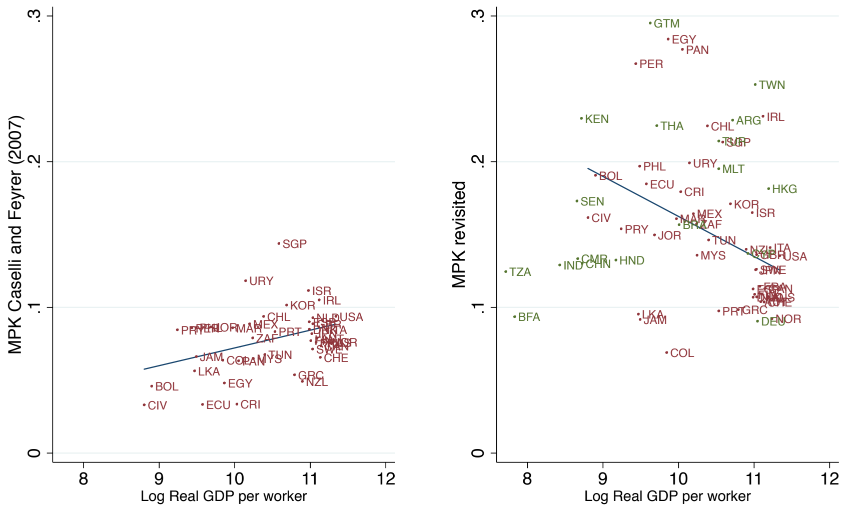
In terms of data quality there are four main improvements which affect the calculations: (i) new data available on the natural resources rents, (ii) new International Price Comparison (IPC) surveys and changes in the estimation methods improve data on relative prices, (iii) the measure of capital stock available in PWT 9.0 accounts for different types and vintages and (iv) revisions of GDP, especially for developing economies in more recent years.

How do the revisions of the data affect the measurement of the MPK? This section compares the revised measure with the original results from Caselli and Feyrer (2007) for the year 1996 using the same set of countries. Figure 1.3 shows the result obtained by Caselli & Feyrer that the MPK is uncorrelated with income and contrasts it with the results of the revised data. We can see that i) on average the revised MPK is higher, ii) the revisions disproportionately affect developing countries and iii) the correlation is negative, although only significant for the original set of countries plotted in red.

Which component of the data revision explains the stark difference? Table 1.1 decomposes the variation of the difference between the original and the revised

Figure 1.3: Revisiting Caselli & Feyrer (2007)

(a) Original result for 1996 (b) Revised MPK for 1996



Notes: (a) This graph uses the original data used by Caselli and Feyrer as presented in their Data Appendix. A linear regression gives a positive slope coefficient which is not significantly different from zero. For more details on the replication see the Appendix A. (b) A linear regression shows a negative slope coefficient which is significantly different from zero for the countries included in CF's original analysis (in red). Additional countries not included in CF are plotted in green. For the total sample of 60 countries, the coefficient is negative, but not statistically significant for the year 1996.

MPK into three components. It finds that revisions to the capital-share and the relative price of capital explain most of the difference. Alterations in the capital share stem from revisions of the labor share and the fact that I use newly available data on natural resource rents from the World Bank. Changes in the output-capital ratio only have a minimal effect. The updated database suggests that the PWT 6.1 underestimated the MPK in emerging economies and understated the variation between countries. This finding is similar to Monge-Naranjo et al. (2015) who use related data on natural resource rents and find significantly more capital misallocation than previously thought.

Discussion of the measure

What can we learn from the returns and what are the limitations of this measure? First of all, it captures the aggregate return to aggregate capital stock on the country-level without considering potential domestic financial frictions. Therefore, it measures the average return and hence will be an imperfect approximation to the marginal return. Recognizing that international investors or policymakers would be likely to face the same constraints, this aggregate measure is still very informative. Note, however, that this prevents us from determining to what extent improvements in domestic financial systems explain changes in MPKs. Until better sector or firm-level data becomes available for developing economies, we can only speculate about the importance of this channel. Commonly used indicators of financial development do not explain the variation in returns once GDP is included, but neither do they capture perfectly the potential domestic frictions at play nor is it clear what correlation one should expect without additional information.

Secondly, the standard way to compute the capital share assumes there are no pure profits. This assumption would be especially problematic if profit shares are large or vary a lot across countries and over time which would overestimate the MPK in countries with large profit shares. Despite the recent rise in research interest on market power, this limitation is unavoidable at the aggregate level. Papers in which the profit share is computed from aggregate data (e.g., Barkai (2016)) need to assume the return. They usually use financial returns from bonds or stock markets, which might be entirely different from the return to physical capital, especially in the presence of financial frictions.

Thirdly, this paper measures the realized or ex-post return, but economic theory normally makes predictions about the expected (ex-ante) return, which is unobserved. Using averages over several years can partly account for this problem if shocks are not very persistent, but nevertheless this measurement approach is better suited for long-run trends, whereas short-run fluctuations will always be determined by a combination of choices and unforeseen realizations of shocks. Finally, using the changes in the price of capital to compute the capital gain implicitly assumes that there is a well-functioning secondary market for capital such that the price of capital truly reflects the price at which capital could be sold on. Since the main results of this paper are related to the MPK, this is less of a concern, but this also highlights the need for better data on capital and investment good prices across countries.

Finally, since this paper focuses on the long-run trends in the return to capital, it ignores differences in returns that stem from differences in risk. David et al. (2016) explain cross-country differences in a similar measure of the return to capital to the one used in this paper by risk-premia in an asset pricing framework. This alternative explanation for differences in returns is not pursued in this paper. As their framework builds on risk-diversification of international investors, they also suppose that international financial markets are integrated. However, it remains unclear whether changes in risk premia would also be a good explanation for convergence. In principle, growth in trade and international financial flows could interact with risk in different ways which could be an additional channel for convergence.

1.3.2 Preview of stylized facts

The empirical analysis shows a set of new facts about the aggregate MPK across countries and over time.

1. **Across countries**
   (a) Negative correlation between the capital-labor ratio and the MPK
   (b) Positive correlation between the capital share and the MPK
2. **Across time**
   (a) Negative correlation between the capital-labor ratio and the MPK
   (b) Positive correlation between the capital share and the MPK
   (c) MPKs have converged over time
   (d) No correlation between capital inflows and changes in MPKs
   (e) Trade has increased over time, especially North-South Trade

1.3.3 Factor Endowments and Factor Income Shares

If the allocation of capital across countries is efficient from a production point of view, the return to capital should be equalized across countries and we would not expect to find any systematic relationship with income. Caselli and Feyrer (2007) find no correlation between the MPK and GDP per capita for 1996, but do not explore further what else explains the dispersion. Revisiting the relationship between GDP and returns using the whole panel shows that (i) countries with higher GDP on average tend to have a lower return and (ii) countries with a higher capital-labor ratio have on average lower returns to capital. The regression results are presented in Table 1.2. An economy with twice as much output per worker than the world average has on average a return 1.1 percentage points below the world average. A similar number is found for capital per worker, where twice the world level reduces the average return by 1.3 percentage points compared to the world average. This finding shows a systematic difference between countries, which suggests gains from reallocation of capital. Nevertheless, a repetition of the exercise in Caselli and Feyrer (2007) would lead to the same result - a regression of the return on GDP per worker in 1996 shows no statistically significant relationship between the two variables.

Moreover, countries with a higher return have a systematically higher capital share. As shown in the last line of Table 1.2 a capital share twice the world

Table 1.2: OLS Regression for Avg. Return 1970-2012

Explanatory variable	Slope coefficient	Constant	R^2	Obs
Avg. $\frac{Y}{L}$	-0.011*	0.041***	0.063	60
	(0.054)	(0.000)		
Avg. $\frac{K}{L}$	-0.013***	0.045***	0.133	60
	(0.004)	(0.000)		
Avg. Capital Share	0.426***	0.005	0.624	60
	(0.000)	(0.182)		

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; p-values in parentheses. Averages are computed over the deviation from the cross-country mean weighted by capital to control for time trends.

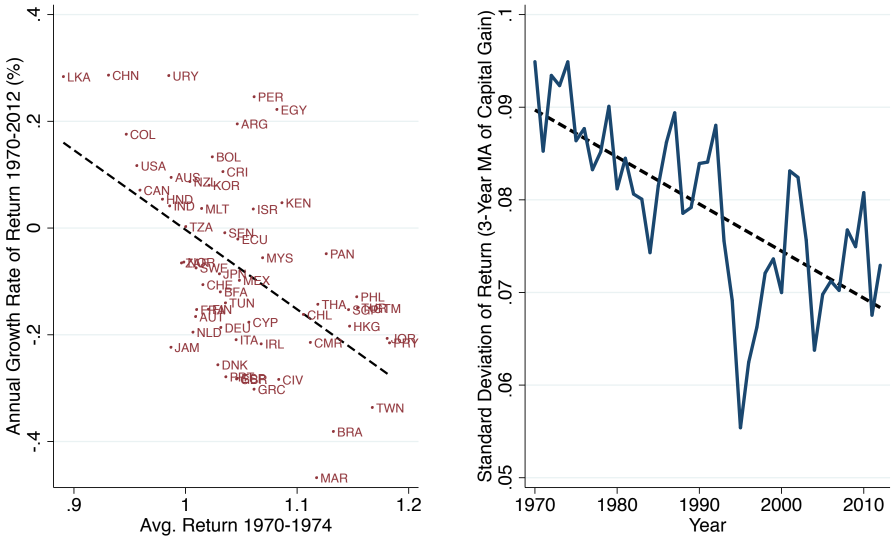
average is associated with a 4.26 percentage point higher return than the world average. This result also highlights that the capital share is far from constant across countries, as is often assumed in macro models. This observation questions the common assumption of a constant labor share not only over time but also across countries.

1.3.4 Convergence in returns across countries

The growth literature differentiates between two related concepts of convergence - β and σ convergence. Here β - convergence is present when countries with a low return show a larger increase over the period and vice versa. This concept of convergence is closely linked to long-run trends that are related to shifts in fundamental variables or convergence dynamics towards some steady state. We usually speak of σ - convergence when the dispersion decreases over time. σ convergence can be the consequence of β convergence, but it may also be the result of a change in the variance of disturbances or a change in the sensitivity to disturbances. Thus β and σ often occur jointly, but observing one does not generally guarantee the other. In this case, the distribution of returns across countries does exhibit both β and σ convergence. Figure 1.4a plots the geometric growth rates of returns against the initial level at the beginning of the sample. Countries with a low initial level increase their returns to capital significantly faster over the sample. Countries with relatively high returns tend to reduce their return relative to the world average. Hence the returns to capital exhibit β convergence. It shows that most OECD economies had already relatively low returns in the early 1970s and their return has not moved much over the whole sample. Many developing economies such as Thailand started with a high return which has reduced over time. It is also

Figure 1.4: Convergence in Returns to Capital

(a) β - Convergence (b) σ - Convergence



Notes: (a) Geometric growth rates and return at the beginning of the sample are calculated relative to the world. OLS coefficient of -1.491 significant at the 1% level. A level of MPK twice the world MPK is associated with a 1.491 percentage points lower growth rate (b) The standard deviation in the return reduces on average by 0.0005 each year or .021 over the whole sample.

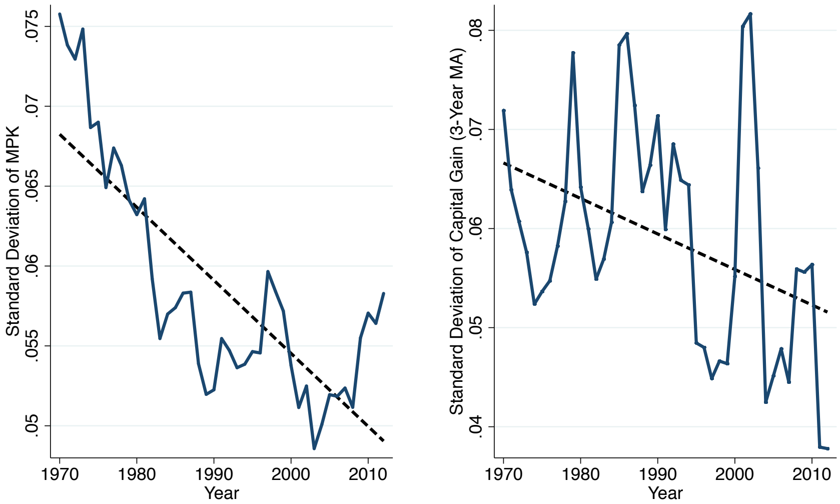
worth highlighting that both India and China did start out with relatively low returns despite their low initial capital-labor ratios. This suggests that for these two economies, the catch-up in productivity over the last decades cannot be ignored to understand their economic success.

Figure 1.4b plots the standard deviation of returns across countries over time. There is also a marked reduction. A robustness check looks at the difference between the 90th and the 10th percentile which decreased from about 0.24 to 0.18 (Figure 1.18b). The reduction suggests that the variation in returns was quite substantial especially prior to 1996, the year studied by Caselli and Feyrer (2007). Then, the dispersion reduced significantly. Thus σ convergence in the return to capital is also present.

Previously, the return to capital has been defined as the sum of the MPK-term and the capital gain. As already shown in Fig. 1.1a and1.1b in the introduction, β and σ convergence are also found for the MPK. Figure 1.5 compares the standard deviation in both components over time. Although the dispersion in both parts shows a downward trend in both components, the convergence in MPKs is particularly strong and universal. The capital gain term is more volatile which reflects

Figure 1.5: Standard deviation across Countries

(a) MPK (b) Capital Gain



Notes: (a) The standard deviation of MPKs reduces on average by .0005 per year. (b) The standard deviation of the capital gain reduces on average by .0004 per year.

the fact that it is mainly computed from price data whereas the MPK is determined by relatively slow-moving economic aggregates. In the subsequent analysis, this paper focuses on the behavior of the MPK which is closely related to factor endowments and factor income shares and which drives most of the convergence in returns. It also corresponds more closely to the measure used by Caselli & Feyrer.

1.3.5 Channels of convergence

This section analyzes through which channels countries adjusted their MPKs over time. For each country, the convergence in the MPK-term can be decomposed into three main components: (i) adjustment in capital share α, (ii) adjustment in relative prices of capital goods and (iii) adjustment in the output-capital ratio. Thus the growth in the MPK relative to the world consists of the growth differentials of each component relative to the world average[1]. These growth differentials reveal the margins of adjustment, for example, to see if countries that increased their

[1]The world-average of a variable is measured as average across countries weighted by the capital stock.

Table 1.3: Growth decomposition 1970-2012

	Contribution (%)		
Time period	g_α	$g_{\frac{Y}{K}}$	$g_{\frac{P^Y}{P^K}}$
1970-2012	31.9***	79.3***	-11.3
1970-1980	25.8***	67.9***	7.1
1980-1990	48.8***	40.1***	13.3
1990-2000	45.5***	36.3*	19.6
2000-2012	22.7***	56.6***	21.1

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

MPK relative to the world also increased their capital share relative to the world.

$$\left(g_{MPK} - g_{MPK}^{World}\right) = \left(g_\alpha - g_\alpha^{World}\right) + \left(g_{\frac{P^y}{P^k}} - g_{\frac{P^y}{P^k}}^{World}\right) + \left(g_{\frac{Y}{K}} - g_{\frac{Y}{K}}^{World}\right) \tag{1.2}$$

Table 1.3 presents the findings of this exercise. The decomposition for the whole sample and different sub-periods shows that the growth in MPKs is predominantly related to changes in the output-capital ratio and the capital share. There does not seem to be a systematic relationship between having experienced changes in relative prices and convergence. Over the whole sample period, roughly 75 % of the change in MPK happens through an adjustment of the output-capital ratio, and the remaining 25 % come from shifts in the capital share. Changes in the relative price of capital are not significantly related to the change in MPK. Changes in the capital-share appear to be particularly important in the 80s and 90s where they explain more of the change than the output-capital ratio.

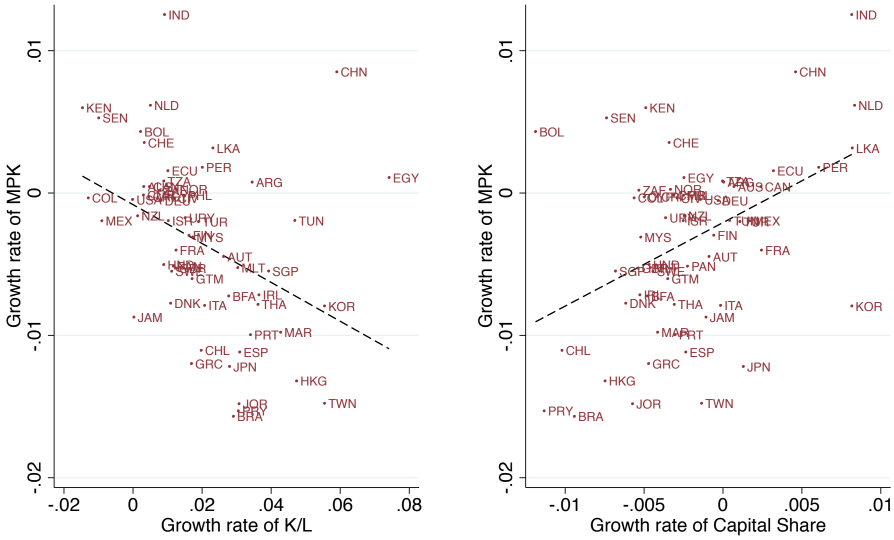
Movements in the output-capital ratio may be the result of capital accumulation or changes in productivities. Figure 1.6a shows a negative correlation between the growth rate of MPKs and the growth rate of capital per worker relative to the world. The quantitative exercise in the last part of the paper finds the factor productivities consistent with the data and decomposes the respective contributions through the lens of the model.

1.3.6 Financial integration

Financial integration is a likely candidate explanation. Since the reduction in MPKs goes along with an increase in the capital-labor ratio, high MPK coun-

Figure 1.6: Channels of convergence

(a) Capital accumulation (b) Capital Share



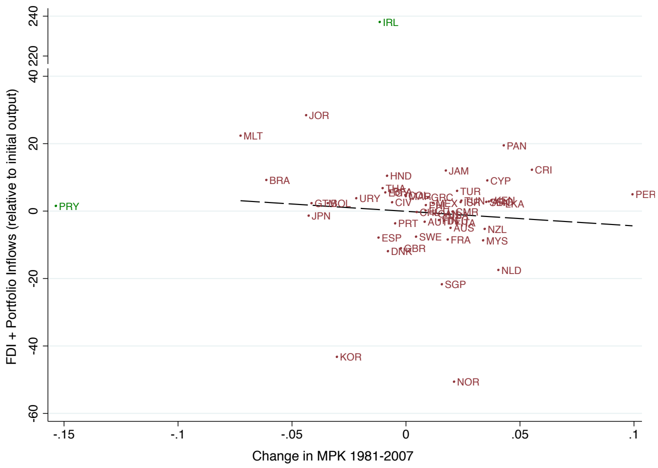
Notes: Geometric growth rates are calculated relative to the world. (a) OLS coefficient of -.137 is significant at the 1% level. (b) OLS coefficient of .587 is significant at the 1% level.

tries may have received capital inflows. However, the evidence does not suggest large-scale reallocation through capital inflows. Figure 1.2 shows capital inflows between 1981-2007 over initial output against the change in MPK. Here we would expect to see a negative correlation, but the coefficient from an OLS regression, although negative, is not significantly different from zero[2].

The measure of capital inflows follows a similar approach as Gourinchas and Jeanne (2013). The database used for international capital flows is an updated and extended version of the External Wealth of Nations Mark II database (EWN) by Lane and Milesi-Ferretti (2007). Net capital inflows are measured in current U.S. dollars using the data on current account deficits. To make this data comparable to the output and capital as reported in the PWT, it has to be converted into constant international dollars. Here the price of traded goods would be the ideal price in-

[2]Singapore and Paraguay are excluded as outliers from the regression, but including them does not alter the result. Capital flows data is not available for all countries for the 70s, but repeating the exercise with all countries with available data also does not yield in any significant result, although the sign of the coefficient is negative. Here the financial crisis is not included in the sample period because the interest is on long-run trends and financial flows were especially affected by the downturn. A robustness check showed that including data up to 2011 does not change the findings

Figure 1.7: FDI + Portfolio Inflows and Changes in MPKs

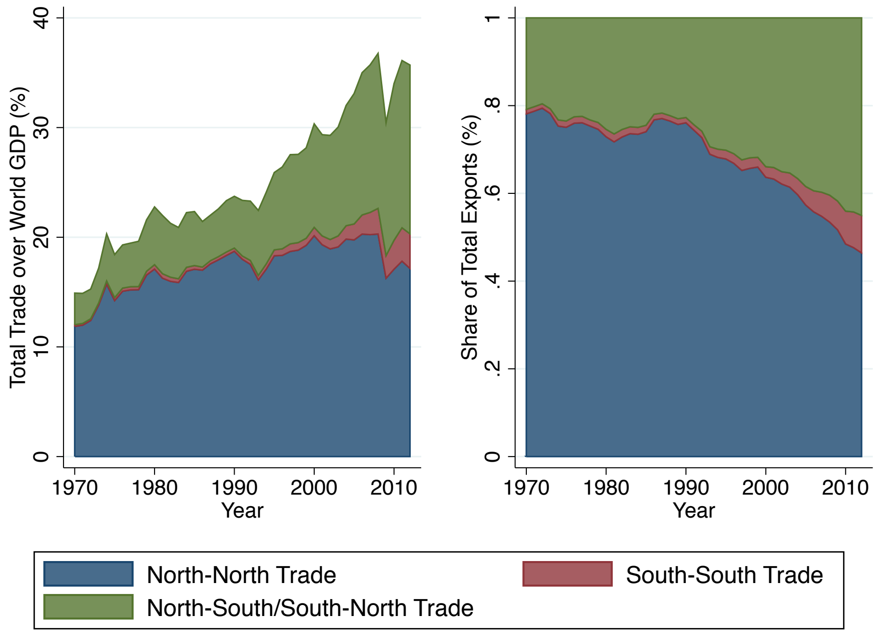


Notes: Using data from Alfaro et al. (2014) net inflows are computed from cumulative net inflows converted into constant international dollars using the price of investment goods provided in the PWT. The sample covers 53 countries from 1981-2007 (excluding the financial crisis). Ireland and Paraguay are excluded as outliers from the regression. For clarity of exposition, Ireland is plotted with a break in scale.

dex, but the PWT does not report it. Instead, the price index of investment goods is chosen. It varies less across countries than the consumption good price index which suggests that investment goods are mostly tradable (Hsieh and Klenow (2007)). However, deflating by the output price index gives very similar results. The deflated net capital inflows are cumulated to measure the capital inflows over the whole period. This measure is preferred over a simple comparison of changes in net foreign asset positions because it is not subject to valuation effects.

Alfaro et al. (2014) found that the allocation puzzle does not apply to private capital flows. To check if this also applies here, Figure 1.7 repeats the exercise described above only for private capital flows, here net Foreign Direct Investment (FDI) and Portfolio flows. Although the omission of the outliers Ireland and Paraguay from the regression yields a negative coefficient, it is far from significant. Thus despite the positive correlation with growth documented by Alfaro et al. (2014), there appears to be no correlation with capital inflows. Repeating this exercise with the average net flows per GDP, the measure preferred by Alfaro et al. (2014) shows no significant correlation either.

Figure 1.8: Increase in Trade integration 1970-2012



Notes: Total trade is calculated as the sum of exports and imports of goods as reported the IMF Direction of Trade statistics (DOTS) of the 60 countries in the sample. The sample captures $\approx$ 70% of world exports. The division into "North" and "South" follows the IMF classification into *Advanced* and *Developing Economies*.

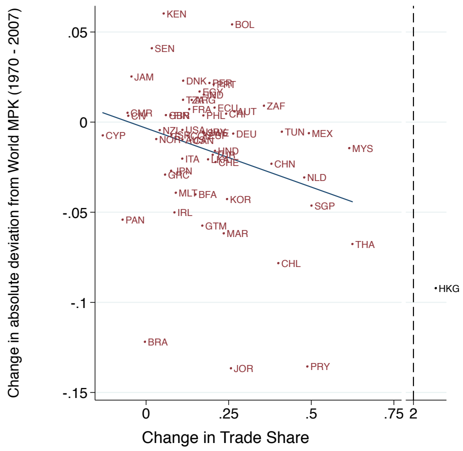
1.3.7 Trade integration

It is well-known that international trade has grown dramatically over the same period. Trade growth has been particularly strong between advanced and developed economies, often also referred to as North and South in the trade literature. Here total trade is the sum of exports and imports of goods over the 60 countries in my dataset[3]. Using the IMF DOTS database and disaggregating the trade between advanced and developed economies shows that North-South and South-North trade together now make up around 40 % of total trade compared to slightly more than 20 % in the 70s. It suggests that trade motivated by differences in factor endowments has made up a substantial part of the total growth in trade. The increase in trade between very different economies highlights the scope for specialization that international trade has created over the last decades.

Figure 1.9 presents suggestive evidence that convergence is linked to trade integration. Economies that reduced the difference with the world MPK on average

[3]These 60 countries capture between 60 and 75 % of all world trade depending on the year.

Figure 1.9: Increase in Trade and Convergence in MPKs



Notes: The linear regression coefficient is -.067 and significant at 5 %. The trade data comes from the IMF DOTS database. The trade share is measured as $\frac{\text{Imports+Exports}}{\text{GDP}}$. Taiwan is missing from the trade data. For clarity of exposition Hongkong is plotted with a break in the x-axis and is excluded from the regression for robustness.

increased trade over GDP more than those that did not converge.

1.4 Section II - Model

This section outlines how the empirical observations documented in this paper can be explained in a dynamic model of costly Heckscher-Ohlin trade combined with monopolistic competition a la Krugman (1980). The purpose of this model is to demonstrate how the fall in trade cost leads to convergence in marginal products of capital. In a world where countries differ in their preferences for savings and productivity, I identify two channels through which trade costs affect MPKs - (i) specialization in production of goods along factor-intensities (classic HO channel) and (ii) aggregate savings (savings channel). The fall in trade cost reduces the relative price of imports to exports. Increasing specialization in the export sector raises the return to the factor of production it uses intensively which directly translates into convergence in MPKs between capital-abundant and labor-abundant economies. In addition, the movement in wages changes life-cycle savings. This stimulates capital accumulation relatively more in low-capital economies. The resulting convergence in capital per effective worker amplifies the convergence in MPKs because terms of trade effects generate decreasing returns to capital.

1.4.1 Description

The world consists of N_C countries where a country is denoted by the subscript c. Labor and capital are the only factors of production. Countries differ in their factor-specific productivity, preferences for savings and population size. Thus each country is defined by a vector $\left(\beta, A^L, A^K, L, \{\tau_{c,c'}\}_{c'=1}^{N_c}\right)$ where β denotes the rate of time preference, A^L and A^K stand for labor and capital productivity, and L is the labor force. $\tau_{c,c'}$ is the iceberg trade cost for exporting to country $c^{'}$. Differences in β translate into differences in the steady-state capital.

Households

Each country is populated by overlapping generations. Each generation lives for two periods. The young work and save part of their income for old age consumption. International financial markets are closed, so households invest all their savings in domestic capital. When retired, they receive the rental income from capital and can sell the un-depreciated capital to the next generation. Formally, every generation faces the following maximization problem:

$$\max_{C_{c,t},C_{c,t+1}} U_{c,t} = \frac{(C_{c,t})^{1-\rho}}{1-\rho} + \beta_c \frac{(C_{c,t+1})^{1-\rho}}{1-\rho} \tag{1.3}$$

subject to the budget constraints:

$$P_{c,t}C_{c,t} = w_{c,t} - s_{c,t} \tag{1.4}$$

$$P_{c,t+1}C_{c,t+1} = \left(\frac{r_{c,t+1}}{P_{c,t}} + (1-\delta)\frac{P_{c,t+1}}{P_{c,t}}\right) s_{c,t} \tag{1.5}$$

Thus the optimal amount of savings of a young household is given by:

$$s_{c,t} = \frac{\beta_c^{\frac{1}{\rho}} w_{c,t}}{\beta_c^{\frac{1}{\rho}} + \left(\frac{r_{c,t+1}}{P_{c,t+1}} + (1-\delta)\right)^{\frac{\rho-1}{\rho}}} \tag{1.6}$$

Preferences are homothetic, so the saving rate of the young is independent of the level of income. Aggregate savings determine next period's capital stock:

$$K_{c,t+1} = \frac{s_{c,t}}{P_{c,t}} L_{c,t} \tag{1.7}$$

Countries with a high β_c have a higher savings rate and hence accumulate more capital. High real wages increase capital accumulation. The effect of changes in the return depends on the intertemporal elasticity of substitution ρ. For $\rho > 1$, the wealth effect dominates the substitution effect and the savings rate decreases as the return increases.

1.4.2 A pure Heckscher-Ohlin model

The special case of a pure Heckscher-Ohlin model without monopolistic competition illustrates the properties of the model in a tractable way and highlights the effect of the fall in trade cost. In particular, this simple model can already explain all five empirical observations documented in section I: (i) Capital abundant countries have lower MPKs, (ii) MPKs converge as trade grows, (iii) the capital share and MPKs can be positively correlated, (iv) movements in MPKs are positively correlated with movements in the capital share and (v) changes in MPKs are negatively correlated with changes in the capital-labor ratio.

There are two layers of production. For consumption and investment households use a final good. Its production requires two intermediate goods which differ in the factor-intensity of their production process. Intermediate goods are tradable at an iceberg cost τ.

Formally, the final good $Q_{c,t}$ is a CES aggregate of two intermediate goods $Q^K_{c,t}$ and $Q^L_{c,t}$ whose prices in country c at time t are $P^K_{c,t}$ and $P^L_{c,t}$. The following unit cost function describes the production technology:

$$B^{Final}_{c,t} = \left({P^K_{c,t}}^{1-\sigma} + {P^L_{c,t}}^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{1.8}$$

where σ is the elasticity of substitution between the capital- and labor-intensive intermediates. Households are price-takers, so the final good price equals the marginal cost. The expenditure on intermediate $i = K, L$ is given by:

$$P^i_{c,t} Q^i_{c,t} = \left(\frac{P^i_{c,t}}{P_{c,t}} \right)^{1-\sigma} P_{c,t} Q_{c,t} \tag{1.9}$$

For $\sigma < 1$ ($\sigma > 1$) the labor and the capital-intensive intermediate are complements (substitutes). As the labor-intensive intermediate becomes cheaper, the expenditure on capital-intensive intermediate rises (falls).

Intermediate good producers rent labor and capital from households. All firms operating in the same sector have access to the same production technology apart from the differences in factor-specific productivity across countries. The factor intensity varies by sector. The following unit cost function describes the production technology of sector i where $i \in K, L$:

Figure 1.10: Conditional Factor Prices in the pure HO model

South		**Autarky**		**North**
$\frac{r_c}{A_c^K} = \tau p_W^K, \frac{w_c}{A_c^L} = p_W^L$	$\tilde{k}^* = \left(\frac{p_W^L}{\tau p_W^K}\right)^\sigma$	$\frac{r_c}{A_c^K}, \frac{w_c}{A_c^L}$	$\tilde{k}^* = \left(\frac{\tau p_W^L}{p_W^K}\right)^\sigma$	$\frac{r_c}{A_c^K} = p_W^K, \frac{w_c}{A_c^L} = \tau p_W^L$

$$B(i)_{c,t} = \left(\frac{r_{c,t}}{A_{c,t}^K}\right)^{z_i} \left(\frac{w_{c,t}}{A_{c,t}^L}\right)^{1-z_i} \tag{1.10}$$

where $z_i \in [0,1]$ denotes the capital-intensity of sector i with $z_K > z_L$ such that z is higher for inputs used to produce $Q_{c,t}^K$. Here I assume extreme factor-intensities such that $z_K = 1$ and $z_L = 0$. This assumption ensures that all countries will always be active in both sectors and never specialize completely. However, the results will be very similar as long as capital-intensities are sufficiently different such that there can be meaningful reallocation between sectors.

For simplicity, assume that the trade cost between each location is the same such that $\tau_{c,\tilde{c}} = \tau \; \forall c \neq \tilde{c}$. We can divide countries into three groups. The first group, let's call it North (N), consists of countries that are exporters in the capital-intensive sector and importers in the labor-intensive sector. On the other hand, exporters in the labor-intensive sector are grouped in South (S). The third group consists of countries in autarky. Intermediates will be priced at marginal cost. By arbitrage, the domestic price of the exported good equals its world market price. Figure 1.10 shows how this sorting depends on the steady-state ratio of efficient capital to efficient labor $\tilde{k}_c^* = \frac{A_c^K K_c^*}{A_c^L L_c}$, where the cutoff depends on world market prices. It also presents the conditional factor prices corresponding to each group.

Market clearing

The competitive equilibrium of this world consists of a sequence of prices and quantities such that factor markets and goods markets clear and producers and households optimize. In each country, total expenditure on final goods cannot exceed total income such that:

$$P_{c,t}Q_{c,t} = w_{c,t}L_{c,t} + r_{c,t}K_{c,t} \tag{1.11}$$

Market clearing requires that world demand equals world supply for each intermediate. By Walras law, it is sufficient if we focus on market clearing for capital-intensive intermediates.

$$\sum_{c\in N} A_c^K K_c + \tau \sum_{c\in S} A_c^K K_c = \sum_{c\in N} Q_c^K + \tau \sum_{c\in S} Q_c^K \tag{1.12}$$

Conditions 1.11 and 1.12 together with the 1.8,1.9 and 1.10 determine the world market prices $p^L_{W,t}$ and $p^K_{W,t}$. The solution up to a normalization is summarized by:

$$\left(\frac{p^K_{W,t}}{p^L_{W,t}}\right)^{\sigma} = \frac{\left(\frac{P_{N,t}}{P_{S,t}}\right)^{1-\sigma} \sum_{c\in S} A_c^L L_c + \tau^{\sigma} \sum_{c\in N} A_c^L L_c}{\sum_{c\in N} A_c^K K_c + \tau^{\sigma} \left(\frac{P_N}{P_S}\right)^{1-\sigma} \sum_{c\in S} A_c^K K_c} \tag{1.13}$$

where the sums are over the endowment of all countries in the corresponding group. Countries in autarky do not matter for world market prices as their markets clear domestically. Finally, exporting the capital-intensive goods will only be optimal for sufficiently capital-abundant economies. Thus, ordering countries along their ratio of efficient capital to efficient labor $\tilde{k}_c = \frac{A_c^K K_c}{A_c^L L_c}$, we can sort them into North, Autarky and South as described in eq. 1.14.

$$\tilde{k}_{South} \leq \left(\frac{p_W^L}{\tau p_W^K}\right)^{\sigma} < \tilde{k}_{Autarky} < \left(\frac{\tau p_W^L}{p_W^K}\right)^{\sigma} \leq \tilde{k}_{North} \tag{1.14}$$

Consequently, the South consists of labor-abundant countries and in the absence of intra-industry trade fairly balanced factor endowments lead to autarky.

Steady State

In the absence of exogenous growth in effective labor, the steady state is characterized by the following system of N_c equations:

$$K_c^* = \frac{\beta_c^{\frac{1}{\rho}}}{\beta_c^{\frac{1}{\rho}} + \left(\frac{r_c}{P_c} + (1-\delta)\right)^{\frac{\rho-1}{\rho}}} \frac{w_c}{P_c} L_c \qquad \forall c \tag{1.15}$$

such that goods and factor markets clear and the final good price of the benchmark country is normalized to one. Which countries are capital-abundant in the long-run? Denoting $k_c = \frac{K_c}{A_c^L L_c}$ and using eq. 1.7 to express the relative ratios of capital to efficiency units of labor highlights what causes steady state levels of capital to be larger in some countries.

$$\frac{k_c^*}{k_{c'}^*} = \left(\frac{\beta_c}{\beta_{c'}}\right)^{\frac{1}{\rho}} \frac{\beta_{c'}^{\frac{1}{\rho}} + \left(\frac{r_{c'}}{P_{c'}} + (1-\delta)\right)^{\frac{\rho-1}{\rho}}}{\beta_c^{\frac{1}{\rho}} + \left(\frac{r_c}{P_c} + (1-\delta)\right)^{\frac{\rho-1}{\rho}}} \frac{w_c}{A_c^L} \frac{A_{c'}^L}{w_{c'}} \frac{P_{c'}}{P_c}$$

In the absence of trade costs when input prices are equalized, the steady state capital is higher for patient countries whose capital-productivity is not too high, such that:

$$\frac{k_c^*}{k_{c'}^*} > 1 \text{ if: } \begin{cases} (1-\delta)\left(\left(\frac{\beta_c}{\beta_{c'}}\right)^{\frac{1}{\rho-1}} - 1\right) + \frac{r}{A^K P}\left(\left(\frac{\beta_c}{\beta_{c'}}\right)^{\frac{1}{\rho-1}} A_{c'}^K - A_c^K\right) & < 0 \text{ for } \rho < 1 \\ \beta_c - \beta_{c'} & > 0 \text{ for } \rho = 1 \\ (1-\delta)\left(\left(\frac{\beta_c}{\beta_{c'}}\right)^{\frac{1}{\rho-1}} - 1\right) + \frac{r}{A^K P}\left(\left(\frac{\beta_c}{\beta_{c'}}\right)^{\frac{1}{\rho-1}} A_{c'}^K - A_c^K\right) & > 0 \text{ for } \rho > 1 \end{cases}$$

In the presence of trade costs, the differences in preferences are amplified by the differences in factor prices. High wages in capital-abundant countries result in an even higher ratio of capital to efficient labor in steady state. Depending on the elasticity of intertemporal substitution, this effect can be amplified, dampened, or stay unaffected by differences in returns. For $\rho = 1$, the savings do not respond to changes in the return to capital. For $\rho > 1$, the convergence in capital will be enhanced as savings reduce more in countries where the return increases. For $\rho < 1$, the effect on total savings dampens the convergence in capital, but the overall effect depends on whether it dominates. In the data, convergence in capital-labor ratios and convergence in MPKs are related which suggests that $\rho \geq 1$ is the empirically more relevant case. Thus the conditional convergence in factor prices induced by the fall in trade cost also leads to convergence in the steady-state level of capital. In the presence of decreasing returns to capital, this is an additional channel through which the fall in trade costs drives convergence in MPKs.

Properties

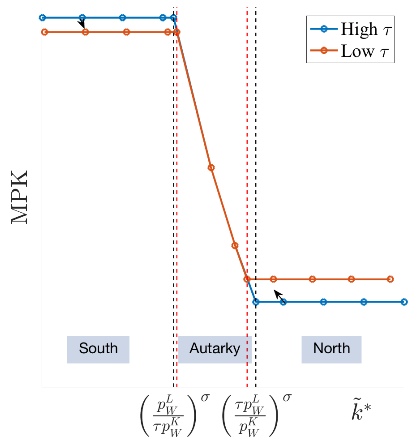
This simple version of the model helps our intuition, can qualitatively already capture the empirical observations (i) - (v), and illustrates how the fall in trade cost can explain the convergence in MPKs.

Property 1: Capital-abundant countries have lower MPKs

Under the prerequisite that factor-specific productivities do not differ too much

Figure 1.11: Effect of a fall in trade cost

(a) MPK



Notes: A fall in trade cost shrinks the region of autarky, equalizes the MPK through the factor-price equalization channel and leads to convergence in capital-labor ratios through the savings channel as highlighted by the arrows.

across countries, this simple model predicts that the capital-abundant Northern countries will have lower MPKs:

$$\frac{\frac{r_N}{P_N}}{\frac{r_S}{P_S}} = \frac{A_N^K}{A_S^K} \frac{\left(\left(p_W^K\right)^{1-\sigma} + \tau^{1-\sigma}\left(p_W^L\right)^{1-\sigma}\right)^{\frac{1}{\sigma-1}}}{\left(\left(p_W^K\right)^{1-\sigma} + \tau^{\sigma-1}\left(p_W^L\right)^{1-\sigma}\right)^{\frac{1}{\sigma-1}}} < 1 \tag{1.16}$$

As long as country N has a lower relative capital-productivity than S, the condition in eq. 1.16 will always be satisfied. When the Northern country has a higher relative capital-productivity, it additionally depends on the level of the trade cost, world market prices and σ.

Property 2: Reduction in trade cost leads to convergence in MPKs
Eq. 1.16 also illustrates how the fall in trade cost leads to convergence in MPKs. A reduction in τ reduces the numerator and increases the denominator. For $\tau = 1$ the differences in MPKs only depend on the differences in capital-productivity. Intuitively, a fall in trade cost raises foreign demand in the export sector which exerts upwards pressure on prices of factors used intensively in export production. Imports become cheaper which reduces factor prices in the import sector. As $\tau \to 1$, the world approaches conditional factor price equalization. Then only

differences in capital productivities can lead to differences in MPKs.

Property 3: Positive correlation between MPKs and capital-shares
In the data, I observe a positive correlation between returns and capital-shares. Here I show that this model allows for two ways to produce this results. Firstly, when capital- and labor-intensive goods are complements ($\sigma << 1$) and the trade cost is sufficiently high ($\tau >> 1$), this model can replicate this positive correlation. The capital-share is lower in the North if:

$$\frac{r_N K_N}{P_N Q_N} < \frac{r_S K_S}{P_S Q_S} \tag{1.17}$$

$$\underbrace{\frac{{Q_N^K}^S}{{Q_N^K}^D}}_{>1} < \underbrace{\frac{{Q_S^K}^S}{{Q_S^K}^D}}_{<1} \frac{P_N^{1-\sigma}}{\tau^{\sigma-1} P_S^{1-\sigma}} \tag{1.18}$$

As Northern (Southern) countries are exporters of the capital- (labor-) intensive good, the quantity produced ${Q_c^i}^S$ exceeds the quantity demanded domestically ${Q_c^i}^D$. A necessary condition for the North to have a lower capital-share is given by:

$$\frac{P_N^{1-\sigma}}{\tau^{\sigma-1} P_S^{1-\sigma}} > 1 \tag{1.19}$$

$$\left(\left(p_W^K\right)^{1-\sigma} + \left(\tau p_W^L\right)^{1-\sigma}\right) > \tau^{\sigma-1}\left(\left(\tau p_W^K\right)^{1-\sigma} + \left(p_W^L\right)^{1-\sigma}\right) \tag{1.20}$$

$$\left(\tau^2\right)^{1-\sigma} > 1 \tag{1.21}$$

which is true if $\sigma < 1$ and a positive trade cost such that $\tau > 1$. For $\sigma << 1$, $\tau >> 1$, and endowments sufficiently different such that countries trade, this model can account for cases where Northern countries have lower capital-shares than Southern ones.

Additionally, low returns would be associated with low capital-shares when the condition stated in eq. 1.16 is violated and a Southern country has such a low capital-productivity that it has a lower MPK than N. In this case, the low capital share in the South would be associated with a low MPK. This explanation may even be consistent with the negative correlation between MPKs and capital-labor ratios if the lack of data on capital-productivity leads a country to be mistaken for capital-abundant. Empirically, one way to directly check which case applies to each country is to look at the factor-content of trade. In the quantitative exercise, I use bilateral trade flows to estimate the trade cost such that these two cases can be distinguished. Moreover, the capital-share can then be used to estimate the capital-productivity directly.

Property 4: Positive correlation between changes in MPKs and capital shares
As in the standard static Heckscher-Ohlin model, the reduction in trade cost leads to sectoral reallocation. As capital-abundant countries specialize in capital-intensive industries, the capital share S^K rises along with the ratio of rental rate to wage. It is easy to see that the direction of the effect is unchanged by the introduction of dynamics. Using the expression for the capital share and plugging in the steady-state level of capital shows that there is a positive correlation between the capital share and the MPK.

$$S_c^K = \frac{r_c K_c^*}{w_c L_c + r_c K_c^*} \tag{1.22}$$

$$= \left(1 + \frac{P_c}{r_c} + \beta_c^{\frac{-1}{\rho}} \left(\left(\frac{r_c}{P_c}\right)^{\frac{1}{1-\rho}} + (1-\delta) \left(\frac{r_c}{P_c}\right)^{\frac{\rho}{1-\rho}} \right)^{\frac{\rho-1}{\rho}} \right)^{-1} \tag{1.23}$$

The right-hand side of eq. 1.23 increases in $\frac{r}{P}$. As the economy reaches a new steady state with lower trade costs, Northern economies have a higher MPK and a higher capital share than before. It it worth mentioning that this effect differs from capital convergence dynamics. The capital-share for a Southern country can be rewritten as a function of world market prices and capital per effective labor:

$$S_S^K = \frac{\tau p_W^K A_S^K k_S}{p_w^L + \tau p_W^K A_S^K k_S} \tag{1.24}$$

This expression illustrates the positive correlation between the capital share and capital accumulation. If the convergence in MPKs were driven mainly by capital convergence instead of changes in trade costs, a Southern economy that accumulates capital to reach its steady state and reduce its MPK would experience an increase in the capital share. This contradicts the positive correlation found in the data. Besides, it does not matter whether the changes in capital occur because of domestic savings or foreign inflows, the result would always be a counterfactual negative correlation between MPKs and capital shares.

Property 5: Negative correlation between changes in MPKs and capital-labor ratios
This simple version of the model also produces a negative correlation between changes in the capital-labor ratio and MPKs. As discussed in detail in section

1.4.2, under reasonable parameter assumptions, convergence in factor prices also leads to convergence in steady-state capital per effective labor. However, note that in the pure H-O version, the convergence in steady states does not amplify the convergence in returns. In this simpler model, all prices are determined by world market prices and the trade costs. Changes in the capital or labor endowment do not affect the MPK as long as the country stays in the same group, i.e., does not switch its export-sector or move into autarky. It is only because of the changes in wages and rental rates that steady-state levels of capital move closer together. The complete absence of decreasing returns to capital on the country-level is an extreme property of the simple model, which is relaxed by the introduction monopolistic competition in the extended version used for quantification.

1.4.3 A Heckscher-Ohlin-Krugman model

The pure Heckscher-Ohlin model highlights the role of interindustry trade in the convergence of MPKs and is a useful tool for exposition, but it misses key features needed to quantify the model. All trade in the simple version is between North and South, a characteristic that besides being unrealistic will hamper the model's ability to correctly match the trade flows observed in the data. Moreover, all countries forming part of the same block will have the same conditional factor prices. However, the high dispersion observed in the data does not suggest that factor prices are overly similar within the North and the South.[4] Finally, countries can transition into autarky in the simple model, which can lead to dynamics which are far away from what we observe in the data. Fortunately, the introduction of monopolistic resolves all of these issues.

In the extended version there is an additional layer of production. For consumption and investment households still use the final good whose production requires two intermediate goods. However, now intermediate and final goods are not tradable. Instead, intermediate good producers assemble a variety of tradable inputs. Monopolistically competitive firms produce the intermediate-specific inputs using labor and capital and supply them to intermediate good producers. Firms located in c can ship goods to $\tilde{c}$ at an iceberg trade cost $\tau_{c,\tilde{c}} \geq 1$ where $\tau_{c,c} = 1$. Free entry endogenously determines the number of firms in each location.

Assemblers domestically produce the intermediate $Q^i_{c,t}$ from a continuum of tradable inputs $q_{c,t}(\omega_i)$ of mass N_i. Inputs are substitutes with elasticity $\epsilon > 1$.

[4]Autarky is not a reasonable explanation since all countries in the sample trade to some extent. Capital-augmenting productivity differences may explain the differences in returns as the model only predicts conditional FPE, but since they are not directly observable and usually thought to be small, this explanation is not entirely satisfying.

Assemblers are price-takers with the unit cost function:

$$B_{c,t}^{Assembly}(i) = \left(\int_0^{N_i} p_{c,t}(\omega_i)^{1-\epsilon} d\omega_i \right)^{\frac{1}{1-\epsilon}} \tag{1.25}$$

The demand for input ω_i is given by:

$$p_{c,t}(\omega_i) {q_{c,t}}^D(\omega_i) = \left(\frac{P_{c,t}^i}{p_{c,t}(\omega_i)} \right)^{\epsilon - 1} P_{c,t}^i Q_{c,t}^i \tag{1.26}$$

Input-producing firms pay a fixed cost f to enter the market and hire workers and rent capital at the domestic wage $w_{c,t}$ and rental rate $r_{c,t}$. All firms operating in the same sector have access to the same production technology apart from the differences in factor-specific productivity across countries. Let $z_i \in [0, 1]$ denote the capital-intensity of sector i with $z_K > z_L$ such that z is higher for inputs used to produce $Q_{c,t}^K$. Again, I assume extreme factor-intensities such that $z_K = 1$ and $z_L = 0$ such that all countries will never specialize completely in one sector. The results will be very similar as long as capital-intensities are sufficiently different such that there can be meaningful reallocation between sectors. Romalis (2004) discusses the properties of a static version of this model with a continuum of industries and factor-intensities.

As I mentioned above, transports of inputs from country c to $\tilde{c}$ are subject to an iceberg trade cost such that $\tau_{c,\tilde{c}}$ -units have to be shipped for every unit delivered. Thus the quantity supplied is larger than the quantity consumed at destination and given by $q_{\tilde{c}}^S(\omega_i) = \tau_{c,\tilde{c}} q_{\tilde{c}}^D(\omega_i)$. The love for variety of assemblers and the fixed cost of entry creates monopolistic competition a la Krugman (1980) in the input market. Firm ω in sector i located in country c sets prices in all destination markets such that profits are maximized subject to the input demand from assemblers given in eq. 1.26:

$$\max_{p_{\tilde{c},t}(\omega_i)} \sum_{\tilde{c}}^{C} \left(\frac{p_{\tilde{c},t}(\omega_i)}{\tau_{c,\tilde{c}}} q_{\tilde{c},t}^S(\omega_i) - \left(\frac{r_{c,t}}{A_{c,t}^K} \right)^{z_i} \left(\frac{w_{c,t}}{A_{c,t}^L} \right)^{1-z_i} q_{\tilde{c},t}^S(\omega_i) \right) - \left(\frac{r_{c,t}}{A_{c,t}^K} \right)^{z_i} \left(\frac{w_{c,t}}{A_{c,t}^L} \right)^{1-z_i} f \tag{1.27}$$

The optimal price charged to households in $\tilde{c}$ is a constant mark-up over marginal cost.

$$p_{\tilde{c},t}(\omega_i) = \frac{\epsilon}{\epsilon - 1} \tau_{c,\tilde{c}} \left(\frac{r_{c,t}}{A_{c,t}^K} \right)^{z_i} \left(\frac{w_{c,t}}{A_{c,t}^L} \right)^{1-z_i} \tag{1.28}$$

It follows from eq. 1.28 that all firms located in the same country-sector will charge the same price $p_{\tilde{c},t}(\omega_i) = p_{\tilde{c},t}^i \; \forall \omega_i$. Free entry implies zero profits in equilibrium. Together with the optimal price eq. 1.28, this implies that each firm produces the same quantity given by:

$$\sum_{c'}^{C} q^S_{c,c',t}(\omega_i) = f(\epsilon - 1) \tag{1.29}$$

Denoting the share of firms located in country c by $n^i_{c,t}$ we can combine eq. 1.28 and eq. 1.25 to express the intermediate good price as a function of factor prices:

$$P^i_{c,t} = \frac{\epsilon}{\epsilon - 1} N_{i,t}^{\frac{1}{1-\epsilon}}$$
$$\times \left[n^i_{c,t} \left(\left(\frac{r_{c,t}}{A^K_{c,t}} \right)^{z_i} \left(\frac{w_{c,t}}{A^L_{c,t}} \right)^{1-z_i} \right)^{1-\epsilon} + \sum_{c' \neq c} n^i_{c',t} \left(\tau_{c,c'} \left(\frac{r_{c',t}}{A^K_{c',t}} \right)^{z_i} \left(\frac{w_{c',t}}{A^L_{c',t}} \right)^{1-z_i} \right)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}} \tag{1.30}$$

It is easy to see that in the absence of trade costs when $\tau_{c,\tilde{c}} = 1 \; \forall c, \tilde{c}$ the intermediate good prices will be the same across all countries.

Market clearing

The competitive equilibrium of this world consists of a sequence of prices, quantities, and numbers of firms such that factor markets and goods markets clear and producers and households optimize. In each country, total expenditure on final goods cannot exceed total income such that:

$$P_{c,t} Q_{c,t} = w_{c,t} L_{c,t} + r_{c,t} K_{c,t} \tag{1.31}$$

The savings of the young determine capital accumulation (eq. 1.7) and hence the dynamics of the model. On the production side, the optimality conditions are laid out in eq. 1.25 to 1.30. Within each country, the total production of inputs and the fixed cost paid have to balance with the total endowment of factors of production. These conditions directly determine the number and allocation of firms across countries:

$$A^K_{c,t} K_{c,t} = \int_0^{n_c N_K} \left(f + \sum_{\tilde{c}}^{C} q^S_{\tilde{c},t}(\omega_K) \right) d\omega_K = n^K_c N_K \epsilon f \tag{1.32}$$

$$A^L_{c,t} L_{c,t} = \int_0^{n_c N_L} \left(f + \sum_{\tilde{c}}^{C} q^S_{\tilde{c},t}(\omega_L) \right) d\omega_L = n^L_c N_L \epsilon f \tag{1.33}$$

Using $\sum n_{c,t}^{K} = 1$, this solves for the total mass of firms in sector K and L.

$$N_{K,t} = \frac{1}{f\epsilon} \sum_{c} A_{c,t}^{K} K_{c,t} \tag{1.34}$$

$$n_{\tilde{c},t}^{K} = A_{\tilde{c},t}^{K} K_{\tilde{c},t} \left(\sum_{c} A_{c,t}^{K} K_{c,t} \right)^{-1} \tag{1.35}$$

$$N_{L,t} = \frac{1}{f\epsilon} \sum_{c} A_{c,t} L_{c,t} \tag{1.36}$$

$$n_{\tilde{c},t}^{L} = A_{\tilde{c},t}^{L} L_{\tilde{c},t} \left(\sum_{c} A_{c,t}^{L} L_{c,t} \right)^{-1} \tag{1.37}$$

The equilibrium wage and rental rates in each country are then pinned down by the market clearing in the goods market such that $q_{\tilde{c}}^{S}(\omega_i) = \tau_{c,\tilde{c}} q_{\tilde{c}}^{D}(\omega_i)\ \forall c, i, t$. Balancing the total supply from eq. 1.29 with total world demand by aggregating eq. 1.26 over countries gives:

$$\left(\frac{r_{\tilde{c},t}}{A_{\tilde{c},t}^{K}} \right)^{-\epsilon} \sum_{c}^{C} \tau_{\tilde{c},c}^{1-\epsilon} \left(P_{c,t}^{K} \right)^{\epsilon} Q_{c,t}^{K} = (\epsilon - 1)^{1-\epsilon} \epsilon^{\epsilon} f \tag{1.38}$$

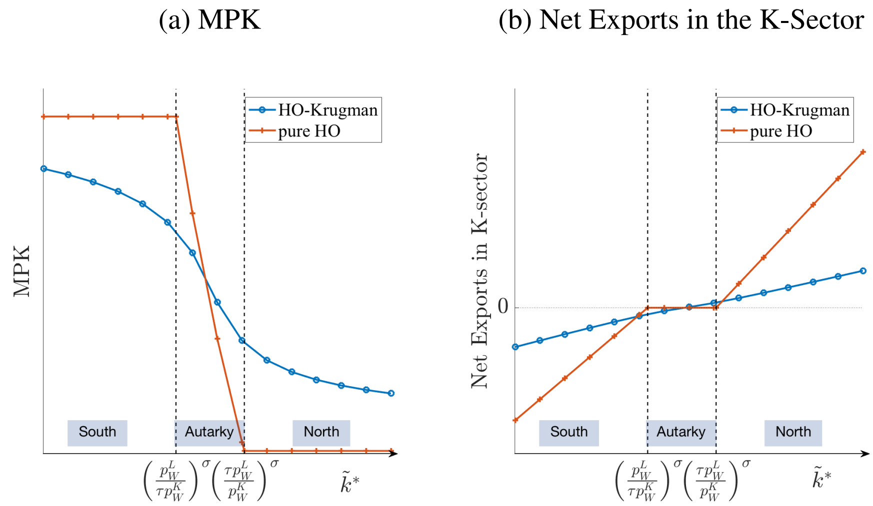
$$\left(\frac{w_{\tilde{c},t}}{A_{\tilde{c},t}^{L}} \right)^{-\epsilon} \sum_{c}^{C} \tau_{\tilde{c},c}^{1-\epsilon} \left(P_{c,t}^{L} \right)^{\epsilon} Q_{c,t}^{L} = (\epsilon - 1)^{1-\epsilon} \epsilon^{\epsilon} f \tag{1.39}$$

where eq. 1.9 determines $Q_{c,t}^{i}$. By Walras Law, prices are determined up to a normalization if $2N_C - 1$ of these equations are satisfied. Prices are normalized to the final good price of a benchmark country to facilitate the quantification where all variables will be expressed in units of U.S. output.

In general, there is no closed form solution for the general equilibrium, but the partial equilibrium can deliver some useful insights. For ease of exposition I present the case of two countries called N and S with symmetric trade cost $\tau_{N,S} = \tau_{S,N} = \tau$. The multi-country case is discussed in appendix B. Using eq. 1.30, the market clearing conditions given in eq. 1.38 and 1.39 can be rewritten as a function of the relative (factory-gate) factor price:

$$\frac{n_N^i}{n_S^i} = \frac{A_N^K K_N}{A_S^K K_S} = \frac{\left(\left(\hat{p}_i^{-\sigma} - \tau^{1-\epsilon} \hat{p}_i^{\epsilon-\sigma} \right) \frac{P_N^{\sigma} Y_N}{P_S{}^{\sigma} Y_S} \right)^{\frac{\epsilon-1}{\epsilon-\sigma}} - \tau^{1-\epsilon} \left(\hat{p}_i^{\epsilon-\sigma} - \hat{p}_i^{-\sigma} \tau^{1-\epsilon} \right)^{\frac{\epsilon-1}{\epsilon-\sigma}}}{\left((1 - \hat{p}_i^{-\epsilon} \tau^{1-\epsilon})^{\frac{\epsilon-1}{\epsilon-\sigma}} - \left(\left(\hat{p}_i^{-\epsilon} - \tau^{1-\epsilon} \right) \frac{P_N^{\sigma} Y_N}{P_S{}^{\sigma} Y_S} \right)^{\frac{\epsilon-1}{\epsilon-\sigma}} (\tau)^{1-\epsilon} \right)} \tag{1.40}$$

Figure 1.12: Comparison of pure HO model with HO-Krugman



Notes: $\tilde{k^*} = \frac{A^K K^*}{A^L L}$, the steady-state, productivity-adjusted relative factor endowment

where $\hat{p}_K = \frac{r_N}{r_S}\frac{A_S^K}{A_N^K}$ and $\hat{p}_L = \frac{w_N}{w_S}\frac{A_S^L}{A_N^L}$.

The relative share of firms is a function of the relative endowment as shown in eq. 1.34 to 1.37. The right-hand-side of eq. 1.40 is decreasing in the relative price. Thus a country with a larger endowment of a factor of production which attracts a high share of firms also has a lower factor price compared to abroad. This also implies that the relative factor price is lower in a country that is abundant in the factor. Hence capital-abundant economies have higher ratios of efficiency wage to efficiency rental rate than a labor-abundant ones.

Note that this model features decreasing returns. Consider the effect of an increase in the capital stock in country N. Consequently, the share of firms located in N in the capital-intensive sector increases. As seen in eq. 1.40 this increase in supply leads to a fall in the relative factor price $\hat{p}_K$.

Results for the full model

Figure 1.12 shows the MPK and net exports in the capital-intensive sector for different countries in the two versions of the model. It illustrates that the introduction of monopolistic competition smooths the kinks at the boundary to autarky. Monopolistic competition a la Krugman (1980) introduces decreasing returns to capital also for trading economies and replaces the area of autarky with a smooth adjustment of sectoral net exports.

As in the pure HO-model, a reduction in trade cost leads to convergence in factor prices. As all countries are both exporters and importers in both sectors, the fall in trade cost does not have a uniform effect across sectors. For example, although imported inputs become cheaper, the increase in foreign demand exerts positive pressure on domestically-produced inputs. However, the upward pressure on prices dominates in the sector in which the country is a net exporter such that both net exports and net imports increase and total trade grows. At the same time, the relative return to the factor used intensively in exports will increase. Furthermore, it is easy to see from eq. 1.38 and 1.39 that in the absence of trade costs, conditional factor prices equalize across countries. As shown in Appendix B, similar to the pure HO-model, there is no conditional FPE in general. But unlike the pure HO-model, even the same relative factor endowment $\frac{A^K K}{A^L L}$ would not lead to FPE. The monopolistic competition also introduces increasing returns to scale such that larger countries benefit from size in the form of a lower overall price level.

Moreover, the introduction of decreasing returns does not alter the essential behavior of the capital share. Eq. 1.23 also applies to the full model, and thus any change in MPK from a steady state with high trade costs to one with low trade costs will be positively correlated with the capital share. To what extent convergence to steady-state capital can deliver a positive correlation between MPKs and capital-shares now depends on the elasticity of substitution between labor and capital. In the pure HO-model when each country is small, this elasticity is infinite. In fact, as long as the elasticity is larger than one, convergence dynamics will generate a negative correlation between MPKs and capital shares. What determines the elasticity between capital and labor in this model? In general, it varies across countries. Small economies with low trade costs will generally have higher elasticities than large, closed economies. For the integrated economy, σ determines how factor prices react to changes in endowment.

Finally, the role of the input elasticity of substitution ϵ can be seen in the gravity equation of this model. Total imports of c from $c^{'}$ are given by:

$$M_{c,c'} = \frac{(\epsilon-1)^{\epsilon-1}}{\epsilon^{\epsilon} f \tau_{c',c}^{\epsilon-1}} \left(A_{c'}^K K_{c'} \left(\frac{P_c^K}{\frac{r_{c'}}{A_{c'}^K}} \right)^{\epsilon-1} \left(\frac{P_c^K}{P_c} \right)^{1-\sigma} + A_{c'} L_{c'} \left(\frac{P_c^L}{\frac{w_{c'}}{A_{c'}}} \right)^{\epsilon-1} \left(\frac{P_c^L}{P_c} \right)^{1-\sigma} \right) P_c Q_c \tag{1.41}$$

This shows that imports from a location are high if the trade cost with that location is low, the location has low factor prices and if the factor endowment is large. Note also that $\epsilon - 1$ governs the elasticity of imports relative to domestic demand commonly referred to as *trade elasticity*.

1.5 Section III - Quantification

1.5.1 Taking the model to the data

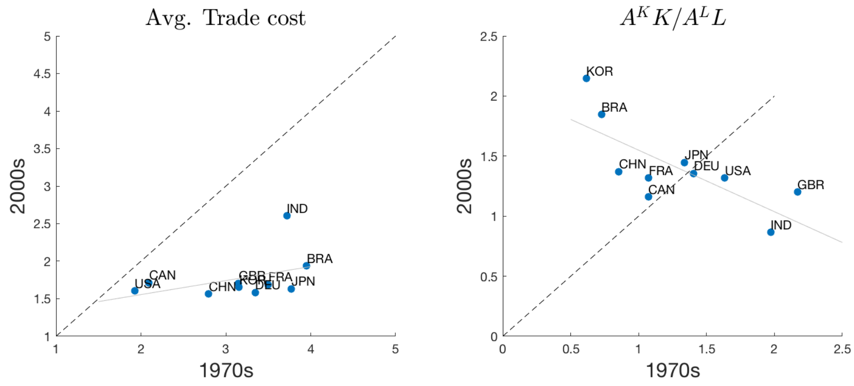
The empirical observations made in Section I are likely to be a result of a combination of changes happening at the same time. In particular, the model illustrates that movements in factor endowments, driven by changes in productivities or preferences for savings, and changes in the trade cost affect the MPK. Which of these accounts for the observed changes in the data? How much of the convergence can be attributed to the fall in trade cost? This section shows the quantitative predictions of the model.

The quantification exercise consists of three steps. Firstly, the model parameters σ, ϵ and ρ are calibrated. Secondly, an accounting exercise shows which fundamental shifts explain the differences between the initial steady state in the 70s (average over 1970-1974) and the more integrated steady state at the end of the sample (average over 2003-2007). The five-year averages limit the influence of short-run fluctuations on the results. I estimate the capital and labor productivities, trade costs and time preferences that make the calibrated model consistent with the data. Thirdly, a counterfactual analysis helps to understand how much changes in the trade cost, the distribution of capital across countries and productivity affect the dispersion of returns across countries.

1.5.2 Calibration

To make this exercise computationally feasible, it only incorporates the ten largest economies in the world which make up roughly 73 % of total GDP in the 60-country sample in 2011. In addition to the PWT 9.0 database, I use the IMF DOTS database for bilateral trade flows in goods trade. The calibration is based on standard values used in the literature. The elasticity of the final good production function σ is chosen such that the U.S. elasticity between K and L in the 1970s is around 0.8, which lies within the range of the literature (Fadinger (2011), Antràs (2004)). The intertemporal elasticity of substitution is such that the income and substitution effect exactly cancel out ($\rho = 1$). For the benchmark calibration, the elasticity between varieties ϵ is fixed at 10. This suggests a markup of 11 %. The ϵ parameter also governs the trade elasticity ($\epsilon - 1$), which the literature often finds to be around 4. This model focuses on long-run adjustments, so the elasticity is likely to be higher. A robustness check shows that although the explanatory power of trade costs declines for lower values of ϵ, it still explains a significant part of the convergence.

Figure 1.13: Sources of convergence



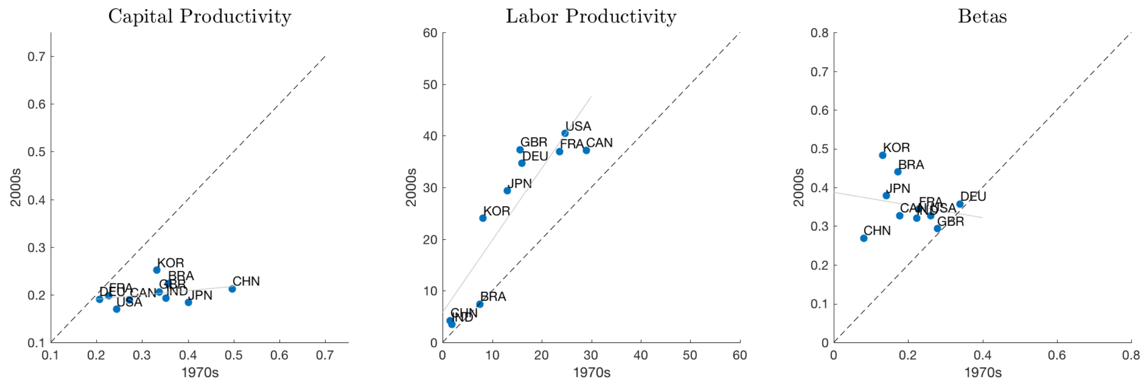
Notes: (a) The model suggests that the relative factor endowments have become more equal over the sample period. (b) The average trade cost is computed over all trading partners. It reduced substantially over the sample period.

1.5.3 Drivers of convergence

Through the lens of the model, what are the fundamental differences between countries in the 1970s compared to the 2000s? To answer this, I determine labor productivities, capital productivities, and bilateral trade costs which match capital-output ratios, capital shares and bilateral trade flows at the beginning and the end of the sample. Then the rate of time preferences is computed such that the observed capital stock is consistent with the steady state. The estimation uses GDP adjusted for natural resources because the model does not incorporate factors of production other than labor and capital. Moreover, the bilateral trade matrix is adjusted such that it accurately reflects the trade to GDP ratio and the assumption of balanced trade in the model. Due to the smaller sample, not all trade is properly captured in this exercise, but trade within these countries already makes up 55 % of total exports. The estimation is fairly precise. In Appendix C, Figure 1.19a and 1.19b show that the model predictions are close to the data. Besides, the labor productivity that comes out of this analysis is highly correlated with the TFP data from the PWT 9.0.

In this model, there are two main drivers of convergence - changes in bilateral trade costs and changes in the relative factor endowment. Figure 1.13 confirms that the trade cost has reduced substantially between the beginning and the end of the sample, especially for emerging markets. In addition, we also see that the relative factor endowments have become more similar. A closer look at the data suggests that mainly capital accumulation and movements in productivities are behind this change. There are no substantial alterations in population shares. Fig-

Figure 1.14: Estimation results



Notes: (a) The model suggests that differences in capital productivity are small and have reduced over time. (b) Differences in labor productivity are large and have widened over the sample period. (c) The rate of time preference has increased and converged over time.

ure 1.14 shows the shifts in estimated productivities and preferences for savings between the two steady states. Differences in capital productivities across countries are small and have reduced over the period. However, differences in labor productivity are substantial and have widened further over time. Finally, the rate of time preference has increased for all countries in the sample and converged over time.

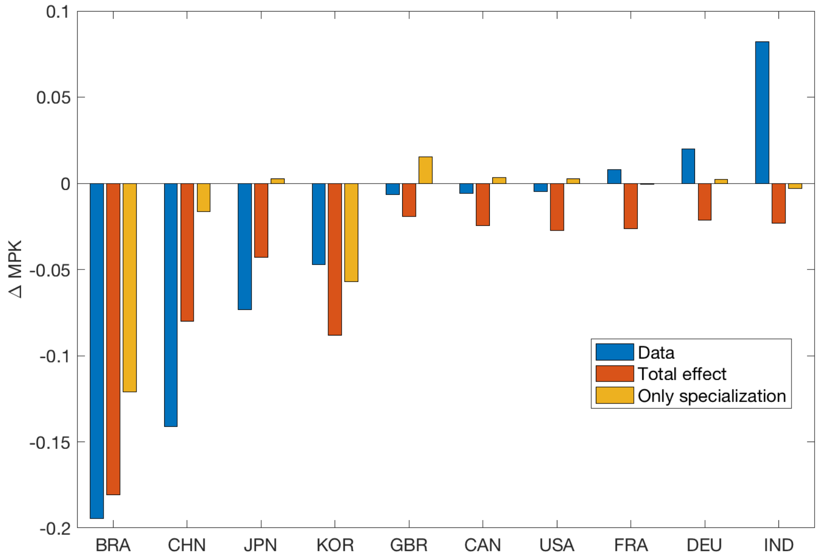
1.5.4 Counterfactual exercise

Which of these fundamental shifts has been the most important for the decline in the dispersion of MPKs? A simple counterfactual exercise provides an estimate of the contribution of the fall in trade cost in the convergence of MPKs. Using the calibrated model for the beginning of the sample, one can compute the counterfactual MPKs predicted by the fall in trade cost from the 1970s to 2000s, keeping everything else constant. Which percentage of the actual changes in MPK can be explained solely by the change in trade cost? Following Święcki (2017), I measure the contribution of the fall in trade cost as:

$$\text{Contribution} = 1 - \frac{\sum_i |\Delta MPK^{cf} - \Delta MPK|}{\sum_i |\Delta MPK|} \tag{1.42}$$

where ΔMPK^{cf} and ΔMPK denote the counterfactual and the actual change in MPK respectively. This simple measure has desirable properties. The contribu-

Figure 1.15: Counterfactual: Reduction of trade cost to level of 2003-2007

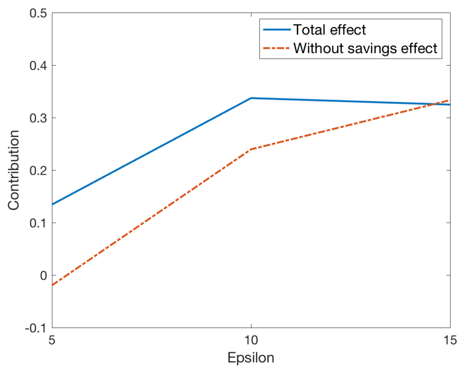


Notes: The graph shows the actual change in MPK compared to the predicted change in the MPKs if only the trade cost had fallen to the level of the end of the sample and productivities and βs had stayed constant. The yellow bars show the predicted change shutting down the savings channel, thus keeping the capital stock constant.

tion would be 1 if the counterfactual change in MPK can fully capture the change observed in the data. Moreover, if the counterfactual exercise predicts movements in the opposite direction of the data, the contribution can be negative. Using this measure, the fall in trade cost explains 33.7 % of the change in MPKs. Shutting down the savings channel and thus keeping the capital endowment constant, the contribution would only be 24.0 %.

Figure 1.15 compares the counterfactual change in MPKs to the data. This shows that the fall in trade cost can explain a large share of the movement in MPKs for Brazil, China, Japan and Korea whereas the change in trade cost is unable to explain the increase in MPK in India. The movements in advanced economies are generally small and trade does not explain the changes in MPK well, which could also be due to fairly high levels of trade integration between developed economies already in the 70s.

Figure 1.16: Robustness: ϵ



Notes: The contribution is measured as described in eq. 1.42.

1.5.5 Robustness

To what extent do the quantitative results depend on the calibration? Here the elasticity of substitution between varieties ϵ governs the size of the effect of a fall in trade cost on factor prices. Figure 1.16 depicts the contribution of the fall in trade cost to the convergence for different parameter choices of ϵ. This highlights that the total effect is higher when substitutability is larger, but the relative importance of the savings channel is larger for low values of ϵ. Intuitively, the parameter ϵ governs the relative importance of intra- vs. inter-industry trade. Low substitutability between inputs makes intra-industry trade more important which lowers the scope of specialization, but it also augments the terms-of-trade effects/ decreasing returns to scale which reinforces the savings channel.

1.6 Conclusion

The contribution of this paper is threefold. First, it uncovers the new fact that marginal products of capital have converged since the 1970s and it shows how this development is linked to changes in capital stock and movements in the capital share. Secondly, it builds a tractable multi-country model of international trade with intertemporal decision-making that contains two channels through which the fall in trade cost can explain the convergence in returns - (i) specialization along factor-intensities and (ii) life-cycle savings. Thirdly, it uses the model to quantify the contribution of trade integration and changes in factor endowments where the fall in trade cost accounts for about 30 % of the convergence in MPKs. Overall, the results suggest that the productive efficiency of the allocation of capital across countries has improved over time, but that this is not the result of large-scale capital reallocation through international financial markets. Trade integration, on the other hand, offers an explanation consistent with the empirical evidence. Thus, international financial frictions may well be substantial and co-exist with equalized MPKs.

1.7 Appendix A - Data

1.7.1 Measurement of the return

Capital share

The capital share is unobserved, but it can be computed as the residual income taking into account other factors of production like labor. The labor share is provided by the Penn World table. Caselli & Feyrer particularly worry about distortions from falsely attributing income to natural resources to produced capital - a potential problem for resource-rich developing countries. Caselli & Feyrer already adjust their measure for natural resources, but they need to make strong assumptions to deduce the income share from wealth data. Here the factor-income share that goes to natural resources is approximated by the rent to natural resources provided by the World Bank. Using these two variables, it is possible to compute the capital share α as:

$$\alpha = 1 - \text{Labor Share} - \text{Natural Resource Rent}$$

This method is also used in Monge-Naranjo et al. (2015) and leads them to find more misallocation than Caselli & Feyrer's original analysis suggests.

Output, capital and prices

The Penn World Table provides data on both real and nominal GDP and capital stock. Since Caselli & Feyrer wrote their paper, this data has been revised substantially and several improvements on measuring the capital stock, output and prices have been introduced. An important adjustment is the correct measurement of the price of capital. There are new waves of the IPC Program available and the Penn World Table has improved the quality of the price data available. These adjustments allows me to analyze the return to capital also over time whereas Caselli & Feyrer focus on the year 1996.

Depreciation

Caselli & Feyrer focus on the MPK-part of the return, assuming that depreciation rates are equal across countries and due to data limitation ignoring movements in the price of capital. The Penn World Table now computes the capital stock from disaggregated investment data which differentiates different types of capital, for example "Structures" and "Transport Equipment". Different types of capital have very different depreciation rates. Stemming from the heterogeneity of the composition of the aggregate capital stock, it is possible to compute country-specific, time-varying depreciation rates. Together with the price data, this data can be used to compute the capital-gain part of the return.

Timing

Note that the timing in the measure does not correspond exactly to the timing in most macro models. In the model, the return is given by $R_t^{\text{Model}} = \frac{r_t}{P_{t-1}^K} + (1 - \delta)\frac{P_t^K}{P_{t-1}^K}$. The advantage of the timing used in my measure is that it ensures an accurate value of capital stock. The price level of capital uses price data on investment goods weighted by the composition of the capital stock. Multiplying the current price of capital with the current capital stock thus gives an accurate estimate of the current value of capital. However, multiplying the current capital stock with past prices only gives an accurate value of current capital at past prices *if the capital composition stayed constant*. Moreover, the timing chosen ensures that the measure is directly comparable to Caselli and Feyrer (2007). A robustness check shows that the measures obtained under the model-consistent timing is very similar. Finally, since the quantification compares two steady-states, the difference in timing does not matter for the quantitative results.

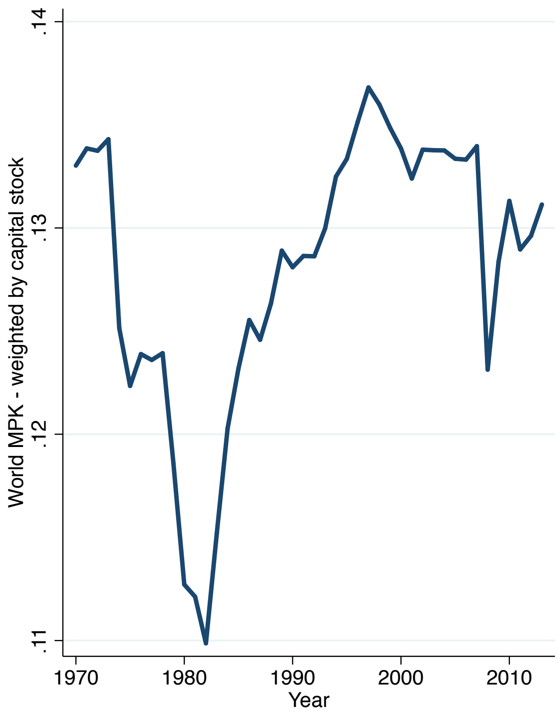
1.7.2 Replicating the results from Caselli & Feyrer (2007)

The replication of the results by Caselli & Feyrer in Fig. 1.3a is based on the Data Appendix to their paper. Their dataset includes 52 countries. Due to missing data, the revised MPK cannot be calculated for Algeria, Burundi, Belgium, Congo, Mauritius, El Salvador and Zambia. Due to data inconsistencies which suggest potential problems with the data, Botswana, Trinidad Tobago and Venezuela are dropped from the revised sample although Caselli & Feyrer use them in their analysis. Hongkong appears in their table of results, but it is not included in the data appendix and thus not included in the figure.

1.7.3 The World Return

This methodology can also be used to compute a measure of the world return and the world MPK where both are weighted averages across countries using capital stocks as weights. As shown in figure 1.17, the world MPK briefly fell in the early 1980s, but quickly rose back to its starting level. Although there is an upward trend since the early 80s, overall the level is relatively constant. The real world return (corrected for U.S. inflation) is also stable over the last 20 years of the sample, although it has decreased from its highest level in the 1970s (see Figure **??**). Here the reduction in volatility catches the eye, but the absence of a decline suggests that lower returns to capital are not the reason for low world interest rates.

Figure 1.17: World Aggregate MPK

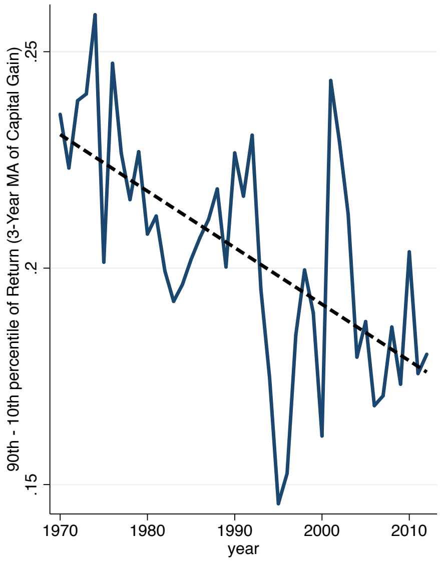


Notes: The World MPK is computed as capital-weighted average across countries.

1.7.4 Additional Figures and Tables

(a) σ - Convergence

(b) Difference between 90th and 10th percentile across Countries



1.8 Appendix B - Model

1.8.1 Proofs

Conditional FPE when $\tau = 1$

Due to the presence of a trade cost, in general there will be no factor price equalization. However, one can show that when $\tau_{c,\tilde{c}} = 1 \; \forall c, \tilde{c}$ then this model exhibits conditional factor price equalization. It is easy to see that in this case the sectoral price index is the same across countries $P_c(z) = P_{c'}(z) = P(z)$ and thus:

$$(\epsilon - 1)^{1-\epsilon}\epsilon^{\epsilon} f = \left(\frac{P^K}{\frac{r_c}{A_c^K}}\right)^{\epsilon} \sum_{c'} Q_{c'}^K$$

$$\frac{r_c}{A_c^K} = \frac{r_{c'}}{A_{c'}^K}$$

$$\frac{w_c}{A_c^L} = \frac{w_{c'}}{A_{c'}^L}$$

In the absence of trade costs, efficient factor prices will be equalized. In the section 1.8.1, I prove that in general there is no factor-price equalization in this model.

$$\left((1-\hat{p}(z)^{-\epsilon})^{\frac{\epsilon-1}{\epsilon-\sigma}} - \left((\hat{p}(z)^{-\epsilon}-1) \frac{P^{\sigma}Y}{P^{*\sigma}Y^{*}} \right)^{\frac{\epsilon-1}{\epsilon-\sigma}} \right) \frac{n}{n_i^*} \tag{1.43}$$

$$= \left((\hat{p}(z)^{-\sigma} - \hat{p}(z)^{\epsilon-\sigma}) \frac{P^{\sigma}Y}{P^{*\sigma}Y^{*}} \right)^{\frac{\epsilon-1}{\epsilon-\sigma}} - (\hat{p}(z)^{\epsilon-\sigma} - \hat{p}(z)^{-\sigma})^{\frac{\epsilon-1}{\epsilon-\sigma}} \tag{1.44}$$

The RHS decrease in p, whereas the LHS increases in p. The only solution is thus p=1, when both have the same factor price. The endowment doesn't matter in this case.

No Conditional Factor Price Equalization when $\tau \geq 1$

Here I show that in general, there will be no (conditional) factor price equalization in this model. This is shown by contradiction. Recall the market clearing conditions. For simplicity, let $\tau_{c',c} = \tau > 1 \, \forall c', c \setminus c = \tilde{c}$.

$$(\epsilon-1)^{1-\epsilon}\epsilon^{\epsilon} f = \left(\frac{r_{c'}}{A^K_{c'}} \right)^{-\epsilon} \left(\left(P^K_{c'}\right)^{\epsilon} Q^K_{c'} + \sum_{c \neq c'} \tau^{1-\epsilon} \left(P^K_c\right)^{\epsilon} Q^K_c \right)$$

$$(\epsilon-1)^{1-\epsilon}\epsilon^{\epsilon} f = \left(\frac{w_{c'}}{A^L_{c'}} \right)^{-\epsilon} \left(\left(P^L_{c'}\right)^{\epsilon} Q^L_{c'} + \sum_{c \neq c'} \tau^{1-\epsilon} \left(P^L_c\right)^{\epsilon} Q^L_c \right)$$

Now assume there is FPE such that $\frac{r_N}{A^K_N} = \frac{r_S}{A^K_S}$ and $\frac{w_N}{A^L_N} = \frac{w_S}{A^L_S}$, then this would imply:

$$\left(\left(P^K_N\right)^{\epsilon} Q^K_N + \sum_{c \neq N} \tau^{1-\epsilon} \left(P^K_c\right)^{\epsilon} Q^K_c \right) = \left(\left(P^K_S\right)^{\epsilon} Q^K_S + \sum_{c \neq S} \tau^{1-\epsilon} \left(P^K_c\right)^{\epsilon} Q^K_c \right)$$

$$(\tau^{1-\epsilon}-1) \left(P^i_S\right)^{\epsilon-1} P^i_S Q^i_S = \left(P^i_N\right)^{\epsilon-1} P^i_N Q^i_N (\tau^{1-\epsilon}-1)$$

$$\left(P^i_S\right)^{\epsilon-1} P^i_S Q^i_S = \left(P^i_N\right)^{\epsilon-1} P^i_N Q^i_N$$

Using the demand for intermediate goods this can be transformed to:

$$
\begin{aligned}
(P_N^i)^{\epsilon-\sigma} P_N^\sigma Q_N &= (P_S^i)^{\epsilon-\sigma} P_S^\sigma Q_S \\
\left(\frac{P_N^K}{P_N^L}\right)^{\epsilon-\sigma} &= \left(\frac{P_S^K}{P_S^L}\right)^{\epsilon-\sigma} \\
\left(\frac{n_N^L N_L \left(\frac{w}{A^L}\right)^{1-\epsilon} + n_S^L N_L \left(\frac{\tau w}{A^L}\right)^{1-\epsilon}}{n_N^K N_K \left(\frac{r}{A^K}\right)^{1-\epsilon} + n_S^K N(K) \left(\frac{\tau r}{A^K}\right)^{1-\epsilon}}\right)^{\frac{\epsilon-\sigma}{\epsilon-1}} &= \left(\frac{n_N^L N_L \left(\frac{\tau w}{A^L}\right)^{1-\epsilon} + n_S^L N_L \left(\frac{w}{A^L}\right)^{1-\epsilon}}{n_N^K N_K \left(\frac{\tau r}{A^K}\right)^{1-\epsilon} + n_S^K N_K \left(\frac{r}{A^K}\right)^{1-\epsilon}}\right)^{\frac{\epsilon-\sigma}{\epsilon-1}} \\
\frac{A_N^L L_N + A_S^L L_S \tau^{1-\epsilon}}{A_N^K K_N + A_S^K K_S \tau^{1-\epsilon}} &= \frac{A_N^L L_N \tau^{1-\epsilon} + A_S^L L_S}{A_N^K K_N \tau^{1-\epsilon} + A_S^K K_S} \\
\frac{A_S^K K_S}{A_S^L L_S}\left(1 - \tau^{2-2\epsilon}\right) &= \frac{A_N^K K_N}{A_N^L L_N}\left(1 - \tau^{2-2\epsilon}\right) \\
\frac{A_S^K K_S}{A_S^L L_S} &= \frac{A_N^K K_N}{A_N^L L_N}
\end{aligned}
$$

The last line contradicts the fact that countries vary in productivity, population and preferences for savings. In general, countries will not have the same efficiency ratio of capital to labor. So only if both countries did have the same capital-efficient labor ratio could there be FPE in this model. This is a necessary, not a sufficient condition. Due to size effects, countries may still have different factor prices even if their efficiency ratios of capital to labor were equal. Moreover, note that I have assumed symmetric trade costs in this derivation. Asymmetric trade costs would be an additional channel that prevents conditional FPE, even when factor endowments are the same.

Deriving equation 1.40

Two countries

$$
p(z)^{-\epsilon}\left((P(z))^{\epsilon-1} P(z)Q(z) + \left(\frac{P^*(z)}{\tau}\right)^{\epsilon-1} P^*(z)Q^*(z)\right) = (\epsilon-1)^{1-\epsilon}\epsilon^\epsilon f
$$
$$
p^*(z)^{-\epsilon}\left(\left(\frac{P(z)}{\tau}\right)^{\epsilon-1} P(z)Q(z) + (P^*(z))^{\epsilon-1} P^*(z)Q^*(z)\right) = (\epsilon-1)^{1-\epsilon}\epsilon^\epsilon f
$$

This can be transformed by combining the two equations and using the Cobb-Douglas property of constant expenditure shares to write as a function of income. Notation: the relative price is denoted by $\hat{p}(z)$.

$$
P_{c,t}^i Q_{c,t}^i = \left(\frac{P_{c,t}^i}{P_{c,t}}\right)^{1-\sigma} (w_{c,t}L_{c,t} + r_{c,t}K_{c,t})
$$

$$\hat{p}(z)^{-\epsilon}\left(\left(\frac{P(z)}{P^*(z)}\right)^{\epsilon-\sigma}\frac{P^\sigma Y}{P^{*\sigma}Y^*}+\tau^{1-\epsilon}\right)=\left(\tau^{1-\epsilon}\left(\frac{P(z)}{P^*(z)}\right)^{\epsilon-\sigma}\frac{P^\sigma Y}{P^{*\sigma}Y^*}+1\right)$$

$$\left(\frac{P(z)}{P^*(z)}\right)^{\epsilon-\sigma}\frac{P^\sigma Y}{P^{*\sigma}Y^*}\left(\hat{p}(z)^{-\epsilon}-\tau^{1-\epsilon}\right)+\hat{p}(z)^{-\epsilon}\tau^{1-\epsilon}=1$$

The relative sectoral price is given by:

$$\frac{P(z)}{P^*(z)}=\left[\frac{\frac{n(z)}{n^*(z)}\hat{p}(z)^{1-\epsilon}+\tau^{1-\epsilon}}{\frac{n(z)}{n^*(z)}\left(\tau\hat{p}(z)\right)^{1-\epsilon}+1}\right]^{\frac{1}{1-\epsilon}}$$

$$\frac{n(z)}{n^*(z)}=\frac{\left(\left(\hat{p}(z)^{-\sigma}-\tau^{1-\epsilon}\hat{p}(z)^{\epsilon-\sigma}\right)\frac{P^\sigma Y}{P^{*\sigma}Y^*}\right)^{\frac{\epsilon-1}{\epsilon-\sigma}}-\tau^{1-\epsilon}\left(\hat{p}(z)^{\epsilon-\sigma}-\hat{p}(z)^{-\sigma}\tau^{1-\epsilon}\right)^{\frac{\epsilon-1}{\epsilon-\sigma}}}{\left(\left(1-\hat{p}(z)^{-\epsilon}\tau^{1-\epsilon}\right)^{\frac{\epsilon-1}{\epsilon-\sigma}}-\left(\left(\hat{p}(z)^{-\epsilon}-\tau^{1-\epsilon}\right)\frac{P^\sigma Y}{P^{*\sigma}Y^*}\right)^{\frac{\epsilon-1}{\epsilon-\sigma}}\tau^{1-\epsilon}\right)}$$

Now for this expression it is clear that an increase in the relative price will decrease the RHS.

Many countries

$$\hat{p}(z)^{-\epsilon}\sum_{c}^{C}\tau_{c',c}\left(\frac{P_c^K}{\tau_{c',c}}\right)^{\epsilon}Q_c^K=\sum_{c}^{C}\tau_{c',\tilde{c}}\left(\frac{P_c^K}{\tau_{c',\tilde{c}}}\right)^{\epsilon}Q_c^K$$

$$\hat{p}(z)^{-\epsilon}\left[\left(\frac{P_{c'}^K}{P_{\tilde{c}}^K}\right)^{\epsilon-\sigma}\frac{P_{c'}^\sigma Q_{c'}}{P_{\tilde{c}}^\sigma Q_{\tilde{c}}}+\tau_{c',\tilde{c}}^{1-\epsilon}+\sum_{c\neq c',\tilde{c}}^{C}\tau_{c',c}^{1-\epsilon}\left(\frac{P_c^K}{P_{\tilde{c}^K}}\right)^{\epsilon-\sigma}\frac{P_c^\sigma Q_c}{P_{\tilde{c}}^\sigma Q_{\tilde{c}}}\right]=$$
$$\left[\tau_{\tilde{c},c'}^{1-\epsilon}\left(\frac{P_{c'}^K}{P_{\tilde{c}}^K}\right)^{\epsilon-\sigma}\frac{P_{c'}^\sigma Q_{c'}}{P_{\tilde{c}}^\sigma Q_{\tilde{c}}}+1+\sum_{c\neq c',\tilde{c}}^{C}\tau_{\tilde{c},c}^{1-\epsilon}\left(\frac{P_c^K}{P_{\tilde{c}^K}}\right)^{\epsilon-\sigma}\frac{P_c^\sigma Q_c}{P_{\tilde{c}}^\sigma Q_{\tilde{c}}}\right]$$

For simplicity, assume that the trade cost is the same between all destinations:

$$\left(\hat{p(z)}^{-\epsilon}-\tau^{1-\epsilon}\right)\left(\frac{P_{c'}^K}{P_{\tilde{c}}^K}\right)^{\epsilon-\sigma}\frac{P_{c'}^\sigma Q_{c'}}{P_{\tilde{c}}^\sigma Q_{\tilde{c}}}+\hat{p(z)}^{-\epsilon}\tau^{1-\epsilon}=1+\left(1-\hat{p(z)}^{-\epsilon}\right)\underbrace{\sum_{c\neq c',\tilde{c}}^{C}\tau^{1-\epsilon}\left(\frac{P_c^K}{P_{\tilde{c}^K}}\right)^{\epsilon-\sigma}\frac{P_c^\sigma Q_c}{P_{\tilde{c}}^\sigma Q_{\tilde{c}}}}_{=G}$$

The relative sectoral price is given by:

$$\frac{P(z)}{P^*(z)}=\left[\frac{\frac{n_{c'}}{n_{\tilde{c}}}(\tau\hat{p}(z))^{1-\epsilon}+1+\sum_{c\neq c',\tilde{c}}\frac{n_c}{n_{\tilde{c}}}(\tau\frac{p_c}{p_{\tilde{c}}})^{1-\epsilon}}{\frac{n_{c'}}{n_{\tilde{c}}}\hat{p}(z)^{1-\epsilon}+\tau^{1-\epsilon}+\sum_{c\neq c',\tilde{c}}\frac{n_c}{n_{\tilde{c}}}\left(\tau\frac{p_c}{p_{\tilde{c}}}\right)^{1-\epsilon}}\right]^{\frac{1}{\epsilon-1}}$$

Using this gives:

$$(\hat{p(z)}^{-\epsilon} - \tau^{1-\epsilon}) \left(\frac{\frac{n_{c'}}{n_{\tilde{c}}}(\tau\hat{p})^{1-\epsilon} + 1 + \sum_{c \neq c', \tilde{c}} \frac{n_c}{n_{\tilde{c}}}(\tau \frac{p_c}{p_{\tilde{c}}})^{1-\epsilon}}{\frac{n_{c'}}{n_{\tilde{c}}}\hat{p}^{1-\epsilon} + \tau^{1-\epsilon} + \sum_{c \neq c', \tilde{c}} \frac{n_c}{n_{\tilde{c}}} \left(\tau \frac{p_c}{p_{\tilde{c}}}\right)^{1-\epsilon}} \right)^{\frac{\epsilon-\sigma}{\epsilon-1}} \frac{P_{c'}^{\sigma} Q_{c'}}{P_{\tilde{c}}^{\sigma} Q_{\tilde{c}}} + \hat{p(z)}^{-\epsilon} \tau^{1-\epsilon} = 1 + (1 - \hat{p(z)}^{-\epsilon})G$$

Solving the equation for the relative share of firms $\frac{n_{c'}}{n_{\tilde{c}}}$ gives:

$$\frac{n_{c'}}{n_{\tilde{c}}} = \frac{(1+H)\left(\hat{p}(z)^{-\sigma} - \tau^{1-\epsilon}\hat{p}^{\epsilon-\sigma}\right)^{\frac{\epsilon-1}{\epsilon-\sigma}} \left(\frac{P_{c'}^{\sigma} Q_{c'}}{P_{\tilde{c}}^{\sigma} Q_{\tilde{c}}}\right)^{\frac{\epsilon-1}{\epsilon-\sigma}} - \left(\hat{p}^{\epsilon-\sigma} - \hat{p}(z)^{-\sigma}\tau^{1-\epsilon} + \hat{p}^{\epsilon-\sigma}(1 - \hat{p(z)}^{-\epsilon})G\right)^{\frac{\epsilon-1}{\epsilon-\sigma}} \left(\tau^{1-\epsilon} + H\right)}{\left(1 - \hat{p}(z)^{-\epsilon}\tau^{1-\epsilon} + \left(1 - \hat{p}(z)^{-\epsilon}\right)G\right)^{\frac{\epsilon-1}{\epsilon-\sigma}} - \tau^{1-\epsilon}\left(\hat{p}(z)^{-\epsilon} - \tau^{1-\epsilon}\right)^{\frac{\epsilon-1}{\epsilon-\sigma}} \left(\frac{P_{c'}^{\sigma} Q_{c'}}{P_{\tilde{c}}^{\sigma} Q_{\tilde{c}}}\right)^{\frac{\epsilon-1}{\epsilon-\sigma}}}$$

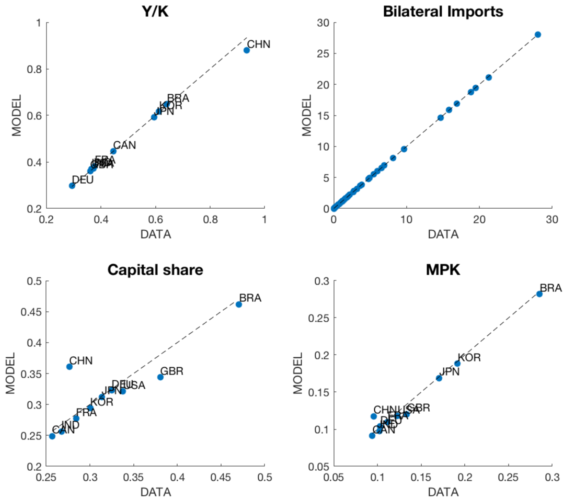
where $H = \sum_{c \neq c', \tilde{c}} \frac{n_c}{n_{\tilde{c}}}(\tau \frac{p_c}{p_{\tilde{c}}})^{1-\epsilon}$. The RHS of this expression clearly decreases in $\hat{p}(z)$.

Appendix C - Quantification

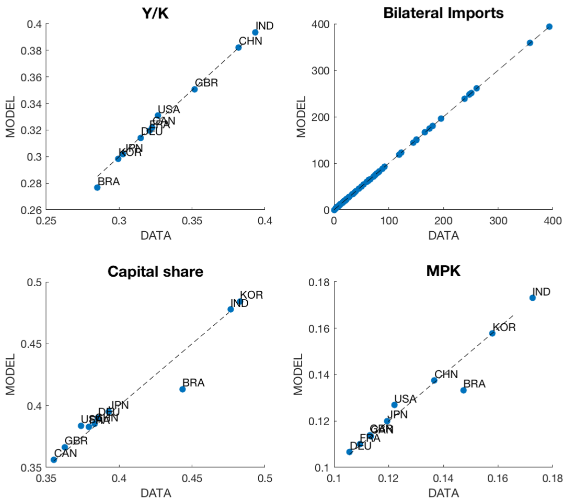
Figures 1.19a and 1.19b present the fit of the calibration for $\epsilon = 10$.

Figure 1.19: Calibration fit

(a) Beginning of sample - 1970s



(b) End of sample - 2000s



Notes: Estimating the productivities and the trade cost allows the model to fit the data almost perfectly.

Chapter 2

WHY HAS THE RETURN TO CAPITAL NOT DECLINED WITH INTEREST RATES?

2.1 Introduction

The U.S. aggregate return to capital has not fallen along with real interest rates over the last 30 years. While the literature has debated the causes and consequences of the fall in real interest rates extensively, this divergence has received less attention. The return to capital and the interest rate should be closely linked. In a world without capital market frictions or taxes, a rational investor must be indifferent between putting his money in the bank at the prevailing interest rate and buying a unit of capital, renting it out at the current rental rate and reselling it next period.

The decrease in real interest rates and the slowdown in productivity growth have raised concerns of secular stagnation, a term coined by Alvin Hansen in 1938 and revived in Summers (2014). It suggests a change in the long-run equilibrium between savings and investment. A key part of the secular stagnation hypothesis is the lack of profitable investment opportunities. However, as Gomme et al. (2015) argue the continued high return to business capital is at odds with a lack of demand for investment. At the same time investment recovered very slowly from the recession and the causes for under-investment constitute a significant policy concern.

Why has the return to capital not fallen along with interest rates? Understanding this development can show if and why investment rates might be too low and how economic policy can respond. To answer this research question, this chapter evaluates three hypotheses that can explain this empirical finding. Firstly, a rise in capital frictions such as borrowing constraints or risk premia may prevent additional funds from flowing into high-return investment projects despite low interest rates. Secondly, pure profit rents might confound factor income attributed to capital. If a decline in competition has allowed concentration, market power and thus markups to rise, measured returns will overestimate the true return. In this case, returns to capital may have fallen along with interest rates, and the difficulty in distinguishing capital from pure profit rents would explain the divergence between measured returns and interest rates. Thirdly, changes in the distribution of both of these frictions across firms can magnify or dampen their effect. Improvements in the allocative efficiency of capital across firms thus resonate on the macro level.

Macroeconomic data provide limited information to disentangle between these different explanations for the aggregate trend and is silent on how firm heterogeneity translates into aggregate effects. Thus to differentiate between these three hypotheses and quantify their contribution, this paper draws on evidence from firm-level data. As opposed to macroeconomic data, the richness of firm-level data opens ways to control for endogeneity in production function estimation and speaks to the underlying structural changes that drive the aggregate movements in returns to capital.

The evidence in this paper suggests that around 70 percent of the divergence between the aggregate return and interest rates can be attributed to a rise in capital frictions, of which around 15 percent are to the changes in risk premia. Although this paper confirms the rise in markups in the U.S. since this 1980s, the decomposition suggests that the profit component explains only around 30 percent of the behavior of the aggregate return. This apparent contradiction is explained by the rise in fixed costs which increasingly requires firms to sell above marginal cost. Thirdly, this exercise shows most of the dynamics occur within-industries and are not driven by sectoral reallocation. There are however important shifts in the composition of firms that contribute to the aggregate trend, with large firms having become more profitable, but also less constraint.

The structure of this chapter is at follows. The first section documents the macro-level evidence on the divergence between the aggregate return to capital and interest rates in the U.S and how it compares to other advanced economies. The second section lays out the theoretical framework to illustrate the different explanations on the firm-level. The third section quantifies the contribution of each hypothesis. It exploits methodological advances to estimate output elasticities of industry-level production functions and translates them into a decomposition of the aggregate return. It then describes how within- and between-firm reallocations explain the movements in the aggregate return and its components.

2.2 Related Literature

This paper is related to three strands of literature. The first strand consists of a number of empirical papers that measure and explain aggregate trends related to interest rates, the return to capital, investment and factor income shares. The second one examines the causes and consequences of some of these aggregate trends drawing on macroeconomic theory and calibrating and estimating macroeconomic models that incorporate a number of potential explanations. Finally, this paper links to the literature that uses firm-level evidence to understand aggregate trends in productivity, which includes a number of papers on production function estimation, on capital misallocation and on financial frictions.

Gomme et al. (2011) compute the return to business capital in the U.S. and find it to be less volatile than stock market returns. Updating their estimates they find a continuously high aggregate return to capital despite the decline in real interest rates. The divergence between the return to (business) capital and real interest rates for the U.S. raises doubts about the secular stagnation hypothesis. High returns indicate profitable investment opportunities and suggest that the economy has not exhausted its full growth potential. Understanding why the return to capital has not fallen along with interest rates can thus shed light on the reasons for the apparent lack of investment or if this trend is the efficient response to fundamental changes in the structure of the economy. Jordà et al. (2018) provide additional historical evidence on rates of returns over time for different assets. Their findings emphasize the increase in premiums across assets with different riskiness.

In addition to the papers that directly compute the return to capital from national accounts data, this paper also links to a large literature measuring factor income shares, facing similar challenges related to data availability and competing theories. For example, many explanations for the fall in the labor share (see Karabarbounis and Neiman (2014)) also affect the return to capital, such as the role of profits (see Barkai (2016)) or the rise of intangible capital (see Koh et al. (2018)). In a recent paper, Karabarbounis and Neiman (2018) coin the term "factorless" income. It denotes the residual income not accounted for by labor or capital income, where capital income is computed as total capital stock times the interest rate. Their paper provides a summary of macroeconomic evidence that support different explanations, but the conclusions are limited by the ability of aggregate data to distinguish between them. The concept of factorless income is closely linked to the divergence between the aggregate return to capital and the real interest rate. The findings in the present paper explain how capital frictions, risk and profits lead capital income to account for the residual income Karabarbounis and Neiman (2018) document.

A few papers show how these trends may fit into macroeconomic theory and use the additional structure to discipline which explanations are consistent and quantitatively important. Caballero et al. (2017) link the relative constancy of the return to productive capital compared to "safe" real rates to the decline in labor shares and additional evidence on corporate spreads. In a macro-accounting framework they compare the ability of changes in markups, the relative price of investment goods, capital-biased technological change and changes in the risk-premium to explain these broad trends for the U.S. They suggest that an increase in the risk-premia between productive capital and government bonds is likely to be important. The shortage of safe assets has been proposed in several papers such as Barro et al. (2017) and Caballero et al. (2016), but an increase in risk-aversion or (perceived) riskiness of investments in business capital as proposed in Marx et al. (2018) has similar effects on the risk-premia. Farhi and Gourio (2018) employ this strategy and extend a neoclassical growth model to account for these developments. They also conclude that rising risk-premia along with rising market power and a larger share of (unmeasured) intangible capital explain the divergence.

While most of the macroeconomic literature focuses on risk and safe assets, empirical evidence from micro studies has emphasized a rise in market power. For example Eeckhout and De Loecker (2017) find that average markups have increased looking at publicly listed firms in the U.S. Comparing across industries, Gutiérrez and Philippon (2017a) link low investment rates relative to Tobin's Q to lower levels of competition/ higher market concentration. On the other hand, Traina (2018) find that Eeckhout and de Loecker's result depends on the definition of variables costs. Karabarbounis and Neiman also ask if higher profit shares can explain factorless income, but they find that it would require an unrealistically high correlation between interest rates and profits and would imply implausibly high markups for the 1960s and 1970s.

In terms of methodology, this paper draws on the large literature on production function estimation using firm-level data, especially Ackerberg et al. (2015) and Loecker and Warzynski (2012). Although this literature has been mainly concerned with productivity, these estimation methods can also be used to uncover markups and marginal products of capital (see Eeckhout and De Loecker (2017)). Moreover, this paper relies on insights of a large literature on the effects of capital misallocation on aggregate trends in productivity, for example Hsieh and Klenow (2009) and Bartelsman et al. (2013). The micro-level evidence in this paper builds on the advances of this literature regarding the measurement of marginal revenue products on the firm-level and the role of capital frictions and risk, for example as in Gopinath et al. (2017) and David et al. (2018). A few papers such as Döttling et al. (2017) and Gutiérrez and Philippon (2017a) have recently examined industry- and firm-level evidence to explain the slump in investment. This highlights again that a firm-level empirical approach provides additional insights into the market structure that remains hidden in aggregate data.

2.3 Section I - Aggregate evidence

This section first shows how frictions and profits can lead to distortions between interest rates and the measured aggregate return to capital. Then it documents the divergence between the measured aggregate return and interest rates.

2.3.1 Aggregate return to capital

The aggregate return to capital measures the average payout obtained from buying and renting out one unit of produced capital in an economy. Let's denote the rental rate paid to capital in period t by r_t and let q_t stand for the relative price of investment goods. The return to capital is composed of the rental income generate by capital over the period and the resale value next period (or capital gain), which depends on the relative price of capital goods. Thus the real, net return to capital can be written as:

$$\text{Net Return}_t = \frac{r_t - \delta q_{t-1}}{q_{t-1}} + \frac{q_t}{q_{t-1}} - 1 \tag{2.1}$$

In a riskless and frictionless world, investor arbitrage implies that the real, net return to capital should be equal to the real, risk-free interest rate. However, the presence of risk or financial frictions can drive a wedge between the interest rate on government bonds and the return to capital. Let RP_t denote the aggregate risk-premium, and τ_t^K a capital friction increasing the average borrowing rate for the private sector, then the relationship between the net return and the risk-free interest rate i_t can be written as:

$$\text{Net Return}_t = i_t + RP_t + \tau_t^K \tag{2.2}$$

Moreover, the rental rate paid to capital is not directly observable. Following Caselli and Feyrer (2007) and Gomme et al. (2011) it can be approximated dividing total rents paid to capital by a measure of the capital stock. Using the accounting identify for national income $Y_t = w_t L_t + r_t K_t + \Pi_t$, capital income is computed from the residual income not attributed to labor $r_t K_t + \Pi_t = Y_t - w_t L_t$. In the presence of pure profits Π_t, this approach overestimates capital income. Then, the measured net return computed from the data lies above the actual net return.

$$\text{Net Return}_t^{Data} = \text{Net Return}_t^{True} + \frac{\Pi_t}{q_t K_t} \tag{2.3}$$

Combining eq. 2.3 with eq. 2.2 shows the three types of wedges that can explain the difference between the risk-free interest rate and the aggregate return to capital.

$$\text{Net Return}_t^{Data} - i_t = \tau_t^K + RP_t + \frac{\Pi_t}{q_t K_t} \tag{2.4}$$

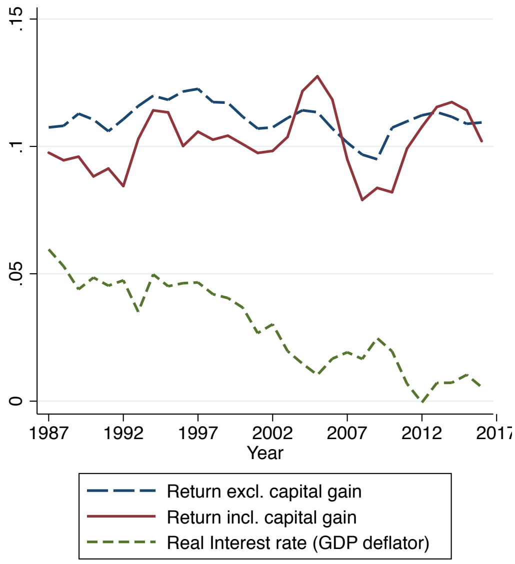
Measurement & Data

Here I compute the returns to capital for the U.S. on the aggregate and industry-level using national accounts data from the U.S. Bureau of Economic Analysis (BEA). These data are preferable compared to other sources because the data is available until 2016, disaggregated at the industry-level and already provides the components of national income. Moreover, BEA also provides estimates of private fixed assets. Similar to Gomme et al. (2011), here capital income is approximated by the gross operating surplus, computed as value added subtracting labor compensation, and taxes and subsidies to production. The capital stock is measured by total private fixed assets which include Private Equipment, Private Structures, and Intellectual Property Products. Data on capital depreciation is also available from BEA. The price index for investment good and for output is used to compute the relative price of investment goods and approximate the capital gain. Due to changes in industry definitions, BEA industry-level returns are available starting in 1987, but data on capital stocks goes back further in time. Also, it is possible to compute the aggregate return for a longer time-series, and when insightful, results are presented going back further in time.

Time trend compared to interest rates

Fig. 2.1 presents the aggregate returns for all private industries over time, showing the effect of including the capital gain term in the measures. It shows that both measures behave similar over time and the divergence between returns and real interest rate is shown

Figure 2.1: U.S. Aggregate return to capital



Sources: U.S. Bureau of Economic Analysis (BEA) and annual macro-economic database of the European Commission's Directorate General for Economic and Financial Affairs (AMECO)

by both. Real interest rates are computed using the GDP deflator. This paper focuses on the measured return excluding the capital gain term, mainly to avoid distortions due to measurement error in investment good prices. The average return over the period is around 10 %, which is higher than what for example Gomme et al. (2011) (5 %) found. The difference is due to the available data on taxation - Gomme et al. (2011) carefully adjust for a number of taxes while in BEA, the only industry-level data on taxes is for production. However, the aggregate trend remains unchanged with further adjustments for taxation. The risk-premium, pure profits and financial frictions affect the aggregate return in a similar way. Without additional information, we cannot distinguish which component drives the divergence in between returns and interest rates. Previous studies have required each potential explanation to jointly generate the divergence and replicate other aggregate trends, for example a decline in the aggregate labor share. Instead, this paper turns to firm-level data to uncover additional empirical evidence that can help to differentiate between the different explanations.

2.3.2 International evidence from Advanced Economies

Is this divergence a common phenomenon across advanced economies? Comparing the trends in the aggregate return across 6 advanced economies - the U.S., the four largest Eurozone economies and Great Britain - reveals considerable heterogeneity across countries. The return is calculated from national accounts data. Apart from the spike during the EU Sovereign debt crisis, the long-run interest rate exhibits a downward trend in all six economies starting in the late 1990s whereas the movements in the return to capital are less homogeneous. Although the aggregate return does not fully match the decline in real interest rates in any economy, there are important differences across countries, partly linked to the behaviour of the real estate sector. The aggregate return to capital in the U.S., the UK and Germany is stable over time and although removing the real estate sectors affects the level of the average return, this does not alter the overall trend. For France, Spain and Italy on the other hand, the aggregate return has declined dramatically since the early 2000s in the non-real estate part of the economy. Further discussion of the international evidence can be found in the appendix. The heterogeneity across countries emphasizes the need to take a closer look at structural changes within economies.

2.4 Section II - Theoretical framework

This section outlines how capital frictions and profits on the firm-level generate an aggregate return to capital different from the interest rate. The purpose of this model is to differentiate between the three competing hypotheses (i) frictions, (ii) competition, and (iii) (mis-)allocation and lay the ground for the empirical decomposition. By allowing firms to vary in productivity, markups, and magnitude of a size constraint, this model illustrates the implications of each hypothesis on the aggregate return to capital, the covariance between firm-characteristics and returns and within- and between-firm components of the aggregate return. Thus the model shows which driving forces behind this aggregate trend are qualitatively and quantitatively in line with the micro evidence.

2.4.1 Description

At any given point in time, the economy is populated by a mass of firms N. Firms vary in their total factor productivity A, a firm-specific size constraint κ and a firm-specific markup $\mu_{i,t}$. I denote the joint distribution of firm-characteristics as $G(A, \kappa, \mu)$.

Each firm hires variable inputs (for example labor) and capital to produce output according to a given production technology $\mathcal{Q}(A, K, L)$. Firms compete in the market for variable inputs such that the wage rate w_t is common across firms. Moreover, each firm can potentially be constrained in its ability to hire capital. This constraint serves as a general way to capture any type of friction (financial or non-financial) that affects the allocation of capital across firms.

Given the production technology, firms minimize costs. Eq. 2.5 summarizes the firm optimization problem:

$$\mathcal{L}(L_{i,t}, K_{i,t}, \Lambda_{i,t}) = w_t L_{i,t} + (i_t + \delta_{i,t})K_{i,t} + f_{i,t} - \Lambda_{i,t}(\mathcal{Q} - Q_{i,t}) - \xi_{i,t}\,(\kappa_i - K_{i,t}) \tag{2.5}$$

The first order conditions imply that firms choose capital and labor optimally if:

$$w_t = \Lambda_{i,t} \frac{\partial \mathcal{Q}(.)}{\partial L_{i,t}} \tag{2.6}$$

$$(i_t + \delta_{i,t}) + \xi_{i,t} = \Lambda_{i,t} \frac{\partial \mathcal{Q}(.)}{\partial K_{i,t}} \tag{2.7}$$

$$(\kappa_i - K_{i,t}) \geq 0 \tag{2.8}$$

$$\xi_i (K_{i,t} - \kappa_i) = 0 \tag{2.9}$$

Eq. 2.7 shows how capital frictions lead to heterogeneity in marginal products of capital across firms. A firm is unconstrained if the optimal amount of capital lies below κ_i such that $\xi_i = 0$.

We can rewrite eq. 2.7 to get an expression of the output elasticity of capital:

$$\theta_K = \frac{\partial \mathcal{Q}(.)}{\partial K_{i,t}} \frac{K_{i,t}}{Q_{i,t}} = \frac{1}{\Lambda_{i,t}} \frac{(i_t + \delta_{i,t} + \xi_{i,t}) K_{i,t}}{Q_{i,t}} \tag{2.10}$$

The Lagrangian parameter $\Lambda_{i,t}$ measures the marginal cost of producing an additional unit of output. Now let us define the markup μ as the ratio of price to marginal cost: $\mu_{i,t} = \frac{P_{i,t}}{\Lambda_{i,t}}$. Hence eq. 2.10 can be rewritten as:

$$(i_t + \delta_{i,t}) + \xi_{i,t} = \frac{\theta_K}{\mu_{i,t}} \frac{P_{i,t} Q_{i,t}}{K_{i,t}} \tag{2.11}$$

Data on prices and marginal costs are generally not available, so the firm-level markup is not directly observable. However, it is possible to infer the markup from expenditure on variable inputs. Note that similar to eq. 2.10, the optimality condition for variable inputs can be stated as:

$$w_t L_{i,t} = \frac{\theta_L}{\mu_{i,t}} P_{i,t} Q_{i,t} \tag{2.12}$$

and rearranging, the markup is given by:

$$\mu_{i,t} = \theta_L \frac{P_{i,t} Q_{i,t}}{w_t L_{i,t}} \tag{2.13}$$

The aggregate return to capital in the empirical section is computed without any assumption on the parameters of the production function. In this framework it is easy to see how frictions and the presence of pure profits explain the difference between interest rates and the measured return for each firm. Let's define pure profit as the amount remaining when all variable and fixed costs have been settled: $\Pi_{i,t} = P_{i,t} Q_{i,t} - w_{i,t} L_{i,t} - (i_t + \delta_{i,t} + \xi_{i,t}) K_{i,t} - f_{i,t}$. We can decompose the return computed from the data as follows:

$$MPK_{i,t}^{Data} = \frac{P_{i,t} Q_{i,t} - w_t L_{i,t} - f_{i,t} - \delta_{i,t} K_{i,t}}{K_{i,t}} = \frac{\Pi + (i_t + \xi_{i,t}) K_{i,t}}{K_{i,t}} \tag{2.14}$$

$$MPK_{i,t}^{Data} = i_t + \underbrace{\xi_{i,t}}_{\text{frictions}} + \underbrace{\frac{\Pi_{i,t}}{K_{i,t}}}_{\text{pure profits}} \tag{2.15}$$

This illustrates the two types of wedges that can cause systematic differences between the measured return and interest rates - frictions and pure profits on the firm-level.[1]

How do these wedges translate into aggregate deviations between the return and interest rates? As a capital-weighted sum of firm-level values, the total return also depends on the distribution of capital across firms. This gives rise to the third potential explanation - changes in (mis-)allocation of capital.

$$AggMPK_t - i_t = \sum_i^N w_{i,t} MPK_{i,t} = \sum_i^N w_{i,t} \left(\xi_{i,t} + \frac{\Pi_{i,t}}{K_{i,t}} \right) \tag{2.17}$$

$$= \mu_\xi + \mu_\pi + \sum_i w_{i,t}(\xi_{i,t} - \mu_\xi) + \sum_i w_{i,t} \left(\frac{\Pi_{i,t}}{K_{i,t}} - \mu_\pi \right) \tag{2.18}$$

where the allocation of capital across firms determines the weights $w_{i.t}$. Eg. 2.17 shows that the aggregate return depends both on the average wedges as well as the covariance between the weight and the wedge. For example reallocation towards high-friction or high-profit firms will move the aggregate return even without significant changes in the average wedge. Ultimately the allocation of capital and labor across firms depends on the distribution of firm characteristics G (productivity, markups and size constraints), the structure of aggregate demand and the functional form of the production function.

In general equilibrium, the allocation of capital across firm is endogenous and depends on the joint distribution of frictions, productivity and mark-ups. For intuition, in the model proposed by Hsieh and Klenow (2009) with constant markups across firms and a joint lognormal distribution of productivity and capital wedge τ_K, the log-difference between the aggregate return and interest rates is described by

$$log(MPK^{Agg}) - log(R) = log(\mu) + log(E(1+\tau_K)) - (\alpha(\epsilon-1)+1)\sigma_K^2 + (\epsilon-1)\rho_{K,A}\sqrt{\sigma_A^2}\sqrt{\sigma_K^2} \tag{2.19}$$

where the CES elasticity $\epsilon > 1$, σ_K^2 denotes the variance of frictions across firms and $\rho_{K,A}$ is the correlation coefficient between firm-level productivity and the capital friction. This decomposition illustrates that more dispersion in capital frictions (higher σ_K^2 misallocation) in frictions reduces the aggregate MPK for a given interest rate, whereas a strong correlation between frictions and productivity amplifies the divergence. High productivity firms attract more capital for a given friction, but a higher capital friction reduces K for a given productivity. Hsieh and Klenow (2009) do not consider heterogeneous markups, but there is a similar interaction with the capital allocation responding to the correlation between markups, productivity and capital wedges.

[1] Profits can also be expressed as:

$$\Pi_{i,t} = P_{i,t} Q_{i,t} \left(1 - \frac{\theta_L}{\mu_{i,t}} - \frac{\theta_K}{\mu_{i,t}} \right) - f_{i,t} \tag{2.16}$$

This framework highlights the different driving forces of the aggregate return. However, the output elasticities of variable inputs and capital are not directly observable from the data. Thus to make progress, further structure is needed in order to differentiate between the different explanations. While the previous literature has focused on macroeconomic data and used estimation and calibration techniques to match certain moments in the aggregate data, using firm-level data quite naturally lends itself to a different approach - production function estimation - to recover an estimate of the output elasticities.

2.5 Section III - Firm-level evidence

First, this section explains how to obtain a firm-level measure of the return to capital and compares it to the aggregate results. Secondly, production function estimation helps to differentiate between the impact of profits compared to capital frictions. Thirdly, the components are decomposed into the movements in within-firm and between-firm effects and show that the increasing correlation between firm-size and returns suggests an important role for capital frictions and misallocation.

2.5.1 Aggregate return from firm-level data

Measurement & Data

A large literature has focused on the *dispersion* in marginal revenue products on the firm-level, but there exists surprisingly little evidence on the *level* of returns in firm-level data over time. This paper uses data from Compustat to compute firm-level returns to capital. Compustat only includes publicly listed firms, which are generally larger than the average firm and not representative of the universe of firms in the economy. However, the database is useful in this application because it is available over a long time-span and includes the key variables needed to compute returns. Moreover, large firms receive considerably more weight in the aggregate return and are thus important to study. Finally, Compustat is widely used in economic research and has high quality standards. Using Compustat, I measure the firm-level marginal revenue product as:

$$MRPK = \frac{\text{Operating Income After Depreciation}}{\text{Last year's end-of-year value of produced capital assets}}$$

where operating income after depreciation is defined as sales (SALE) minus cost of production and depreciation. The cost of production includes variable costs measured by the cost of goods sold (COGS) and fixed costs measured by Selling, General Administrative Expenses (SGA).

The measurement of capital deserves more discussion. Many studies use the book-value of fixed tangible assets recorded under Property, Plants and Equipment (Net) as a measure for physical capital stock. However, numerous papers have highlighted the changing nature of capital as the economy becomes more knowledge-intensive. Thus a second measure of capital includes intangible assets. In principle, there are two ways to control for this. Either one could exclude the factor income attributed to intangible capital from the operating income. Alternatively, one can construct a measure of the total return that includes both types of produced capital. Because the factor income share of intangible assets is unobserved, this paper follows the second approach. It requires the implicit assumptions that by arbitrage investors are indifferent between additional investments in either tangible or intangible capital such that returns equalize across these two types of capital. Then the return on total capital should also be equal to the return on each individual type of asset and one can recover the return to physical capital with this adjustment.

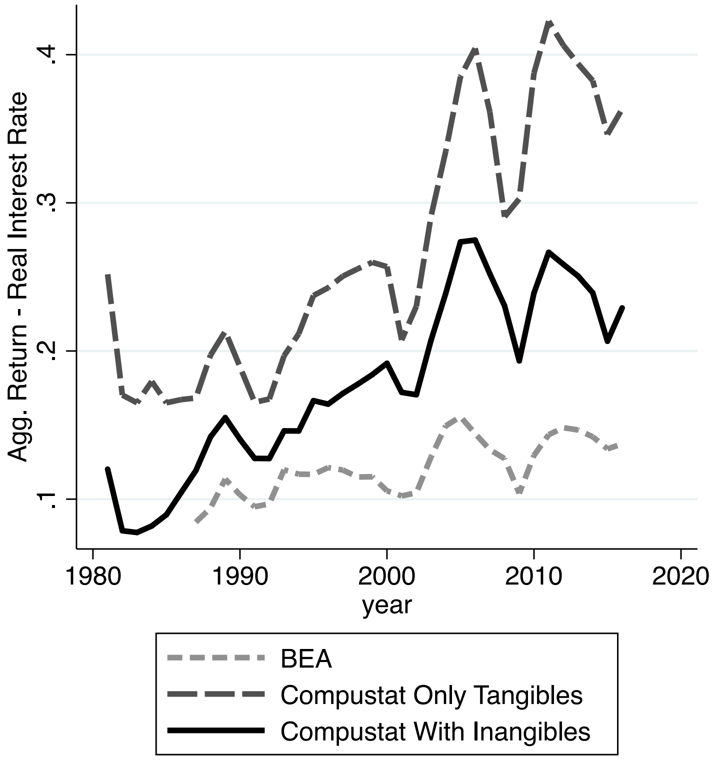
Measurement of intangible capital is still an open research area with several alternative methodologies being proposed. The preferred specification follows the recent papers by Gutiérrez and Philippon (2017b) and Gutiérrez and Philippon (2017a) to compute internally produced intangible capital in the form of knowledge and organizationial capital[2]. Internally produced intangible capital is not recorded as an asset on the balanced sheet. However, a measure can be constructed using information on research expenditure and expenditure on brands and management improvements. Research expenditure is directly recorded in Compustat. Organizational capital is approximated as 20 % of Selling, General Administrative Expenses. Deflating each series with the GDP deflator and using the perpetual inventory method over time gives an estimate of knowledge and organizational capital. Following the literature, the depreciation rate for both types of capital is 20 %. The initial values are computed from the first observed expenditure divided by the depreciation rate. Note that research expenses are normally included in Selling, General Administrative Expenses, so I subtract research expenses from the costs whenever research expenses are smaller. Finally, research expenses and expenses on organizational capital are added back to operating income. This is important because including them in intangible capital changes their definition from a cost of production to an investment expense.

Including intangible capital for firms is also in line with the aggregate data. With the change to the System of National Accounts 2008, intangible capital in the form of Intellectual Property Products (IPP) is included in the measure for private fixed assets. Firm-level data originate from companies' financial statements and the underlying accounting rules differ from national accounts, so we should not generally expect to find exactly the same result. So it is important to confirm if the firm-level data produces the same aggregate trends as we see in national accounts.

Finally, I trim the data by 1 percent, removing firms with the 1 percent highest and lowest average return. Moreover, I exclude firms in the real estate or the financial sector as well as those with a missing industry classification. The nature of capital in the real estate sector differs from its uses in many others and financial firms are subject to special accounting rules that could distort the results. I am left with 12,802 individual firms.

[2]Some papers also add purchased intangible assets recorded under "Intangibles" in Compustat. However, this measure includes the balance sheet item Goodwill which should not be included in intangible capital. Therefore, intangible capital here only consists of internally produced intangible capital.

Figure 2.2: Comparing measured aggregate return using firm-level data vs. national accounts data



Sources: U.S. Bureau of Economic Analysis (BEA) and Compustat Database

Because not all firms report all variables over the whole period, I end up with 157,935 firm-year observations between 1980 to 2016.

Comparison with national accounts data

Fig 2.2 compares the aggregate return obtained from BEA data with the capital-weighted average[3] of firm-level returns. It confirms that the firm-level data replicate the trend found in the aggregate data. Interestingly, the level of the aggregate return computed from firm-level data is substantially higher than the estimate obtained from BEA. Intangible capital explains part of this gap. Broadening the measure of capital to include a proxy for knowledge and organization capital significantly reduces the difference. However, the diver-

[3]Compustat is not representative of the U.S. industry composition. To address this the aggregation takes industry weights from BEA. Thus industry composition does not explain the difference between data sources.

gence between interest rates and the aggregate return remains[4]. What else could explain the gap? Firstly, Compustat data does not capture the universe of U.S. firms. Depending on the definition of capital, the sample from Compustat used for this comparison accounts for 30-50 % of capital in the U.S. private sector. If small and medium-sized firms (not publicly listed) have lower measured returns, e.g. due to higher pure profits, this could explain why the Compustat returns are higher. Secondly, attrition could be a concern. Firms with low or negative returns on their capital stock leave the sample and thus the aggregate return is overestimated. However, the sample composition is relatively stable and on average only around 0.02 percent of the capital stock will leave the sample in the next period due to firm exit - too small to explain the large aggregate return measured here. Finally, there are differences in the accounting rules applied at the firm-level and in aggregate data compilations. This could lead to estimates of capital being lower in the firm-level data or measures of operating surplus being higher, contributing to the gap between the measures of the aggregate return. A further consistency check shows that despite the different accounting methods and the use of a subsample of the economy, there is significant positive correlation between average industry returns computed from BEA and from Compustat.

2.5.2 Production Function Estimation

Methodology

As shown in the theoretical framework, more structure is needed to distinguish between frictions and the effect of pure profits. Production function estimation techniques as proposed in Ackerberg et al. (2015) and Loecker and Warzynski (2012) provide progress in this direction. By obtaining an estimate of the output elasticity of the variable input θ_L, the markup - the ratio between price and marginal cost - can be recovered from the ratio of sales to cost of variable inputs. This procedure gives an estimate of the marginal product and the profit component of the measured return.

Production function estimation requires a choice of functional form, here Cobb-Douglas. Eg. 2.20 shows the production function with lower case letters denoting natural logarithms.

$$q_t = \theta_L l_t + \theta_k k_t + \omega_{i,t} + \epsilon_t \tag{2.20}$$

Output q is measured by sales (SALE), variable inputs by the cost of goods sold (COGS), capital is measured by a firm's total property, plant, and equipment (PPENT) adjusted to include intangible capital as described in the previous subsection, and $\omega_{i,t}$ and ϵ_t denote total factor productivity and an error term respectively. The estimation technique chosen uses a control function. This approach helps to address the endogeneity problem created by both output and input demand responding to unobserved productivity shocks.

[4]This suggests that the omission of intangible capital in the measure of capital is not the reason for the divergence (assuming the adjustment was sufficient to capture this unobserved type of capital). However, intangible capital may still explain the trend if contributes to increases in capital frictions, for example by lowering collateral (see Caggese and Perez-Orive (2017)).

Because input demand is endogenous, we cannot use an OLS regression to recover the parameters of eq. 2.20. However, by exploiting that input demand increases monotonically with productivity, inverting the input demand function can recover the unobserved productivity shocks:

$$v_t = f(\omega_t, k_t) \tag{2.21}$$

$$\omega = f^{-1}(l_t, k_t) \tag{2.22}$$

The estimation proceeds in two stages. The first step purges sales of measurement errors and unanticipated shocks to production captured by ϵ. In the second step, the parameters of the production function θ_K and θ_L are estimated using a standard GMM procedure.

To purge output of the measurement error, substitute productivity by the inverted demand function in the production function:

$$q_t = \theta_L l_t + \theta_K k_t + f^{-1}(l_t, k_t) + \epsilon_t = \phi(l_t, k_t) + \epsilon_t \tag{2.23}$$

where $\phi(.)$ can now be estimated using any consistent non-parametric estimator. Let us denote the fitted output by $\hat{\phi}$.

For the second stage, I assume that the process for productivity is given by $\omega_t = \rho\omega_{t-1} + \eta_t$, where η_t is an innovation to the firm's productivity process. Conditional on the parameters of the production function, productivity can be obtained from the first stage as $\omega_t = \hat{\phi}_t - \theta_L l_t - \theta_K k_t$. Projecting productivity on its lagged values then recovers the shocks to productivity η .

Eq. 2.24 shows the moment conditions for a standard GMM procedure to estimate the parameters of the production function:

$$E_t \begin{pmatrix} \eta_t l_{t-1} \\ \eta_t k_t \end{pmatrix} = 0 \tag{2.24}$$

This approach identifies the parameters of the production function under the assumption that variable input use responds to current productivity shocks, but that lagged variable input demand does not. Moreover, the persistence in the productivity process guarantees that variable input demand is correlated with its lagged values.

What are the limitations of this approach? Production function estimation faces many challenges such as measurement issues (measurement errors, unobserved prices), specification problems when choosing the functional form, simultaneous determination of output and input choices in response to unobserved shocks, and selection bias as firms exit and enter the sample. The control function approach is particularly useful to prevent bias due to endogeneity of input choice. However, especially the choice of functional form is central in determining the results. The use of a Cobb-Douglas function facilitates the estimation as it is linear in logarithms and a parsimonious specification widely used in the literature and hence easily comparable with other results, i.e., Eeckhout and De Loecker

(2017) also use this functional form and find results to be similar using a translog production function. On the downside, this choice implicitly assumes an elasticity of substitution between labor and capital equal to one. Thus it is unable to capture movements stemming from factor-biased technological change.

Estimation Results

The estimation uses Compustat data from 1980 to 2016 using information from 12,802 firms from 51 industries (3-digit NAICS). Output elasticities are estimated separately for each industry. The median values for the estimated output elasticities are .844 and .155 for variable inputs and capital respectively.

As described in the theory section, combining the cost share of variables inputs with the estimated elasticities gives a time-varying, firm-specific markup. Note that the baseline estimation uses fixed elasticities across time and sub-sectors, but differences in the cost share translate into variation in markups. However, note that the estimation takes into account measurement error, by removing variation in output not related to variables impacting input demand. We observe $\tilde{Q}_{i,t} = Q_{i,t} exp(\epsilon_{i,t})$, but the first stage provides us with an estimate for $\epsilon_{i,t}$. This adjustment follows Loecker and Warzynski (2012). Thus the markup is computed from:

$$\mu_{i,t} = \alpha \frac{COGS_{i,t}}{exp(\hat{\phi}_{i,t})} \tag{2.25}$$

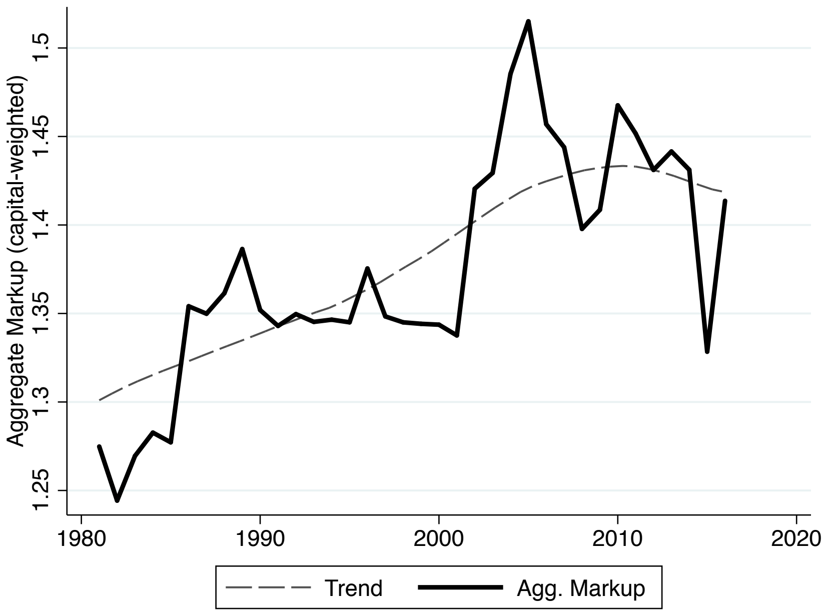
As in Eeckhout and De Loecker (2017), the aggregate markup (computed as a capital-weighted average) has increased since the 1980s. Fig. 2.3 shows that the increase is more pronounced for the second half of the sample.

Next, for each firm I compute the implied marginal product of capital as $MPK_{i,t} = \frac{\theta_K}{\mu_{i,t}} \frac{exp(\hat{\phi}_{i,t})}{K_{i,t}}$. Pure profits consist of the remaining gain after variable inputs and capital has been paid for: $\Pi_{i,t} = exp(\hat{\phi}_{i,t}) - COGS_{i,t} - MPK_{i,t} - SGA_{i,t}$. Here Selling, General and Administrative Expenses (SGA) approximate the fixed cost of production which are excluded from profits. The presence of fixed cost can explain positive markups despite close to zero profits. If fixed costs such as advertisement or product development increase, firms need to charge prices over marginal cost to recover them. Given the estimate of the marginal product, interest rates and the depreciation rate, the friction ξ can be computed as a residual. Following eq. 2.15 the deviation between the return (adjusted for measurement error) and interest rates consists of the two wedges:

$$Return_{i,t}^{\text{ME adjusted}} - i = \xi_{i,t} + \frac{\Pi_{i,t}}{K_{i,t}} \tag{2.26}$$

To understand how much each component explains of the aggregate trend, the firm-level wedges are aggregated by sub-sector and for the whole economy. Again, to ensure that the measure represents the U.S. economy accounting for biased representation of

Figure 2.3: Aggregate markup in the U.S.



industries in the sample, the industry results are weighted by BEA capital shares. Moreover, to understand better what explains the capital friction, the contribution of risk, an important components of capital frictions, is estimated separately (see next subsection).

Figure 2.4 separates the aggregate return into the capital friction and the profit component. The figure illustrates that the capital friction makes up about $\frac{3}{4}$ of the measured return. A simple variance decomposition, summarized in Table 2.1, shows that the capital friction explains around 70 percent of the movements in the measured aggregate return. On average a quarter of the measured return is due to profits, which explain around 30 percent of the variation.

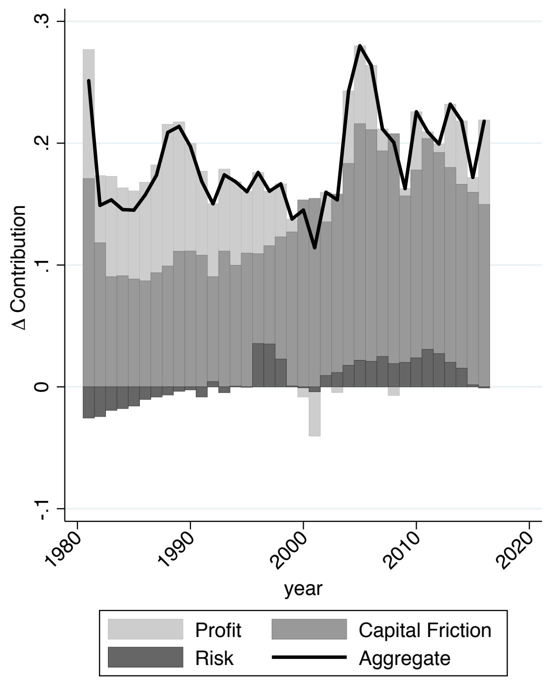
Risk and the capital friction

Although there are many potential distortions that could lead to capital frictions, risk premia are considered to be one of the most important reasons why firms cannot borrow unlimited amounts of money. How much of the capital friction computed here is due to risk? To get a sense of the importance of risk, we can follow the standard Capital-Asset-Pricing-Model (CAPM) which says that a firm i's risk-premium depends on the aggregate risk-aversion coefficient λ and a firm-specific covariance with the aggregate economy σ_i such that:

$$\text{risk premium}_{i,t} = \lambda_t \sigma_i$$

As a simple benchmark, the risk-premium is set to be constant over time with a

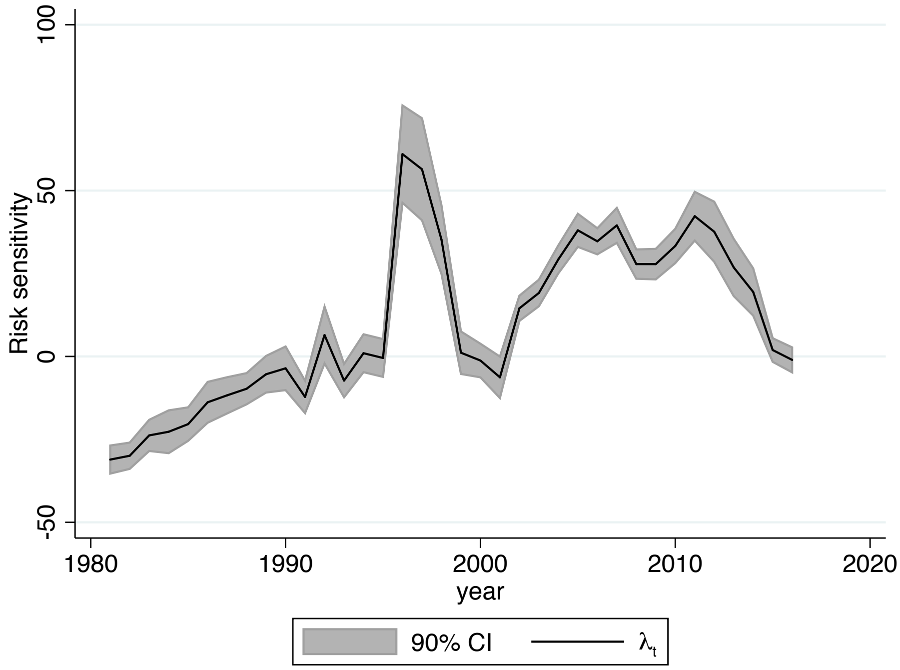
Figure 2.4: Decomposition of the Aggregate Return: Marginal Product versus Profits



Notes: Contribution to the difference between the aggregate return and interest rates are computed from capital-weighted sums across the firm-level decomposition described in eq. 2.26.

risk-aversion coefficient of 12 (see for example Farhi and Gourio (2018)). The risk premium only makes up a small portion of the overall result due to low covariances of firms with the aggregate. However, this benchmark cannot capture any time-variation in risk-aversion. Estimating the sensitivity to risk (higher correlation with the aggregate) for each year, it is possible to capture a higher explanatory power of risk over time. Figure 2.5 shows the coefficient and its 90 percent confidence interval of the regression $MPK_{i,t} = a_t + \lambda_t \sigma_{i,t} + u_{i,t}$. This result suggest an overall increase in the sensitivity to riskiness of the return of a firm. Table 2.1 shows that a time-varying risk aversion increases the explanatory power of risk from almost zero to 15 percent. Nevertheless the explanatory power of risk remains low. This may be because the CAPM model is overly simplistic. This risk-premium measures non-diversifiable risk within a U.S. portfolio, but relies on strict assumptions. For example, the utility function only depends on the first and second moment, a condition to make variance a sufficient measure of riskiness, or that there are no transaction costs or borrowing constraints which prevent investors form holding a perfectly diversified portfolio. Moreover, the market return is defined as the aggregate return measured here, which is an imperfect measure as other types of investment

Figure 2.5: Time-varying Risk sensitivity



Notes: The coefficient is obtained from the following regression model
$MPK_{i,t} = a_t + \lambda_t \sigma_{i,t} + u_{i,t}$ and plotted with 90 % confidence interval

Table 2.1: Decomposition of the Trend in the Aggregate Return

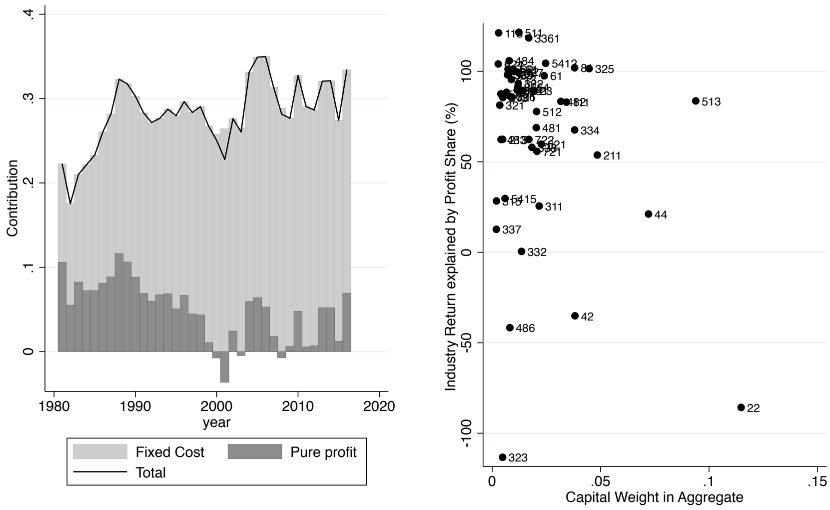
Component		**Contribution (%)**
Profit		29.1
Capital Friction		70.9
of which risk	$\lambda = 12$	-0.04
	λ_t	15.11

are not included. Finally, riskiness of each firm is measured by the covariance in historic data which provides relatively few data points for each firm and is constant over-time. However, risk is likely to vary over time and may depend on factors not capture by the variance. Nevertheless, the CAPM approach is useful as a first pass of the role of risk without imposing more structure on the sources of risk and risk preferences.

Figure 2.6: Why is the explanatory power of profits low in the aggregate?

(a) Decomposition of profits

(b) Explanatory power of profits by industry



Markups and the explanatory power of pure profits

Why does the aggregate markup increase so dramatically, but the profit share explains a relatively small part of the difference between returns and the interest rate? Here there are two main explanations, (i) the role of fixed costs and (ii) the sectoral composition. Note that the unaccounted part of profits in this framework is given by:

$$\Pi^{pure} = exp(\hat{\phi}_{i,t}) \left(1 - \frac{\theta_l}{\mu i, t} - \frac{\theta_k}{\mu i, t}\right) - f_{i,t}$$

In the presence of fixed costs, part of the markup charged over marginal cost goes to cover the fixed costs of production, here measured as Selling and General Administrative Expenses. Fig 2.6a illustrates that although total profits have increased over the period, this results from an increase in fixed costs. The remaining component of pure profit has diminished over time. These findings indicate an underlying shift in the market structure towards a world with higher fixed costs. Although this result naturally depends on the accounting conventions that lead to the classification of variable costs versus fixed costs, this finding is line with the emergence of increasing returns to scale generated by information technology where many products require high investment in research and development, but can be produced at relatively low marginal costs per additional customer.

Moreover, the extend to which the higher return is driven by profits varies across industries. Figure 2.6b shows the share of variation in the industry-level return that is explained by variation in the profit components. Although movements in the profit share explain most of the divergence between returns and interest rates in many industries, the explanatory power of profit is particularly low in some capital-rich industries such as utilities, retail, and telecommunications which reduces the role of profit in the aggregate trend.

Therefore, this paper concludes that although the aggregate markup in the U.S. economy has increased over the last 30 years, this trend can only account for about a third of the gradual divergence between returns to capital and interest rates. Instead, the evidence points towards capital frictions, i.e. explanations that increase the user cost of capital for firms or changes in their distribution across firms as the main driver. These frictions, partly linked to risk premia, have prevented the fall in real interest rates from translating into higher investment and a matching fall in the marginal product of capital.

2.5.3 Composition effects

Firm-level data shows to what extend the aggregate return to capital moves in response to *within firm* or *between firm* changes, for example due to capital reallocation. The aggregate return to capital in sub-sector s is a weighted sum of individual firm returns:

$$MPK_{s,t} \equiv \sum_{i \in N} w_{i,s,t} MPK_{i,s,t} \tag{2.27}$$

where $w_{i,s,t} = \frac{K_{i,s,t}}{\sum_{i \in N} K_{i,s,t}}$ is determined by the share in total capital owned by firm i. There exist two standard decompositions widely used in the literature on firm dynamics to understand the role of across-firm reallocation[5]. These decompositions are usually applied to productivity, but they are equally informative for the aggregate return to capital. Firstly, the decomposition proposed by Olley and Pakes (1996) - henceforth OP - splits the aggregate return, a capital-weighted average of firm-level returns, into an unweighted firm-level average $M\bar{P}K_{s,t}$ and a covariance term (eq. 2.28). This covariance term summarizes the within-industry cross sectional covariance between size (measured by capital stock) and the return to capital.

$$MPK_{s,t} = M\bar{P}K_{s,t} + \sum_{i \in N} w_{i,s,t}(MPK_{i,s,t} - M\bar{P}K_{s,t}) \tag{2.28}$$

The decomposition of the aggregate return consists of a sum, weighted by industry capital shares $\omega_{s,t}$, as stated in eq. 2.29.

[5]These are account decomposition that give a sense of the importance of capital reallocation without making further assumptions on what caused the reallocation. However, capital reallocation is endogenous and depends on the underlying distribution of productivities, frictions and markups. Without further structure on the general equilibrium of the model, the underlying changes in the distribution cannot be identified.

$$MPK_t = \sum_s \omega_{s,t} MPK_{s,t} = \underbrace{\sum_s \omega_{s,t} M\bar{P}K_{s,t}}_{\text{Mean}} + \underbrace{\sum_s \omega_{s,t} \sum_{i \in N} w_{i,s,t}(MPK_{i,s,t} - M\bar{P}K_{s,t})}_{\text{Covariance}} \tag{2.29}$$

The OP decomposition shows that a positive covariance between size and returns increases the aggregate return by over 5 percentage points. However, the OP method decomposes the aggregate return period-by-period and thus is limited with regards to across-firm reallocation dynamics and firm churning. To decompose the change in the aggregate return into within and between firm components, I extend the standard decomposition for example used in Haltiwanger (1997) to incorporate sectoral reallocation as a potential explanation.

Eq. 2.30 first states the aggregate change as a sum of the within sector, the between sector and the sectoral covariance components. For example, if sectoral reallocation towards high return sectors were to explain the rise in aggregate returns relative to interest rates, we would expect the between sector component to be large and positive.

$$\Delta MPK_t = \sum_s \underbrace{\omega_{s,t-1} \Delta MPK_{s,t}}_{\text{within sector}} + \sum_s \underbrace{\Delta\omega_{s,t} MPK_{s,t-1}}_{\text{Between sector}} + \sum_s \underbrace{\Delta\omega_{s,t} \Delta MPK_{s,t}}_{\text{Sector covariance}} \tag{2.30}$$

Secondly, eq. 2.31 decomposes the change in sector-level returns into within firm and between firm components. It differentiates between the sets of continuing firms C, entering firms E, and exiting firms X. Substituting the change in sector-level returns in eq. 2.30 gives the final decomposition whose results are presented in Figure 2.7b.

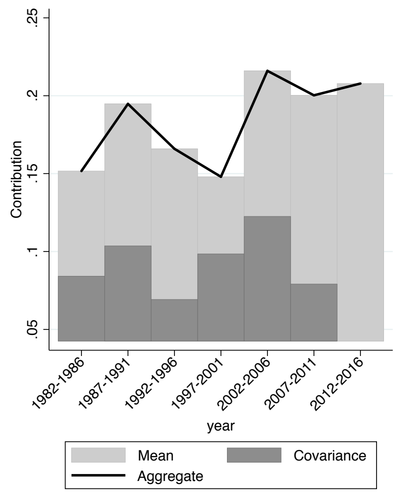
$$\Delta MPK_{s,t} = \sum_{i \in C} \underbrace{\Delta w_{i,t} \Delta MPK_{i,t}}_{\text{Firm Covariance}} + \sum_{i \in C} \underbrace{w_{i,t-1} \Delta MPK_{i,t}}_{\text{within firm}} + \sum_{i \in C} \underbrace{\Delta w_{i,t}(MPK_{i,t-1} - M\bar{P}K)}_{\text{Between firm}} \tag{2.31}$$

$$+ \sum_{i \in E} \underbrace{w_{i,t}(MPK_{i,t} - M\bar{P}K)}_{\text{Entry}} - \sum_{i \in X} \underbrace{w_{i,t-1}(MPK_{i,t-1} - M\bar{P}K)}_{\text{Exit}} \tag{2.32}$$

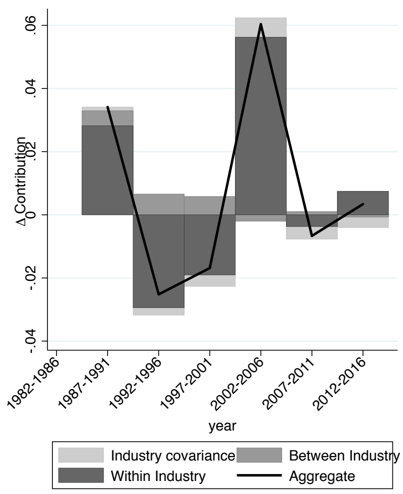
Figure 2.7 presents the results of both decomposition exercises for 5-year averaged data. The OP-decomposition (Fig. 2.7a) reveals a positive correlation between size and the return to capital on the firm-level which explains about a third of the difference between the aggregate return and interest rates. On the right, figure 2.7b illustrates the decomposition of the change in the aggregate return. The increase in the aggregate return is not explained by sectoral reallocation to high-return sectors - both the sectoral composition and the sectoral covariance component are small. Further decomposing the within-industry effect in figure 2.7c shows that it is mostly driven by within-firm movements. The reallocation of capital towards high return firms explains most of the change

Figure 2.7: Decomposition: Aggregate Return

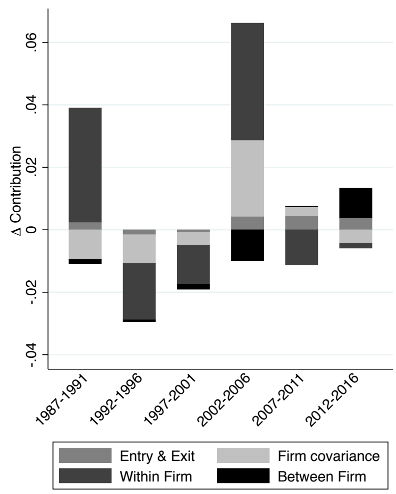
(a) OP decomposition - 5-year averages



(b) Dynamic decomposition - 5-year averages



(c) Disaggregating the within-industry component



only in the last decade. Also, I find a negative firm covariance component which confirms that firm returns fall in response to an increase in capital. Entry into and exit out of the sample explains only a small part of the change. In general, it seems that exiting firms tend to have low returns and thus contribute positively to the change in aggregate MPK.

Entering firms on the other hand often also have below average returns, thus offsetting most of the positive contribution of exit. The Compustat database of large, publicly-listed firms is not well-suited to study exit and entry dynamics, but studying the role of firm churning with different data could help to discover additional channels that contribute to this trend.

The decomposition reveals within-firm changes as the source of most of the dynamics. The next subsection takes a closer look at the main components and the importance of within versus between firm movements. Have large firms become more profitable? Has capital shifted to more risky firms? Does this evidence suggests that small firms face higher capital frictions?

Marginal product of capital

Isolating the marginal product of capital from the decomposition and repeating the decomposition exercise shows (i) a negative correlation between the capital weight and the marginal product and (ii) reallocation of capital towards high return firms throughout the sample. Figure 2.8a shows that smaller firms (less capital) tend to have higher returns, a finding that is in line with the common observation that small firms are more (financially) constraint. This correlation has become more negative over time, but the overall increase in the average marginal product has offset this. The negative covariance can either indicate a larger dispersion in the capital friction across firms or tighter capital constraints for small firms. Again almost all of the change in the MPK component comes from within-industry dynamics. However, in the late 1990s and early 2000s, the between-industry component also contributed positively to the increase in the aggregate MPK. Although within-firm changes explain around two-thirds of the within-industry component, between-firm movements, although largely offset by the negative covariance, are also important (Fig. 2.8c)

Profit

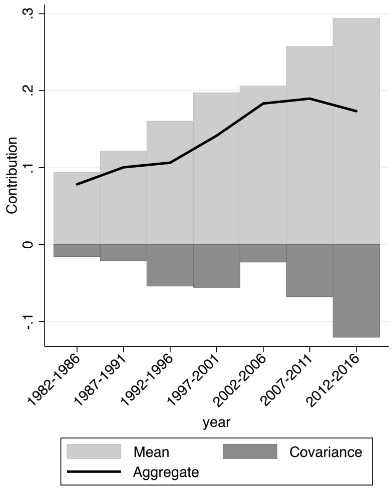
Fig 2.9a highlights that large firms (with a large capital stock) make higher pure profits. This positive covariance explains most of the profit component in the measured aggregate return (especially in the second half of the sample). Figure 2.9b shows that within-industry changes explain most of the movements in the profit component. These are mainly driven by within-firm movements. Interestingly, reallocation of capital is associated with a higher profit component as indicated by a positive covariance effect. This suggests that as a firm's market share grows, its pure profits increase more than proportionally.

Capital Friction

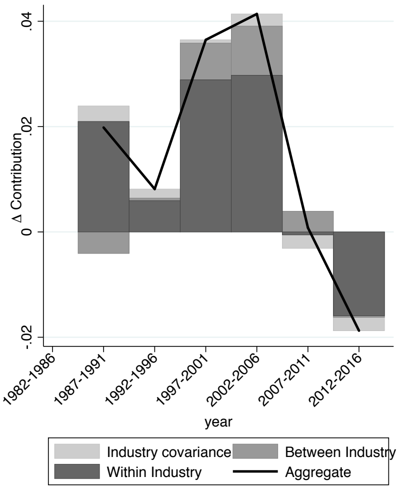
The movements in the marginal product can be further divided into changes in the risk component and the remaining capital friction. The remaining capital friction is the main driver of the marginal product and the OP decomposition shows that smaller firms tends to face larger friction - a trend that increased over time (see figure 2.10b). The negative

Figure 2.8: Decomposition: Marginal product of capital

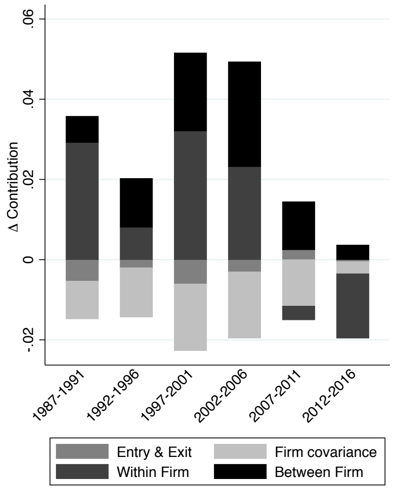
(a) OP



(b) Dynamic



(c) Disaggregating the within-industry component

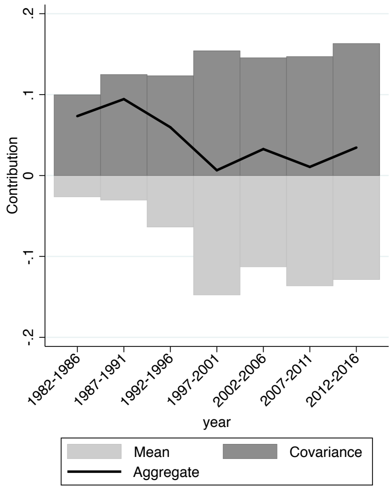


Notes: (a) Olley & Pakes decomposition of the marginal product into its mean and the covariance with the capital weights (b) Dynamic decomposition of the change in the marginal product over 5-year averages (c) Further disaggregation of the within-industry component.

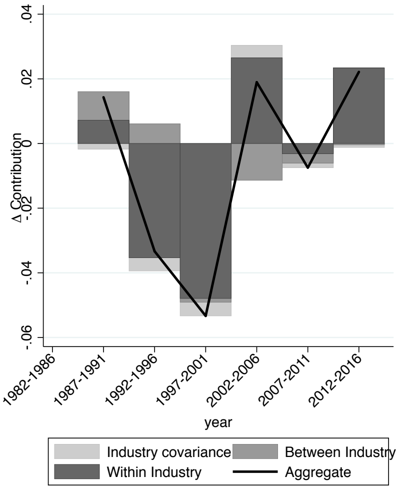
covariance component can be due to either greater dispersion in capital frictions or shifts

Figure 2.9: Decomposition: Pure profit component

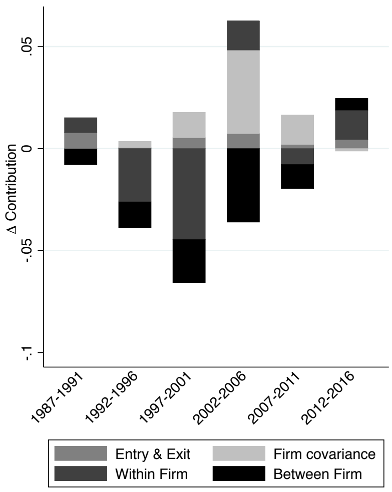
(a) OP decomposition



(b) Dynamic decomposition



(c) Disaggregating the within-industry component



Notes: (a) Olley & Pakes decomposition of the profit component into its mean and the covariance with the capital weights (b) Dynamic decomposition of the change in the profit component over 5-year averages (c) Further disaggregation of the within-industry component.

in the correlation between frictions and underlying factors that determine firm size (for example productivity). Changes in the capital friction component are again explained by within-industry movements, in this case both between- and within-firm components explain an almost equal share, but as reductions in frictions naturally lead to more investment, the negative covariance offsets most of the between-firm component (see figure 2.10c).

Risk

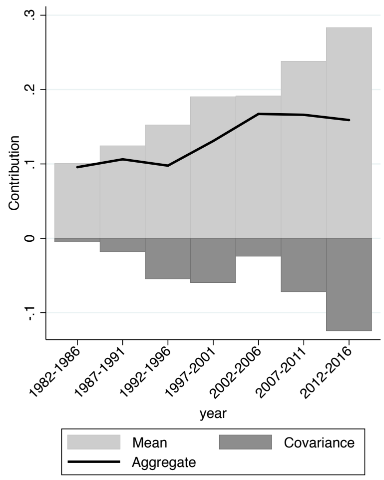
Figure 2.11a shows that larger firms tend to be more correlated with the aggregate return and suggests some reallocation towards riskier firms at the end of the sample. Even though the risk-aversion parameters is chosen to give the highest possible explanatory power to risk, the overall contribution is small. The dynamic decomposition shows that within-industry movements drive almost all the change in this component. The data does not suggest any reallocation towards riskier sectors. However, most of the risk dynamics here is driven by the time-varying parameter λ which increases the sensitivity to risk over time. This also leads within-firm dynamics to account for most of the within-industry effect.

Conclusion

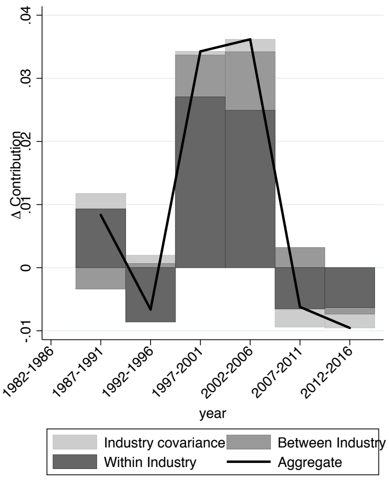
This second chapter first documents the divergence between the aggregate return and interest rate in aggregate data for the U.S. since the 1980s and shows that it stands out compared to other advanced economies. It then exploits firm-level evidence on returns to capital to understand why the aggregate return has diverged from interest rate since the 1980s. Utilizing methodological advances in production function estimation, it disentangles the contribution of rising markups versus marginal products. The findings suggest that around 70 percent of the divergence can be attributed to the level and distribution of capital frictions. Although the results confirm that markups have increased over the period, the contribution of pure profit to the overall return is limited by increasing fixed costs and steady accumulation of capital. Moreover, this paper finds that shifts in the distribution of produced capital across firms have amplified the aggregate trend and suggest significant structural changes behind the absence of a strong comovement between interest rates and the aggregate return to capital in the U.S. The results of this paper raise further questions. Despite the growth of the financial sector and new financial instruments, capital frictions appear to have risen over time. What is the nature of these frictions? Which mechanism are behind the increase in markups? Future research is also needed to better understand the policy implications of the decoupling of marginal products from interest rates. Finally, there are other potential channels such as factor-biased technological change or firm exit and entry dynamics which could be worth exploring in future research.

Figure 2.10: Decomposition: Friction component

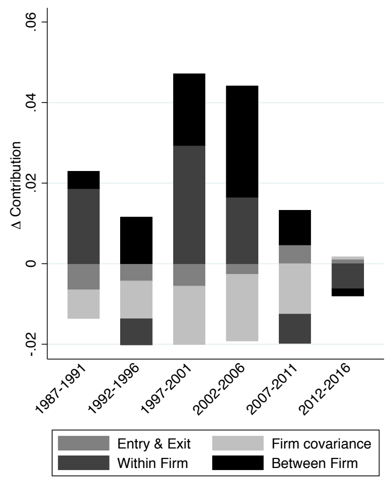
(a) OP decomposition



(b) Dynamic decomposition



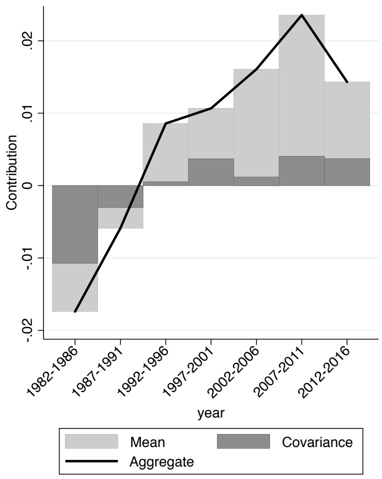
(c) Disaggregating the within-industry component



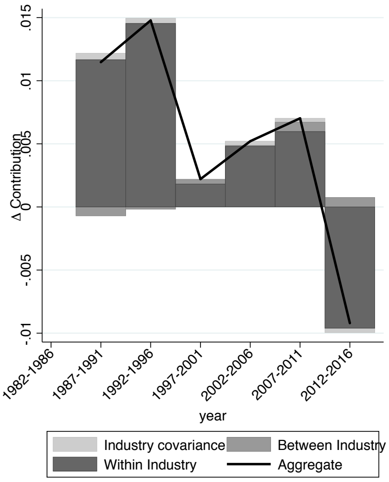
Notes: (a) Olley & Pakes decomposition of the residual capital friction into its mean and the covariance with the capital weights (b) Dynamic decomposition of the change in the residual capital friction over 5-year averages (c) Further disaggregation of the within-industry component.

Figure 2.11: Decomposition: Risk component (time-varying λ)

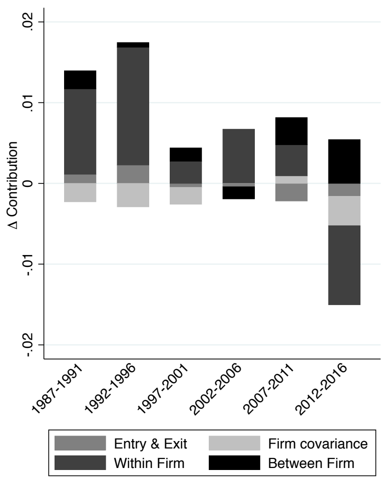
(a) OP decomposition



(b) Dynamic decomposition



(c) Disaggregating the within-industry component



Notes: (a) Olley & Pakes decomposition of the risk component into its mean and the covariance with the capital weights (b) Dynamic decomposition of the change in the risk component over 5-year averages (c) Further disaggregation of the within-industry component.

2.6 Appendix

2.6.1 Aggregate evidence from other Advanced Economies

Is this divergence a common phenomenon across advanced economies? Fig. 2.12a - 2.12f summarize the aggregate return to capital for 6 advanced economies - the U.S., the four largest Eurozone economies and Great Britain. The return is calculated from national accounts data. The dashed line shows the path of the long-run real interest rate (deflated by the GDP deflator). Apart from the spike during the EU Sovereign debt crisis, the long-run interest rate exhibits a downward trend in all six economies starting in the late 1990s. The movements in the return to capital are less homogeneous. Although the aggregate return does not fully match the decline in real interest rates in any economy, there are important differences across countries. The return to capital in the U.S., the UK and Germany is stable over time and although the omission of the real estate sectors affects the level of the average return, this does not alter the overall trend. For France, Spain and Italy on the other hand, the return has declined dramatically since the early 2000s in the non-real estate part of the economy. This decline is dampened substantially when the real estate sector is included in the measure.

The divergence between returns and interest rates is most pronounced for the U.S., but the German experience appears similar since the 2000s. In the UK, the decline in the aggregate return to capital is in line with real interest rates - both have fallen by around 4 percentage points between 2000 and the end of the sample in 2015. Analyzing the correlation between long-run real interest rates and the return to capital, the U.S. and Germany show no significant correlation for both the whole sample and the subperiod since 1998. The UK and France on the other hand show a positive correlation at least since 1998. Interestingly, for the UK it is the strong link between returns in the real estate sector and interest rates that drives this correlation. For France, it is a strong link between the market sector and interest rates that matters. Spain even shows a negative correlation between the aggregate return and interest rates since 2000. Here the market sector is negatively linked to interest rates, whereas there is a positive correlation for the real estate sector. Italy's relationship is also negative, but here only the real estate sector shows a significant coefficient. However, both Spain and Italy were heavily affected by the sovereign debt crisis where low output coincided with high long-term interest rates.

What is the role of the real estate sector? First of all, the real estate sector makes up a very large share of capital in all of the economies. In France it covers over 60%, in the UK and in the US it is not as high, but still quite important. In general, the return in the real estate sector is relatively low, so to offset a fall in the return in the rest of the economy we need that either the share of RE falls over time or the return in the real estate sector increases over time.

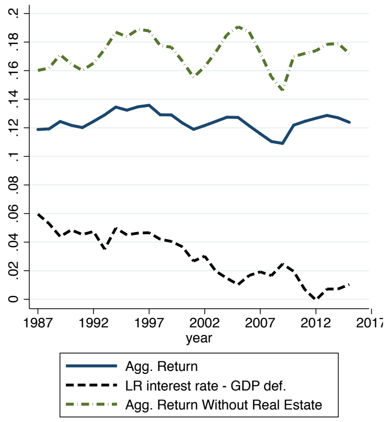
Computing the aggregate return excluding the real estate sector shows a downwards trend in returns for France, Spain and Italy. This decline partially matched the fall in interest rates, suggesting that the European experience is driven by other factors than the U.S. Moreover, the heterogeneity across advanced economies motivates a closer look at structural changes within the economy as opposed to global macro trends.

2.6.2 A Partial equilibrium model

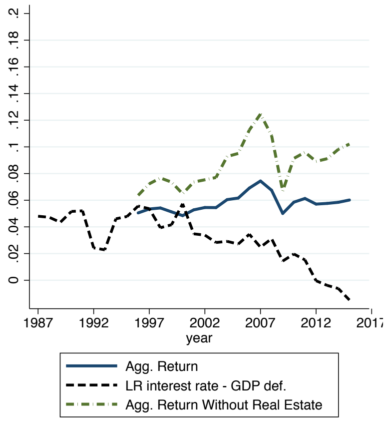
As simple partial equilibrium model illustrates how the underlying distribution of firm characteristics generates movements in the aggregate return. Each firm hires labor and

Figure 2.12: Aggregate return to capital in other advanced economies

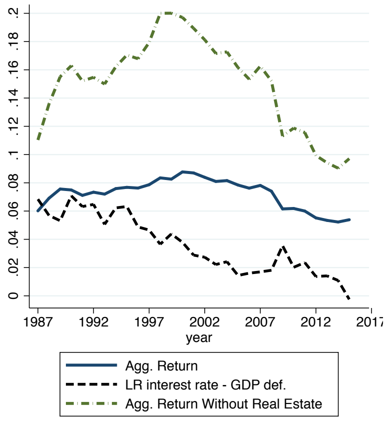
(a) USA



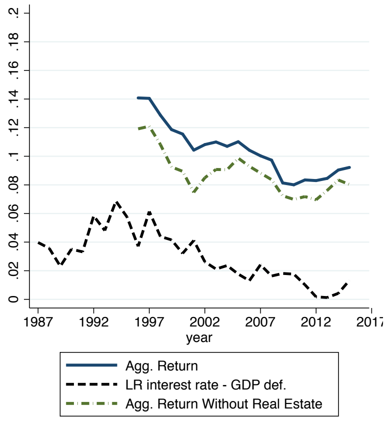
(b) Germany



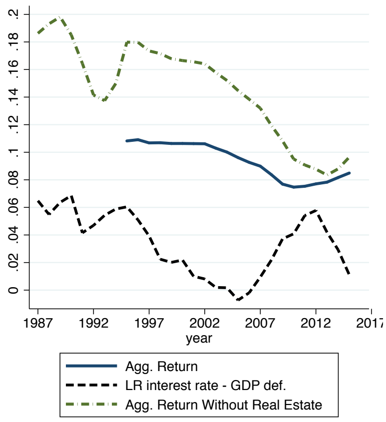
(c) France



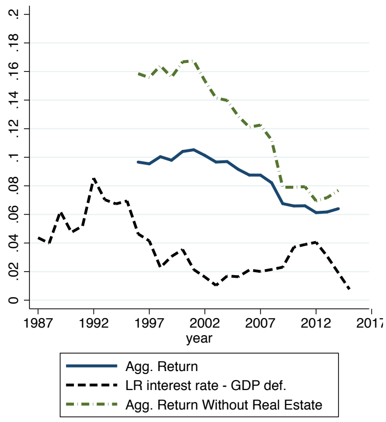
(d) UK



(e) Spain



(f) Italy



Sources: U.S. Bureau of Economic Analysis (BEA) and annual macro-economic database of the European Commission's Directorate General for Economic and Financial Affairs (AMECO) capital to produce output according to a Cobb-Douglas production technology. Firms compete in the labor market such that the wage rate w_t is common across firms. However, firms can face different rental rates on capital because of heterogeneity in risk-premia. A firm's risk-premium depends on the aggregate risk-aversion coefficient λ and a firm-specific covariance with the aggregate economy σ_i. This simple functional form is taken from the Capital-Asset-Pricing-Model (CAPM). Thus, the interest rates faced by firm i is given by the sum of the aggregate interest rate i and the firm-specific risk-premium. Moreover, each firm can potentially be constrained in its ability to hire capital. This constraint is a simple way to capture frictions that affect the allocation of capital across firms.

Given the production technology, firms maximize profits subject to a downwards sloping demand given by $Q_{i,t} = P_{i,t}^{-\epsilon}$. Eq. 2.33 summarizes the firm maximization problem:

$$\max_{L,K} P_{i,t}Q_{i,t} - w_t L_{i,t} - (i_t + \lambda\sigma_i)K_{i,t} \tag{2.33}$$

$$\text{s.t.} \quad Q_{i,t} = A_{i,t}K_{i,t}^{\alpha}L_{i,t}^{1-\alpha} \tag{2.34}$$

$$K_{i,t} \leq \kappa_i \tag{2.35}$$

$$Q_{i,t} = P_{i,t}^{-\epsilon} \tag{2.36}$$

The first order conditions imply that firms choose capital and labor optimally if:

$$w_t = \frac{\epsilon - 1}{\epsilon}(1-\alpha)\frac{P_{i,t}Q_{i,t}}{L_{i,t}} \tag{2.37}$$

$$(i_t + \lambda\sigma_i) + \xi_{i,t} = \frac{\epsilon - 1}{\epsilon}\alpha\frac{P_{i,t}Q_{i,t}}{K_{i,t}} \tag{2.38}$$

$$0 = \xi_i(K_{i,t} - \kappa_i) \tag{2.39}$$

$$(\kappa_i - K_{i,t}) \geq 0 \tag{2.40}$$

A firm is unconstrained if the optimal amount of capital lies below κ_i such that $\xi_i = 0$.

Each firm charges a mark-up $\frac{\epsilon}{\epsilon-1}$ above its marginal cost. Eq. 2.41 shows how the price depends on firm characteristics.

$$P_i = \frac{1}{\tilde{\alpha}}\frac{\epsilon}{\epsilon - 1}\left(\frac{i + rp_i + \xi_i}{A_i}\right)^{\alpha}\left(\frac{w_t}{A_i}\right)^{1-\alpha} \tag{2.41}$$

where $\tilde{\alpha} = \alpha^{\alpha}(1-\alpha)^{1-\alpha}$.

The relative wage determines the capital-labor ratio used in production shown in eq. 2.42.

$$\frac{K_{i,t}}{L_{i,t}} = \frac{w}{i + \lambda\sigma_i + \xi_{i,t}}\frac{\alpha}{1-\alpha} \tag{2.42}$$

Using eq. 2.42 in eq. 2.34 gives the labor demand for each firm:

$$
\begin{aligned}
L_{i,t} &= (1-\alpha)\frac{P_{i,t}^{-\epsilon}}{A_{i,t}}\frac{1}{\tilde{\alpha}}\left(\frac{w_t}{i_t+\lambda\sigma_i+\xi_{i,t}}\right)^{-\alpha} && (2.43)\\
&= (1-\alpha)A_{i,t}^{\epsilon-1}\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}\tilde{\alpha}^{\epsilon-1}\left(w_t\right)^{-\epsilon}\left(\frac{w_t}{i_t+\lambda\sigma_i+\xi_{i,t}}\right)^{\alpha(\epsilon-1)} && (2.44)
\end{aligned}
$$

When is a firm unconstrained? $\xi_{i,t} = 0$ holds when the optimal amount of capital inputs in production remains below the maximum value κ_i.

$$
\xi_i = 0 \quad \text{iff} \quad K_{i,t}^{U} = \alpha A_{i,t}^{\epsilon-1}\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}\tilde{\alpha}^{\epsilon-1}\frac{w_t^{(1-\epsilon)(1-\alpha)}}{\left(i_t+\lambda\sigma_i\right)^{1+\alpha(\epsilon-1)}} < \kappa_i \tag{2.45}
$$

$$
\text{Otherwise:} \quad \xi_{i,t} = \left(\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}\frac{\alpha}{\kappa_i}\tilde{\alpha}^{\epsilon-1}A_{i,t}^{\epsilon-1}w_t^{(1-\alpha)(1-\epsilon)}\right)^{\frac{1}{1+\alpha(\epsilon-1)}} - \left(i_t+\lambda\sigma_i\right) \tag{2.46}
$$

Eq. 2.45 demonstrates that firms with a tight constraint κ_i, a high productivity $A_{i,t}$ and a low risk-premium $\lambda\sigma_i$ are more likely to be constraint. Eq. 2.46 shows that the value of relaxing the constrained $\xi_{i,t}$ increases if the constraint is tight (low κ), productivity $A_{i,t}$ is high, the wage is high or the risk-premium is low.

Market clearing

The partial equilibrium of this economy consists of a sequence of quantities and prices such that firms optimize and the labor market clears. Here the market for labor is perfectly competitive. This captures the fact that labor supply is local and wages are determined within countries but can vary substantially between countries. On the other hand, the interest rates is taken as exogenous. This captures an economy where the aggregate supply of capital is flexible for example because of open international capital markets. Given a fixed supply of labor $\bar{L}$, the aggregate wage w_t clears this factor market if:

$$
\bar{L} = \sum_{i\in N} L_i = \sum_{i\in \text{Unconst.}} L_i^U + \sum_{i\in \text{Const.}} L_i^C \tag{2.47}
$$

where the number and composition of firms in the set of unconstrained and constrained firms depends on the joint distribution of firm characteristics. The labor demand is given in eq. 2.43 where ξ_i is determined by 2.45 and 2.46.

Firm composition and a fall in interest rates

The elasticity of the capital-labor ratio depends on the firm-specific financial friction. Here we can see that firms with a lower friction will expand their capital-labor ratio relatively more in response to a fall in interest rates:

$$\log(\frac{K_i}{L_i}) = \log(\frac{\alpha}{1-\alpha}) + \log(w) - \log(i + rp_i) \quad = \log(\frac{\alpha}{1-\alpha}) + \log(w) - \log(\exp^{log(i)} + rp_i)$$
$$\frac{d\log(\frac{K_i}{L_i})}{dlog(i)} = -\frac{i}{i + rp_i}$$

A lower risk-premium leads to a higher elasticity of the capital-labor ratio to the interest rate. Similarly, we can also look directly of the response of capital:

$$log(K_i) = log(\alpha A_{i,t}^{\epsilon-1}\left(\frac{\epsilon-1}{\epsilon}\right)^{\epsilon}\tilde{\alpha}^{\epsilon-1}) + (1-\epsilon)(1-\alpha)log(w_t) - (1+\alpha(\epsilon-1))log\left(i_t + rp_i\right)$$
$$\frac{dlog(K_i)}{dlog(i)} = -(1+\alpha(\epsilon-1))\frac{i_t}{i_t + rp_i}$$

Here we also see that the expansion in capital is larger for firms with a low rp_i. Moreover, we see that the response in capital is larger than the response of the capital-labor ratio because the demand for labor in unconstrained firms also rises relatively more. This framework quite naturally produces the result of Gopinath et. al. that if unconstrained firms are large and constrained firms are small, then a fall in interest rate increases the dispersion in log returns. On the other hand, a fall in interest rate leads to a general increase in the output-capital ratio, the profit component becomes less important, which has a big effect on the measured return. If epsilon falls at the same time, the profit component falls less (1/eps increases), and the higher mark-up works against the fall in value-added per capital. So the markup has to increase quite substantially to generate and increase in return-i. On the other hand, the financial friction would have to increase just enough to offset the fall in the interest rate and increase the measured return. The effect is amplified by the positive effect on value-added per capital, but it is dampened by the composition effect, i.e. higher friction for some firms decreases their weight in the aggregation.

Bibliography

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

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6):2411–2451.

Aizenman, J., Pinto, B., and Radziwill, A. (2007). Sources for financing domestic capital - Is foreign saving a viable option for developing countries? *Journal of International Money and Finance*, 26(5):682–702.

Alfaro, L., Kalemli-Ozcan, S., and Volosovych, V. (2014). Sovereigns, upstream capital flows, and global imbalances. *Journal of the European Economic Association*, 12(5):1240–1284.

Antràs, P. (2004). Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution. *Contributions in Macroeconomics*, 4(1).

Barkai, S. (2016). Declining Labor and Capital Shares. *Stigler Center New Working Paper Series*, (2).

Barro, R. J., Fernández-Villaverde, J., Levintal, O., and Mollerus, A. (2017). Safe Assets. *NBER Working Paper Series*.

Bartelsman, E., Haltiwanger, J., and Scarpetta, S. (2013). Cross-country differences in productivity: The role of allocation and selection. *American Economic Review*, 103(1):305–334.

Caballero, R. J., Farhi, E., and Gourinchas, P. O. (2016). Safe Asset Scarcity and Aggregate Demand. *American Economic Review: Papers & Proceedings*, 106(5):513–518.

Caballero, R. J., Farhi, E., and Gourinchas, P. O. (2017). Rents, technical change, and risk premia accounting for secular trends in interest rates, returns on capital, earning yields, and factor shares. *American Economic Review*, 107(5):614–620.

Caggese, A. and Perez-Orive, A. (2017). Capital Misallocation and Secular Stagnation. *Finance and Economics Discussion Series*, 2017(009).

Caselli, F. and Feyrer, J. (2007). The Marginal Product of Capital. *The Quarterly Journal of Economics*, 122(2):535–568.

Chatterjee, S. and Naknoi, K. (2010). The marginal product of capital, capital flows, and convergence. *American Economic Review*, 100(February):73–77.

Chirinko, R. S. and Mallick, D. (2008). The Marginal Product of Capital : A Persistent International Puzzle. *CESifo WORKING PAPER NO. 2399*.

Cuñat, A. and Maffezzoli, M. (2007). Can Comparative Advantage explain the Growth of US Trade? *The Economic Journal*, (117):583–602.

David, J. M., Schmid, L., and Zeke, D. (2018). Risk-Adjusted Capital Allocation and Misallocation. *Available at SSRN: https://ssrn.com/abstract=3255111 or http://dx.doi.org/10.2139/ssrn.3255111*.

David, J. M., Simonovska, I., and Henriksen, E. (2016). The Risky Capital of Emerging Markets. *Society for Economic Dynamics 2016 Meeting Papers*.

Döttling, R., Gutierrez Gallardo, G., and Philippon, T. (2017). Is There an Investment Gap in Advanced Economies? If So, Why? *ECB Conference Paper*, pages 1–44.

Eeckhout, J. and De Loecker, J. (2017). The Rise of Market Power and the Macroeconomic Implications. *NBER Working Paper*.

Elsby, M. W. L., Hobijn, B., and Sahin, A. (2013). The Decline of the U.S. Labor Share. *Brookings Papers on Economic Activity*, 2013(2):1–63.

Fadinger, H. (2011). Productivity differences in an interdependent world. *Journal of International Economics*, 84(2):221–232.

Farhi, E. and Gourio, F. (2018). Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia. *Brookings Papers on Economic Activity (Conference Draft)*, page 61.

Feenstra, R. C., Inklaar, R., and Timmer, M. (2013). The Next Generation of the Penn World Table. *American Economic Review*, 105(10):1–42.

Gomme, P., Ravikumar, B., and Rupert, P. (2011). The return to capital and the business cycle. *Review of Economic Dynamics*, 14(2):262–278.

Gomme, P., Ravikumar, B. F. R. B. o. S. L., and Rupert, P. (2015). Secular Stagnation and Returns on Capital. *Economic Synopses*, (19):6–8.

Gopinath, G., Kalemli-Özcan, Å., Karabarbounis, L., and Villegas-Sanchez, C. (2017). Capital Allocation and Productivity in South Europe. *Quarterly Journal of Economics*, 132(4):1915–1967.

Gourinchas, P. O. and Jeanne, O. (2013). Capital flows to developing countries: The allocation puzzle. *Review of Economic Studies*, 80:1484–1515.

Gutiérrez, G. and Philippon, T. (2017a). Declining Competition and Investment in the U.S. *NBER Working Paper Series*, (July).

Gutiérrez, G. and Philippon, T. (2017b). Investmentless Growth: An Empirical Investigation. *Brookings Papers on Economic Activity*.

Haltiwanger, J. C. (1997). Measuring and Analyzing Aggregate Fluctuations: The Importance of Building From Microeconomic Evidence. *Federal Reserve Bank of St. Louis Review*, May/June:55–78.

Heathcote, J. and Perri, F. (2013). Assessing International Efficiency. *Federal Reserve Bank of Minneapolis, Research Department Staff Report 480*, (April).

Hsieh, C.-T. and Klenow, P. J. (2007). Relative Prices and Relative Prosperity. *American Economic Review*, 97(3):562–585.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics*, CXXIV(4):1–55.

Jin, K. (2012). Industrial structure and capital flows. *American Economic Review*, 102(5):2111–2146.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. (2018). The Rate of Return on Everything, 1870-2015. *CESifo Working Paper Series 6899*.

Karabarbounis, L. and Neiman, B. (2014). The Global Decline in the Labor Share. *Quarterly Journal of Economics*, pages 61–103.

Karabarbounis, L. and Neiman, B. (2018). Accounting for Factorless Income. *NBER Working Paper Series*, (24404).

Koh, D., Santaeulàlia-Llopis, R., and Zheng, Y. (2018). Labor Share Decline and Intellectual Property Products Capital. *Working Papers 873, Queen Mary University of London, School of Economics and Finance.*

Krugman, P. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *The American economic review*, 70(5):950–959.

Lane, P. R. and Milesi-Ferretti, G. M. (2007). The external wealth of nations mark II: Revised and extended estimates of foreign assets and liabilities, 1970-2004. *Journal of International Economics*, 73(2):223–250.

Loecker, B. J. D. and Warzynski, F. (2012). Markups and Firm-Level Export Status. *American Economic Review*, 102(6):2437–2471.

Lowe, M., Papageorgiou, C., Perez-sebastian, F., and Sebastian, F. (2012). The Public and Private MPK. *DEGIT Conference Papers*.

Lucas, R. E. J. (1990). Why Doesn't Capital Flow From Rich To Poor Countries ? *The American Economic Review*, 80(2):92–96.

Marx, M., Mojon, B., and Velde, F. R. (2018). Why Have Interest Rates Fallen Far Below the Return on Capital. *Federal Reserve Bank of Chicago Working Paper*.

Monge-Naranjo, A., Sánchez, J. M., and Santaeulàlia-Llopis, R. (2015). Natural Resources and Global Misallocation. *Working Papers 2015-36, Federal Reserve Bank of St. Louis, revised 13 Jul 2016*.

Olley, G. S. and Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64(6):1263–1297.

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

Samuelson, P. A. (1971). Ohlin Was Right. *The Swedish Journal of Economics*, 73(4):365–384.

Summers, L. H. (2014). U.S. economic prospects: Secular stagnation, hysteresis, and the zero lower bound. *Business Economics*, 49(2):65–73.

Święcki, T. (2017). Determinants of structural change. *Review of Economic Dynamics*, 24:95–131.

Traina, J. (2018). Is Aggregate Market Power Increasing? Production Trends Using Financial Statements. *Stigler Center New Working Paper Series*, 17(February):1–21.

Ventura, J. (1997). Growth and interdependence. *Quarterly Journal of Economics*, 112(1):57–84.

Zymek, R. (2015). Factor proportions and the growth of world trade. *Journal of International Economics*, 95(1):42–53.