Topics in Macroeconomics: Mortgage Default, Demographic Change and Factor Misallocation

Thomas Schelkle

Thesis submitted for the degree of

Doctor of Philosophy in Economics at the

London School of Economics and Political Science

June 2012

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The second chapter draws on work that was carried out jointly with equal share by Alexander Ludwig, Edgar Vogel and me. This chapter contains material from our article published in the *Review of Economic Dynamics* (Ludwig, Schelkle, and Vogel 2012) and the online appendix to that article. This work has an early foundation in my Diploma thesis at the University of Mannheim. Prior to publication our article has been edited for conventions of language, spelling and grammar by a professional editing company called American Journal Experts.

The third chapter expands on my MRes paper at the London School of Economics.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 45,000 words.

Abstract

This thesis analyzes three topics in macroeconomics: Mortgage default, demographic change and factor misallocation.

The first chapter asks which theories of mortgage default are quantitatively consistent with observations during the U.S. mortgage crisis. Different default models are simulated for the path of observed house prices and their predictions are compared to observed default rates. The double-trigger hypothesis attributing mortgage default to the joint occurrence of negative equity and a life event like unemployment explains this data well. A structural partial-equilibrium model with liquidity constraints and unemployment risk provides micro-foundations for this hypothesis. The model implies that subsidizing homeowners can mitigate a mortgage crisis at a lower cost than bailing out lenders.

The second chapter investigates the macroeconomic effects of population aging in the United States during the coming decades. In particular we analyze the role of endogenous human capital formation during this process. We build a large-scale overlapping generations model with endogenous human capital accumulation and calibrate it such that it replicates observed life-cycle earnings profiles. We then simulate a realistic demographic transition. Our key finding is that human capital adjustments may act as a quantitatively important mitigation mechanism that dampens the macroeconomic and adverse welfare effects of demographic change.

The third chapter estimates the degree of capital and labor misallocation between the agricultural and non-agricultural sector in different countries. The framework employs the flexible Translog production function and performs non-linear regressions on cross-country panel data observed during 1967-1992. The findings are that in developing countries marginal products of labor are higher in non-agriculture than in agriculture.

The reverse holds for capital allocation. Industrialized countries are closer to an efficient factor allocation. A sensitivity analysis reveals that using Cobb-Douglas production functions leads to much higher estimates of misallocation.

Acknowledgements

This thesis would not have been possible without the support of many people to whom I am greatly indebted.

First, I would like to thank my supervisor, Francesco Caselli, for invaluable guidance and advice throughout the past years. I have learnt a tremendous amount from his questions, comments and suggestions. I am also very grateful to my advisor, Rachel Ngai, for helpful discussions and advice in particular during the run-up to the job market.

The work presented in this thesis has also benefitted greatly from discussions with Albert Marcet, Wouter den Haan, Stephan Seiler, James Hansen, Piotr Zurawski and Sebastian Stumpner. I also thank participants at various seminars and conferences for their comments. I am especially grateful to Alexander Ludwig and Edgar Vogel for our joint work on the research project presented in the second chapter.

For financial support during my PhD, I thank the Economic and Social Research Council, the Economics Department at LSE and Stiftung Besinnung und Ordnung.

I also thank my friends and fellow students for the great years during the PhD.

My special thanks go to my parents, Christa and Erich, for their continued love and support, and to Madeleine for all her love and patience.

Contents

List of Figures

List of Tables

Preface

This thesis analyzes three topics in macroeconomics: Mortgage default, demographic change and factor misallocation.

The first chapter “Mortgage Default during the U.S. Mortgage Crisis” asks which theories of mortgage default are quantitatively consistent with observations in the United States during 2002-2010. Theoretical models are simulated for the observed time-series of aggregate house prices and a realistic microeconomic house price distribution. Their predictions are then compared to actual default rates on prime fixed-rate mortgages. An out-of-sample test discriminates between estimated reduced forms of the two most prominent theories. The test reveals that the double-trigger hypothesis attributing mortgage default to the joint occurrence of negative equity and a life event like unemployment outperforms a frictionless option-theoretic default model. Based on this finding a structural partial-equilibrium model with liquidity constraints and idiosyncratic unemployment shocks is presented to provide micro-foundations for the double-trigger hypothesis. In this model borrowers with negative equity are more likely to default when they are unemployed and have low liquid wealth. The model explains most of the observed strong rise in mortgage default rates. A policy implication of the model is that subsidizing homeowners can mitigate a mortgage crisis at a lower cost than bailing out lenders.

The second chapter “Demographic Change, Human Capital and Welfare” investigates the macroeconomic effects of population aging in the United States during the coming decades. In particular we analyze the quantitative role that endogenous human capital formation may play as an adjustment mechanism to demographic change. We build a large-scale overlapping generations model with endogenous human capital accumulation and calibrate it such that it replicates past observed life-cycle earnings profiles. In this model we then simulate a realistic demographic transition based on observed demographic data and projections for the future. The projected demographic changes will reduce the share of the working-age population. Analyses based on standard models with a fixed human capital profile predict that these changes will increase the capital-labor ratio. Hence, rates of return to capital decrease and wages increase, which has adverse welfare consequences for current cohorts who will be retired when the rate of return is low. This chapter argues that adding endogenous human capital accumulation to the standard model dampens the macroeconomic and adverse welfare effects of demographic change. We find that this adjustment channel is quantitatively important. The standard model with exogenous human capital predicts welfare losses up to 12.5% (5.6%) of lifetime consumption, when contribution (replacement) rates to the pension system are kept constant. These numbers reduce to approximately 8.7% (4.4%) when human capital can endogenously adjust.

The third chapter "Factor Misallocation in Dual Economies" estimates the degree of capital and labor misallocation between the agricultural and non-agricultural sector in different countries. First it is shown how the observed average product ratio between the two sectors is linked to the unobserved marginal product ratio. The degree of misallocation is then estimated by non-linear regressions using a panel data set of developed and developing countries observed during 1967-1992. The econometric approach allows for general production functions, so the paper employs the flexible Translog form. The findings are that in developing countries marginal products of labor are higher in non-agriculture than in agriculture. The reverse holds for capital allocation. Industrialized countries are closer to an efficient factor allocation. A sensitivity analysis reveals that using the more restrictive Cobb-Douglas form would lead to much higher estimates of misallocation.

1 Mortgage Default during the U.S. Mortgage Crisis

1.1 Introduction

After the collapse of the house price boom in the United States residential mortgage delinquencies of both prime and subprime loans have increased substantially. The widespread rise in default rates and resulting losses of mortgage-backed-securities marked the onset of the recent financial and economic crisis. These events highlight key research questions on mortgage default. What are the economic mechanisms driving mortgage default? And what explains the strong rise in mortgage default rates in recent years?

This paper examines how well theoretical models of mortgage default can quantitatively explain the rise in default rates in the Unites States between 2002 and 2010. Theoretical models are simulated for the observed time-series of aggregate house prices and a realistic microeconomic house price distribution. Their predictions are then compared to data on default rates of prime fixed-rate mortgages. In the first part of the paper the observed variation in default rates and aggregate house prices is used to discriminate between the two major mortgage default theories - the frictionless option-theoretic default model and the "double-trigger" hypothesis.

The traditional frictionless option-theoretic literature, sometimes also called the "ruthless" default model, assumes that borrowers default on their mortgage in order to maximize their financial wealth. In this framework negative equity is a necessary, but not sufficient, condition for default. Instead there exists a threshold level of negative equity or the house price such that a rational wealth-maximizing agent will exercise the default option as in Kau, Keenan, and Kim (1994), among others. This theory assumes that the borrower has access to a perfect credit market for unsecured credit such that default is unaffected by liquidity considerations and income fluctuations. Quercia and Stegman (1992) and Vandell (1995) provide a survey and further references.

Another prominent idea on mortgage default is the double-trigger hypothesis. This theory also views negative equity as a necessary condition for default. But it attributes default to the joint occurrence of negative equity and a life event like unemployment or divorce. The double-trigger hypothesis is well-known among mortgage researchers. But it is usually discussed only in words or stylized models as in Gerardi, Shapiro, and Willen (2007), Foote, Gerardi, and Willen (2008) and Foote, Gerardi, Goette, and Willen (2009), among others, and has not been presented as a structural dynamic stochastic model.

These two microeconomic theories are tested on their aggregate predictions. The procedure specifies reduced form models of the two theories, estimates them on part of the data and then tests the estimated models on out-of-sample predictions. The result of this test is that the double-trigger hypothesis outperforms the frictionless default model. The frictionless theory is excessively sensitive to changes in aggregate house prices and predicts a far too strong rise in default rates. In contrast, the double trigger hypothesis is consistent with the evidence. The economic reason is that default rates have increased roughly in proportion to the number of borrowers who experience any level of negative equity as predicted by the double-trigger theory. In contrast, the predictions of the frictionless theory are based on the number of homeowners experiencing extreme levels of negative equity and this has increased by much more than actual default rates. This is an important result in itself given the disagreement in the literature. It is also an important step towards developing mortgage default models that can be used for policy and risk analysis because such analysis needs to be based on models that are empirically accurate.

Based on this finding the second part of the paper aims at providing a micro-foundation for the double-trigger hypothesis. A structural dynamic stochastic partial-equilibrium model of mortgage default featuring liquidity constraints and idiosyncratic unemployment shocks is presented. The liquidity constraint forces unemployed borrowers who have exhausted their buffer stock savings to make painful cuts to consumption. This magnifies the cost of servicing the mortgage such that unemployment becomes a trigger event for default. In addition the model includes a direct utility flow of owning a house.

This is an important feature to generate double-trigger behavior because it prevents employed agents from defaulting after a strong fall of house prices. The model then attributes default to the joint occurrence of negative equity and the liquidity problems caused by unemployment. The model is calibrated, estimated and assessed on its power to predict out-of-sample. A comparison to observed default rates reveals that the model can quantitatively explain most of the rise in mortgage default as a consequence of falling aggregate house prices.

One benefit of the structural model is that it can be used for policy analysis. This is exemplified by analyzing two possible policies in a mortgage crisis that neutralize the losses lenders incur from mortgage default. One could either bail out the lenders or mitigate the liquidity problems of homeowners who would otherwise default such that they stay in their houses. An implication of the structural model is that a subsidy policy to homeowners is the cheaper option when liquidity problems play a key role in default decisions.

From a macroeconomic perspective the finding that the structural and reduced-form double-trigger model can explain the rise in default rates by the dynamics of aggregate house prices is important. This points towards the existence of systematic macroeconomic risk in the mortgage market. The main alternative explanation is that lending standards and loan quality deteriorated sharply before the crisis. This paper presents evidence that at least in my data set on prime mortgages this is an unlikely explanation for the rise in default rates because average loan characteristics are fairly stable over time.

As a background to the paper it is important to know that loan-level data that links an individual borrower's repayment history to the history of individual house prices and employment status does not exist. This makes it difficult to distinguish empirically between different theories at the individual level. This paper takes a different approach and tests the aggregate predictions of different theories. Along this line the key and unique feature of the paper is that it includes a realistic microeconomic house price distribution around the aggregate trend. This means that an empirically successful model is required to be consistent with both the aggregate house price trends and the moments of the observed microeconomic distribution. In contrast, the prior empirical literature relies on regional house price indices as explanatory variables and thus very likely omits part of the microeconomic house price variation from its regressions.

One of the main contributions of the paper is to compare simulations from theoretical mortgage default models directly to empirical observations. Most of the prior literature has in contrast been divided into theoretical work that does not discuss the explanatory power of the theories on the one hand, and reduced-form regressions on the other.

The structural model of the paper builds on previous work by Campbell and Cocco (2003, 2011) and Corradin (2009) who also model liquidity constraints in a mortgage framework.[1] These models are similar to the structural model presented in this paper, but these papers do not compare the models to the data. Their focus is also different, for example Campbell and Cocco (2011) are mainly concerned with theoretical differences between fixed- and variable rate mortgages. In contrast, my paper adds the macroeconomic perspective. It shows how variation in the time-series of aggregate house prices can explain the rise in default rates during recent years within a structural model. My analysis also reveals that in addition to liquidity constraints it is important to allow for a direct utility flow from owning a house as explained above. Otherwise the model remains too close to a ruthless default model and cannot match the data well.

There is also a recent literature that uses equilibrium models to examine the role various institutional features like bail-out guarantees or mortgage product innovation and falling house prices played for the mortgage crisis. Examples include Chatterjee and Eyigungor (2011), Corbae and Quintin (2011), Jeske, Krueger, and Mitman (2011) and Garriga and Schlagenhauf (2009). In contrast to this line of research, my paper focusses in more detail on the household decision to default and discriminating between the two default theories. While the other papers simulate the effect of a relatively stylized fall in house prices on aggregate foreclosures, I feed detailed time-series of observed house prices into the simulation and try to explain differences in default behavior of different loan cohorts. Another important advantage of my simulation framework for house prices is that it is based closely on the procedures and estimates of the FHFA and thus requires the theories to be also consistent with the dynamics observed in microeconomic house price data.

A vast number of empirical papers have studied the determinants of mortgage default typically estimating hazard models on loan-level data. The pre-crisis literature is surveyed by Quercia and Stegman (1992) and Vandell (1995) and an example is the study

[1]In modeling liquidity constraints the structural model also builds on the buffer-stock saving framework of Zeldes (1989), Deaton (1991) and Carroll (1997).

by Deng, Quigley, and Van Order (2000). The U.S. mortgage crisis has then caused an enormous increase in empirical work on mortgage default.[2] These papers present a wealth of evidence that negative equity or falling house prices are strong determinants of default. Some studies have also investigated the role of life events as triggers for default and found that state unemployment or divorce rates can explain default. My paper is motivated by these empirical results. But it uses a very different methodology and thus provides complementary evidence on the relative merit of the two theories.

The empirical literature also finds a great heterogeneity in default behavior for borrowers with the same level of negative equity (Quercia and Stegman 1992). The structural model I present here can rationalize this fact because in that model the default threshold of negative equity depends on liquid wealth and employment status. Individual heterogeneity in these variables, which are unobserved in all standard mortgage data sets, may then account for the heterogeneity in default behavior of borrowers with the same level of negative equity. The theoretical model also suggests that interaction effects between negative equity and variables measuring liquidity are of key importance for default as has been found empirically by Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010).

The paper is structured as follows. Section 1.2 describes the data and empirical facts on mortgages and house prices. The test between the two theories based on reduced-form models is presented in section 1.3. The structural model is developed in section 1.4 and parameterized in section 1.5. The results of the structural model are presented in section 1.6. Section 1.7 discusses the alternative explanation for the rise in default rates that loan quality deteriorated sharply and shows that there is no strong evidence for this in my data set. The structural model is applied for policy analysis in section 1.8. For reasons explained in the data section most of the paper concentrates on loans with a high loan-to-value ratio, but section 1.9 discusses an extension to lower loan-to-value ratios. Section 1.10 concludes.

[2]Studies within this extensive literature differ by research question, estimation method, analyzed data set and results. A detailed literature review that would do justice to these different contributions is unfortunately beyond the scope of this paper. Examples of this empirical research include Amromin and Paulson (2009), Bajari, Chu, and Park (2010), Demyanyk and Van Hemert (2011), Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010), Foote, Gerardi, Goette, and Willen (2008), Foote, Gerardi, Goette, and Willen (2009), Foote, Gerardi, and Willen (2008), Gerardi, Lehnert, Sherlund, and Willen (2008), Gerardi, Shapiro, and Willen (2007), Ghent and Kudlyak (2010), Guiso, Sapienza, and Zingales (2011), Jagtiani and Lang (2010), Mayer, Pence, and Sherlund (2009) and Mian and Sufi (2009), among others.

1.2 Data and Empirical Facts

This section presents the data on mortgages default rates and house prices and the key facts the paper attempts to explain. It also describes how the simulation procedure for house prices is based on empirical evidence.

1.2.1 Mortgage Data

In this paper, I use aggregate data on mortgage characteristics and payment histories in the United States. The data set contains information that was aggregated from the large loan-level data base of Lender Processing Services (LPS), also known as McDash data. "Aggregate" here simply means that my data contain the average value of a certain characteristic for all loans in the data base that satisfy a set of conditions that I can specify. These conditioning variables allow the selection of sub-samples from the full data base and tracking different loan cohorts over time.

The data cover the time period from January 2002 until June 2010 at a monthly frequency and the analysis is focussed on loans originated between 2002 and 2008. I restrict the sample to prime, first, fixed-rate, 30-years mortgages that have a standard amortization schedule (are not balloon mortgages). I focus on only one mortgage type because the structural model would have to be recomputed for each different mortgage contract. The selection is motivated by the fact these are the most common mortgage contracts. The data base contains around 23 million loans with these characteristics in 2010.[3]

I further focus the analysis on loans with a loan-to-value ratio (LTV) above 95%, which depending on the year represents about $20-30\%$ of all loans that satisfy the above restrictions. Looking at loans with different LTVs separately allows to generate a more accurate home equity distribution in the model. This is important due to the highly non-linear relationship between default decisions and negative equity. Furthermore, the loans with a high LTV default most frequently, so it makes sense to focus an analysis of mortgage default on them. But the main reason for concentrating on this group is a

[3]Amromin and Paulson (2009) estimate that the LPS data cover about 60% of the prime market between 2004 and 2007. Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010) report that the LPS data cover about 70% of all mortgage originations in 2005 and 2006. But coverage varies by year with lower coverage in earlier years.

data problem. In the LPS data only the LTV of the first mortgage is observed, but not the combined LTV of the first and a possible second mortgage.[4] Since the combined mortgage amount should be relevant for a borrower's decision to default the fact that second mortgages are unobserved is a problem for empirical work. This is a particular concern for structural models because of the strong role that theoretical approaches place on negative equity. In order to mitigate this data problem I thus focus on first mortgages with a very high LTV because these borrowers should be least likely to have a second mortgage on their home. However I also investigate whether and how the conclusions of the paper generalize to loans with a LTV of the first mortgage between 75% and 84% in section 1.9.

The data set contains aggregate information on contract characteristics by month of origination like the mortgage rate or credit scores. Furthermore, aggregate statistics on payment behavior are observed each month and broken down by the age of the loan. This allows tracking the payment behavior of different cohorts of loans (defined by month of origination) over time. Specifically, each of these cells (defined by time period and loan age) contains how many active loans are delinquent or in foreclosure and how many are terminated through foreclosure or prepayment. Following much of the recent empirical literature cited in the introduction, I define a loan to be in default when it is 60 days or more past due, i.e. two payments have been missed. Accordingly, cumulative default rates for a loan cohort are constructed as the share of active loans that are 60 days or more delinquent or in foreclosure times the share of initial loans that are still active plus the share of initial loans where foreclosure has already been completed.

1.2.2 House Prices

Information on house prices comes from the Federal Housing Finance Agency (FHFA). The monthly national and census division level repeat-purchase house price indices between 1991 and 2010 deflated by the Consumer Price Index (CPI) are used as measures of aggregate real house price movements. Estimates of the moments of the microeconomic house price distribution within a census division around the respective aggregate

[4]Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010) provide evidence that second mortgages are frequent and significantly affect the combined loan-to-value ratio. They report that on average 26% of all borrowers have a second mortgage and this adds on average 15% to the combined LTV. Unfortunately, they do not report a break-down of these statistics by the LTV of the first mortgage.

trend are used to generate a realistic house price distribution in the simulation. This is important because otherwise theoretical models cannot explain any default during times of positive aggregate house price growth.

Throughout the paper the evolution of the real house price P_{it} of an individual house i in period t is modeled as

$$\ln(P_{it}) = \ln(P_{i,t-1}) + g_t^{agg} + g_{it}^{ind} \tag{1.1}$$

where the house price growth rate has two components, an aggregate component g_t^{agg} that is common to all houses and an individual component g_{it}^{ind} specific to the individual house. Such a formulation is consistent with the approach used by the FHFA to estimate the house price index, cf. the description in Calhoun (1996).[5] The general aim is to base the simulation framework for house prices as directly as possible on the empirical procedures and estimates of the FHFA.

In equation (1.1) a census division index was suppressed for convenience. But the aggregate trend represented by g_t^{agg} and the moments of g_{it}^{ind} are in fact specific to the census division in which the house is located. Thus, this paper uses data at the census division level and information on the regional composition of loan cohorts in the mortgage data. When drawing house prices the simulation draws are allocated across census divisions such that in each cohort the simulated sample has the same regional composition as in the mortgage data. The aggregate component g_t^{agg} represents the growth rate of the census division real house price index. In the simulation this component is taken directly from the data.

The individual component g_{it}^{ind} is unobserved. But the FHFA provides estimates of the variance that are used to simulate a realistic microeconomic house price distribution. Specifically, it is assumed that the individual component g_{it}^{ind} is independent over time and individuals and normally distributed with mean zero and variance V_t. The variance of g_{it}^{ind} depends on the time since the house was bought. This is a realistic feature of the data and based on estimates of the FHFA. Using my own notation, cf. footnote 5,

[5] I use a slightly different notation relative to the FHFA because I want to use this equation in a dynamic optimization problem and simulations. In order to see how it is related, rewrite equation (1.1) as

$$\ln(P_{it}) = \ln(P_{i,0}) + \sum_{\tau=1}^{t} g_\tau^{agg} + \sum_{\tau=1}^{t} g_{i\tau}^{ind}$$

where $\ln(P_{i,0}) + \sum_{\tau=1}^{t} g_\tau^{agg} = \beta_t + N_i$ and $\sum_{\tau=1}^{t} g_{i\tau}^{ind} = H_{it}$ give equation (1) in Calhoun (1996).

the FHFA specifies a quadratic formula in time for the variance of the total individual part of the house price change since purchase given by

$$\text{Var}\left(\sum_{\tau=1}^{t} g_{i\tau}^{ind}\right) = \frac{\kappa}{3}t + \frac{\lambda}{9}t^2. \tag{1.2}$$

where an adjustment has been made for the fact that this paper operates at a monthly instead of a quarterly frequency. By the independence assumption the variance of g_{it}^{ind} is then given by

$$V_t = \text{Var}\left(g_{it}^{ind}\right) = \text{Var}\left(\sum_{\tau=1}^{t} g_{i\tau}^{ind}\right) - \text{Var}\left(\sum_{\tau=1}^{t-1} g_{i\tau}^{ind}\right) = \frac{\kappa}{3} + \frac{\lambda}{9}(2t-1).$$

The FHFA provides estimates of κ and λ at the census division level that I use to generate realistic distributions around the division level aggregate trends. The estimates of κ are positive and those of λ are negative and small in absolute magnitude. This implies that the variance of $\sum_{\tau=1}^{t} g_{i\tau}^{ind}$ increases less than linearly with time and the variance of a single g_{it}^{ind} is decreasing over time. On average across census divisions the estimates of κ and λ imply that the shock in the first month g_{i1}^{ind} has a standard deviation of about 2.49%, while after five years the standard deviation of g_{i60}^{ind} is around 2.37%. Hence the standard deviation of g_{it}^{ind} decreases relatively slowly over time.

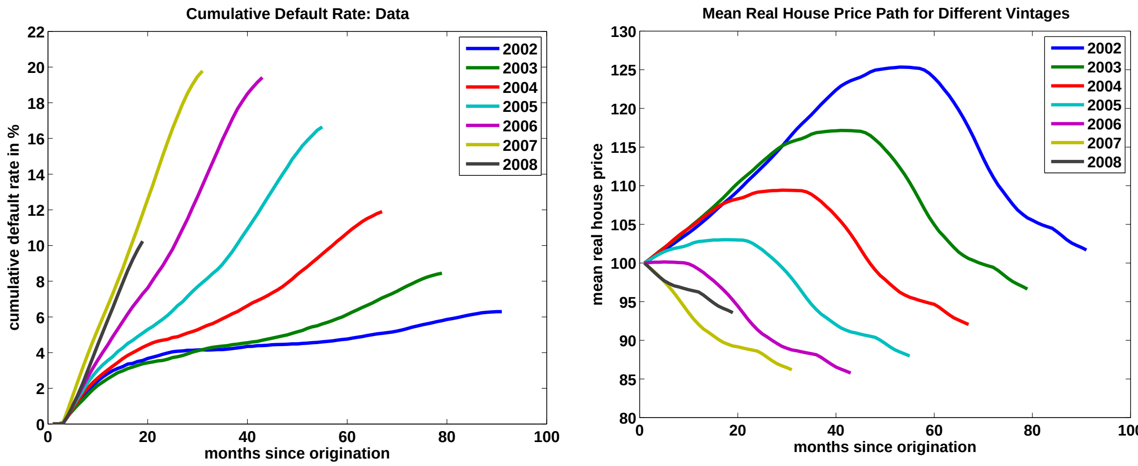
1.2.3 Empirical Facts on Default Rates and House Prices

The key empirical facts on mortgage default rates and house prices are presented in figure 1.1. Figure 1.1(a) shows the average cumulative default rates for loan cohorts originated between 2002 and 2008 grouped by the year of origination in my data set. The data show clearly that loan cohorts originated later during this period defaulted much more frequently at the same time since origination. This increase constitutes part of the US mortgage default crisis and shows that the rise in default rates was in no way restricted to the subprime market and adjustable rate or hybrid mortgages.

Figure 1.1(b) presents the mean real house price paths (normalized to 100 at origination) for the cohorts of loans originated between 2002 and 2008. Borrowers of loans originated between 2002 and 2005 experienced on average rising real house prices during the immediate time after origination and falling house prices later during the course

Figure 1.1: Cumulative Default Rates and House Prices for Different Origination Years

(a) Cumulative Default Rates (b) Mean Real House Price Paths



of the loan. In contrast, the real value of a home of a borrower who took out a mortgage between 2006 and 2008 decreased sharply immediately after origination.

The key research questions of the paper are motivated by the facts in figure 1.1. Can the variation in house price paths quantitatively explain the variation in mortgage default rates across cohorts within a structural economic model? What features should such an empirically successful model have? Does this variation allow discrimination between different theoretical models of mortgage default?

1.3 Reduced Form Models

This section presents evidence on mortgage default from estimating and simulating two highly stylized models. But these models are motivated by economic theory and represent the simplest possible reduced forms of a frictionless option-theoretic model and the double-trigger hypothesis. The aim is to discriminate between these different theories in a relatively general way that is independent of the exact specification of the respective structural model. Building on these results the following section then develops a structural economic model that has hope to be empirically successful.

1.3.1 Model Setup

The paper considers individual borrowers who took out a fixed-rate 30-years mortgage. Each loan cohort defined by origination date consists of many borrowers who are indexed by $i = 1, \ldots, N$ and observed in periods $t = 1, \ldots, T$ after loan origination. Borrowers take a single decision each period and can either service the mortgage or default on the loan and "walk away" from the house. Denote the default decision of an individual borrower i in month t after origination by a set of dummy variables d_{it}. The variables d_{it} take the value 1 once the borrower has defaulted, and the value 0 in all periods prior to default. Thus it is sufficient to present default decision rules in period t for situations when the borrower has not defaulted yet.

The next two sections present the two models, the "threshold" and "shock" model. Both models view negative equity as a necessary, but not sufficient, condition for default. The individual decision rules in the two models differ in how default exactly depends on house equity, and hence the house price and the mortgage balance.

For a fixed-rate mortgage the nominal mortgage balance M_{it} of borrower i evolves deterministically over time according to

$$M_{i,t+1} = (1 + r^m)M_{it} - m_i \tag{1.3}$$

where r^m is the monthly mortgage rate which is constant across individuals. m_i are fixed nominal monthly payments covering mortgage interest and principal. These payments are determined at the beginning of the contract and satisfy

$$m_i = \left[\sum_{t=1}^{T} \frac{1}{(1 + r^m)^t} \right]^{-1} M_{i0} \tag{1.4}$$

where M_{i0} is the initial loan amount and the loan has a maturity of $T = 360$ months. The initial loan amount is a function of the initial loan to value ratio LTV_i and initial house price P_{i0} and given by $M_{i0} = LTV_i \times P_{i0}$. Here borrowers are heterogenous with respect to the LTV. It is assumed that agents take decisions based on real variables. Thus it is useful to define the real mortgage balance as $M_{it}^{real} = \frac{M_{it}}{\Pi_t}$ where Π_t is the CPI and $\Pi_0 = 1$. This assumption does not affect the results and the conclusions are identical when decisions are based on nominal variables.

The real house price P_{it} of an individual homeowner evolves according to equation (1.1). House price growth has an aggregate and individual component as described in section 1.2.2. P_{i0} is normalized to 100. This involves no loss of generality as seen below.

Due to the simplicity of the presented models I also add the constraint that default is only allowed from the fourth month since origination onwards. This is completely ad-hoc, but provides a better fit of both models to the data in the early periods after origination when default rates are essentially zero. But the comparison across models and the conclusions drawn below do not depend on this assumption.

1.3.2 The Threshold Model

The first model assumes that borrowers with negative equity default on their mortgage at the first time that the real value of equity falls below a certain threshold value. Therefore I call this the "threshold model". Here, I adopt the simplest possible specification with a threshold that is proportional to the initial house price and constant over time given by ϕP_{i0} where $\phi < 0$. If in period $t \geq 4$ the borrower has not defaulted yet then the default decision in that period is described by

$$d_{it} = \begin{cases} 1, & \text{if } P_{it} - M_{it}^{real} < \phi P_{i0} \\ 0, & \text{otherwise} \end{cases} \tag{1.5}$$

This is a simple reduced-form of a frictionless option model. The corresponding structural model would derive the threshold parameter ϕ from optimizing behavior. For example the borrower might trade off the expected future capital gains on the house for the mortgage payments in excess of rents. Here I remain agnostic about the exact trade-off and the value of ϕ and instead estimate it from the data.

1.3.3 The Shock Model

The second model assumes that borrowers with any level of negative equity only default on their mortgage when they also receive a default shock in that period. I call this the "shock model". Again I adopt the simplest possible specification. The probability to receive a default shock ψ is constant and satisfies $0 \leq \psi \leq 1$ and default shocks are independently and identically distributed over time. If the borrower has not defaulted yet, the default decision in period $t \geq 4$ is determined by

$$d_{it} = \begin{cases} 1, & \text{if } P_{it} - M_{it}^{real} < 0 \text{ and the default shock occurs} \\ 0, & \text{otherwise} \end{cases} \tag{1.6}$$

This is a reduced-form of a double-trigger model. Here the default shock represents the life event like unemployment or divorce that combined with negative equity triggers default. The parameter ψ represents the probability that the life event occurs. Again ψ needs to be estimated from the data.

1.3.4 Model Simulation, Estimation and Test

Conditional on the respective model parameters ϕ and ψ both models can be simulated for subsequent cohorts of loans originated each year between 2002 and 2008. For each cohort I draw $100,000$ individual histories of house prices and default shocks with the same length in months as the respective cohort is observed in the data.[6] When computing the mortgage balance the mortgage rate is kept constant within a cohort and set equal to the respective cohort average. But borrowers within a cohort are heterogenous with respect to the LTV which varies in steps of one percentage point between 95% and 104%.[7] The frequency of these different loan-to-value ratios at origination is varied across cohorts as observed in the mortgage data. This means that possible changes to the average mortgage rate and the LTV distribution across cohorts are taken into account in the simulation. Data on the path of inflation rates from the CPI is used to compute the real mortgage balance. The decision rules are then applied to these shock histories and paths of the real mortgage balance.

The idea of the test procedure is to use only the default data of the cohort originated in 2002 to estimate the unknown parameters ϕ and ψ. The test of the models is then based on out-of-sample predictions. Conditional on the parameter values estimated from the 2002 cohort, default rates for the cohorts 2003 to 2008 are simulated from the

[6] The simulation procedure for individual house prices is explained in detail in section 1.2.2. For the shock model I also draw histories from an i.i.d. uniform distribution on the interval $[0,1]$. For a given parameter ψ the default shock occurs for the respective individual and month if the uniform draw is smaller or equal to ψ.

[7] The few loans with a LTV above 104% are subsumed under the 104% LTV group.

models. The test constitutes in comparing simulated and empirically observed default rates and checking which estimated model gives a better fit to the data.[8]

The model parameters are estimated by a simulated method of moments procedure. Let θ stand in for the parameter to be estimated in the respective model. The idea of the estimation is to choose θ such that the cumulative default rates for the 2002 cohort simulated from the model match as well as possible those observed in the data. Collect the variables d_{it} in one vector $D_i = [d_{i1}, \ldots, d_{iT}]'$ for each individual. The mean of this vector $\overline{D} = \frac{1}{N}\sum_{i=1}^{N} D_i$ represents the empirically observed cumulative default rate. The expected value of D_i is $E[D_i] = D(\theta)$ and denote the expected value evaluated by simulation of S individuals from the model by $\widetilde{D}(\theta)$. The deviation of the model from the data is then given by $G(\theta) = \overline{D} - \widetilde{D}(\theta)$. The simulated method of moment estimator of θ minimizes $G(\theta)'WG(\theta)$ where W is a weighting matrix. I weight all moments equally by using an identity matrix as the weighting matrix. θ is then estimated by minimizing a least squares criterion function given by

$$\frac{1}{T}\sum_{t=1}^{T}\left(\overline{d_t} - \widetilde{d}_t(\theta)\right)^2 \tag{1.7}$$

where $\overline{d_t}$ and $\widetilde{d}_t(\theta)$ are the t-th element in the vectors $\overline{D}$ and $\widetilde{D}(\theta)$, respectively. Here $\widetilde{d}_t(\theta)$ is evaluated using a frequency simulator such that $\widetilde{d}_t(\theta) = \frac{1}{S}\sum_{j=1}^{S}\widetilde{d_{jt}}(\theta)$ and $\widetilde{d_{jt}}(\theta)$ represents the outcome for period t of applying the decision rules to the drawn history j of the underlying shocks. The minimization problem is solved by a grid search algorithm.

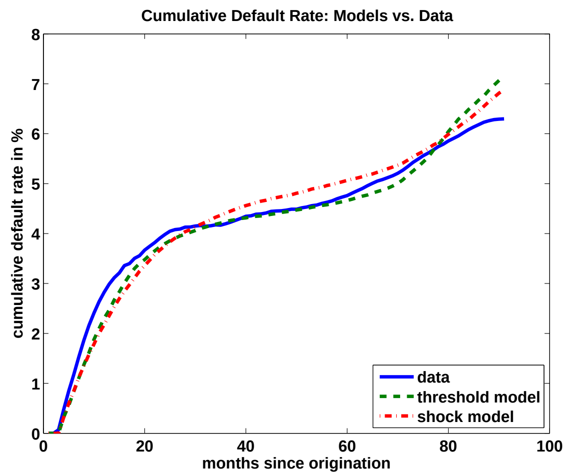
1.3.5 Results

For the threshold model the negative equity default threshold ϕ is estimated as -11.0%. This means borrowers default as soon as they have a real value of negative equity of 11% of the initial house price. In contrast, for the shock model the default shock probability ψ is estimated to be 1.3% such that each period 1.3% of those borrowers with negative equity default on their loan. The fit of the two models to the cumulative

[8] I use a formal estimation approach to derive point estimates of the two model parameters as explained in the next paragraph. But the testing procedure I adopt is relatively informal. Neither do I compute standard errors for the estimated parameters, nor confidence bands for the model predictions due to parameter uncertainty. I also do not provide a formal test statistic to evaluate the different models. This is an interesting and important area to improve the paper in a future revision.

default rate of the 2002 cohort is shown in figure 1.2. Both models are able to fit this data very well.

Figure 1.2: Cumulative Default Rate for 2002 Cohort: Models vs. Data

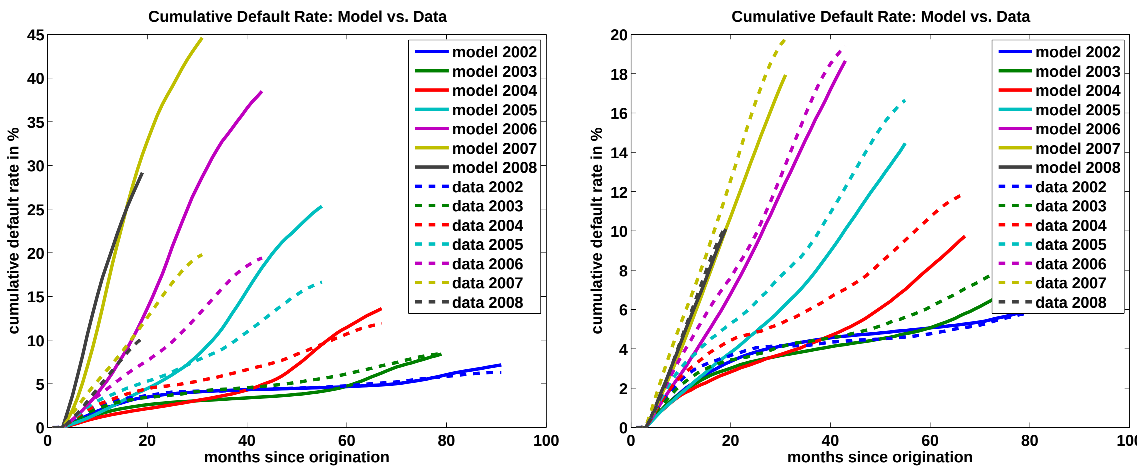


The next step is to test the two estimated models by checking how well they perform in predicting out-of-sample. Figure 1.3(a) shows the fit of the threshold model to the full sample of all cohorts between 2002 and 2008. The equivalent fit of the shock model is presented in figure 1.3(b). It turns out that the threshold model has severe empirical problems. When it is forced to match default rates of the 2002 cohort, it over-predicts default rates for the later cohorts in the simulation period by at least one order of magnitude. The threshold model is excessively sensitive to the shifts in the mean of the house price distribution observed in the data. In contrast, the shock model gives a very good fit to the broad dynamics in the data. However, the shock model predicts too few defaults especially for the 2004 cohort and to some extent also for the 2003 and 2005 cohorts. This could imply that these cohorts were in fact composed of somewhat more risky borrowers though they appear to be similar based on observed characteristics discussed later in section 1.7.

The explanation for the difference between models is the following. The shock model predicts that a fraction ψ of borrowers with negative equity default each period. When the whole equity distribution shifts left due to the fall in aggregate house prices, the shock model predicts that the default rate should increase in proportion to the number of borrowers who experience negative equity. It turns out that observed default rates exhibit this pattern. But the threshold model is concerned with the (far left) tail of

Figure 1.3: Cumulative Default Rates for Loans originated in 2002-2008: Models vs. Data

(a) Threshold Model (b) Shock Model



the equity distribution. It predicts that all borrowers with an extreme level of negative equity below ϕ times the initial house price default. When the equity distribution shifts left the number of borrowers with such an extreme level of negative equity increases faster than the observed default rate. This generates the inconsistency with the data.

Two conclusions can be drawn from these results. First, an empirically successful structural model cannot rely on a single-trigger or negative equity threshold mechanism alone. Instead some shocks other than house price shocks must play a role. Second, in a double-trigger model the increase in the fraction of borrowers with negative equity caused by the mean shift in house prices is sufficient to explain the broad rise in default rates. Together with the evidence on the stability of loan characteristics presented in section 1.7 this supports a hypothesis featuring a strong explanatory role of the macroeconomic house price movements for the rise in default rates and against the pool of borrowers becoming more risky per se.

Motivated by these results, the next main section presents a structural model featuring idiosyncratic unemployment risk and liquidity constraints. This serves several purposes. First, the model aims at providing micro-foundations for the double trigger hypothesis. This means to provide conditions under which a rational agent exhibits double-trigger default behavior. Second, it allows a check of whether unemployment shocks can quantitatively play the role of the trigger events. One can also check whether the strong explanatory role of aggregate house prices survives in such a structural framework. Third, such a model can be used for policy analysis.

1.3.6 Robustness Checks

This section reports a battery of robustness checks that were performed to scrutinize these results. I find that the results are robust across all the modifications considered here. Graphs equivalent to figure 1.3 for each of the performed scenarios are available upon request.

Instead of estimating the models on the 2002 cohort with low default rates, I also estimate them on the 2008 cohort with very high default rates. This does not affect the good fit of the shock model. But now the threshold model greatly undershoots the default rates of early cohorts and also still overshoots the 2006 and 2007 cohort. Thus the comparison across models is unaffected.

Another robustness check replaces the out-of-sample test with an in-sample test. Here I estimate the two models on all cohorts and then examine the fit within that sample. The threshold model still has considerable problems to match the data. It generally undershoots earlier cohorts and the early months after origination for all cohorts and at the same time still overshoots the late months of the 2006 and 2007 cohorts. In contrast, the shock model gives an excellent fit to the data. The conclusions across models are essentially unchanged.

I also examine the role of the variation in mortgage rates and the distribution of loan-to-value ratios across cohorts in three alternative specifications. In the first specification, I keep the within cohort LTV distribution fixed across cohorts according to the average frequency. The second specification abstracts from within cohort heterogeneity such that everyone has the same LTV according to the respective within cohort average. The third specification is the same as the second except that the LTV and mortgage rate are not varied across cohorts. All these changes have very modest effects on both models and leave the conclusions across models unaffected. This implies that the double-trigger model attributes the rise in default rates to the variation in aggregate house prices and not the changes in contract characteristics across cohorts. It also suggests that abstracting from this heterogeneity across cohorts in the structural model is not too restrictive.

In section 1.2.2 it was assumed that the individual house price shocks are normally distributed. The major argument supporting this choice is that by the central limit theorem the sum of individual shocks converges asymptotically to a normal distribution anyway. But since the analysis also covers periods where t is still small, I perform an additional check here. Instead of using a normal distribution for the individual shocks I specify them as being uniformly distributed on the interval $[-b_t, b_t]$. The parameter b_t is then chosen such that the variance of the uniformly distributed shock in period t in the respective census division is identical to the one used in the standard framework. I find that the results are almost identical.

Another potential concern is that the simplicity of the presented reduced-form models with only one constant parameter somehow biased the results against the frictionless option model. There is also no strong reason why the default threshold parameter ϕ and default shock probability ψ should be constant over the course of a loan. It turns out that the results are robust to changing this assumption. As a check I have performed a scenario where the respective default parameter depends fully on the month since origination t. The constant parameters in the model are then replaced with ϕ_t and ψ_t that are allowed to differ each period from $t = 1, \ldots, T$ when fitting the models to the 2002 cohort. Under these circumstances both models use all degrees of freedom of the data and perfectly match the 2002 cohort. The cumulative default rates simulated for the other cohorts then inherit the non-smoothness of the first differences of the cumulative default rate of the 2002 cohort. But subject to that qualification the conclusions on the out-of-sample fit remain essentially unchanged. The threshold model still greatly overshoots. The shock model generates default rates of the right magnitude, but predicts slightly lower default rates for some months compared to the benchmark specification.

1.4 Structural Model

This section introduces a theoretical model of the repayment decision of a homeowner who financed the home purchase with a fixed-rate mortgage. Each period the borrower chooses non-housing consumption and whether to stay in the house and service the mortgage or leave the house and terminate the mortgage. The mortgage can be terminated either by selling the house and repaying the mortgage or defaulting on the loan by "walking away". The homeowner faces uncertainty on the future price of the house, unemployment shocks and a borrowing constraint for unsecured credit. One period corresponds to one month. Throughout this section an individual index i is suppressed for convenience.

1.4.1 Mortgage Contract

The household took out a fixed rate mortgage with outstanding nominal balance M_0 and nominal mortgage rate r^m to finance the purchase of a house of price P_0 in period 0. Mortgage interest and principal have to be repaid over T periods in equal instalments of nominal value m that are fixed at the beginning of the contract and satisfy equation (1.4). Over time the outstanding nominal mortgage balance M_t evolves according to equation (1.3) as long as the household services the mortgage.

1.4.2 Preferences and Choices

Preferences are specified as in Campbell and Cocco (2003), but allow for a direct utility benefit of owning a house. Household decisions over the length of the mortgage contract are determined by maximizing expected utility given by

$$U = \mathrm{E}_0 \sum_{t=1}^{T} \beta^{t-1} \left(\frac{C_t^{1-\gamma}}{1-\gamma} + \theta \mathcal{I}(own_t) \right) + \beta^T \frac{W_{T+1}^{1-\gamma}}{1-\gamma} \qquad (1.8)$$

which is derived from consumption C_t in periods 1 to T and remaining wealth W_{T+1} at the end of the contract. The flow utility function is assumed to be of the CRRA form where γ denotes the parameter of relative risk aversion and the inverse of the intertemporal elasticity of substitution. β is the time discount factor. $\mathcal{I}(own_t)$ is an indicator variable that is one if the agent owns a home in period t and zero otherwise. θ is a direct utility benefit from being a homeowner. This could reflect for example an emotional attachment to the house or the benefit that an owner cannot be asked to move out by a landlord as may happen to a renter. The specification of the utility function implicitly assumes that consumption and the size of the house are separable in the homeowner's utility function.[9]

[9] Following Campbell and Cocco (2003), the specification in equation (1.8) implicitly assumes that the borrower maximizes utility only over the course of the mortgage contract because the continuation value is largely arbitrary. Ideally the model should be extended to the full life-cycle and include a period after the end of the mortgage contract. This will be implemented in a future revision.

In each period the homeowner has to decide how much to consume and on staying or leaving the house. If the agent wants to leave this can be done by either selling the house (and repaying the current mortgage balance) or defaulting on the loan by "walking away".[10] It is assumed that a homeowner who leaves the house will rent a house of the same size for the rest of life.

1.4.3 Constraints

The dynamic budget constraint depends on the borrower's house tenure choice. For a homeowner who stays in the house it is given by

$$A_{t+1} = (1+r)(A_t + Y_t - \frac{m}{\Pi_t} + \tau r^m \frac{M_t}{\Pi_t} - C_t) \tag{1.9}$$

where A_t denotes real asset holdings and Y_t real net labor income in period t. The real interest rate on savings r is assumed to be constant over time. m is the nominal payment to service the mortgage. But the nominal mortgage interest $r^m M_t$ is tax deductable and τ is the tax rate. All nominal variables need to be deflated by the current price level for consumption goods Π_t to arrive at a budget constraint in terms of real variables. The presence of Π_t generates the "mortgage tilt effect". This means that due to inflation the real burden of the mortgage is highest at the beginning of the contract and then declines over time. It is assumed that the inflation rate π is constant over time and Π_t thus evolves according to $\Pi_{t+1} = (1+\pi)\Pi_t$.

In case the house is sold at the current real price P_t, the homeowner needs to repay the current outstanding nominal mortgage balance M_t and can pocket the rest. The budget constraint then reads as

$$A_{t+1} = (1+r)(A_t + Y_t - R + P_t - \frac{M_t}{\Pi_t} - C_t). \tag{1.10}$$

Here R is the real rent for a property of the same size. It is assumed that an agent who terminates the mortgage through prepayment or default needs to rent for the rest of life.[11] Real rents are assumed to be proportional to the initial house price and then

[10]The model does not include a mortgage termination through refinancing for computational reasons. Otherwise the mortgage balance becomes a separate state variable. This is unlikely to be a major limitation because refinancing is only feasible when the borrower has positive equity in the house. Thus it does not directly compete with the default decision in a negative equity situation.

[11]Thus a change of housing status from owning to renting is irreversible. This assumption simplifies the computational solution of the model, but could also be a potential limitation. The assumption

constant over time as

$$R = \alpha P_0. \tag{1.11}$$

This specification involves both a highly realistic feature of rents and an approximation. The realistic feature is that during the period of study real rents remained almost constant, while real house prices first increased and then decreased enormously. The specification implies that after origination the rent-price ratio decreases when real house prices increase. Such a negative relationship between the rent-price ratio and real house prices exists in the data provided by Davis, Lehnert, and Martin (2008) not only during the recent period, but at least since 1975. In this paper I take these observations as given and specify the exogenous variables of the model accordingly. But explaining this pattern is an important area for future research. However a fully realistic specification would also require to make α cohort-specific. But I use an approximation for computational reasons such that α is constant across cohorts and calibrated to a suitable average.

In contrast, if the agent decides to default on the mortgage by "walking away" or is already a renter the budget constraint is given by

$$A_{t+1} = (1 + r)(A_t + Y_t - R - C_t). \tag{1.12}$$

It is assumed that for reasons not explicitly modeled here the household faces a borrowing constraint for unsecured credit given by

$$A_{t+1} \geq 0. \tag{1.13}$$

Together with the budget constraints above this implies that the amount of resources available for consumption in a period depend on the house tenure choice.

Remaining wealth at the end of the contract for a homeowner is given by $W_{T+1} = A_{T+1} + Y_{T+1} + P_{T+1}$ and for a renter by $W_{T+1} = A_{T+1} + Y_{T+1}$.

that after selling or defaulting the household rents a property of the same size is also a possible limitation. This prevents downsizing of the house after a default which could play an important role in the default decision of borrowers in the real world. However I have also experimented with a fixed and exogenous downsizing factor after mortgage termination and this left the explanatory power of the model largely unchanged or even improved it a bit.

1.4.4 Labor Income Process

The household's real net labor income Y_t is subject to idiosyncratic unemployment shocks and exogenously given by

$$Y_t = \begin{cases} (1-\tau)Y_0 & \text{if employed} \\ \rho(1-\tau)Y_0 & \text{if unemployed} \end{cases} \tag{1.14}$$

where Y_0 is initial real gross income, τ is the tax rate and ρ is the net replacement rate of unemployment insurance. Over time employment status evolves according to a Markov transition process with the two states "employed" and "unemployed" and constant job separation and finding probabilities. Employed agents lose their job with probability s and stay employed with probability $(1-s)$. Unemployed agents find a job with probability f and stay unemployed with probability $(1-f)$.

There are several reasons why I focus on income fluctuations due to unemployment risk here. First, unemployment involves a severe fall in labor income from one month to another. This makes it a very plausible cause for short run liquidity problems. Second, other frequently used specifications of income processes as for example in Campbell and Cocco (2003) are typically calibrated for yearly frequencies. Thus, they are not directly applicable to a monthly framework. In any case, most of the income variation from month to month probably comes from unemployment spells and it therefore seems preferable to use such a process explicitly. Third, this allows the model to be related more closely to the double-trigger hypothesis and the empirical literature that has provided evidence that default is correlated with state unemployment rates.

I also abstract from deterministic changes to labor income like a life-cycle profile and keep the labor income of employed and unemployed agents constant over time. The reason is that I do not have any demographic information on the borrowers in my data set.

1.4.5 House Price Process

Real house prices are exogenous and evolve over time as specified in section 1.2.2 and equation (1.1). It is assumed that homeowners view the aggregate component g_t^{agg} of house price appreciation to be stochastic and distributed according to an i.i.d. normal distribution with mean μ and variance σ^2. This process for the aggregate house price component is only used for forming agents' expectations. In the simulation the realizations of g_t^{agg} are those observed in the data. For the individual component agents know that g_t^{ind} is distributed normally with mean zero and time-varying variances that depend on the parameters κ and λ as specified in section 1.2.2. In order to reduce the computational burden when computing policy functions the parameters μ, σ, κ and λ are not varied across the nine census divisions. Instead they are set equal to national averages, cf. section 1.5.2 on the calibration. But the realizations in the simulation of the model of course come from the division specific data and distributions.

1.4.6 Initial Conditions

The homeowner solves the dynamic stochastic optimization problem conditional on initial asset holdings A_0, initial employment status, an initial loan-to-value ratio $LTV = \frac{M_0}{P_0}$ and a debt to (gross) income ratio $DTI = \frac{m}{Y_0}$.[12] The initial house price P_0 is normalized to 100. LTV and DTI then uniquely determine M_0 and Y_0.

1.4.7 Computation

The borrower's optimization problem is characterized by four state variables (liquid wealth $X_t = A_t + Y_t$, employment status L_t, house price P_t and time t) and two choice variables (consumption C_t and the mortgage termination choice). Note that for a fixed-rate mortgage the mortgage balance M_t evolves deterministically over time and is thus captured by the state variable t. The solution proceeds backwards in time. The continuous state and control variables are discretized and the utility maximization problem in each period is solved by grid search. Expected values of future variables are computed by Gaussian Quadrature. Between grid points the value function is evaluated using cubic interpolation.

1.4.8 Model Simulation

The model presented above is a dynamic stochastic partial-equilibrium model that maps contract characteristics at origination and realizations of the stochastic processes

[12]The name debt to income ratio is part of standard mortgage terminology, but can be easily misunderstood. It means the ratio of the monthly mortgage payment to gross income.

for house prices and employment status into default decisions. I simulate the model for subsequent cohorts of loans originated each month between January 2002 and December 2008 from the respective origination month until June 2010. For each cohort I draw $20,000$ individual house price and employment histories with the same length in months as the respective cohort is observed in the data. House price histories are drawn as explained in section 1.2.2 and employment histories are drawn from the two-state Markov process specified in section 1.4.4.

Accordingly, within a cohort borrowers face the same aggregate house price movements (except for the differences between census divisions), but different individual house price and employment shocks. Differences between cohorts are generated from different paths of aggregate house prices depending on the date of origination.

1.5 Parametrization

The structural model is parameterized in two steps. First the mortgage contract, house price expectations, rents, labor income, interest and inflation rates are calibrated to data on the respective variables, i.e. to data other than default rates. Then due to identification concerns the preference parameters are divided into a set that is calibrated ad-hoc and another that is estimated such that the model fits the cumulative default rates of the 2002 loan cohort. All parameter values are summarized in table 1.1 below. The model is solved at a monthly frequency. But a few parameters are presented at their yearly values if it is more convenient for comparison.

1.5.1 Contract Characteristics

This paper restricts attention to 30-years ($T = 360$ months) fixed-rate mortgages. I use average characteristics at origination of the loans in my data set to determine the loan-to-value ratio, mortgage rate and debt-to-income ratio. The average initial loan-to-value ratio of these loans is 98.2%, so I set $LTV = 98.2\%$. The nominal mortgage rate r^m is set to 6.4% per annum which is the average mortgage rate for newly originated loans in my data set. The debt-to-income ratio DTI is set to 40% as in the data.[13] Naturally, all of these parameters could be changed in order to model different mortgage contracts.

1.5.2 House Price Expectations

As explained before, when computing policy functions the parameters μ, σ, κ and λ are not varied across the nine census divisions. Instead they are set according to national averages in order to reduce the computational burden. The monthly house price index from the FHFA at the national level between 1991 and 2010 deflated by the Consumer Price Index (CPI) is used to estimate the parameters μ and σ of the aggregate component. I find that at a monthly frequency $\mu = 0.065\%$ and $\sigma = 0.55\%$. These values imply expected yearly aggregate real house price growth of 0.8% and a yearly standard deviation of 1.9%. This calibration procedure implies that agents in the model have expectations on real aggregate house price growth that on average were correct in the years 1991 to 2010 as far as the mean and standard deviation are concerned.

The parameters κ and λ are determined as a simple average of the ones estimated by the FHFA for each of the nine census divisions. This gives $\kappa = 0.00187$ and $\lambda = -4.51E{-}6$ and implies that the individual house price growth shock g_{it}^{ind} in the first month after house purchase is expected to have a standard deviation around 2.5%.

1.5.3 Income Process

The average tax rate τ is set to 16% and the net replacement rate of unemployment insurance ρ to 62%. This is based on the OECD Tax-Benefit calculator for the United States. Specifically, the average loan amount, mortgage rate and debt-to-income ratio are used to determine the average gross income of the borrowers in the data set. Based on gross income the calculator reports the net income in work and out of work which

[13] The data on the DTI is the only mortgage variable in the whole paper that is based on a somewhat different loan selection. The reason is that the DTI was not available in the tool that was used to aggregate and extract information from the LPS loan-level data set. Instead LPS provided me with a separate tabulation where it was not possible to use the same selection criteria. Specifically, the DTI information is for the same LTV class as the rest of the data, but it does not only cover prime, fixed-rate, 30-years mortgages. However the vast majority of loans in the LPS data are prime, fixed-rate mortgages and the modal maturity of these loans is 30 years, so this information should at least be a good approximation to the actual loan pool I consider.

then determine the average tax and net replacement rates. These calculations take taxes, social security contributions, in-work and unemployment benefits into account. Precise numbers especially for the tax rate also depend on the demographics of the household. I have used the average values for a married couple with one earner and no children.

Data from the Bureau of Labor Statistics on the national unemployment rate and median unemployment duration are used to compute time-series of monthly job finding and separation probabilities. This is done using steady state relationships. Since the data on median unemployment duration is reported in weeks, I first transform it to months by multiplying the weekly value by 12/52. The resulting median duration d in months is of course in general not an integer value. Given that I operate in discrete time I use an approximation to the relationship between median duration and the monthly finding probability f in steady state given by

$$(1-f)^{\underline{d}-1}f + (d-\underline{d})(1-f)^{\underline{d}}f = 0.5 \tag{1.15}$$

where $\underline{d}$ is the next integer number lower than or equal to d. If the median duration in months is an integer value then the second term in equation (1.15) is zero. If it is not an integer value then the second term gives an approximation to the number of unemployed who find a job between month $\underline{d}$ and $\underline{d}+1$ for a given finding probability f.

The steady state relationship between the unemployment rate u and job finding probability f and job separation probability s in the flows approach to unemployment is well known and given by

$$u = \frac{s}{s+f}. \tag{1.16}$$

Equations (1.15) and (1.16) are then solved for the time-series of f_t and s_t implied by the time-series of the unemployment rate u_t and median duration d_t.[14]

I then set $s = 1.8\%$ and $f = 31\%$ which are the average values of the computed monthly finding and separation probabilities for the period from 1990 to 2010. These values imply a steady state unemployment rate around 5.7%.

[14] As a check on this procedure I predict the unemployment rate from the dynamic equation of unemployment $u_{t+1} = u_t + s_t(1-u_t) - f_t u_t$ using the computed time series of finding and separation probabilities as inputs. It turns out that this gives an excellent fit to the path of the actual unemployment rate.

1.5.4 Other Prices

Nominal interest rates for 1-year Treasuries and changes to the Consumer Price Index (CPI) are used to compute real interest rates and inflation rates. Based on this data between 1990 and 2010 the real interest rate r is set equal to 1.4% and the inflation rate π equal to 2.7% on an annual basis. The initial rent-price ratio parameter α is set equal to 4.0% on a yearly basis which is the average rent-price ratio between 2002 and 2008 in the data provided by Davis, Lehnert, and Martin (2008).

Table 1.1: Model Parameters

Contract characteristics	Contract length in months	T	360
	Mortgage rate (yearly)	r^m	6.4%
	Initial loan-to-value ratio	LTV	98.2%
	Initial debt-to-income ratio	DTI	40%
House price process	Mean of aggregate component	μ	0.065%
	Standard deviation of aggregate component	σ	0.55%
	Linear coefficient in individual component	κ	0.00187
	Quadratic coefficient in individual component	λ	-4.51E-6
Income process	Job separation probability	s	1.8%
	Job finding probability	f	31%
	Tax rate	τ	16%
	Net replacement rate of unemployment insurance	ρ	62%
Other prices	Real interest rate (yearly)	r	1.4%
	Inflation rate (yearly)	π	2.7%
	Rent-price ratio (yearly)	α	4.0%
Preferences	CRRA coefficient	γ	4
	Discount factor (yearly)	β	0.9
	Utility benefit of owning	θ	0.18

1.5.5 Initial Conditions

Initial assets and employment status are unobserved. But it seems reasonable that borrowers were employed when they got their loan, so I assume that. With respect to initial assets A_0, I use the computed policy functions to set initial assets equal to the buffer-stock desired by a borrower in period 1 who is employed and faces a house value equal to P_0. Thus I shut down possible effects from borrowers first converging to their desired buffer-stock and being more vulnerable to income shocks during the time immediately after origination.

1.5.6 Preferences

Ideally the three preference parameters β, γ and θ would all be estimated such that the model gives the best fit to the data on default rates. But it is well known that dynamic discrete choice models are not fully identified, cf. the discussion and references in Magnac and Thesmar (2002). Furthermore, given the complexity of the model estimating several parameters would be computationally costly. Faced with this situation I decide to calibrate the parameters β and γ ad-hoc and estimate only θ. I also investigate how much the results depend on the specific choice of β and γ.

The parameters β and γ appear in most dynamic economic models and estimating them is the aim of a vast empirical literature. But unfortunately these empirical studies have not produced reliable estimates. For the discount factor β on a yearly basis the survey of Frederick, Loewenstein, and O'Donoghue (2002) shows that empirical estimates cluster over the full range between 0 and 1. For the intertemporal elasticity of substitution, which is the inverse of γ, Guvenen (2006) reviews empirical estimates ranging from around 1 to 0.1, which implies values of γ ranging from 1 to 10.

My impression is that many economists regard values of β below 1, but not too much below 1, and values of γ between 1 and 4, possibly even up to 10, as reasonable. But strong views on specific parameter values are probably not warranted given the empirical evidence. The large variation in estimates could also reflect that preferences are not stable across choice situations and individuals. With respect to the intertemporal elasticity of substitution Guvenen (2006) argues that conflicting estimates can be reconciled if the rich have a high and the poor have a low elasticity. I follow his argument and since the average borrowers in my data set belong to the lower half of the income distribution, I set a relatively high value of $\gamma = 4$. This implies an intertemporal elasticity of substitution of 0.25. For β I choose a value of 0.9 at a yearly frequency in order to be below, but still close to 1. Compared with assumptions in many macroeconomic studies this might appear as a low value. But adapting Guvenen's argument to β, this does not necessarily conflict with other studies. The reason is that I am analyzing a particular pool of borrowers who are not rich and were only able to make a very small down-payment. This could be due to the fact that they are very impatient. The other agents in the economy who are net savers and lenders could then have a higher discount factor more in line with the macro literature. In any case these are only the benchmark values and I also investigate the sensitivity of the main results to these parameter choices.

Given values of β and γ, the preference parameter θ representing the direct utility benefit from owning the house is estimated by the simulated method of moments. The procedure is identical to the one used earlier for the reduced-form models, cf. section 1.3.4. Again the parameter is chosen such that cumulative default rates simulated from the model match those observed in the data using only information from loans originated in 2002. This yields an estimated value for θ of 0.18. The remaining data is used to test the ability of the estimated model to predict out of sample.

1.6 Results

This section explains the repayment policy function of a homeowner and the basic mechanism generating default over the life-cycle of a loan in the model. Then the main results how well the model fits the rise in default rates across loan cohorts are presented. Finally, a sensitivity analysis explores how the model depends on certain preference parameters.

1.6.1 The Repayment Policy Function

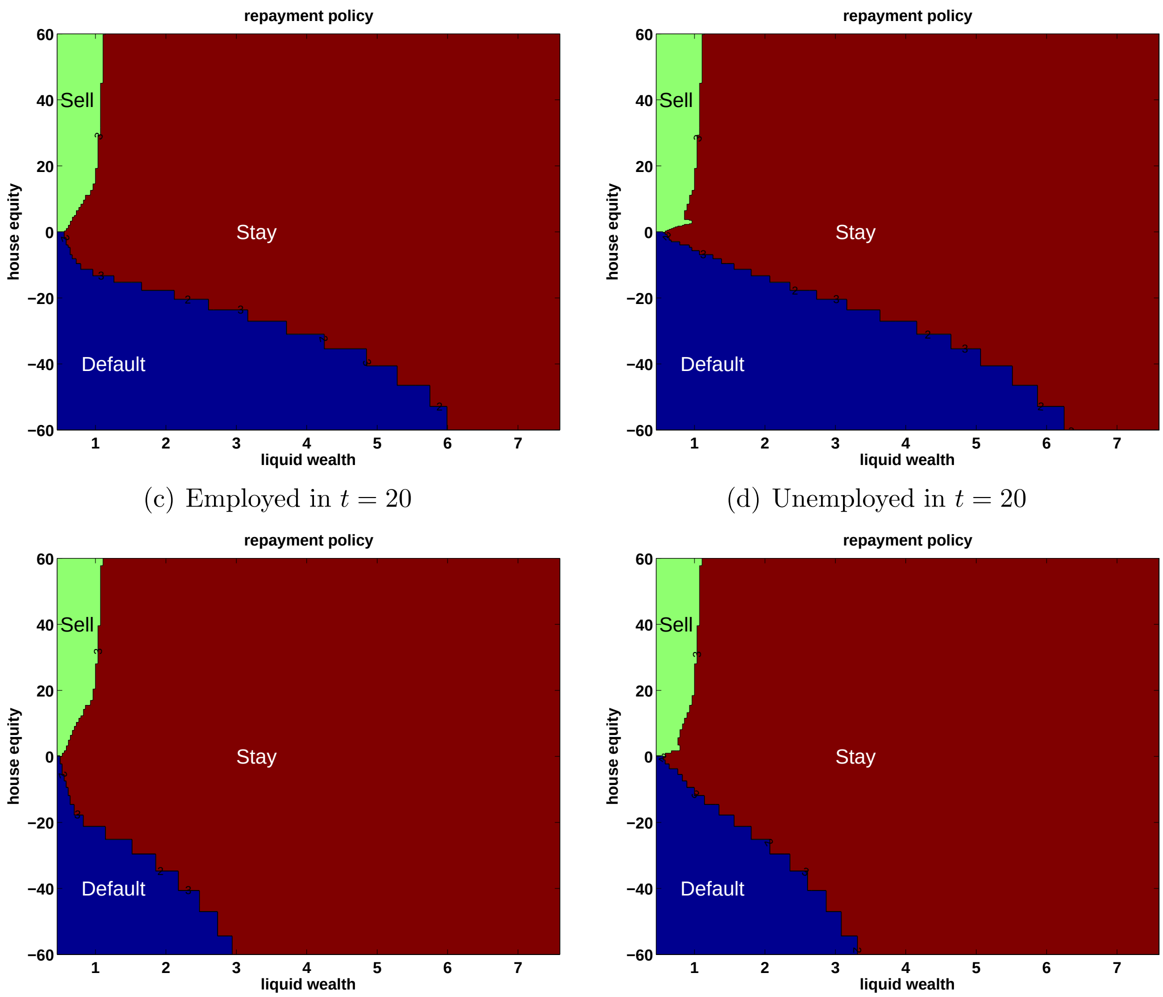
The repayment policy function of a borrower in the model is presented in figure 1.4 as a function of house equity, liquid wealth, employment status and time. Several features are noteworthy. First, negative equity is a necessary condition for default. Instead, with positive equity selling is strictly preferred to defaulting because the borrower is the residual claimant of the house value after the mortgage balance has been repaid.

Second, negative equity is not sufficient for default. There are many combinations of state variables where a borrower with negative equity prefers to stay in the house and service the mortgage. In a negative equity situation the basic trade-off of the borrower is the following (postponing the role of the borrowing constraint until the next paragraph). The cost of staying in the house is that the borrower needs to make the mortgage payment, which is higher than the rent for an equivalent property. The

Figure 1.4: Repayment Policy Function

(a) Employed in $t = 1$ (b) Unemployed in $t = 1$

(c) Employed in $t = 20$ (d) Unemployed in $t = 20$



Notes: Repayment choice as a function of the state variables liquid wealth, house equity, employment status and time. Blue region: Default. Green region: Sell. Red region: Stay.

benefit of staying is that the borrower receives the utility benefit of owning a house and keeps the option to default, sell or stay later. Specifically, there are possible future states of the world with positive equity. But the probability of reaching these states depends on the current house price. This establishes a default threshold level of the house price. Of course, when making this decision the rational borrower will also need to discount these future gains and take risk aversion into account.

Third and importantly, the level of negative equity at which the borrower exercises the default option depends on non-housing state variables: liquid wealth and employment status. Specifically, a borrower who is unemployed and/or has low liquid wealth will default at lower levels of negative equity. There are two reasons for terminating the mortgage in these states. One is that current borrowing constraints may bind and the borrower terminates the mortgage to increase current consumption. The other reason is that in these states it becomes very likely that borrowing constraints bind in the future and the agent is forced to terminate the mortgage then. But an anticipated future mortgage default creates an incentive to default already today to save the difference between the mortgage payment and the rent in the meantime. This also explains why unemployment, which is persistent, shifts the default frontier to the right.

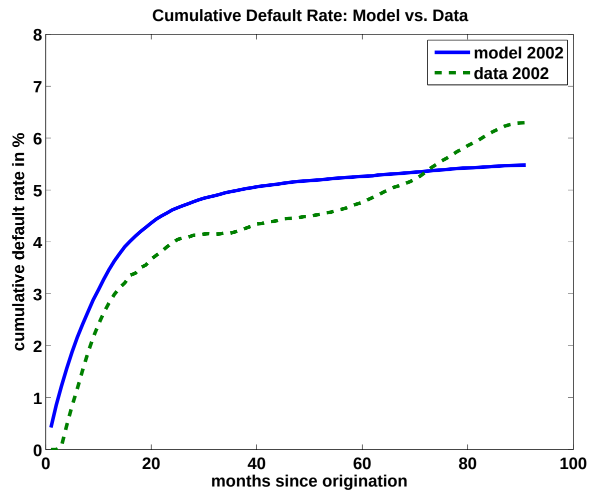
Fourth, over time the default region shrinks. This is mainly due to the effect of inflation that diminishes the real difference between the effective mortgage payments and rents. This has two implications. First, a liquidity constrained borrower cannot increase current consumption much by a mortgage default. Second, staying in the home eventually dominates renting in all states because the real value of the mortgage payment falls below the real rent.

1.6.2 Default over the Loan Life-Cycle

In this section I compare model results and data on the cohort of loans for which I have the longest time dimension in order to get an impression of default behavior over the life-cycle of a loan. Figure 1.5 presents the average cumulative default rate for loans originated in 2002. This is the cohort on which the model is estimated. Accordingly, the dynamics of default over the life-cycle of this cohort are captured relatively well by the model. But the model predicts too many defaults in the first months after origination and too few in the very late months. I will discuss the reasons for this in more detail in the next section.

Though this cohort faces growing average house prices during the immediate time after origination as seen in figure 1.1(b), some individuals experience falling house prices and negative equity as a consequence of individual house price shocks. Households with negative equity default when prolonged stretches of unemployment have exhausted their buffer stock savings, cf. the default region of the state space in figure 1.4. Eventually, the cumulative default rate levels off due to two reasons. First, borrowers who are still active have amortized their mortgages sufficiently such that most have positive equity. Second, due to the mortgage tilt effect the difference between the real mortgage payment and real rents shrinks over time such that a default becomes less appealing.

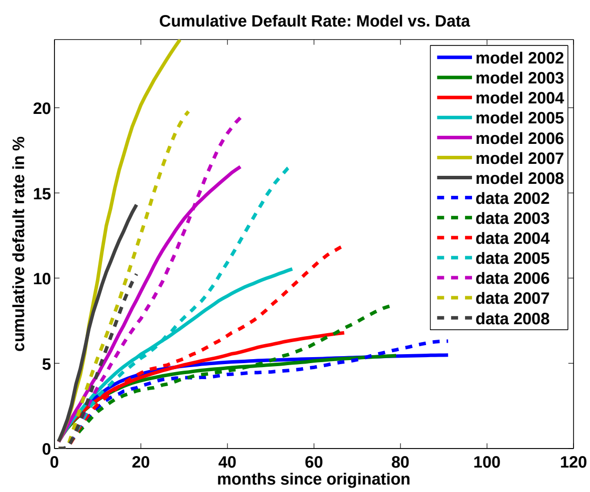
Figure 1.5: Cumulative Default Rates of 2002 Cohort: Structural Model vs. Data



1.6.3 The Rise in Cumulative Default Rates

The next step is to compare the default behavior of different cohorts during the time period of the U.S. mortgage crisis. Figure 1.6 presents average cumulative default rates for cohorts of loans originated each year between 2002 and 2008.

Figure 1.6: Cumulative Default Rates of 2002-2008 Cohorts: Model vs. Data



When average house price appreciation slows down and eventually becomes negative as witnessed in figure 1.1(b) a higher fraction of borrowers experience negative equity which translates into more frequent default. The model can explain the broad pattern in the data and attributes the rise in cumulative default rates across cohorts to the different aggregate house price paths. The model is particularly successful in the early months after loan origination, but has problems to explain default in later months. In the model this is due to the effect of inflation, the mortgage tilt effect. This effect diminishes the difference between real mortgage payments and rents over time. The model is sensitive to this difference and reacts too strongly compared to the data. It is also noteworthy that the model inflation rate is constant and calibrated to the average inflation rate between 1990 and 2010 which is 2.7%. But in the final years of the simulation period inflation was much lower. For example on average between 2008 and 2010 it was 1.4% with 0.1% in 2008, 2.7% in 2009 and 1.5% in 2010. It is likely that the model would perform better for these actual inflation rates.

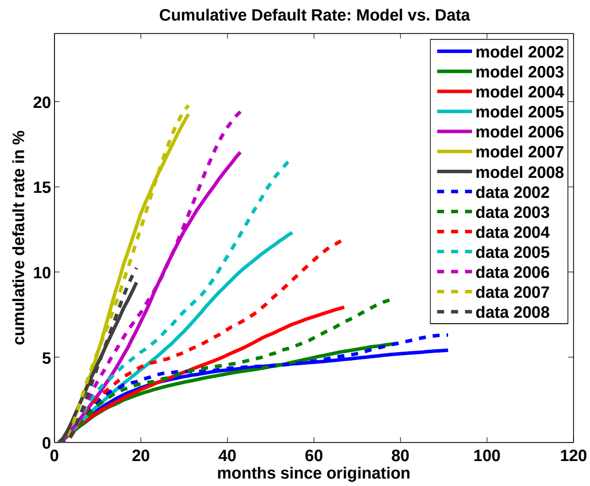
1.6.4 Role of Inflation

In this section I confirm that the role of inflation in the model and how I calibrated it are responsible for the poor performance of the model during periods long after origination. I simply change the inflation rate π ad-hoc to 1% instead of 2.7% in the benchmark calibration. All other parameters are unchanged, but θ is reestimated at a value of 0.33 to fit the 2002 cohort. Figure 1.7 presents these results. The fit of the model improves and is now comparable to the one of the reduced-form double-trigger model, cf. figure 1.3(b).

There are at least two possible ways to interpret the results in sections 1.6.3 and 1.6.4 on the role of inflation for the fit of the model. One possibility is that in the real world borrowers do not fully understand or underestimate the effect of inflation. This could be the reason why the model with a rational agent does not explain default so well in periods long after origination. It could also be that moving away from policy functions that are conditional on a constant inflation rate would improve the fit of the model.

The other possible interpretation is that unemployment and liquidity problems are not able to explain default in periods long after loan origination. Instead other reasons like marital break-up that were excluded from the structural model could be responsible for default in these periods. This paper only analyzes whether and how unemployment shocks could act as the trigger event in a structural model and found that they could definitely play an important role. But assessing the role of other life events and a

Figure 1.7: Performance of the Model for a low Inflation Rate



decomposition of actual default rates into the different causes within the double-trigger paradigm is an important area for future research.

1.6.5 Dependence on Preference Parameters

All results from theoretical models depend in some way on parameters and the model presented here is no exception. Unfortunately, it is not easy to provide an exact characterization of the parameter space for which the agents in the model exhibit double-trigger default behavior because of the lack of a closed-form solution. But this section computes results for some examples of alternative parameter values for β and γ in order to get an idea how the model behaves in different parts of the parameter space.

The benchmark preference parameter values are $\beta = 0.9$ and $\gamma = 4$. Here I consider all combinations of $\beta \in \{0.85, 0.9, 0.95\}$ and $\gamma \in \{2, 4, 6\}$. For each of these (β, γ)-combinations the parameter θ is reestimated in order to fit the 2002 cohort. All other parameters are as in the benchmark calibration. The resulting values of θ for all combinations of β and γ are presented in table 1.2.

The results for the different parameter combinations are presented in figure 1.8. The graphs are ordered such that γ increases horizontally from 2 (left) to 6 (right) and β increases vertically from 0.85 (top) to 0.95 (bottom). These results show that the model

Table 1.2: Dependence of the estimated Value of θ on β and γ

	$\gamma = 2$	$\gamma = 4$	$\gamma = 6$
$\beta = 0.85$	0.09	0.27	0.50
$\beta = 0.90$	0.05	0.18	0.36
$\beta = 0.95$	-0.02	0.07	0.20

works as well or better than in the benchmark calibration for higher values of γ and/or lower values of β. These parameter changes make the agent less willing to substitute intertemporally and/or more impatient to consume today. This worsens the liquidity problem caused by unemployment. The model can only feature double-trigger behavior when being employed and being unemployed are sufficiently different. In contrast, for lower values of γ and higher values of β temporary income reductions can more easily be smoothed out. The model then implies that a sizeable portion of employed agents default in all cohorts. This brings the model close to a frictionless option model and the model then inherits all the problems of such a specification witnessed already in section 1.3.

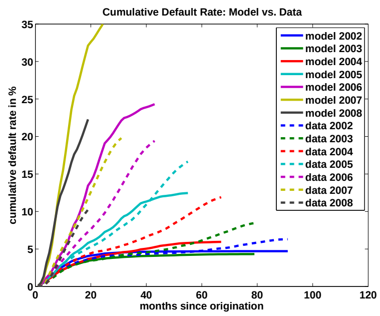
1.7 Discussion of an Alternative Explanation

All mortgage default theories hypothesize that default by a borrower is a function of the house price. This paper has presented further evidence that supports this view. However there is a competing explanation in the public and academic debate for the rise in default rates observed in figure 1.1(a). This explanation is that lending standards deteriorated sharply before the mortgage crisis. If this were true then the increase in mortgage default rates across cohorts could be due to a worsening of the loan quality. This would then also confound the empirical relationship between default rates and house prices that I use to test mortgage default theories. Thus, this section presents evidence that loan quality is fairly stable across cohorts in my data set.

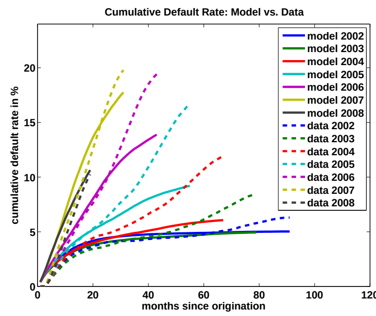
First of all I only look at data on prime fixed-rate mortgages. Therefore a shift towards more risky lending as far as it manifests itself in a shift from prime to subprime lending or from fixed to variable rate or hybrid mortgages is ruled out by construction. These compositional effects might or might not be significant contributors to the overall mortgage crisis, but they do not affect my analysis. We see clearly from

Figure 1.8: Sensitivity to Preference Parameters β and γ

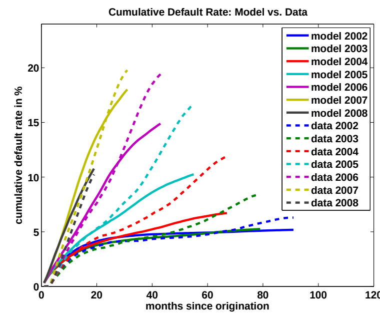
(a) $\beta = 0.85$ & $\gamma = 2$



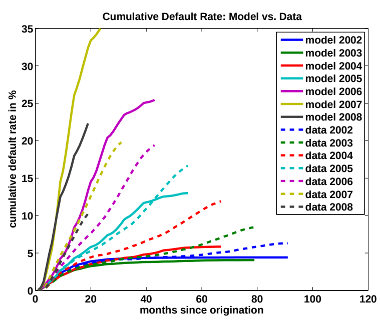
(b) $\beta = 0.85$ & $\gamma = 4$



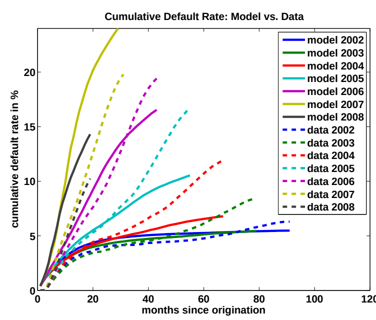
(c) $\beta = 0.85$ & $\gamma = 6$



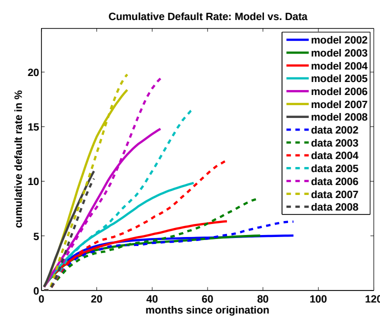
(d) $\beta = 0.9$ & $\gamma = 2$



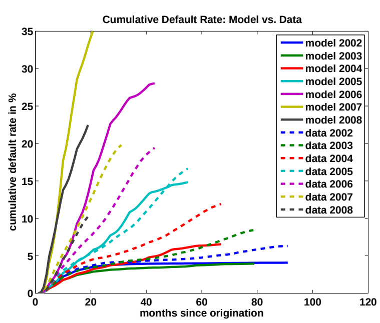
(e) $\beta = 0.9$ & $\gamma = 4$



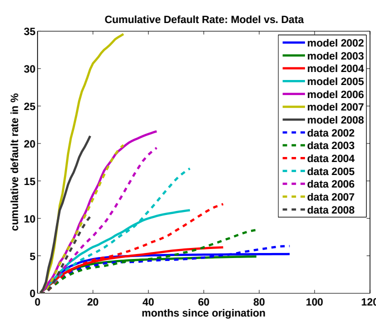
(f) $\beta = 0.9$ & $\gamma = 6$



(g) $\beta = 0.95$ & $\gamma = 2$



(h) $\beta = 0.95$ & $\gamma = 4$



(i) $\beta = 0.95$ & $\gamma = 6$

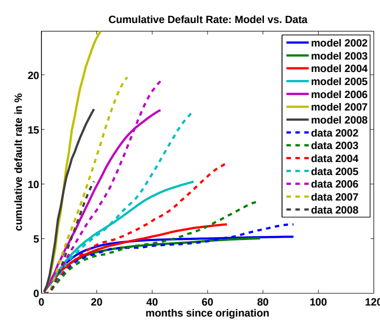


figure 1.1(a) that even without such compositional effects mortgage default rates have increased substantially.

Another concern is that the loan-to-value ratio (LTV) might have increased over time leaving a smaller buffer before borrowers experience negative equity. I only consider loans that have a LTV above 95% and thus limit this possibility to shifts within that class of loans. Within this class the average LTV is basically constant across cohorts and only fluctuates mildly around the average value of 98.2% as seen in the first row of table 1.3. In the reduced-form models I even controlled for changes to the distribution of LTVs and found that the observed changes are irrelevant for the models considered here.

The second row of table 1.3 reports the average FICO credit score at origination of the different loan cohorts. These are very stable as well. To the extent that these credit scores are good measures of creditworthiness a significant deterioration in loan quality is not observable here.

Table 1.3 also contains information on the average mortgage rate that different cohorts face. A higher mortgage rate might make the loan as such less attractive to the borrower. There is some variation in this variable across cohorts. But the mortgage rate and default rates seem to be fairly uncorrelated across cohorts.

The average debt-to-income (DTI) ratio representing the share of the required mortgage payment in gross income is presented in the last row of table 1.3.[15] This has increased over time indicating that borrowers in later cohorts need to devote more of their gross income to service the mortgage. But the increase was quite modest.

Table 1.3: Average Loan Characteristics at Origination by Loan Cohort

Cohort	2002	2003	2004	2005	2006	2007	2008	Average
Loan-to-value ratio in %	98.2	98.3	98.2	98.3	98.4	98.1	97.8	98.2
FICO credit score	676	673	669	670	668	670	678	672
Mortgage rate in %	6.9	6.0	6.1	6.0	6.6	6.7	6.2	6.4
Debt-to-income ratio in %	39	40	40	40	40	42	42	40

These statistics show that there is no strong evidence in favor of a deterioration of lending standards over time in my data set of prime fixed-rate mortgages with a LTV above 95%.[16] I conclude that this loan pool and time period indeed constitute a good testing ground for mortgage default theories.

[15] Footnote 13 also applies here.

[16] This conclusion might be specific to the prime market. For example Demyanyk and Van Hemert (2011) present evidence that loan quality deteriorated in the subprime market. But Amromin and Paulson (2009) also note that it is less obvious that a similar deterioration was present in the prime market. A particular advantage of my descriptive statistics is that they are based on all loans in the LPS data base satisfying my sample selection criteria. In contrast, other empirical studies using LPS data typically work with a 1% random sample such that their descriptive statistics are based on far fewer observations.

One limitation of the paper is that it does not fully control for variation in contract characteristics across and within cohorts for computational reasons and the fact that I only have aggregate data. The evidence presented here suggests that this is not a major limitation because the different origination characteristics are quite stable. The reduced-form models also took variation of the mortgage rate and LTV distribution across cohorts into account and found that it cannot explain the rise in default rates. It would be interesting to extend my framework in future research such that one can analyze how contract characteristics affect default rates within cohorts.

1.8 Analysis of two Bailout Policies

This section discusses an application of the presented structural model for policy analysis. I study a situation where the government is concerned about a destabilization of the financial system due to the losses that mortgage lenders incur from mortgage default. Assume that the government decides to neutralize all these losses by a suitable bailout policy. The question is then: Should the government bail out lenders or homeowners?

In case lenders are bailed out the government needs to cover the negative equity of defaulters, i.e. by how much the outstanding mortgage balance exceeds the value of the collateral. In contrast, the government could also give subsidies to homeowners who would otherwise default such that they continue to service the mortgage. This policy might well be cheaper because homeowners are willing to accept some negative equity and thus bear some of the losses on the house value unless they face severe liquidity problems. The subsidies then only have to overcome the temporary liquidity shortage to neutralize the losses for lenders. However it is also possible that subsidizing homeowners simply delays default to a later period such that the subsidy policy ends up being more expensive in the long run. These opposing effects make a quantitative analysis desirable.

The two policies are compared by calculating the average cost per borrower who would default in absence of an intervention. For the bailout of lenders this simply amounts to the average negative equity of a defaulter which can readily be computed during the simulation. For the subsidy to homeowners one needs to modify the standard simulation procedure. Each period default decisions of borrowers given their liquid wealth, negative equity and other state variables are determined. Then for each potential defaulter the subsidy required to make the borrower stay in the house is computed. When doing this the standard policy functions are used. This means borrowers will consume out of the subsidy, but further negative incentive effects are ruled out. The total sum of all subsidies to a cohort is divided by the number of defaulters without any intervention to make it comparable to the other bailout policy. The required real payment streams of both policies are compared by calculating present discounted values using the real interest rate r.

In order to account for the delayed default effect of the subsidy policy it is important to follow a cohort up to the point where the model does not predict any more default. Therefore this analysis will only be done for the 2002 cohort with the longest time horizon. Of course, this calculation can only be as accurate as the model captures actual default behavior. Since by construction the model explains the 2002 cohort relatively well this is an additional reason to focus on it. I find that bailing out lenders implies average real present discounted costs of 5.82% of the initial house price per borrower who defaults. In contrast subsidizing homeowners on average only costs 0.52% of the initial house price in real present discounted value terms. Bailing out lenders is thus 11 times more expensive than subsidizing homeowners. This is a huge difference.

A couple of comments on these results are in order. First, these are partial-equilibrium results. But it seems that general equilibrium effects of subsidizing homeowners would also be more favorable because keeping borrowers in their houses avoids downward pressure on house prices due to foreclosure sales. Second, both homeowners and lenders would probably prefer the subsidy to homeowners because borrowers like to stay in their houses and lenders do not have to deal with foreclosures and housing sales which will cause additional administrative costs for them. Finally, in reality one would of course need to take negative incentive effects into account. While both policies have negative incentive effects on lenders, the bailout of homeowners would also have negative incentive effects on borrowers. The subsidy could for example make unemployed borrowers more reluctant to accept a new job and prolong their unemployment spells which would make the subsidy less favorable. Investigating the quantitative role of this effect by including an explicit job search with endogenous job acceptance into the model is an interesting avenue for future research. There might also exist practical problems of implementing a subsidy to homeowners in a fashion as assumed here.

But one feasible policy could be to increase unemployment benefits for unemployed mortgage borrowers during a mortgage crisis such that they have enough resources to continue their mortgage payments. In any case these calculations show that there is potential for improving on policies that simply bail out the lenders both in terms of costs to taxpayers, but possibly also in terms of what lenders and borrowers would prefer.

1.9 Extension to lower Loan-to-Value Ratios

So far the paper focussed on loans with a LTV above 95% because these borrowers should be least likely to have a second mortgage on their home, cf. the discussion in section 1.2.1. The question arises whether the results of the paper also generalize to loans with a lower LTV. This section provides some evidence on this by repeating the reduced-form analysis of section 1.3 for loans with a LTV of the first mortgage between 75% and 84%. Due to the discussed data problems this section is necessarily somewhat tentative. Nevertheless, some very interesting results emerge.

First I take the data for the loans with a LTV of the first mortgage between 75% and 84% at face value and assume that no one has a second mortgage. Accordingly the LTV varies within cohorts in steps of one percentage point between 75% and 84%. Changes to the distribution of loans over this support across cohorts observed in the mortgage data are again taken into account. The mortgage rate is again kept constant within a cohort and set equal to the respective cohort average. When estimating the models on the 2002 cohort I find that neither of the two models can capture this data well. Both models undershoot the cumulative default rate even for the most extreme parameter values where $\phi = 0$ and $\psi = 1$. The reason is that the equity buffer generated by the down-payment is substantial for these borrowers. Because the 2002 cohort faced strongly increasing average house prices immediately after origination, too few borrowers in the simulation experience negative equity compared to observed default rates. It is important that both models fail if we take this data at face value. One can draw two possible conclusions from these results. Either we need a completely new theory of default for these loans or it is crucial to take second mortgages into account. I present evidence on the second explanation next.

Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010) report that 26% of all borrowers have a second mortgage and this adds on average 15% to the combined LTV. But they neither report a break-down of these statistics by the LTV of the first mortgage nor when borrowers take out the second mortgage. Faced with this situation I model a very simple form of intra-cohort heterogeneity taking these estimates of the frequency and size of second mortgages into account. I assume that 74% of borrowers have only one mortgage with a distribution of LTVs as in the mortgage data. But 26% of borrowers in each cohort independently of the LTV of the first mortgage also have a second mortgage adding 15% to the combined LTV. This implies that the support of the LTV distribution is expanded and also includes values between 90% and 99%. It is assumed that borrowers got the second mortgage at the same time as the first one and pay the same mortgage rate on both. Admittedly, these are very crude assumptions. This exercise can only provide preliminary evidence until better data is available and should be regarded with considerable caution.

For this setup the reduced-form models are estimated again on the 2002 cohort. This yields estimates of $\phi = -7.7\%$ and $\psi = 2.4\%$. The estimated models are again tested on their ability to predict out-of-sample. Figure 1.9 presents the results for all cohorts. The threshold model overshoots the data again. In contrast, the shock model provides an excellent fit to the data. Thus the double-trigger theory also provides a better explanation for this data under the maintained assumptions on second mortgages. Due to the discussed data problems I would personally put a lower weight on these results compared to the benchmark results. But these results are at least suggestive that the main conclusions on the relative merit of the two theories may well extend to loans with a lower LTV.

1.10 Conclusions

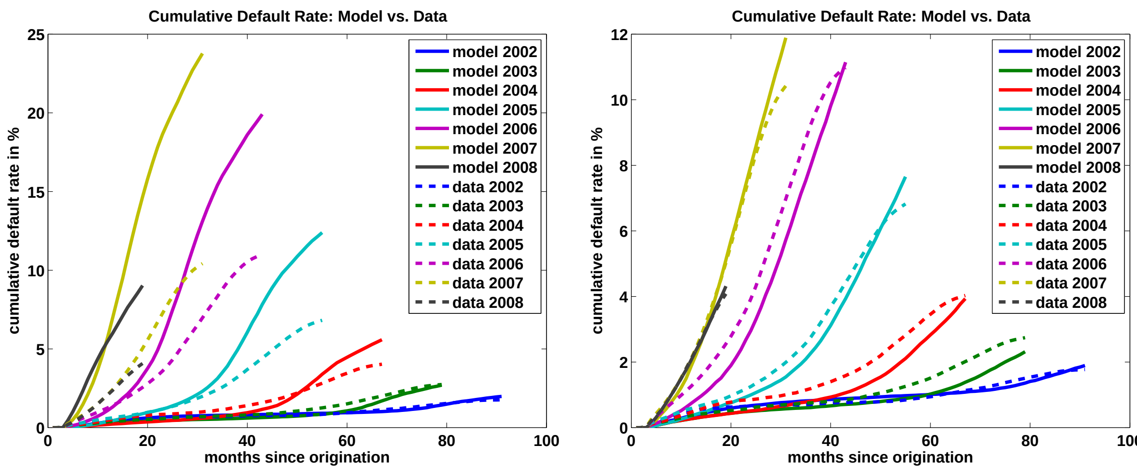
This paper has presented simulations of theoretical default models for the observed path of aggregate house prices and a realistic microeconomic distribution. Theoretical predictions were then compared to data on default rates on prime fixed-rate mortgages to assess the explanatory power of the theories.

A test has been developed that examined whether estimated reduced forms of the frictionless option model and the double trigger hypothesis are able to predict out-of-

Figure 1.9: Reduced-Form Results for Borrowers with a First Mortgage LTV of 75−84% taking Second Mortgages into account

(a) Threshold Model

(b) Shock Model



sample. This test revealed that the frictionless default theory is too sensitive to the mean shifts in the house price distribution observed in recent years. In contrast, the double-trigger hypothesis attributing default to the joint occurrence of negative equity and a life event is consistent with the data.

Based on this finding a structural dynamic stochastic model with liquidity constraints and unemployment shocks was presented to provide micro-foundations for the double-trigger hypothesis. In this model the liquidity problems associated with unemployment can act as a trigger event for default. Accordingly, the level of negative equity at which individual borrowers default on their mortgage depends on non-housing state variables: liquid wealth and employment status. The model is broadly consistent with the data and explains most of the rise in mortgage default rates as a consequence of aggregate house price dynamics.

The structural model was used to analyze two bailout policies in a mortgage crisis. This revealed that in order to neutralize losses for lenders subsidizing homeowners is much cheaper than bailing out lenders when liquidity problems are a key determinant of mortgage default. A related policy question to which the model can be applied is how the design of unemployment insurance can help to prevent mortgage default.

The results of the reduced form and structural model as well as further supporting evidence on loan characteristics show that mortgage default has a strong macroeconomic component resulting from aggregate house price dynamics. This suggests that the recent events should not be attributed entirely to a deterioration of loan quality. Instead, they hint at the existence of systematic macroeconomic risk in the mortgage market.

An important goal for future research is to develop an explanation of the house price boom and bust and the mortgage crisis in general equilibrium. This paper has presented a model where default rates match the data reasonably well taking house prices as given. It remains to provide a model that matches house prices as well as quantities in the housing and mortgage market. Obviously this represents a great challenge. But the model presented here may serve as a building block for that more general model.

2 Demographic Change, Human Capital and Welfare[1]

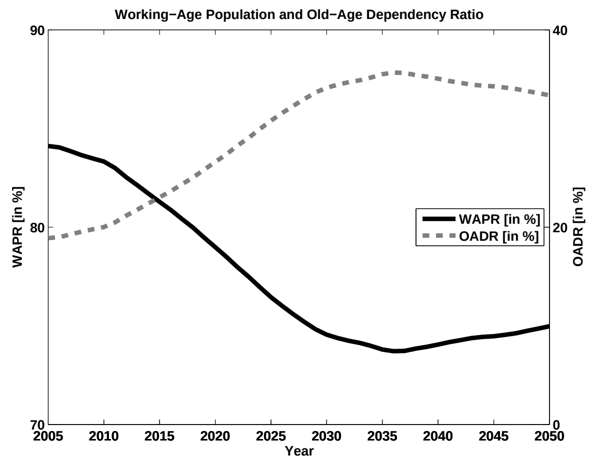
2.1 Introduction

As in all major industrialized countries, the population of the United States is aging over time. This process is driven by increasing life expectancy and a decline in birth rates from the peak levels of the baby boom. Consequently, the fraction of the working-age population will decrease, and the fraction of elderly people will increase. Figure 2.1 presents two summary measures of these demographic changes: the working-age population ratio is predicted to decrease from 84% in 2005 to 75% in 2050, while the old-age dependency ratio will increase from 19% in 2005 to 34% in 2050. These projected changes in the population structure will have important macroeconomic effects on the balance between physical capital and labor. Specifically, labor is expected to be scarce relative to physical capital, with an ensuing decline in real returns on physical capital and increases in gross wages. These relative price changes have adverse welfare effects, especially for individuals close to retirement because they receive a lower return on their assets accumulated for retirement and cannot profit from increased wages.

This paper argues that a strong incentive to invest in human capital emanates from the combined effects of increasing life expectancy and changes in relative prices, particularly if social security systems are reformed such that contribution rates remain constant. In general equilibrium, such endogenous human-capital adjustments substantially mitigate the effects of demographic change on macroeconomic aggregates and individual welfare.

[1]This chapter draws on work that was carried out jointly with equal share by Alexander Ludwig, Edgar Vogel and me. The chapter contains material from our article published in the *Review of Economic Dynamics* (Ludwig, Schelkle, and Vogel 2012) and the online appendix to that article.

Figure 2.1: Working-Age and Old-Age Dependency Ratio



Notes: Working-age population ratio (WAPR, left scale): ratio of population of age $16-64$ to total adult population of age $16-90$. Old-age dependency ratio (OADR, right scale): ratio of population of age $65-90$ to working-age population.
Source: Own calculations based on Human Mortality Database (2008).

The key contribution of our paper is to show that the human-capital adjustment mechanism is quantitatively important. We add endogenous human-capital accumulation to an otherwise standard large-scale OLG model in the spirit of Auerbach and Kotlikoff (1987). The central focus of our analysis is then to work out the differences between our model, with endogenous human capital adjustments and endogenous labor supply, and the "standard" models in the literature, with fixed (exogenous) productivity profiles.

We find that the decrease of the return to physical capital induced by demographic change in a model with endogenous human capital is only one-third of that predicted in the standard model. Welfare consequences from increasing wages, declines in rates of return, changes to pension contributions and benefits induced by demographic change are substantial. When human capital cannot adjust, some of the agents alive in 2005 will experience welfare losses up to 12.5% (5.6%) of lifetime consumption with constant pension contribution (replacement) rates. However, importantly, we find that these maximum losses are only 8.7% (4.4%) of lifetime consumption when the human capital adjustment mechanism is taken into account. Ignoring this adjustment channel thus leads to quantitatively important biases in the welfare assessment of demographic change.

Our work relates to a vast number of papers that have analyzed the economic consequences of population aging and possible adjustment mechanisms. Important examples in closed economies with a focus on social-security adjustments include Huang et al. (1997), De Nardi et al. (1999) and, with respect to migration, Storesletten (2000). In open economies, Börsch-Supan et al. (2006), Attanasio et al. (2007) and Krüger and Ludwig (2007), among others, investigate the role of international capital flows during a demographic transition. We add to this literature by highlighting an additional mechanism through which households can respond to demographic change.

Our paper is closely related to the theoretical work on longevity, human capital, taxation and growth[2] and to Fougère and Mérette (1999) and Sadahiro and Shimasawa (2002), who also quantitatively investigate demographic change in large-scale OLG models with individual human-capital decisions. In contrast to their work, we focus our analysis on relative price changes during a demographic transition and therefore consider an exogenous growth specification.[3] We also extend their analysis along various dimensions. We use realistic demographic projections instead of stylized scenarios. More importantly, our model contains a labor supply-human capital formation-leisure trade-off. It can thus capture effects from changes in individual labor supply, i.e., human capital utilization, on the return to human-capital investments. As has already been stressed by Becker (1967) and Ben-Porath (1967), it is important to model human-capital and labor supply-decisions jointly in a life-cycle framework. Along this line, a key feature of our quantitative investigation is to employ a Ben-Porath (1967) human-capital model and calibrate it to replicate realistic life-cycle wage profiles.[4] Furthermore, we place particular emphasis on the welfare consequences of an aging population for households living through the demographic transition.

---

[2]See, for example, de la Croix and Licandro (1999), Boucekkine et al. (2002), Kalemli-Ozcan et al. (2000) Echevarria and Iza (2006), Heijdra and Romp (2008), Ludwig and Vogel (2009) and Lee and Mason (2010). Our paper is also related to the literature emphasizing the role of endogenous human-capital accumulation for the analysis of changes to the tax or social-security system, as in Lord (1989), Trostel (1993), Perroni (1995), Dupor et al. (1996) and Lau and Poutvaara (2006), among others.

[3]Whether the trend growth rate endogenously fluctuates during the demographic transition or is held constant is of minor importance for the questions we are interested in. This is shown in our earlier unpublished working paper. The results are available upon request.

[4]The Ben-Porath (1967) model of human capital accumulation is one of the workhorses in labor economics used to understand such issues as educational attainment, on-the-job training, and wage growth over the life cycle, among others. See Browning, Hansen, and Heckman (1999) for a review. Extended versions of the model have been applied to study the significant changes to the U.S. wage distribution and inequality observed since the early 1970s by Heckman, Lochner, and Taber (1998) and Guvenen and Kuruscu (2009).

The paper is organized as follows. In Section 2.2, we present our quantitative model. Section 2.3 describes the calibration strategy and our computational solution method. Our results are presented in Section 2.4. Finally, Section 2.5 concludes the paper. An appendix contains robustness checks, a description of our demographic model and technical details.

2.2 The Model

We employ a large-scale OLG model à la Auerbach and Kotlikoff (1987) with endogenous labor supply and endogenous human-capital formation. The population structure is exogenously determined by time- varying demographic processes for fertility and mortality, the main driving forces of our model.[5] In a perfectly competitive environment, firms produce with a standard constant returns to scale production function. We assume that the U.S. is a closed economy.[6] Agents contribute a share of their wages to the pension system, and retirees receive a share of their average indexed past yearly earnings as pensions. Technological progress is exogenous.

2.2.1 Timing, Demographics and Notation

Time is discrete, and one period corresponds to one calendar year t. Each year, a new generation is born. Birth in this paper refers to the first time households make their own decisions and is set to the age of 16 (model age $j = 0$). Agents retire at an exogenously given age of 65 (model age $jr = 49$). Agents live at most until age 90 (model age $j = J = 74$). At a given point in time t, individuals of age j survive to age $j+1$ with probability $\varphi_{t,j}$, where $\varphi_{t,J} = 0$. The number of agents of age j at time t is denoted by $N_{t,j}$, and $N_t = \sum_{j=0}^{J} N_{t,j}$ is the total population in t.

[5]We do not model endogenous life expectancy, fertility or endogenous migration and assume that all exogenous migration is completed before agents begin making economically relevant decisions (cf. Appendix 2.B). Thus, we also abstract from potential feedback effects of social-security policies on fertility, as studied by Ehrlich and Kim (2007).

[6]For our question, the assumption of a closed economy is a valid approximation. As documented in Krüger and Ludwig (2007), demographically induced changes in the return to physical capital and wages from the U.S. perspective do not differ much between closed- and open-economy scenarios. The reason is that demographic processes are correlated across countries and, in terms of speed of the aging processes, the U.S. is somewhere in the middle with respect to all OECD countries.

2.2.2 Households

Each household comprises one representative agent who makes decisions regarding consumption and saving, labor supply and human-capital investment. The household maximizes lifetime utility at the beginning of economic life ($j = 0$) in period t,

$$\max \sum_{j=0}^{J} \beta^j \pi_{t,j} \frac{1}{1-\sigma} \{c_{t+j,j}^{\phi}(1 - \ell_{t+j,j} - e_{t+j,j})^{1-\phi}\}^{1-\sigma}, \quad \sigma > 0, \ \phi \in (0,1), \tag{2.1}$$

where the per-period utility function is a function of individual consumption c, labor supply ℓ and the time invested in formation of human capital, e. The agent is endowed with one unit of time, thus, $1 - \ell - e$ is leisure time. β is the pure time-discount factor, ϕ determines the weight of consumption in utility, and σ is the inverse of the inter-temporal elasticity of substitution with respect to the Cobb-Douglas aggregate of consumption and leisure time. $\pi_{t,j}$ denotes the (unconditional) probability to survive until age j, $\pi_{t,j} = \prod_{i=0}^{j-1} \varphi_{t+i,i}$, for $j > 0$ and $\pi_{t,0} = 1$.

Agents earn labor income (pension income when retired) as well as interest payments on their savings and receive accidental bequests. When working, they pay a fraction τ_t from their gross wages to the social-security system. The net wage income in period t of an agent of age j is given by $w_{t,j}^n = \ell_{t,j} h_{t,j} w_t (1 - \tau_t)$, where w_t is the gross wage per unit of supplied human capital at time t. There are no annuity markets, and households leave accidental bequests. These are collected by the government and redistributed in a lump-sum fashion to all households. Accordingly, the dynamic budget constraint is given by

$$a_{t+1,j+1} = \begin{cases} (a_{t,j} + tr_t)(1 + r_t) + w_{t,j}^n - c_{t,j} & \text{if } j < jr \\ (a_{t,j} + tr_t)(1 + r_t) + p_{t,j} - c_{t,j} & \text{if } j \geq jr, \end{cases} \tag{2.2}$$

where $a_{t,j}$ denotes assets, tr_t are transfers from accidental bequests, r_t is the real interest rate, the rate of return to physical capital, and $p_{t,j}$ is pension income. Initial household assets are zero ($a_{t,0} = 0$), and the transversality condition is $a_{t,J+1} = 0$.

2.2.3 Formation of Human Capital

The key element of our model is the endogenous formation of human capital. Households enter economic life with a predetermined and cohort invariant level of human capital $h_{t,0} = h_0$. Afterwards, they can invest a fraction of their time into acquiring additional human capital. We adopt a version of the Ben-Porath (1967) human-capital technology[7] given by

$$h_{t+1,j+1} = h_{t,j}(1-\delta^h) + \xi(h_{t,j}e_{t,j})^{\psi} \quad \psi \in (0,1),\ \xi > 0,\ \delta^h \geq 0, \tag{2.3}$$

where ξ is a scaling factor, the average learning ability, ψ determines the curvature of human-capital technology, δ^h is the depreciation rate of human capital, and $e_{t,j}$ is time invested in human-capital formation.

The costs of investing in human capital in this model are only the opportunity costs of foregone wage income and leisure. We understand the process of accumulating human capital to be a mixture of knowledge acquired by formal schooling and on-the- job training programs after schooling is complete. Human capital can be accumulated until retirement age, but an agent's optimally chosen time investment converges to zero some time before retirement.

2.2.4 The Pension System

The pension system is a simple balanced-budget, pay-as-you-go system that resembles key features of the U.S. system. Workers contribute a fraction τ_t of their gross wages, and pensioners receive a fraction ρ_t of their average indexed past yearly earnings.[8] The level of pensions in each period is given by $p_{t,j} = \rho_t w_{t+jr-j}\bar{h}_{t+jr-j}\frac{s_{t,j}}{jr-1}$, where $w_{t+jr-j}\bar{h}_{t+jr-j}\frac{s_{t,j}}{jr-1}$ are average indexed past yearly earnings (AIYE)[9], $w_{t+jr-j}\bar{h}_{t+jr-j}$ are average earnings of all workers in the period when a retiree of current age j reaches retirement age jr, and $\bar{h}_t$ is defined as $\bar{h}_t = \frac{\sum_{j=0}^{jr-1}\ell_{t,j}h_{t,j}N_{t,j}}{\sum_{j=0}^{jr-1}N_{t,j}}$. We refer to $\bar{h}_t$ as the average (hours weighted) human-capital stock. The sum up to age j of past individual earnings of an agent relative to average economy-wide earnings in the respective year is given by $s_{t,j} = \sum_{i=0}^{j}\frac{\ell_{t-j+i,i}h_{t-j+i,i}}{\bar{h}_{t-j+i}}$. This links pensions to individuals' past earnings.

[7] This functional form is widely used in the human-capital literature, cf. Browning, Hansen, and Heckman (1999) for a review.

[8] The U.S. system applies an additional bend-point formula to pensions, which results in intra-generational redistribution. However, in our model, without intra-cohort heterogeneity, we do not take this feature of the actual system into account. For descriptions of the current U.S. system, see Diamond and Gruber (1999) and Geanakoplos and Zeldes (2009).

[9] Our concept of AIYE is an approximation to the "average indexed monthly earnings" (AIME) in the current U.S. system where only the 35 years of working life with the highest individual earnings relative to average earnings are counted for the calculation of AIME. We ignore this feature for computational reasons and count all years of working life.

Using the above formula for $p_{t,j}$, the budget constraint of the pension system is given by

$$\tau_t w_t \sum_{j=0}^{jr-1} \ell_{t,j} h_{t,j} N_{t,j} = \rho_t \sum_{j=jr}^{J} N_{t,j} w_{t+jr-j} \bar{h}_{t+jr-j} \frac{s_{t,j}}{jr-1} \quad \forall t. \tag{2.4}$$

Below, we consider two opposite scenarios of parametric adjustment of the pension system to demographic change. In our first scenario ("const. τ"), we hold the contribution rate constant, $\tau_t = \bar{\tau}$, and endogenously adjust the replacement rate to balance the budget of the pension system. In the other extreme scenario ("const. ρ"), we hold the replacement rate constant, $\rho_t = \bar{\rho}$, and endogenously adjust the contribution rate.

2.2.5 Firms

Firms operate in a perfectly competitive environment and produce one homogeneous good, according to the Cobb-Douglas production function

$$Y_t = K_t^{\alpha} (A_t L_t)^{1-\alpha}, \tag{2.5}$$

where α denotes the share of capital used in production. K_t, L_t and A_t are the stocks of physical capital, effective labor and the level of technology, respectively. Output can be either consumed or used as an investment good. We assume that labor inputs and human capital of different agents are perfect substitutes, and effective labor input L_t is accordingly given by $L_t = \sum_{j=0}^{jr-1} \ell_{t,j} h_{t,j} N_{t,j}$. Factors of production are paid their marginal products, i.e., $w_t = (1-\alpha)\frac{Y_t}{L_t}$ and $r_t = \alpha \frac{Y_t}{K_t} - \delta_t$, where w_t is the gross wage per unit of efficient labor, r_t is the interest rate, and δ_t denotes the depreciation rate of physical capital. Total factor productivity, A_t, is growing at the exogenous rate of g_t^A: $A_{t+1} = A_t(1 + g_t^A)$.

2.2.6 Equilibrium

Denoting current period/age variables by x and following period/age variables by x', a household of age j solves the maximization problem at the beginning of period t

$$V(a, h, t, j) = \max_{c,\ell,e,a',h',s'} \{u(c, 1-\ell-e) + \varphi\beta V(a', h', s', t+1, j+1)\} \tag{2.6}$$

subject to equations (2.2) and (2.3), and the constraints $\ell \in [0,1)$, $e \in [0,1)$.

Definition 1. *Given the exogenous population distribution and survival rates in all periods $\{\{N_{t,j}, \varphi_{t,j}\}_{j=0}^{J}\}_{t=1}^{T}$, an initial physical capital stock and an initial level of average human capital, $\{K_0, \bar{h}_0\}$, and an initial distribution of assets and human capital, $\{a_{t,0}, h_{t,0}\}_{j=0}^{J}$, a competitive equilibrium of the economy is defined as a sequence of individual variables $\{\{c_{t,j}, \ell_{t,j}, e_{t,j}, a_{t+1,j+1}, h_{t+1,j+1}, s_{t+1,j+1}\}_{j=0}^{J}\}_{t=1}^{T}$, sequences of aggregate variables $\{L_t, K_{t+1}, Y_t\}_{t=1}^{T}$, government policies $\{\rho_t, \tau_t\}_{t=1}^{T}$, prices $\{w_t, r_t\}_{t=1}^{T}$ and transfers $\{tr_t\}_{t=1}^{T}$ such that*

1. *given prices, bequests and initial conditions, households solve their maximization problem, as described above,*
2. *physical capital and efficiency units of labor are paid their marginal products, i.e., $w_t = (1-\alpha)\frac{Y_t}{L_t}$ and $r_t = \alpha\frac{Y_t}{K_t} - \delta$,*
3. *per-capita transfers are determined by*

$$tr_t = \frac{\sum_{j=0}^{J} a_{t,j}(1-\varphi_{t-1,j-1})N_{t-1,j-1}}{\sum_{j=0}^{J} N_{t,j}}, \tag{2.7}$$

4. *government policies are such that the budget of the social-security system is balanced every period, i.e., equation (2.4) holds $\forall t$, and household pension income is given by $p_{t,j} = \rho_t w_{t+jr-j}\bar{h}_{t+jr-j}\frac{s_{t,j}}{jr-1}$,*
5. *markets clear every period:*

$$L_t = \sum_{j=0}^{jr-1} \ell_{t,j} h_{t,j} N_{t,j} \tag{2.8a}$$

$$K_{t+1} = \sum_{j=0}^{J} a_{t+1,j+1} N_{t,j} \tag{2.8b}$$

$$Y_t = \sum_{j=0}^{J} c_{t,j} N_{t,j} + K_{t+1} - (1-\delta)K_t. \tag{2.8c}$$

Definition 2. *A stationary equilibrium is a competitive equilibrium at which per-capita variables grow at the constant gross rate of $1+\bar{g}^A$ and aggregate variables grow at the constant gross rate $(1+\bar{g}^A)(1+n)$.*

2.2.7 Thought Experiments

We take as an exogenous driving process a time-varying demographic structure. Computations begin in year 1750 ($t = 1$), assuming an artificial initial steady state[10]. We then compute the model equilibrium from 1750 to 2500 ($t = T = 751$) when the new steady state is assumed and reached[11] and report simulation results for the main projection period of interest, from 2005 ($t = 256$) to 2050 ($t = 301$). We use data during our calibration period, 1960 − 2004 (from $t_0 = 211$ to $t_1 = 255$), to determine several structural model parameters (cf. section 2.3).

Our main objective is to compare the time paths of aggregate variables and welfare across two model variants for two social-security scenarios. Our first model variant is one in which households adjust their human capital, and our second variant is one in which human capital is held constant across cohorts. Therefore, our strategy is to first solve for transitional dynamics using the model described above. Next, we use the endogenously obtained profile of time invested in human-capital formation to compute an average time investment and associated human-capital profile, which is then fed into our alternative model in which agents are restricted with respect to their time-investment choice. We do so separately for the two opposite social-security scenarios described in subsection 2.2.4. The average time investment is computed as $\bar{e}_j = \frac{1}{t_1 - t_0 + 1} \sum_{t=t_0}^{t_1} e_{t,j}$ for our calibration period ($t_0 = 211$ and $t_1 = 255$). In the alternative model, we then add the constraint $e_{t,j} = \bar{e}_j$. The human-capital profile is then obtained from (2.3) by iterating forward on age.[12]

[10] The artificial initial steady state and long phase-in period are only used to generate suitable starting values for our main projection period. Bar and Leukhina (2010) provide an explicit model of the demographic transition and economic development that began in 17th Century England.

[11] In fact, changes in variables that are constant in steady state are already numerically irrelevant circa the year 2400.

[12] By imposing the restriction of identical time-investment profiles for all cohorts (instead of, e.g., imposing the restriction only on cohorts born after 2005), we shut down direct effects from changing mortality on human capital and indirect anticipation effects of changing returns. This alternative model is a "standard" model of endogenous labor supply and an exogenously given age-specific productivity profile—as used in numerous studies on the consequences of demographic change—with the only exception being that the time endowment is age-specific. By setting the time endowment to $1 - \bar{e}_j$ rather than 1, we avoid re-calibration across model variants. For details, see below.

2.3 Calibration and Computation

To calibrate the model, we choose model parameters such that simulated moments match selected moments in NIPA data and the endogenous wage profiles match the empirically observed wage profile in the U.S. during the calibration period 1960−2004.[13] The calibrated parameters are summarized in Table 2.1.

Table 2.1: Model Parameters

Preferences	σ	Inverse of Inter-Temporal Elasticity of Substitution	2.00
	β	Pure Time Discount Factor	0.993
	ϕ	Weight of Consumption	0.401
Human Capital	ξ	Scaling Factor	0.16
	ψ	Curvature Parameter	0.65
	δ^h	Depreciation Rate of Human Capital	0.8%
	h_0	Initial Human Capital Endowment	1.00
Production	α	Share of Physical Capital in Production	0.33
	$\bar{\delta}$	Depreciation Rate of Physical Capital	3.8%
	$\bar{g}^A$	Exogenous Growth Rate	1.8%

2.3.1 Demographics

Actual population data from 1950 − 2004 are collected from the Human Mortality Database (2008). Our demographic projections beyond 2004 are based on a population model that is described in detail in Appendix 2.B.[14] Prior to 1950, we keep the population structure constant, as in 1950.

[13]We perform this moment matching in the endogenous human-capital model and the constant contribution-rate scenario. We do not recalibrate model parameters across social-security scenarios or for the alternative human-capital model because simulated moments do not differ much. Furthermore, we are interested in how our welfare conclusions are affected by imposing various constraints on the model—either through our social-security scenarios or by restricting human-capital formation—and any parametric change in this comparison would confound our welfare analysis.

[14]The key assumptions of this model are as follows: First, the total fertility rate is constant at 2004 levels of 2.0185, until 2100, when fertility is adjusted slightly to keep the number of newborns constant for the remainder of the simulation period. Second, life expectancy monotonically increases from a current (2004) average life expectancy at birth of 77.06 years to 88.42 years in 2100, when it is held constant. Third, total migration is constant at the average migration for 1950 − 2004 for the remainder of the simulation period. These assumptions imply that a stationary population is reached in about 2200.

2.3.2 Household Behavior

The parameter σ, the inverse of the inter-temporal elasticity of substitution, is set to 2. In Appendix 2.A, a sensitivity analysis shows that our main quantitative results are robust when we change the predetermined parameter σ to 1 or 3, respectively. The time-discount factor β is calibrated to match the empirically observed capital-output ratio of 2.8 which requires $\beta = 0.993$. The weight of consumption in the utility function, ϕ, is calibrated such that households spend one-third of their time working, on average, which requires $\phi = 0.401$.

2.3.3 Individual Productivity Profiles

We choose values for the parameters of the human-capital production function such that average simulated wage profiles resulting from endogenous human-capital formation replicate empirically observed wage profiles. Data for age-specific productivity are collected from Huggett et al. (2010)[15]. We first normalize $h_0 = 1$ and then determine the values of the structural parameters $\{\xi, \psi, \delta^h\}$ by indirect inference methods (Smith 1993; Gourieroux et al. 1993). To this end, we run separate regressions of the data and simulated wage profiles on a third-order polynomial in age, given by

$$\log w_j = \eta_0 + \eta_1 j + \eta_2 j^2 + \eta_3 j^3 + \epsilon_j. \tag{2.9}$$

Here, w_j is the age-specific productivity, and ϵ_j is a residual. Denote the coefficient vector determining the slope of the polynomial estimated from the actual wage data by $\overrightarrow{\eta} = [\eta_1, \eta_2, \eta_3]'$ and the one estimated from simulated human capital profiles during $1960 - 2004$ by $\overrightarrow{\hat{\eta}} = [\hat{\eta}_1, \hat{\eta}_2, \hat{\eta}_3]'$. The latter coefficient vector is a function of the structural model parameters $\{\xi, \psi, \delta^h\}$. Finally, the values of our structural model parameters are determined by minimizing the distance $\| \overrightarrow{\eta} - \overrightarrow{\hat{\eta}} \|$. See subsection 2.3.6 for further details.

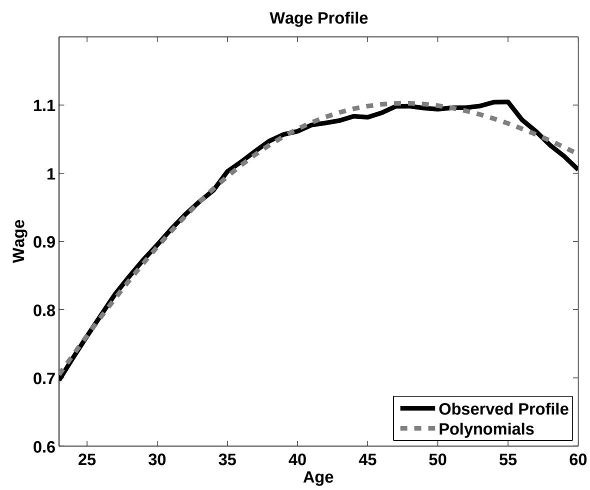
Figure 2.2 presents the empirically observed productivity profile and the estimated polynomials. Our coefficients[16] and the shape of the wage profile are in line with others reported in the literature, especially with those obtained by Hansen (1993) and

[15] We thank Mark Huggett for sending us the data.

[16] The coefficient estimates from the data are η_0: -1.6262, η_1: 0.1054, η_2: -0.0017 and η_3: 7.83e-06. We do not display the polynomial profile estimated from simulated data in Figure 2.2 because it perfectly tracks the polynomial obtained from the data.

Altig et al. (2001). The estimate of $\delta^h = 0.008$ is reasonable (Arrazola and de Hevia (2004), Browning, Hansen, and Heckman (1999)), and the estimate of $\psi = 0.65$ is in the middle of the range reported in Browning, Hansen, and Heckman (1999).[17]

Figure 2.2: Wage Profiles



Notes: Observed profile: average life-cycle wage profiles collected from Huggett, Ventura, and Yaron (2010). Polynomials: predicted wage profile based on estimated polynomial coefficients of (2.9). Both profiles were normalized by their respective means.

2.3.4 Production

We calibrate the capital share in production, α, to match the income share of labor in the data, which requires that $\alpha = 0.33$. We estimate a series of TFP and actual depreciation using NIPA data. We HP-filter these data series and then feed them into the model for the period 1950 to 2004. Thereafter, both parameters, g and δ, are held constant at their respective means. The average growth rate of total factor productivity, $\bar{g}^A$, is calibrated to match the growth rate of the Solow residual in the data. Accordingly, $\bar{g}^A = 0.018$. Finally, we calibrate $\bar{\delta}$ (and thereby scale the exogenous time path of depreciation, δ_t) such that our simulated data match an average investment-output ratio of 20%, which requires $\bar{\delta} = 0.038$.

[17]In a sensitivity analysis, we have shown that the estimate of the average time-investment productivity, $\xi = 0.16$, depends on the predetermined value of h_0, whereas the other parameters are rather insensitive to this choice. We have also found that parameterizations with a different value for h_0 yield the same results for the effects of demographic change on aggregate variables and welfare.

2.3.5 The Pension System

In our first social-security scenario ("const. τ"), we fix contribution rates and adjust replacement rates of the pension system. We calculate contribution rates from NIPA data for $1960 - 2004$ and freeze the contribution rate at the 2004 level for all following years. When simulating the alternative social-security scenario with constant replacement rates ("const. ρ"), we feed the equilibrium replacement rate obtained in the "const. τ" scenario into the model and hold it constant at the 2004 level for all remaining years. Then the contribution rate endogenously adjusts to balance the budget of the social-security system.

2.3.6 Computational Method

For a given set of structural model parameters, the solution of the model is determined by outer- and inner-loop iterations. On the aggregate level (outer loop), the model is solved by guessing initial time paths of four variables: the capital intensity, the ratio of bequests to wages, the replacement rate (or contribution rate) of the pension system and average human capital for all periods from $t = 1$ until T. On the individual level (inner loop), we begin each iteration by guessing the terminal values for consumption and human capital. We then proceed by backward induction and iterate over these terminal values until the inner-loop iterations converge. In each outer loop, disaggregated variables are aggregated each period. We then update aggregate variables until convergence, using the Gauss-Seidel-Quasi-Newton algorithm developed in Ludwig (2007).

To calibrate the model in the "const. τ" scenario, we consider additional "outer outer" loops to determine structural model parameters by minimizing the distance between the simulated average values and their respective calibration targets for the calibration period $1960 - 2004$. To summarize the description above, the parameter values determined in this way are β, ϕ, δ, ξ, ψ and δ^h.

2.4 Results

Before using our model to investigate the effects of future demographic change, we show how well it can replicate observed individual life-cycle profiles of the past. Next, we turn to the analysis of the transitional dynamics for the period 2005 to 2050, whereby we focus especially on the developments of major macroeconomic variables and the welfare effects of demographic change.

2.4.1 Backfitting

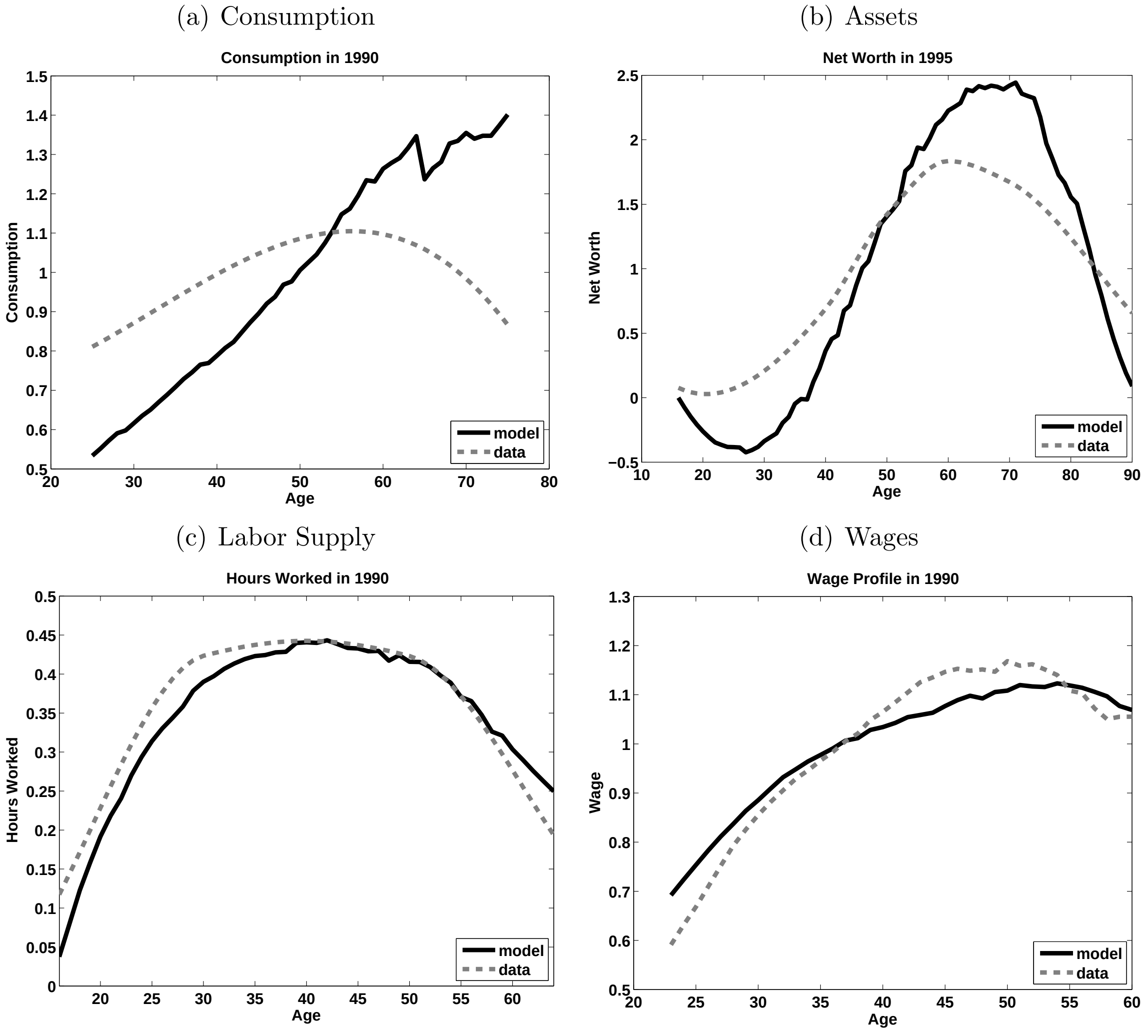
We first examine consumption profiles. We recognize that our model fails to replicate the empirically observed cross-sectional consumption profile in the 1990 Consumer Expenditure Survey[18], cf. Figure 2.3(a). The increase of consumption over the life cycle is too steep, and the peak is too late compared to the data. Because the decrease of consumption after the peak is solely caused by falling survival rates in a model without idiosyncratic risk, we cannot expect to match this dimension of the data (cf. Hansen and İmrohoroğlu (2008), Fernández-Villaverde and Krüger (2007)). As shown in Ludwig et al. (2007), in a model without human-capital adjustments, omitting idiosyncratic risk has only a negligible effect on welfare calculations. This is because welfare calculations are based on differences in consumption profiles, and the exact shape of the consumption profile is therefore less important. However, verifying the robustness of this finding in a model with endogenous human capital such as ours requires the introduction of idiosyncratic risk. We leave this extension for future research, mainly for technical reasons.[19]

We next examine asset profiles. Figure 2.3(b) shows household net worth data from the Survey of Consumer Finances for a cross-section in 1995, obtained from Bucks et al. (2006), and the corresponding cross-sectional asset profile in the model. Our

[18]The empirical profile is based on the observations on non-durable consumption for 1990 in the data set of Aguiar and Hurst (2009). We equalize the data using the traditional OECD scale that attributes weights of $(1.0, 0.7, 0.5)$ to the first adult, further adults (above age 16) and children, respectively. We then estimate a third-order polynomial in age on the adult-equivalent consumption data and show the predicted profile in the figure.

[19]Introducing idiosyncratic risk into our large model with two continuous-state variables would render computation of the transition path practically infeasible. However, to address the sensitivity of our welfare results with respect to the consumption profile, we have performed an additional sensitivity check, whereby we introduce lump-sum transfers that redistribute resources from aged individuals to young individuals within a household, such that the present value of lifetime resources is unaffected. This increases savings at younger ages. Our calibration then offsets this increase in savings via a lower β. This yields a flatter consumption profile. The total effect of lump-sum transfers on the consumption profile, therefore, mimics the effects of precautionary savings. This sensitivity analysis shows that our findings continue to hold in a model that achieves a better fit of the consumption profile and is otherwise as close as possible to our benchmark model. The results are available upon request.

Figure 2.3: Cross-Sectional Profiles



Notes: Model and data profiles for consumption, assets, labor supply and wages. All profiles are cross-sectional profiles for 1990, except for the asset profile, which is for 1995. Consumption, asset and wage profiles are normalized by their respective means. Hours data are normalized by 76 total hours per week.
Data Sources: Based on CEX consumption data collected from Aguiar and Hurst (2009), SCF net worth data obtained from Bucks et al. (2006), hours worked data from McGrattan and Rogerson (2004) and PSID wage data.

model matches the broad pattern in the data. Observed discrepancies are threefold: First, as borrowing constraints are absent from our model, initial assets are negative, whereas they are positive in the data. Second, the run-up of wealth until retirement age is stronger in our model than it is in the data.[20] Third, decumulation of assets is stronger as well. This last fact is due to the fact that our model neither has health risks, as in De Nardi et al. (2009), nor explicit bequest motives, cf., e.g., Attanasio (1999).

Our model does a good job of matching the cross-sectional hours profile observed in 1990 Census data collected from McGrattan and Rogerson (2004); see figure 2.3(c).[21] Given our preference specification, the inverse u-shape of hours worked translates into a u-shaped pattern of Frisch labor-supply elasticities over the life-cycle. This implicitly captures higher elasticities at the extensive margin at the beginning and end of the life-cycle, cf., e.g., Rogerson and Wallenius (2009). Using a Frisch elasticity concept with constant (variable) time invested in human-capital formation[22], we find that agents of age 30-50 have an average elasticity of 0.8 (1.3). The hour-weighted average Frisch elasticity across all ages, a "macro" elasticity, is approximately 1.1 (1.9).

[20] The asset profiles are generated by normalizing actual assets at each age by mean assets. However one could object that this may not be the most natural approach when assets can sometimes be negative as is the case in our model. Taking this effect into account would improve the fit of our model to the data around the middle of the life-cycle.

[21] The hours data are normalized, with total hours per week equal to 76. This might appear to be a low number for total available hours, but such a magnitude is needed to make the McGrattan and Rogerson (2004) hours data broadly consistent with the common belief that agents spend about one-third of their time working and the standard practice of macroeconomists to calibrate their models (which we have followed). The McGrattan and Rogerson (2004) data only contain average hours for certain age bins, e.g., average hours for persons of age 15-24. We associate the average hours with the age mean of that age bin, e.g., associate the value for ages 15-24 with age 20 and then use cubic interpolation to construct the empirical hours profile for all other ages. A similar procedure is used to construct the empirical asset profile.

[22] The Frisch (or λ-constant) elasticity of labor supply holds the marginal utility of wealth constant. First we compute this elasticity using the standard formula. In the context of our model, this means holding time invested in human-capital formation constant. It is then given by

$$\epsilon^j_{\ell,w} = \frac{1-\phi(1-\sigma)}{\sigma}\frac{1-\ell_j-e_j}{\ell_j},$$

see Browning, Hansen, and Heckman (1999) for a derivation. But we also report a Frisch labor-supply elasticity that allows time invested in human-capital formation to vary. In the spirit of the Frisch elasticity concept, we hold the marginal utility of human capital constant in addition to the marginal utility of wealth. This Frisch elasticity is then given by

$$\widetilde{\epsilon}^j_{\ell,w} = \frac{1-\phi(1-\sigma)}{\sigma}\frac{1-\ell_j-e_j}{\ell_j} + \frac{1}{1-\psi}\frac{e_j}{\ell_j}.$$

As usual, an interior solution is assumed here. If we use this concept, then the labor-supply elasticity is higher because the second term is positive, i.e., agents invest less in human-capital formation when they face a higher wage today, and the marginal utility of human capital remains unchanged.

These numbers are higher than the standard microeconomic estimates reported in the literature, which are typically approximately 0.5. See, e.g., Domeij and Flodén (2006). However, these standard estimates are based on prime-age, full-time employed, male workers. In contrast, Browning, Hansen, and Heckman (1999) report that the scarce empirical estimates for females are much higher than for males. Our model is a unisex model and should, accordingly, represent both sexes. Furthermore, Imai and Keane (2004) argue that standard estimates are downward-biased by not considering endogenous human-capital accumulation explicitly and thereby not correctly accounting for the true opportunity cost of time.[23] We therefore regard our value of the Frisch elasticity as very reasonable. In Section 2.A of the appendix, a sensitivity analysis further shows that our main quantitative results are robust to using a higher value of σ, which implies a lower Frisch elasticity.

Finally, Figure 2.3(d) shows the cross-sectional wage profile observed in PSID data in 1990. Our model matches the broad pattern observed in the data.[24] We have also investigated the fit of our model to cross-sectional data on wages and hours in the years 1970, 1980, 1990 and 2000. The model profiles are broadly consistent with the data at all those points in time.

2.4.2 Transitional Dynamics

We divide our analysis of transitional dynamics into two parts. First, we analyze the behavior of several important aggregate variables from 2005 to 2050. Second, we investigate the welfare consequences of demographic change for generations already alive in 2005 and for households born in the future. Throughout, we demonstrate how the design of the social-security system affects our results.

Aggregate Variables

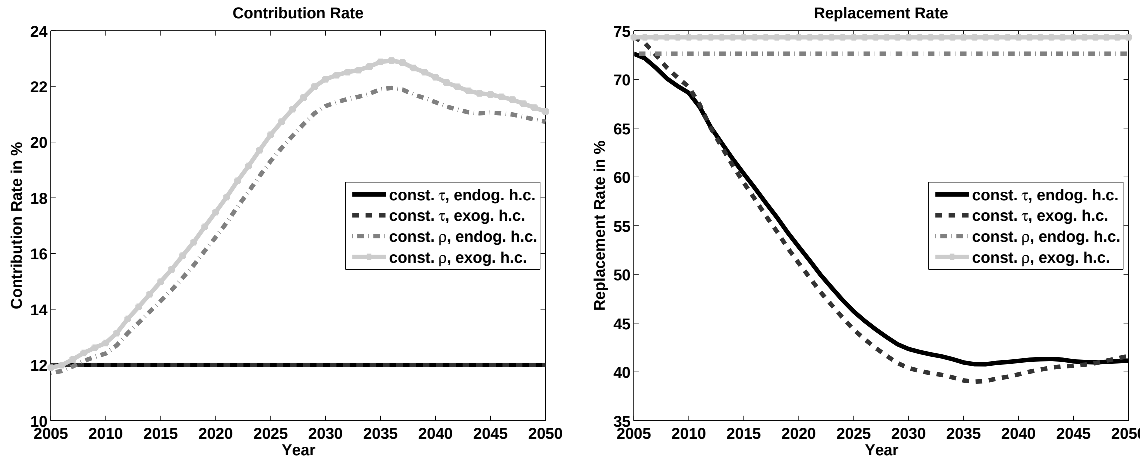
The evolution of policy variables in the two social-security scenarios is presented in Figure 2.4. In the "const. τ" scenario, pensions become less generous over time, which

[23] Imai and Keane (2004) make this argument in the context of a learning-by-doing model, but similar biases might be present in our model. We are unaware of any attempt to estimate the Frisch elasticity with varying time invested empirically in a framework such as ours, which would require inclusion of the marginal utility of human capital in the set of conditioning variables.

[24] The wage data were selected using the same criteria as in Huggett, Ventura, and Yaron (2010). To smooth the data, we show a centered average of five subsequent PSID samples.

Figure 2.4: Evolution of Policy Variables

(a) Contribution Rate (b) Replacement Rate



Notes: Pension system contribution and replacement rates for the two social-security scenarios. "const. τ": constant contribution-rate scenario. "const. ρ": constant replacement-rate scenario. "endog. h.c.": endogenous human-capital model. "exog. h.c.": exogenous human-capital model.

is represented by a decrease in the replacement rate, from approximately 73% (74%) in 2005 to 41% (42%) in 2050 for the endogenous (exogenous) human-capital model. In contrast, in the "const. ρ" scenario, the generosity of the pension system remains at the 2005 level, implying that contribution rates have to increase from approximately 12% in 2005 to 21% in 2050 in both human-capital models.[25]

Figure 2.5 reports the dynamics of four major macroeconomic variables for the two model variants—with endogenous and exogenous human capital—in the "const. τ" social-security scenario, and Figure 2.6 does so in the "const. ρ" scenario.

In Figures 2.5(a) and 2.6(a), we show the evolution of the rate of return to physical capital for the different models.[26] In the "standard" models with endogenous labor

[25] As explained in Section 2.2.4, our model of the pension system abstracts from the fact that in the United States, only the 35 years of working life with the highest individual earnings relative to average earnings are counted for the calculation of average indexed past earnings. This leads us to overstate the replacement rate but does not directly affect the level of pensions. Furthermore, we assume a balanced budget, whereas the U.S. system runs a social-security trust fund that collects excess paid-in contributions. This biases upward the replacement rate and the level of pensions.

[26] There are two reasons for the small level differences in 2005 across the various scenarios. First, our calibration targets are averages for the period $1960-2004$. Second, as already discussed in Section 2.3, we do not recalibrate across scenarios. Such level differences in initial values can be observed in all of the following figures.

supply only, the rate of return decreases from an initial level of approximately 8.1% in 2005 to 7.1% in the "const. τ" scenario and to 7.7% in the "const. ρ" scenario in 2050.[27] This magnitude is in line with results reported elsewhere in the literature, cf., e.g., Börsch-Supan et al. (2006) and Krüger and Ludwig (2007), whereas Attanasio et al. (2007) find slightly larger effects. On the contrary, in the two models with endogenous human-capital adjustment, the rate of return is expected to fall by only 0.4 (0.2) percentage points in the "const. τ" ("const. ρ") scenario. This difference in the decrease of the rate of return until 2050 between the exogenous and endogenous human-capital models is large, at a factor of about 2.5.

In Figures 2.5(b) and 2.6(b), we depict the evolution of average hours worked by all working-age individuals. Average hours worked increase both for the endogenous and exogenous human-capital models. Observe that there are level differences between the two model variants. This is mainly caused by differences in time invested in human-capital formation.

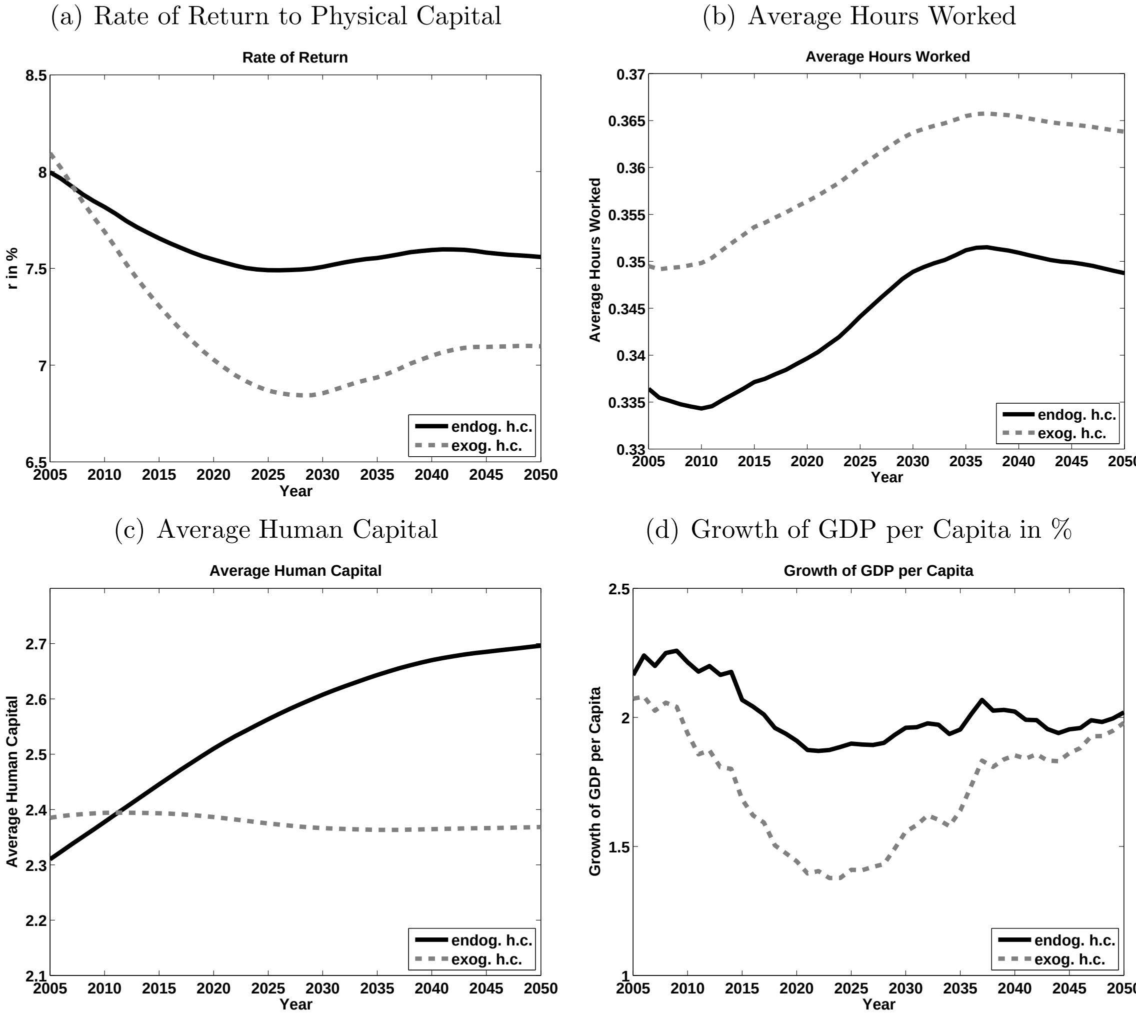
Figures 2.5(c) and 2.6(c) show that time invested in human-capital formation increases when agents are allowed to adjust their human capital. The reasons for this adjustment are that both higher wage growth and a lower rate of return to physical capital strengthen the incentive to accumulate human capital. Higher wage growth increases the benefit of a higher earning ability in the future relative to the opportunity cost of investing in human capital today in the form of foregone labor income. A lower rate of return to physical capital implies lower discounting of the future benefits of human capital and increases incentives to invest in human capital today. Specifically, with endogenous human capital in the "const. τ" ("const. ρ") scenario, average human capital per working hour increases by approximately 17% (11%) until 2050.

Finally, we focus on the evolution of the growth rate of GDP per capita, as shown in Figures 2.5(d) and 2.6(d). When the U.S. aging process peaks in 2025 (cf. figure 2.1), the growth rate of per-capita GDP falls in all scenarios to its lowest level. The drop is least pronounced for the endogenous human capital model with a fixed contribution rate. There, the growth rate gradually declines from 2.2% in 2005 to 1.9% in 2025.[28] Comparing the two "const. τ" scenarios, it can be observed that not adjusting the human-capital profile entails a large drop in the growth rate. The maximum difference

[27] The high initial level of the rate of return is caused by the previous baby boom, which increased the labor force and hence decreased capital intensity.

[28] The high initial growth rate is a consequence of the past baby boom, cf. footnote 27.

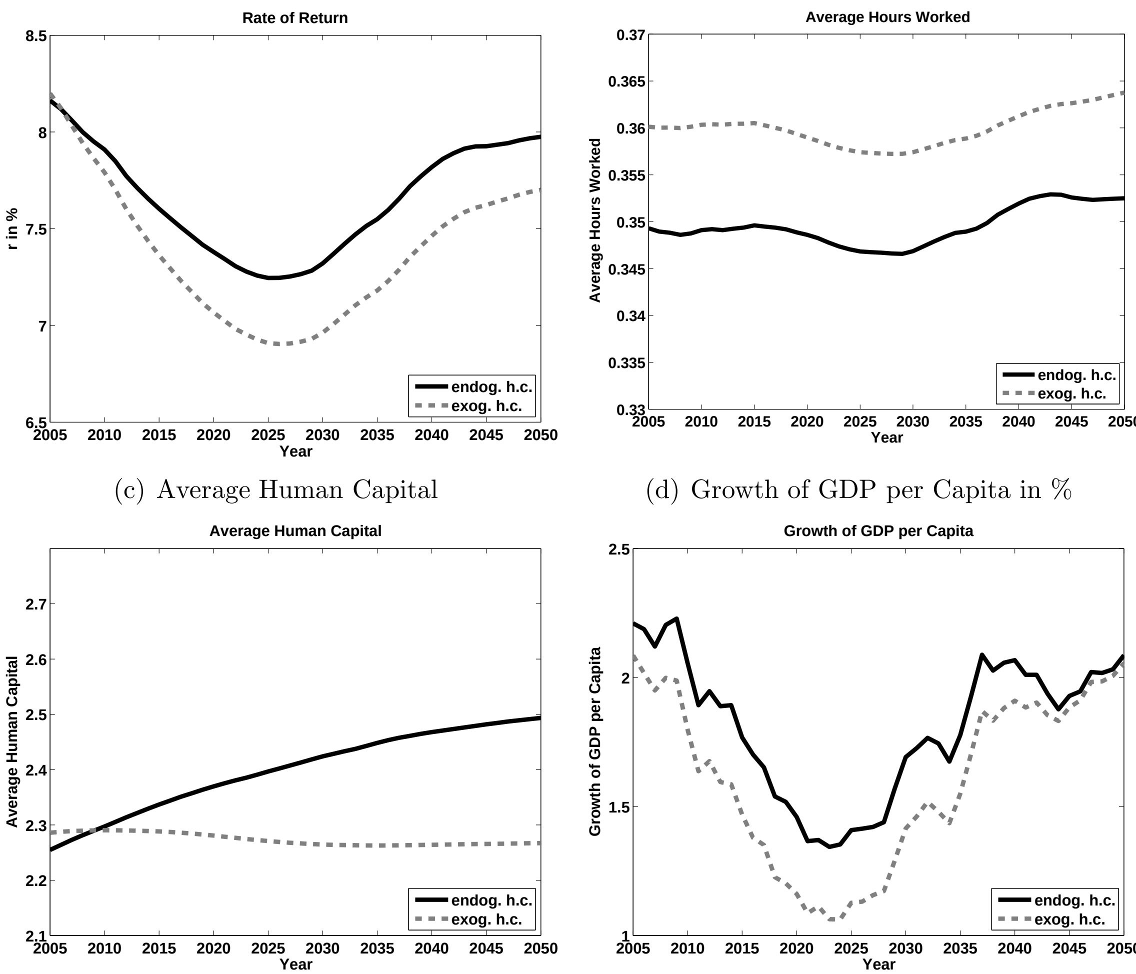
Figure 2.5: Aggregate Variables for Constant Contribution-Rate Scenario



Notes: Rate of return to physical capital, average hours worked of the working-age population, average human capital per working hour and growth of GDP per capita in the constant contribution-rate social-security scenario for two model variants. "endog. h.c.": endogenous human-capital model. "exog. h.c.": exogenous human capital model.

circa 2025 is 0.4 percentage points. Although the difference across human-capital models is only 0.2 percentage points in the case that the replacement rate is held constant ("const. ρ" scenarios), the same conclusion applies. The aging process induces relative price changes, such that agents increase their time invested in human-capital formation and thereby cushion the negative effects of demographic change on growth. The cumulative effect of the growth rate differences between the endogenous and exogenous human-capital models on the level of GDP per capita is large. With human-capital

Figure 2.6: Aggregate Variables for Constant Replacement-Rate Scenario



(a) Rate of Return to Physical Capital (b) Average Hours Worked

(c) Average Human Capital (d) Growth of GDP per Capita in %

Notes: Rate of return to physical capital, average hours worked of the working-age population, average human capital per working hour and growth of GDP per capita in the constant replacement-rate social-security scenario for two model variants. "endog. h.c.": endogenous human-capital model. "exog. h.c.": exogenous human-capital model.

adjustments, the detrended level of GDP per capita will increase by approximately 14% (10%) more until 2050 in the "const. τ" ("const. ρ") scenario than without these adjustments.

Welfare Effects

In our model, a household's welfare is affected by two consequences of demographic change. First, her lifetime utility changes because her own survival probabilities increase. Second, households face a path of declining interest rates, increasing gross wages and decreasing replacement rates (increasing contribution rates) relative to the situation without a demographic transition.

We want to isolate the welfare consequences of the second effect. To this end, we compare for an agent born at time t and of current age j her lifetime utility when she faces equilibrium factor prices, transfers and contribution (replacement) rates, as documented in the previous section, with her lifetime utility when she instead faces prices, transfers and contribution (replacement) rates that are held constant at their 2005 value. For both of these scenarios, we fix the households' individual survival probabilities at their 2005 values.[29] Following Attanasio et al. (2007) and Krüger and Ludwig (2007), we then compute the consumption-equivalent variation $g_{t,j}$, i.e., the percentage increase in consumption that needs to be given to an agent with characteristics t, j at each date in her remaining lifetime at fixed prices to make her as well off as she would be in the situation with changing prices. With our assumptions on preferences, $g_{t,j}$ can be calculated as

$$g_{t,j} = \left(\frac{\bar{V}_{t,j}}{\bar{V}_j^{2005}} \right)^{\frac{1}{\phi(1-\sigma)}} - 1, \tag{2.10}$$

where $\bar{V}_{t,j}$ denotes lifetime utility at changing prices and $\bar{V}_j^{2005}$ at fixed 2005 prices. Positive numbers of $g_{t,j}$ thus indicate that households obtain welfare gains from the general-equilibrium effects of demographic change, and negative numbers indicate welfare losses.

Welfare of Generations Alive in 2005

Of particular interest is how the welfare of all generations already alive in 2005 will be affected by demographic change. This analysis allows for an inter-generational welfare comparison of the consequences of demographic change in terms of well-being that

[29] Of course, they fully retain their age dependency. Welfare calculations based on varying survival probabilities according to the underlying demographic projections leave our conclusions on welfare in the comparison across the two models essentially unchanged.

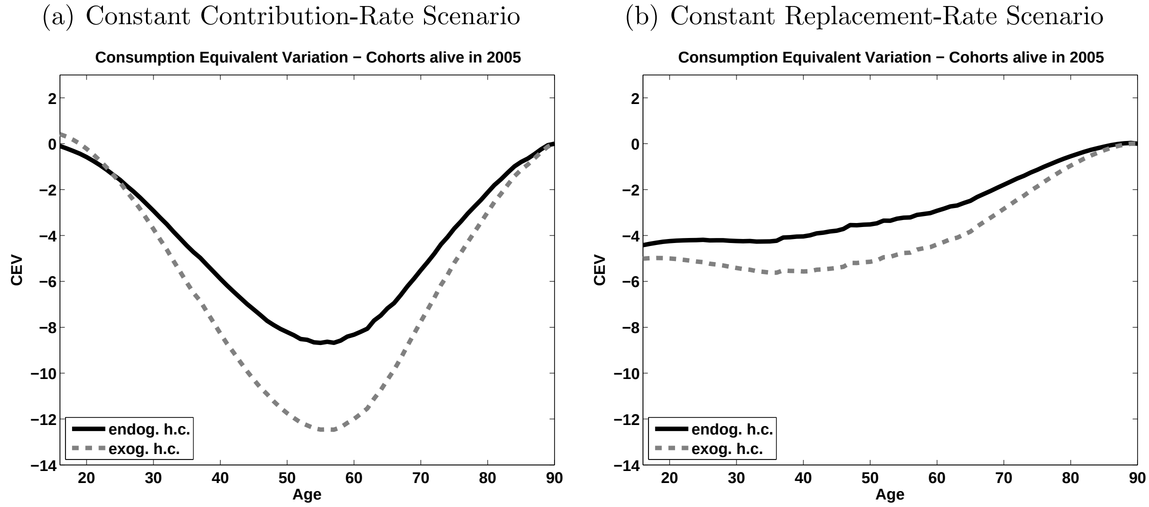
would not be possible using aggregate statistics such as per-capita GDP. Newborns and young generations benefit from increasing wages as well as decreasing returns, if they borrow to finance their human-capital formation. However, older—and thus asset-rich—generations are expected to lose lifetime utility. First, they benefit less from increasing wages because they do not significantly adjust their human capital and because their remaining working period is short. Second, falling returns diminish their capital income. Third, retirement income decreases in our scenario with constant contribution rates.

The results shown in Figure 2.7 can be summarized as follows: First, the welfare of newborn agents is essentially unchanged in the "const. τ" scenarios, whereas in the "const. ρ" scenario, newborns experience welfare losses of roughly 4.4% (5.0%) in the endogenous (exogenous) human-capital model. The fact that these welfare changes are almost identical in the two human-capital models is due to a complex interaction between the value of human-capital adjustments, which is positive, and differential general-equilibrium effects, which partially offset this interaction.[30] Second, middle-aged agents incur the highest losses in the "const. τ" scenarios: the maximum loss of agents is much larger compared to a scenario with fixed replacement rates. Clearly, constant replacement rates decrease net wages of the young but keep pensions more generous. This decreases lifetime utility of the young but narrows the loss of utility of the old (compared to a situation with falling replacement rates). The redistribution through the pension system shifts the balance somewhat in favor of the old. This also explains why the maximum of the losses occurs at a much higher age in the "const. τ" scenario in which agents close to retirement lose interest income *and* receive lower pensions. Third, independent of future pension policy, agents lose relatively less in the endogenous human-capital model. Younger agents can adjust their human capital in response to higher wages, whereas older (asset-rich) households benefit from a smaller drop in the interest rate (cf. Figures 2.5(a) and 2.6(a)) and higher pension payments.

Table 2.2 finally provides numbers on the maximum welfare loss displayed in Figure 2.7 as a summary statistic. In the exogenous human-capital model, the maximum welfare loss is approximately 12.5% (5.6%) of lifetime consumption in the "const. τ" ("const.

[30]Specifically, the increase of wages and the associated decrease of interest rates is much stronger in the exogenous human-capital model. As newborn households generally benefit from the combined effects of increasing wages and decreasing returns, welfare gains from these general-equilibrium effects are higher in the exogenous human-capital model. This explains why the overall welfare consequences for newborns across models do not differ much, despite the fact that the value of human-capital adjustments is positive.

Figure 2.7: Consumption Equivalent Variation of Agents Alive in 2005



Notes: Consumption equivalent variation (CEV) in the two social-security scenarios.

ρ") scenario, while it is only 8.7% (4.4%) of lifetime consumption in the endogenous human-capital model. This exemplifies that ignoring the adjustment channel through human-capital formation leads to quantitatively important biases in the welfare assessments of demographic change.

Table 2.2: Maximum Utility Loss for Generations alive in 2005

	Human Capital	
	Endogenous	Exogenous
Const. τ ($\tau_t = \bar{\tau}$)	-8.7%	-12.5%
Const. ρ ($\rho_t = \bar{\rho}$)	-4.4%	-5.6%

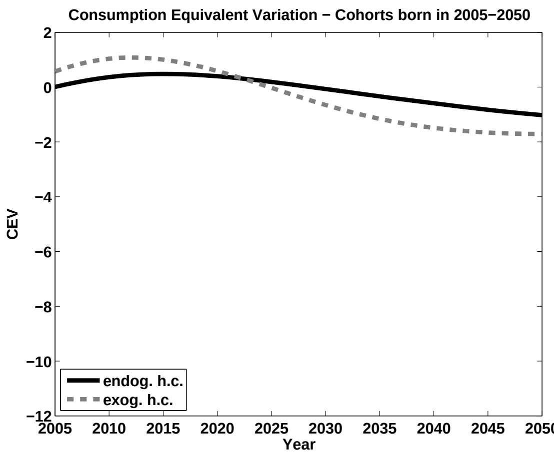
Welfare of Future Generations

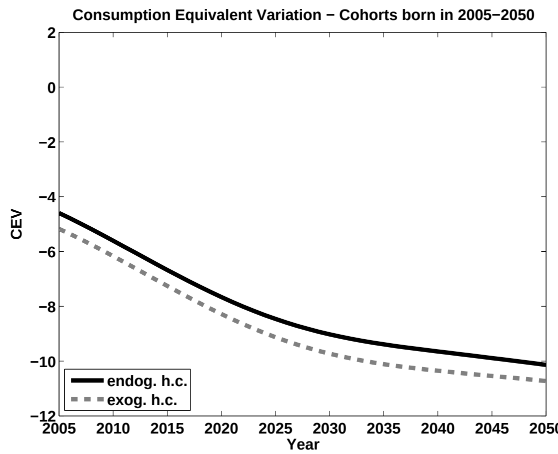
We next examine the welfare consequences for all future newborns. Figure 2.8 shows the consumption-equivalent variation for the two models and the two social-security scenarios. Agents born into a "const τ"-world experience welfare gains of up to 1.1% and losses of up to 1.7% of lifetime consumption, depending on how soon after 2005 they are born. However, welfare losses for future generations may be quite large, despite the human-capital channel, if the social-security system is not reformed ("const ρ"). These losses are between 5.2% and 10.7% of lifetime consumption with exogenous human capital and not much lower with endogenous human-capital adjustments. Notice, again, that in our comparison across models, differences are not large because the positive value of human-capital adjustments is offset by the more beneficial general-equilibrium effects in the exogenous human-capital model. For this reason, welfare gains for some cohorts may even be slightly higher in the exogenous human-capital model when the contribution rate is held constant.

Figure 2.8: Consumption Equivalent Variation of Agents born in 2005-2050

(a) Constant Contribution-Rate Scenario

(b) Constant Replacement-Rate Scenario





Notes: Consumption-equivalent variation (CEV) in the two social-security scenarios.

2.5 Conclusion

This paper finds that increased investments in human capital may substantially mitigate the macroeconomic impact of demographic change, with profound implications for individual welfare. As labor will be relatively scarce and capital will be relatively abundant in an aging society, interest rates will fall. As we emphasize, these effects will be much smaller once we account for changes in human-capital formation. For the U.S., our simulations predict that if contribution rates (replacement rates) are held constant, then the rate of return will fall by only 0.5 (0.9) percentage points until 2025 with endogenous human capital, compared to 1.2 (1.3) percentage points in the standard model with a fixed human-capital profile.

We also document that the increase in wages, declines in rates of return and changes to pension contributions and benefits induced by demographic change have substantial welfare consequences. When human capital cannot adjust, some of the agents alive in 2005 will experience welfare losses of up to 12.5% (5.6%) of lifetime consumption with constant contribution (replacement) rates. However, importantly, we find that these maximum losses are only 8.7% (4.4%) of lifetime consumption when the human-capital adjustment mechanism is taken into account.

However, we have operated in a frictionless environment, where all endogenous human-capital adjustments are driven by relative price changes. If, instead, human-capital formation is affected by market imperfections, such as borrowing constraints, then these automatic adjustments will be inhibited. In this case, appropriate education and training policies in aging societies are an important topic for future research and the policy agenda.

Appendix 2.A Sensitivity Analysis

We now provide a sensitivity analysis with respect to the parameter σ, the inverse of the inter-temporal elasticity of substitution. In our benchmark model, we set $\sigma = 2$, but we now also explore the cases of $\sigma = 1$ (log-utility) and $\sigma = 3$. We recalibrate the model when we vary σ, such that we match the same calibration targets as in the main text.

This exercise serves two purposes. First, because σ is a predetermined parameter in our calibration procedure, it is interesting to observe how much our results depend on our choice of σ. Second, we want to investigate how sensitive our results are to changes of the theoretical Frisch labor-supply elasticities in our model. Table 2.3 shows how varying σ generates variation in these elasticities in the differently calibrated versions of our model.[31] We observe that these experiments generate substantial variation in labor-supply elasticities. With $\sigma = 3$ and a constant (variable) time investment, the "macro" elasticity is approximately 15% (6%) lower than in the benchmark calibration, while with $\sigma = 1$, it is approximately 46% (21%) higher than in the benchmark. A limitation of this sensitivity check is, of course, that we cannot separately identify the effects of the inverse of the inter-temporal elasticity of substitution and the Frisch labor-supply elasticity on our results.

Table 2.3: Sensitivity Analysis with respect to σ: Mean Frisch Labor-Supply Elasticities during 1960-1995

	Time Investment Constant			Time Investment Variable		
	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
Age 30 to 50	1.2	0.8	0.7	1.6	1.3	1.2
Age 20 to 60	1.5	1.0	0.9	2.2	1.8	1.7
All Ages	2.0	1.3	1.2	3.3	2.8	3.0
"Macro"	1.6	1.1	0.9	2.3	1.9	1.8

The fit of our model to the observed cross-sectional profiles of consumption, assets, hours and wages is very similar for all values of σ considered here, so detailed figures are omitted. Unfortunately, the failure of our model to match the observed consumption profile cannot be fixed by varying σ in this way.

[31] Footnote 22 explains how the Frisch labor-supply elasticity depends on σ.

We next turn to the implications of the different parameterizations on the transitional dynamics of the macroeconomic variables. We omit the figures for the contribution and replacement rates of the pension system because the dynamics for the alternative values of σ are almost identical to the benchmark model.

Figure 2.9 presents the evolution of the four major macroeconomic variables for the "const. τ" social-security scenario, and Figure 2.10 does so for the "const. ρ" scenario. We observe that for the alternative values of σ, the broad dynamics of these variables are very similar to the benchmark model. The most significant differences are that for log-utility ($\sigma = 1$), when human capital is exogenous, average hours increase by more, and the interest rate decreases by less than in the benchmark calibration with $\sigma = 2$ in the years after 2020.

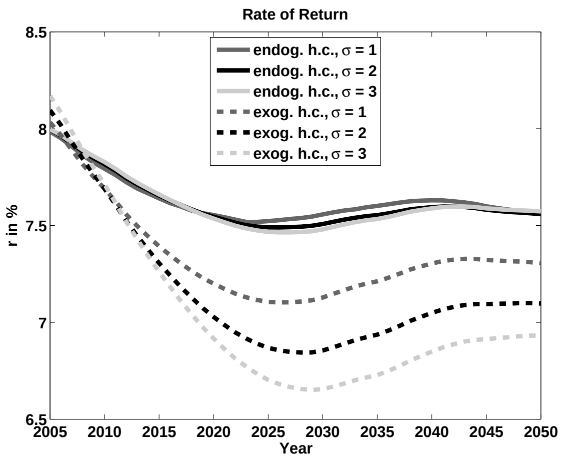
For the different values of σ, the welfare analysis of demographic change for agents alive in 2005 is presented in Figure 2.11 and Table 2.4. The welfare results can be viewed as an important and convenient summary measure of all of the differences between differently parameterized models. We find that the welfare assessment of demographic change does not depend much on the value of σ and the comparison across models with endogenous and exogenous human capital is largely unaffected. We thus conclude that our main quantitative result that human-capital adjustments mitigate the macroeconomic and welfare effects of demographic change is robust to the changes of σ we considered here.

Table 2.4: Sensitivity Analysis with respect to σ: Maximum Utility Loss for Generations Alive in 2005

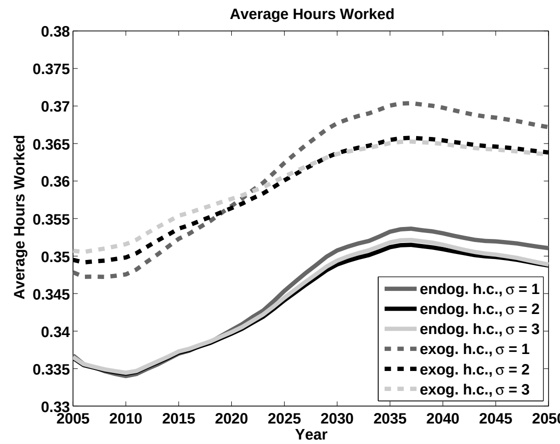
	$\sigma = 1$		$\sigma = 2$		$\sigma = 3$	
	Human Capital		Human Capital		Human Capital	
	Endog.	Exog.	Endog.	Exog.	Endog.	Exog.
Const. τ ($\tau_t = \bar{\tau}$)	-8.4%	-11.4%	-8.7%	-12.5%	-8.7%	-13.3%
Const. ρ ($\rho_t = \bar{\rho}$)	-4.4%	-5.8%	-4.4%	-5.6%	-4.3%	-5.6%

Figure 2.9: Sensitivity Analysis with respect to σ: Aggregate Variables for the Constant Contribution-Rate Scenario

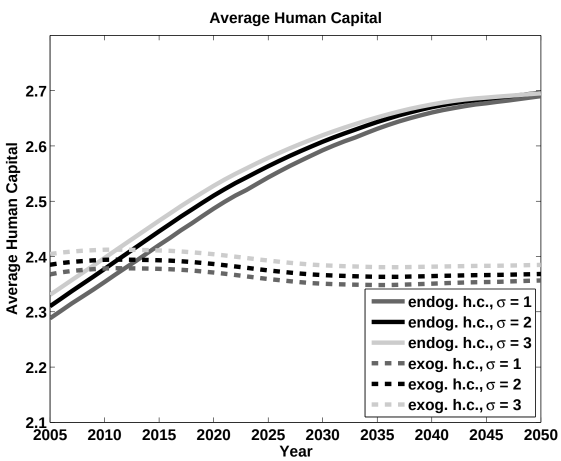
(a) Rate of Return to Physical Capital



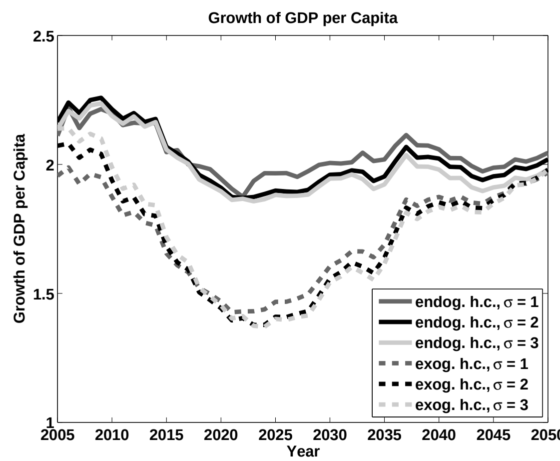
(b) Average Hours Worked



(c) Average Human Capital



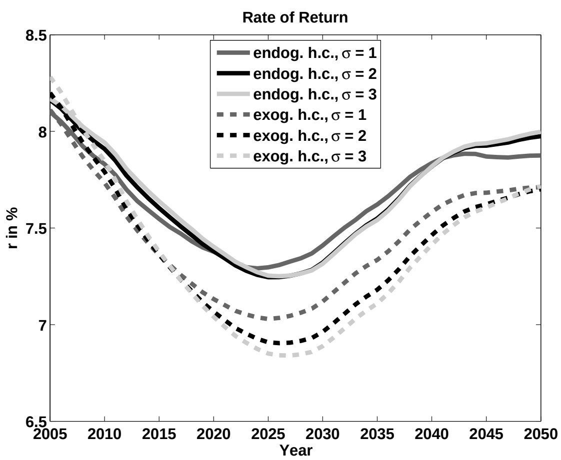
(d) Growth of GDP per Capita in %



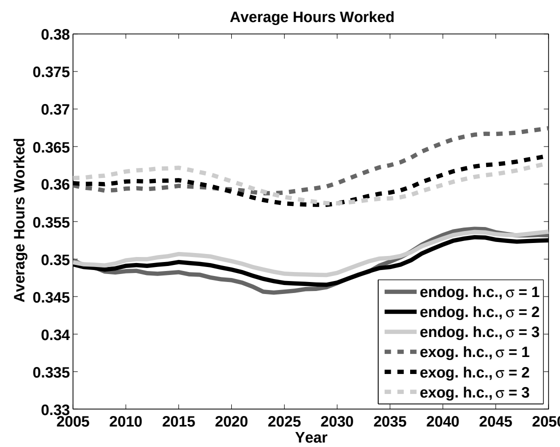
Notes: Rate of return to physical capital, average hours worked of the working-age population, average human capital per working hour and growth of GDP per capita in the constant contribution-rate social-security scenario for two model variants and three different values of σ. "endog. h.c.": endogenous human-capital model. "exog. h.c.": exogenous human-capital model.

Figure 2.10: Sensitivity Analysis with respect to σ: Aggregate Variables for the Constant Replacement-Rate Scenario

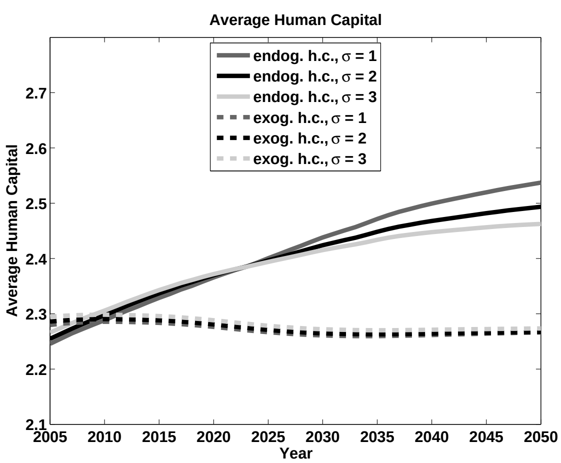
(a) Rate of Return to Physical Capital



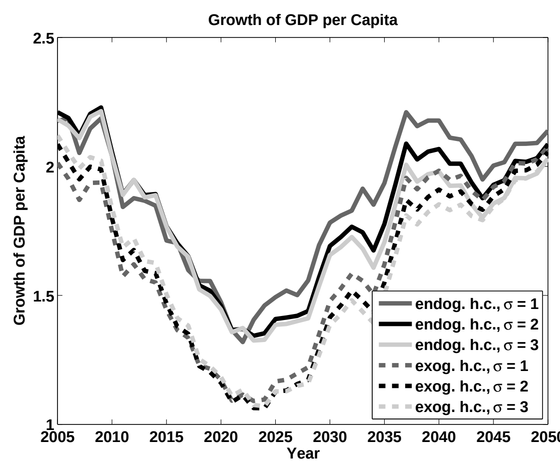
(b) Average Hours Worked



(c) Average Human Capital



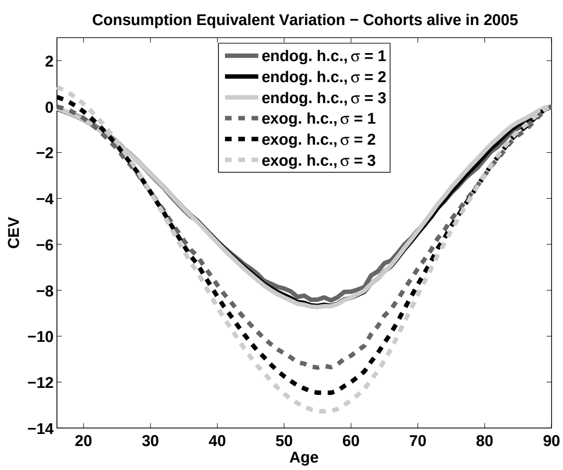
(d) Growth of GDP per Capita in %



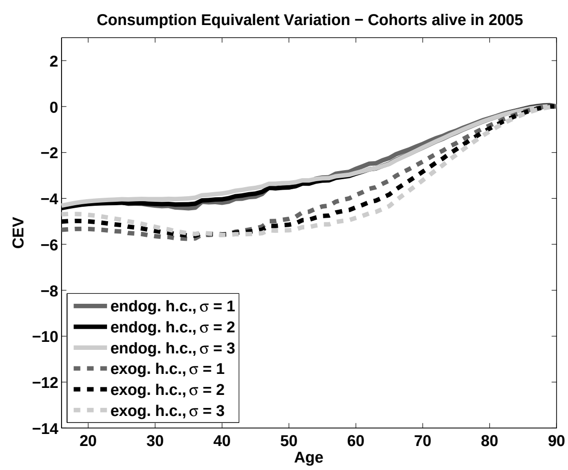
Notes: Rate of return to physical capital, average hours worked of the working-age population, average human capital per working hour and growth of GDP per capita in the constant replacement-rate social-security scenario for two model variants and three different values of σ. "endog. h.c.": endogenous human-capital model. "exog. h.c.": exogenous human-capital model.

Figure 2.11: Sensitivity Analysis with respect to σ: Consumption-Equivalent Variation of Agents Alive in 2005

(a) Constant Contribution-Rate Scenario



(b) Constant Replacement-Rate Scenario



Notes: Consumption-equivalent variation (CEV) in the two social-security scenarios for three different values of σ. "endog. h.c.": endogenous human-capital model. "exog. h.c.": exogenous human-capital model.

Appendix 2.B Demographic Model and Data

Our demographic data are based on the Human Mortality Database (2008). Population of age j in year t is determined by four factors: (i) an initial population distribution in year 0, (ii) age- and time-specific mortality rates, (iii) age- and time-specific fertility rates and (iv) age- and time-specific migration rates. We describe here how we model all of these elements and then briefly compare results of our demographic predictions with those of United Nations (2007).

Initial Population Distribution

We collect the age- and time-specific population data for the period $1950 - 2004$.

Mortality Rates

Our mortality model is based on sex-, age- and time-specific mortality rates. To simplify notation, we suppress a separate index for sex. Using data from $1950 - 2004$, we apply a Lee-Carter procedure (Lee and Carter 1992) to decompose mortality rates as

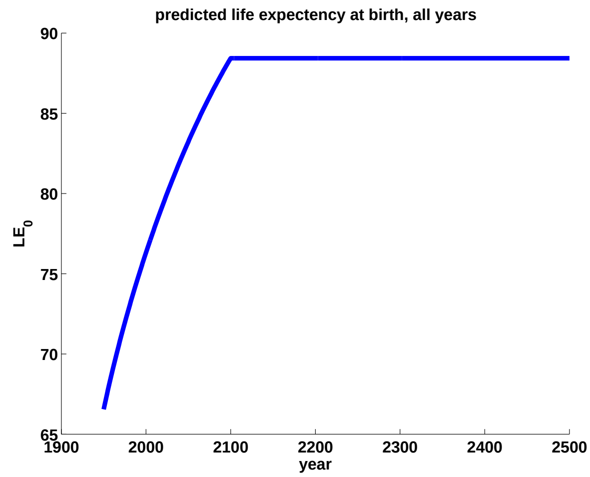
$$\ln(1 - \varphi_{t,j}) = a_j + b_j d_t, \tag{2.11}$$

where a_j and b_j are vectors of age-specific constants, and d_t is a time-specific index that equally affects all age groups. We assume that the time-specific index, d_t, evolves according to a unit-root process with drift,

$$d_t = \chi + d_{t-1} + \epsilon_t. \tag{2.12}$$

This implies that d_t is a linear function of time. The estimate of the drift term is $\hat{\chi} = -1.2891$. We then predict mortality rates into the future (until 2100) by holding $\hat{a}_j$, $\hat{b}_j$ and $\hat{\chi}$ constant and setting $\epsilon_t = 0$ for all t. For all years beyond 2100, we hold survival rates constant at their respective year 2100 values. Figure 2.12 shows the corresponding path of life expectancy at birth.

Figure 2.12: Life Expectancy at Birth



Notes: Our own predictions of life expectancy at birth based on Human Mortality Database (2008).

Fertility Rates

Fertility in our model is age and time specific. For our predictions, we assume that age-specific fertility rates are constant at their respective year 2004 values for all periods $2005, \ldots, 2100$. For periods after 2100, we assume that the number of newborns is constant. Because the U.S. reproduction rate is slightly above replacement levels, this implies that the total fertility rate is slightly decreasing each year from 2100 onwards, until approximately year 2200, when the population converges to a stationary distribution.

Population Dynamics

We use the estimated fertility and mortality data to forecast future population dynamics. The transition of the population is accordingly given by

$$N_{t,j} = \begin{cases} N_{t-1,j-1}\varphi_{t-1,j-1} & \text{for } j > 0 \\ \sum_{i=0}^{J} f_{t-1,i} N_{t-1,i} & \text{for } j = 0, \end{cases} \tag{2.13}$$

where $f_{t,j}$ denotes age- and time- specific fertility rates. Population growth is then given by $n_t = \frac{N_{t+1}}{N_t} - 1$, where $N_t = \sum_{j=0}^{J} N_{t,j}$ is total population in t.

Migration

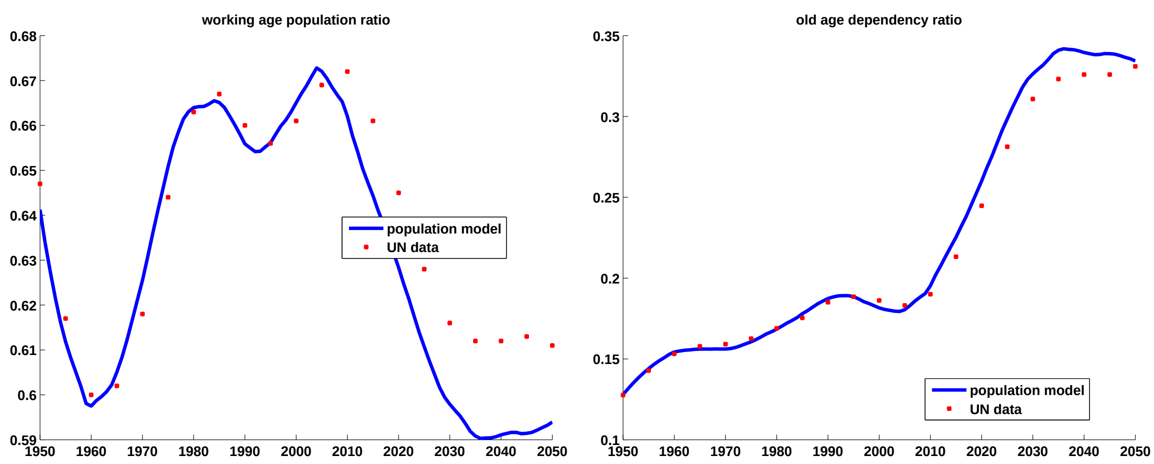
Migration is exogenous in our economic model. Setting migration equal to zero would lead us to overestimate future decreases in the working-age population ratio and to overstate the increases in old-age dependency. We therefore restrict migration to ages $j \leq 15$, such that migration plays a similar role as fertility in our economic model. This simplifying assumption allows us to treat newborns and immigrants alike. We compute aggregate migration from United Nations (2007) and distribute age-specific migrants in each year equally across all ages $0, \ldots, 15$.

Evaluation

Figures 2.13(a) and 2.13(b) display the predicted working-age population and old-age dependency ratios, according to our population model and according to United Nations (2007). Compared to this benchmark, our population model is close to the UN but predicts a slightly stronger decrease of the working-age population ratio and a correspondingly stronger increase of the old-age dependency ratio until 2050.

Figure 2.13: Comparison to United Nations Population Data and Predictions

(a) Working-Age Population Ratio (b) Old-Age Dependency Ratio



Notes: Population model: own predictions of the working-age population ratio and old-age dependency ratio based on Human Mortality Database (2008). UN data: working-age population ratio and old-age dependency ratio according to United Nations (2007).

Appendix 2.C Computational Procedures

2.C.1 Household Problem

To simplify the description of the solution of the household model for given prices (wages and interest rates), transfers and social-security payments, we focus on steady states and therefore drop the time index t. Furthermore, we focus on a de-trended version of the household problem in which consumption c, assets a, wages w and transfers, tr, are transformed into $\tilde{c} = \frac{c}{A}$, $\tilde{a} = \frac{a}{A}$, $\tilde{w} = \frac{w}{A}$ and $\tilde{tr} = \frac{tr}{A}$, where A is the technology level growing at the exogenous rate g.[32] Other variables are not transformed because they are already stationary.

To understand our transformations of the value functions, notice that utility in the last period, period J, takes the form

$$u(c_J, 1 - e_J - \ell_J) = u(c_J, 1) = A^{\phi(1-\sigma)} u(\tilde{c}), \tag{2.14}$$

Observe that the homotheticity of the utility function is inherited by the value function in period J and in all other periods. We consequently adjust the discount factor to $\tilde{\beta} = \beta \varphi (1 + g)^{\phi(1-\sigma)}$.

To understand the transformation of the budget constraint, notice that during the retirement period, the budget constraint is

$$a_{t+1,j+1} = (a_{t,j} + tr_t)(1 + r_t) + \rho_t w_{t+jr-j} \bar{h}_{t+jr-j} \frac{s_{t,j}}{jr - 1} - c_{t,j}. \tag{2.15}$$

Division by the trend component A then gives

$$\tilde{a}_{t+1,j+1} = \frac{1}{1+g} \left((\tilde{a}_{t,j} + \tilde{tr}_t)(1 + r_t) + \rho_t \tilde{w}_{t+jr-j} (1 + g)^{jr-j} \bar{h}_{t+jr-j} \frac{s_{t,j}}{jr - 1} - \tilde{c}_{t,j} \right).$$

The term $(1 + g)^{jr-j}$ reflects the fact that pension income in the US is only indexed to inflation and not to growth of nominal wages.

[32]These transformations are made for convenience, to simplify the structure of our computer code.

Taking corresponding adjustments to the budget constraint during the working period, the de-trended version of the household problem is then given by

$$
\begin{aligned}
V(\tilde{a}, h, s, j) = \max_{\tilde{c}, \ell, e, \tilde{a}', h', s'} & \left\{ u(\tilde{c}, 1 - \ell - e) + \tilde{\beta} V(\tilde{a}', h', s', j+1) \right\} \\
\text{s.t.} & \\
\tilde{a}' &= \frac{1}{1+g} \left((\tilde{a} + \tilde{tr})(1+r) + \tilde{y} - \tilde{c} \right) \\
\tilde{y} &= \begin{cases} \ell h \tilde{w}(1-\tau) & \text{if } j < jr \\ \rho \tilde{w}_{jr} (1+g)^{jr-j} \bar{h}_{jr} \frac{s_{jr}}{jr-1} & \text{if } j \geq jr \end{cases} \\
h' &= g(h, e) \qquad (2.16) \\
s' &= s + \ell \frac{h}{\bar{h}} \qquad (2.17) \\
\ell &\in [0,1], \quad e \in [0,1].
\end{aligned}
$$

Here, $g(h, e)$ is the human-capital technology.

Using the budget constraints, now rewrite the above as

$$
V(\tilde{a}, h, s, j) =
$$

$$
\max_{\tilde{c}, \ell, e, \tilde{a}', h'} \left\{ u(\tilde{c}, 1 - \ell - e) + \tilde{\beta} V \left(\frac{1}{1+g} \left((\tilde{a} + \tilde{tr})(1+r) + \tilde{y} - \tilde{c} \right), g(h, e), s + \ell \frac{h}{\bar{h}}, j+1 \right) \right\}
$$

s.t.

$\ell \geq 0.$

In the above, we have also replaced the bounded support of time invested and leisure with a one-side constraint on ℓ because the upper constraints, $\ell = 1$, respectively $e = 1$, and the lower constraint, $e = 0$, are never binding due to Inada conditions on the utility function and the functional form of human-capital technology (see below). Recall that $\ell = 0$ for $j \geq jr$.

Denoting by μ_ℓ the Lagrange multiplier on the inequality constraint for ℓ, we can write the first-order conditions as

$$\tilde{c}: \quad u_{\tilde{c}} - \tilde{\beta}\frac{1}{1+g}V'_{\tilde{a}'}(\cdot) = 0 \tag{2.18a}$$

$$\ell: \quad -u_{1-\ell-e} + \tilde{\beta}\left[h\tilde{w}(1-\tau)\frac{1}{1+g}V'_{\tilde{a}'}(\cdot) + V'_{s'}(\cdot)\frac{h}{\bar{h}}\right] + \mu_\ell = 0 \tag{2.18b}$$

$$e: \quad -u_{1-\ell-e} + \tilde{\beta}g_e V'_{h'}(\cdot) = 0 \tag{2.18c}$$

and the envelope conditions as

$$\tilde{a}: \quad V_{\tilde{a}}(\cdot) = \tilde{\beta}\frac{1+r}{1+g}V'_{\tilde{a}'}(\cdot) \tag{2.19a}$$

$$h: \quad V_h(\cdot) = \begin{cases} \tilde{\beta}\left(\ell\tilde{w}(1-\tau)\frac{1}{1+g}V'_{\tilde{a}'}(\cdot) + g_h V'_{h'}(\cdot) + V'_{s'}(\cdot)\ell\frac{1}{\bar{h}}\right) & \text{if } j < jr \\ \tilde{\beta}V'_{h'}(\cdot)g_h & \text{if } j \geq jr \end{cases} \tag{2.19b}$$

$$s: \quad V_s(\cdot) = \begin{cases} \tilde{\beta}V'_{s'}(\cdot) & \text{if } j < jr \\ \tilde{\beta}\left(V'_{s'}(\cdot) + \rho\tilde{w}_{jr}(1+g)^{jr-j}\bar{h}_{jr}\frac{1}{jr-1}\frac{1}{1+g}V'_{\tilde{a}'}\right) & \text{if } j \geq jr \end{cases} \tag{2.19c}$$

Note that for the retirement period, i.e., for $j \geq jr$, equations (2.18b) and (2.18c) are irrelevant.

From (2.18a) and (2.19a) we obtain

$$V_{\tilde{a}} = (1+r)u_{\tilde{c}} \tag{2.20}$$

and, using the above in (2.18a), the familiar inter-temporal Euler equation for consumption follows as

$$u_{\tilde{c}} = \tilde{\beta}\frac{1+r}{1+g}u_{\tilde{c}'}. \tag{2.21}$$

From (2.18a) and (2.18b) we get the intra-temporal Euler equation for leisure,

$$u_{1-\ell-e} = u_{\tilde{c}}h\left(\tilde{w}(1-\tau) + (1+g)\frac{V'_{s'}}{V'_{\tilde{a}'}}\frac{1}{\bar{h}}\right) + \mu_\ell. \tag{2.22}$$

From the human capital technology (2.3) we further have

$$g_e = \xi\psi(eh)^{\psi-1}h \tag{2.23a}$$

$$g_h = (1-\delta^h) + \xi\psi(eh)^{\psi-1}e. \tag{2.23b}$$

We loop backwards on j from $j = J-1, \ldots, 1$ by taking an initial guess of $[\tilde{c}_J, h_J]$ as given and by initializing $V_{\tilde{a}'}(\cdot, J) = V_{h'}(\cdot, J) = V_{s'}(\cdot, J) = 0$. During retirement, that is, for all ages $j \geq jr$, our solution procedure is standard backward shooting using the first-order conditions. However, during the period of human-capital formation, that is, for all ages $j < jr$, the first-order conditions would not be sufficient if the problem is not a convex-programming problem. Thus, our backward-shooting algorithm will not necessarily find the true solution. In fact, this may be the case in human-capital models such as ours because the effective wage rate is endogenous (it depends on the human-capital investment decision). For a given initial guess $[\tilde{c}_J, h_J]$, we therefore first compute a solution and then consider variations of initial guesses of $[\tilde{c}_J, h_J]$ on a large grid and check whether we converge to the same unique solution. In all of our scenarios, we never find any multiplicities. The details of our steps are as follows:

1. In each j, $h_{j+1}, V_{\tilde{a}'}(\cdot, j+1), V_{h'}(\cdot, j+1)$, and $V_{s'}(\cdot, j+1)$ are known.

2. Compute $u_{\tilde{c}}$ from (2.18a).

3. For $j \geq jr$, compute h_j from (2.3) by setting $e_j = \ell_j = 0$ and by taking h_{j+1} as given. Compute $\tilde{c}_j$ directly from Equation (2.26) below.

4. For $j < jr$:

   a) Guess h_j

   b) Compute e_j from (2.3) as

   $$e_j \quad = \tfrac{1}{h_j} \left(\tfrac{h_{j+1} - h_j(1-\delta^h)}{\xi} \right)^{\frac{1}{\psi}} . \tag{2.24}$$

   c) Compute $lcr_j = \frac{1-e_j-\ell_j}{\tilde{c}_j}$, the leisure-to-consumption ratio, from (2.22), as follows: From our functional-form assumption on utility, marginal utilities are given by

   $$u_{\tilde{c}} = \left(\tilde{c}^{\phi}(1-\ell-e)^{1-\phi}\right)^{-\sigma} \phi \tilde{c}^{\phi-1}(1-\ell-e)^{1-\phi}$$
   $$u_{1-\ell-e} = \left(\tilde{c}^{\phi}(1-\ell-e)^{1-\phi}\right)^{-\sigma} (1-\phi)\tilde{c}^{\phi}(1-\ell-e)^{-\phi}$$

   Hence, we obtain from (2.22) the familiar equation:

   $$\frac{u_{1-\ell-e}}{u_{\tilde{c}}} = h\left(\tilde{w}(1-\tau) + (1+g)\frac{V'_{s'}}{V'_{\tilde{a}'}}\frac{1}{\bar{h}}\right) = \frac{1-\phi}{\phi}\frac{\tilde{c}}{1-\ell-e},$$

and therefore:

$$lcr_j = \frac{1 - e_j - \ell_j}{\tilde{c}_j} = \frac{1-\phi}{\phi}\left(h\left[\tilde{w}(1-\tau) + (1+g)\frac{V'_{s'}}{V'_{\tilde{a}'}}\frac{1}{\bar{h}}\right]\right)^{-1}. \tag{2.25}$$

d) Next, compute $\tilde{c}_j$ as follows. Notice first that one may also write marginal utility from consumption as

$$u_{\tilde{c}} = \phi\tilde{c}^{\phi(1-\sigma)-1}(1-\ell-e)^{(1-\sigma)(1-\phi)}. \tag{2.26}$$

Using (2.25) in (2.26), we then obtain

$$\begin{aligned} u_{\tilde{c}} &= \phi\tilde{c}^{\phi(1-\sigma)-1}(lcr \cdot \tilde{c})^{(1-\sigma)(1-\phi)} \\ &= \phi\tilde{c}^{-\sigma} \cdot lcr^{(1-\sigma)(1-\phi)}. \end{aligned} \tag{2.27}$$

Because $u_{\tilde{c}}$ is given from (2.18a), we can now compute $\tilde{c}$ as

$$\tilde{c}_j = \left(\frac{u_{\tilde{c}_j}}{\phi \cdot lcr_j^{(1-\sigma)(1-\phi)}}\right)^{-\frac{1}{\sigma}}. \tag{2.28}$$

e) Given $\tilde{c}_j$, e_j, compute labor, ℓ_j, as

$$\ell_j = 1 - lcr_j \cdot \tilde{c}_j - e_j.$$

f) If $\ell_j < 0$, set $\ell_j = 0$ and recompute $\tilde{c}_j$ from (2.26) as

$$\tilde{c} = \left(\frac{u_{\tilde{c}}}{\phi(1-e)^{(1-\sigma)(1-\phi)}}\right)^{\frac{1}{\phi(1-\sigma)-1}}.$$

g) Finally, use (2.23a) in (2.18c) and define the resulting equation as a distance function $f(h)$. We solve for the root of f to obtain h_j by a non-linear solver iterating steps 4a through 4g until convergence. The following proposition establishes that this solution is unique.

Proposition 1. *For given values of human capital next period, h_{j+1}, and marginal values next period, $V'_{\tilde{a}'}$, $V'_{h'}$ and $V'_{s'}$, a solution h_j to the first-order conditions (2.18a), (2.18b), (2.18c), and the human-capital constraint (2.3) exists and is unique.*

We present the proof of Proposition 1 after the description of our algorithm.

5. Update as follows:

   a) Update $V_{\tilde{a}}$ using either (2.19a) or (2.20).

   b) Update V_h using (2.19b).

   c) Update V_s using (2.19c).

Next, loop forward on the human-capital technology (2.3) for given h_0 and $\{e_j\}_{j=1}^J$ to compute an update of h_J denoted by h_J^n. Compute the present discounted value of consumption, PVC, and, using the previously computed values $\{h_j^n\}_{j=1}^J$, $\{\ell_j^n\}_{j=1}^J$, and $\{p_j^n\}_{j=jr}^J$ compute the present discounted value of income, PVI. Use the relationship

$$\tilde{c}_0^n = \tilde{c}_0 \cdot \frac{PVI}{PVC} \tag{2.29}$$

to form an update of initial consumption, $\tilde{c}_0^n$, and next use the Euler equations for consumption to form an update of $\tilde{c}_J$, denoted as $\tilde{c}_J^n$. Define the distance functions

$$g_1(\tilde{c}_J, h_J) = \tilde{c}_J - \tilde{c}_J^n \tag{2.30a}$$

$$g_2(\tilde{c}_J, h_J) = h_J - h_J^n. \tag{2.30b}$$

In our search for general-equilibrium prices, constraints of the household model are occasionally binding. Therefore, solution of the system of equations in (2.30) using Newton-based methods, e.g., Broyden's method, is unstable. We solve this problem by a nested Brent algorithm, that is, we solve two nested univariate problems, an outer one for $\tilde{c}_J$ and an inner one for h_J.

Check for uniqueness: Observe that our nested Brent algorithm assumes that the functions in (2.30) exhibit a unique root. What is computed above is a candidate solution under the assumption that the first-order conditions are necessary and sufficient. As a consequence of potential non-convexities of our programming problem, first-order conditions may, however, not be sufficient, and our procedure may therefore not give the unique global optimum. To systematically check whether we also always converge to the unique optimum, we fix, after convergence of the household problem, a large box around the previously computed $[\tilde{c}_J, h_J]$. Precisely, we choose as boundaries for this box $\pm 50\%$ of the solutions in the respective dimensions. For these alternative starting values, we then check whether there is an additional solution to the system of equations

(2.30). For all of these combinations, our procedure always converged, and we never detected any such multiplicities.

Proof of Proposition 1. Consider the cases with and without a binding constraint on labor supply separately.

1. Consider an interior solution for labor supply, i.e., $\ell_j > 0$ and $\mu_\ell = 0$. In this case, one can find the values of e_j and h_j satisfying the first-order conditions independently of c_j and l_j. Combining the first-order conditions for labor supply (2.18b) and human-capital investment (2.18c) yields

$$e_j h_j = \left(\frac{\xi\psi(1+g)V_{h'}(\cdot)}{\tilde{w}(1-\tau)V_{\tilde{a}'}(\cdot) + (1+g)V'_{s'}\frac{1}{\bar{h}}} \right)^{\frac{1}{1-\psi}}. \tag{2.31}$$

   Note that the term on the right-hand side does not depend on h_j. Finally, substituting $e_j h_j$ in Equation (2.3) gives

$$h_{j+1} = h_j(1-\delta^h) + \xi \left(\frac{\xi\psi(1+g)V_{h'}(\cdot)}{\tilde{w}(1-\tau)V_{\tilde{a}'}(\cdot) + (1+g)V'_{s'}\frac{1}{\bar{h}}} \right)^{\frac{\psi}{1-\psi}}. \tag{2.32}$$

   Clearly, this equation has a unique solution for h_j. Given this value for h_j, Equation (2.31) determines a unique value for e_j.

2. Consider a binding constraint on labor supply, i.e., $\ell_j = 0$ and $\mu_\ell > 0$. In this case, the values of e_j and h_j satisfying the first-order conditions depend on $\tilde{c}_j$. The first-order condition for human-capital investment (2.18c) reads as

$$(1-\phi)\tilde{c}_j^{\phi(1-\sigma)}(1-e_j)^{(1-\phi)(1-\sigma)-1} = \tilde{\beta}\xi\psi e_j^{\psi-1} h_j^{\psi} V'_{h'}, \tag{2.33}$$

   and the first-order condition for consumption is given by

$$\phi\tilde{c}_j^{\phi(1-\sigma)-1}(1-e_j)^{(1-\phi)(1-\sigma)} = \tilde{\beta}\frac{1}{1+g}V'_{\tilde{a}'}. \tag{2.34}$$

   Combining these two equations to eliminate $\tilde{c}_j$ yields

$$h_j = e_j^{\frac{1-\psi}{\psi}}(1-e_j)^{\frac{1}{\psi}\frac{\sigma}{\phi(1-\sigma)-1}}\Phi^{\frac{1}{\psi}}, \tag{2.35}$$

   where

$$\Phi = (1-\phi)\left(\phi\tilde{\beta}\frac{1}{1+g}V'_{\tilde{a}'}\right)^{\frac{\phi(1-\sigma)}{\phi(1-\sigma)-1}}(\tilde{\beta}\xi\psi V'_{h'})^{-1}. \tag{2.36}$$

   The (e_j,h_j) combination we are seeking needs to satisfy Equation (2.35) and the human-capital constraint (2.3). This means we have a system of two equations in the two unknowns, e_j and h_j. Because both equations (2.35) and (2.3) are continuous, and the admissible values of e_j are in the range $0 \leq e_j \leq 1$,

- existence of a solution follows from the fact that in Equation (2.35), $h_j = 0$, if $e_j = 0$, and $h_j \to \infty$, if $e_j \to 1$ because $\phi(1-\sigma) < 1$; and in Equation (2.3), $h_j = h_{j+1}/(1-\delta^h) > 0$, if $e_j = 0$, and h_j is finite if $e_j = 1$.
- uniqueness of the solution to these two equations follows because in the relevant range of e_j, $\frac{\partial h_j}{\partial e_j} < 0$ in Equation (2.3) by the implicit-function theorem, and $\frac{\partial h_j}{\partial e_j} > 0$ in Equation (2.35) because the derivative of h_j w.r.t. e_j in Equation (2.35) is given by

$$e_j^{\frac{1-\psi}{\psi}-1}(1-e_j)^{\frac{1}{\psi}\frac{\sigma}{\phi(1-\sigma)-1}}\left(\frac{1-\psi}{\psi} - \frac{1}{\psi}\frac{\sigma}{\phi(1-\sigma)-1}(1-e_j)^{-1}\right)\Phi^{\frac{1}{\psi}},$$

  and the second term in brackets is positive for $\phi(1-\sigma) < 1$.

□

2.C.2 The Aggregate Model

For a given $r \times 1$ vector $\vec{\Psi}$ of structural model parameters, we first solve for an "artificial" initial steady state in period $t = 1$, which gives initial distributions of assets and human capital. We thereby presume that households assume prices to remain constant for all periods $t \in \{1, \dots, T\}$ and are then surprised by the actual price changes induced by the transitional dynamics. Next, we solve for the final steady state of our model, which is reached in period T and supported by our demographic projections (see Appendix 2.B). For both steady states, we solve for the equilibrium of the aggregate model by iterating on the $m \times 1$ steady-state vector $\vec{P}_{ss} = [p_1, \dots, p_m]'$. In our case, $m = 4$. p_1 is the capital intensity, p_2 are transfers (as a fraction of wages), p_3 are social-security contribution (or replacement) rates, and p_4 is the average (hours weighted) human-capital stock. Notice that all elements of $\vec{P}_{ss}$ are constant in the steady state.

The solution for the respective initial and final steady states of the model involves the following steps:

1. In iteration q for a guess of $\vec{P}_{ss}^q$ solve the household problem.
2. Update variables in $\vec{P}_{ss}$ as follows:

a) Aggregate across households to obtain aggregate assets and aggregate labor supply to form an update of the capital intensity, p_1^n.

b) Calculate an update of bequests to get p_2^n.

c) Using the update of labor supply, update social-security contribution (or replacement) rates to get p_3^n.

d) Use labor supply and human-capital decisions to form an update of the average human-capital stock, p_4^n.

3. Collect the updated variables in $\vec{P}_{ss}^n$ and notice that $\vec{P}_{ss}^n = H(\vec{P}_{ss})$ where H is a vector-valued non-linear function.

4. Define the root-finding problem $G(\vec{P}_{ss}) = \vec{P}_{ss} - H(\vec{P}_{ss})$, and iterate on $\vec{P}_{ss}$ until convergence. We use Broyden's method to solve the problem and denote the final approximate Jacobi matrix by B_{ss}.

Next, we solve for the transitional dynamics by the following steps:

1. Use the steady-state solutions to form a non-linear interpolation to obtain the starting values for the $m(T-2)\times 1$ vector of equilibrium prices, $\vec{P} = [\vec{p}_1', \dots, \vec{p}_m']'$, where $p_i, i = 1, \dots, m$ are vectors of length $(T-2)\times 1$.

2. In iteration q for guess $\vec{P}^q$, solve the household problem. We do so by iterating backwards in time for $t = T-1, \dots, 2$ to obtain the decision rules and forward for $t = 2, \dots, T-1$ for aggregation.

3. Update variables as in the steady-state solutions, and denote by $\tilde{\vec{P}} = H(\vec{P})$ the $m(T-2)\times 1$ vector of updated variables.

4. Define the root-finding problem as $G(\vec{P}) = \vec{P} - H(\vec{P})$. Because T is large, this problem is substantially larger than the steady-state root-finding problem, and we use the Gauss-Seidel-Quasi-Newton algorithm suggested in Ludwig (2007) to form and update guesses of an approximate Jacobi matrix of the system of $m(T-2)$ non-linear equations. We initialize these loops with a scaled-up version of B_{ss}.

2.C.3 Calibration of Structural Model Parameters

We split the $r \times 1$ vector of structural model parameters, $\vec{\Psi}$, as $\vec{\Psi} = \left[(\vec{\Psi}^e)', (\vec{\Psi}^f)'\right]'$. $\vec{\Psi}^f$ is a vector of predetermined (fixed) parameters, whereas the $e \times 1$ vector $\vec{\Psi}^e$ is estimated by minimum distance (unconditional matching of moments using e moment conditions). Denote by

$$u_t(\vec{\Psi}^e) = y_t - f(\vec{\Psi}^e) \text{ for } t = 1, \ldots, T_0 \tag{2.37}$$

the GMM error as the distance between actual values, y_t, and model-simulated (predicted) values, $f(\vec{\Psi}^e)$.

Under the assumption that the model is correctly specified, the restrictions on the GMM error can be written as

$$E[u_t(\vec{\Psi}_0^e)] = 0, \tag{2.38}$$

where $\tilde{\Psi}_0^e$ denotes the vector of true values. Denote sample averages of u_t as

$$g_{T_0}(\vec{\Psi}^e) \equiv \frac{1}{T_0 + 1} \sum_{t=0}^{T_0} u_t(\vec{\Psi}^e). \tag{2.39}$$

We estimate the elements of $\vec{\Psi}^e$ by setting these sample averages to zero (up to some tolerance level).

In our economic model, only two parameters are pre-determined, and we therefore have

$$\vec{\Psi}^f = [\sigma, h_0]'. \tag{2.40}$$

The vector $\vec{\Psi}^e$ is given by

$$\vec{\Psi}^e = \left[g, \alpha, \delta, \beta, \phi, \psi, \xi, \delta^h\right]'. \tag{2.41}$$

We estimate the structural model parameters using data from various sources for the period $1960, ..., 2004$. Hence $T_0 = 45$. The parameters $\vec{\Psi}_1^e = [g, \alpha]'$ are directly determined using NIPA data on GDP, fixed assets, wages and labor supply. The remaining structural model parameters, $\vec{\Psi}_2^e = [\delta, \beta, \phi, \psi, \xi, \delta^h]'$ are estimated by simulation. Our calibration targets are summarized in Table 2.5.

Determining the subset of parameters $\vec{\Psi}_2^e$ along the transition is a computationally complex problem that we translate into an equivalent simple problem. The point of

Table 2.5: Calibration Targets

Parameter	Target	Moment
$\vec{\Psi}^f$		
σ	predetermined parameter	
h_0	predetermined parameter	
$\vec{\Psi}_1^e$		
g^A	growth rate of Solow residual	0.018
α	share of wage income	0.33
$\vec{\Psi}_2^e$		
δ	investment output ratio	0.2
β	capital output ratio	2.8
ϕ	average hours worked	0.33
ψ, ξ, δ^h	coefficients of wage polynomial (from PSID)	

departure of our procedure is the insight that calibrating the model for a steady state is easy and fast. However, simulated steady-state moments may differ quite substantially from simulated averages along the transition, even when the steady state is chosen to lie in the middle of the calibration period, in our case, year 1980. We therefore proceed as follows.

1. Initialization: Choose a vector of scaling factors, $\vec{sf}$, of length e_2 that appropriately scales the steady-state calibration targets (see below).

2. Calibrate the model in some steady-state year, e.g., 1980, by solving the system of equations
$$\frac{\bar{y}_{2,i}^e}{sf_i} - f_{2,i}^{e,ss}(\vec{\Psi}) \tag{2.42}$$
for all $i = 1, \ldots, e_2$ to get $\hat{\vec{\Psi}}_2^e$. Here, $\bar{y}_{2,i}^e$ is the average of moment i in the data for the calibration period (1960-2004), e.g., the investment-output ratio for $i = 1$.

3. For the estimated parameter vector, $\hat{\vec{\Psi}}_2^e$, solve the model along the transition.

4. Compute the relevant simulated moments for the transition, $f_2^e(\vec{\Psi})$.

5. Update the vector of scaling vectors as
$$sf_i = \frac{f_{2,i}^e(\vec{\Psi})}{f_{2,i}^{e,ss}(\vec{\Psi})} \tag{2.43}$$

for all $i = 1, \ldots, e_2$.

6. Continue with step 2 until convergence on scaling factors (fixed-point problem).

We thereby translate a complex root-finding problem into a combination of a simple root-finding problem (steady-state calibration) and a fixed-point iteration on scaling factors. Because scaling factors are relatively insensitive to Ψ_2^e, convergence is fast and robust. The resulting scaling factors range from 0.94 to 1.29, which means that differences between simulated moments in the artificial steady-state year (1980) and averages during the transition are large (up to 30%). This also implies that calibrating the model in some artificial steady-state year would lead to significantly biased estimates of structural model parameters.

3 Factor Misallocation in Dual Economies

3.1 Introduction

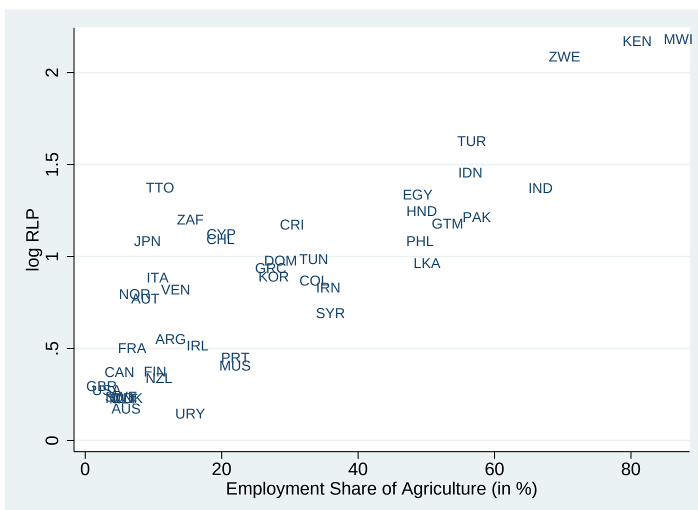
The idea of a dual economy, understood as a dualism between the agricultural and non-agricultural sector, has played a central role in economic thought on the development process and the characteristics of developing countries. A particular asymmetry between sectors might be caused by factor market imperfections or distortionary policies that drive a wedge between the marginal products in agriculture and non-agriculture, thus leading to factor misallocation. The research question of this paper is, how severely are physical and human capital misallocated between the agricultural and non-agricultural sector in different economies? Thus, this paper is concerned with estimating the cumulative effect of markets, institutions and transaction costs on allocative efficiency by investigating whether marginal value products are equalized across the two sectors.

A stylized fact, that motivates the question, is that the ratio of the average product of labor in non-agriculture to the one in agriculture is larger in poor, dominantly agricultural countries. Figure 3.1 plots the logarithm of this ratio, sometimes also called relative labor productivity, RLP[1], against the employment share in agriculture for a cross-section of countries in 1985. It is striking that countries with the largest

[1]The ratio of average products, RLP, can be calculated as $RLP = \frac{p_m Y_m / L_m}{p_a Y_a / L_a} = \frac{1-s}{s} \frac{l}{1-l}$, where p refers to prices, Y to output and L to the labor force in the two sectors a and m, and s denotes the output share of agriculture at domestic prices and l is the employment share of agriculture. While this paper always uses output in domestic prices, Caselli (2005) provides purchasing power parity (PPP) adjusted output numbers for agriculture and non-agriculture in 1985. Interestingly, a comparison between these two reveals that RLP is even higher in developing countries when it is evaluated at PPP prices. But for the purpose of this paper using domestic prices seems appropriate since factor allocation is determined by domestic prices.

average product differential employ most of their labor force in a sector where they have a relatively low productivity. However, these facts do not directly imply that factor allocation is inefficient in poor countries since an efficient allocation is characterized by equating marginal value products and not average products. Additional assumptions are necessary to justify such a conclusion. For example, the finding is inconsistent with marginal product equalization and Cobb-Douglas production functions in both sectors that are identical across countries. The reason is that for the Cobb-Douglas function average products are proportional to marginal products. This property implies that if marginal products are equalized between sectors in all countries then the average product ratio will be a constant and should not vary across countries. In contrast, for more general production functions one cannot directly draw such strong conclusions. Another possible explanation for the *RLP* variation is that skill differences between workers in the two sectors could be greater in poor than in rich countries.[2]

Figure 3.1: Logarithm of the ratio of non-agricultural to agricultural output per worker in domestic prices versus the agricultural employment share in 1985



Notes: Based on data on nominal value added shares from the World Development Indicators of the World Bank (2007) and sectoral employment shares obtained from the Food and Agriculture Organization of the United Nations (2004). For the calculation see footnote 1.

[2]Another possible explanation is that output and labor input are mismeasured in poor countries and particularly in the agricultural sector. But in the absence of hard evidence on the magnitudes of such measurement problems, I take the data at face value in this paper and try to explain the variation across countries observed in figure 3.1. Gollin, Parente, and Rogerson (2004), Temple (2005) and Temple and Wößmann (2006) provide further discussion of the average product ratio, its potential weaknesses, how it evolves over time and differs across world regions.

This paper contributes to the literature by providing a novel way to identify the degree of factor misallocation between the agricultural and non-agricultural sector based on formal econometric methods and cross-country panel data on output and factor allocations. The analytical framework addresses the different points of criticism of RLP as a measure of factor misallocation discussed above and clarifies the relation between the two. The approach is independent of the assumption that factor prices are equal to marginal products, which has formed the basis of a prior literature that identifies misallocation by factor price differentials.[3] The derivation of a meaningful relationship between RLP, the production functions in both sectors and the marginal value product differential forms the basis of the analysis. This relation as well as one for relative capital productivity, RKP, can be used to estimate the relevant parameters of the production functions in both sectors and the marginal product differentials by non-linear regressions. For the production functions this paper employs a very flexible functional form, the Translog production function, instead of a simple Cobb-Douglas. The possibility that human capital differences are responsible for the observed pattern of RLP is addressed by attempting to appropriately account for the labor input in efficiency units. A sensitivity analysis investigates the role of the critical assumptions concerning technologies and human capital differences.

The empirical findings of the paper are that there is evidence for significant capital and labor misallocation. In most developing countries, the marginal product of labor is around 1.5 to 5 times higher in non-agriculture than in agriculture, while the one of capital is around 1.2 to 3 times higher in agriculture than in the non-agricultural sector. Most industrialized countries are closer to marginal product equalization, but also tend to have a higher marginal product of labor in non-agriculture, while the conclusions for capital vary more across countries. A sensitivity analysis reveals that technological assumptions play an important role for the identification of misallocation. In particular, when Cobb-Douglas functions are used then estimated marginal product of labor differentials in developing countries are between 50 and 200 percent higher compared to the Translog case. The use of Cobb-Douglas functions can thus lead

[3]This approach is for instance prominent in the literature on implied output losses such as Dougherty and Selowsky (1973), de Melo (1977) and Williamson (1987), among others. The central drawbacks of this approach are the obvious need to control for skill differences and the possibility that factor prices depart from marginal value products at least in one of the sectors. The latter possibility has traditionally played an important role in development economics as discussed by Rosenzweig (1988) in his survey on agricultural and non-agricultural labor markets in developing countries.

to quantitatively important biases. It is also shown that the assumptions concerning human capital stocks in the two sectors affect the results, but in a more limited way.

The question and analytical framework of this paper relate to and are motivated by several strands of the literature. First, it is related to a long history of thought in development economics on economic dualism, the structural transformation and the role of agriculture for economic development. Recent studies in this vein that abstract from marginal product differentials or misallocation are Caselli and Coleman (2001), Caselli (2005), Gollin, Parente, and Rogerson (2002, 2004, 2007) and Duarte and Restuccia (2010), among others. Some papers that investigate the role of inefficient factor allocations are Chanda and Dalgaard (2008), Córdoba and Ripoll (2006), Hayashi and Prescott (2008), Restuccia, Yang, and Zhu (2008), Temple (2002), Temple and Wößmann (2006) and Vollrath (2009), also see Temple (2005) for a survey. Most of these studies employ calibrated models and virtually all of them are based on Cobb-Douglas production functions. This paper contributes to that literature by providing an econometric framework for the estimation of marginal product differentials and by allowing for more general production functions.

The most closely related paper is the one of Vollrath (2009). Based on a cross-country calibration exercise of Cobb-Douglas production functions he finds substantial marginal product differentials between agriculture and non-agriculture particularly in developing countries. He also shows that this can explain a large fraction of the aggregate efficiency variation found by the development accounting literature, see for example Hall and Jones (1999) and Caselli (2005). In contrast to Vollrath's study, my econometric procedure can allow for general production functions and thus employs the more flexible Translog instead of a Cobb-Douglas production function. Accordingly, I find much lower labor misallocation in developing countries.

A second line of related research is concerned with misallocation on the microeconomic level between different firms in the economy or within an industry. There is extensive evidence on a great heterogeneity of rates of return to the same factor between different firms in the microeconomic development literature surveyed by Banerjee and Duflo (2005). Furthermore, the role of misallocation between firms for aggregate efficiency is studied by Restuccia and Rogerson (2008) and within the manufacturing sector in China and India by Hsieh and Klenow (2009). My paper complements this strand of research by providing macroeconomic evidence on economic dualism and factor misallocation between different broad sectors of the economy.

The next section develops the framework and estimation procedure for the analysis of factor misallocation. Section 3.3 describes the data set of a panel of developed and developing countries. The results are presented in section 3.4 and section 3.5 concludes.

3.2 Analytical Framework

3.2.1 The Relationship between the Average and Marginal Product Ratio

Consider an economy comprised of two sectors, agriculture and manufacturing[4]. Total Output in domestic prices Y is given by $Y = p_a Y_a + p_m Y_m$, where p_a, Y_a and p_m, Y_m refer to prices and output in agriculture and manufacturing. Let K_a, L_a and K_m, L_m denote capital and labor employed in the two sectors. Of prime interest to this paper is the question whether marginal value products are equalized across sectors, i.e. whether the marginal product ratios $d_k \equiv \frac{p_m \partial Y_m/\partial K_m}{p_a \partial Y_a/\partial K_a}$ and $d_l \equiv \frac{p_m \partial Y_m/\partial L_m}{p_a \partial Y_a/\partial L_a}$ are equal to one. The following key equations of the paper show that a meaningful relationship between the *average product ratio* of a factor and the corresponding *marginal product ratio* exists as[5]

$$RKP \equiv \frac{p_m \frac{Y_m}{K_m}}{p_a \frac{Y_a}{K_a}} = \frac{p_m \frac{\partial Y_m}{\partial K_m}}{p_a \frac{\partial Y_a}{\partial K_a}} \frac{\frac{\partial Y_a}{\partial K_a}\frac{K_a}{Y_a}}{\frac{\partial Y_m}{\partial K_m}\frac{K_m}{Y_m}} = \frac{\frac{\partial Y_a}{\partial K_a}\frac{K_a}{Y_a}}{\frac{\partial Y_m}{\partial K_m}\frac{K_m}{Y_m}} d_k \tag{3.1}$$

$$RLP \equiv \frac{p_m \frac{Y_m}{L_m}}{p_a \frac{Y_a}{L_a}} = \frac{p_m \frac{\partial Y_m}{\partial L_m}}{p_a \frac{\partial Y_a}{\partial L_a}} \frac{\frac{\partial Y_a}{\partial L_a}\frac{L_a}{Y_a}}{\frac{\partial Y_m}{\partial L_m}\frac{L_m}{Y_m}} = \frac{\frac{\partial Y_a}{\partial L_a}\frac{L_a}{Y_a}}{\frac{\partial Y_m}{\partial L_m}\frac{L_m}{Y_m}} d_l. \tag{3.2}$$

The equations show that one can only draw conclusions from data on RKP or RLP on the marginal product differential if one also uses information and assumptions on the ratio of the output elasticity in agriculture to the one in manufacturing at the present allocation. From here, there are at least two ways to proceed with the analysis - an econometric and a calibration approach. Both of them have their respective advantages and require different identifying assumptions.

[4] I follow the convention in the literature and name the second sector manufacturing even though it is meant to represent the total non-agricultural part of the economy.

[5] The equation uses the simple fact that average products are equal to marginal products times the inverse of the elasticity.

This paper employs an econometric approach that rests on assumptions ensuring that the output elasticity of a factor does only depend on observable factors and not on unobservable technology terms. Two well-known cases that have these properties are general production functions with factor-neutral (Hicks-neutral) technological change and the Cobb-Douglas production function irrespective of whether technology is factor-neutral, labor or capital augmenting since the output elasticity of the Cobb-Douglas is in all cases simply a constant. The first case will be used for the following derivation since it essentially also covers the second one. Assume output in the two sectors is produced according to

$$Y_a = A_a F(K_a, L_a, T_a) \tag{3.3}$$

$$Y_m = A_m G(K_m, L_m), \tag{3.4}$$

where T_a denotes land, which is only used in agricultural production. A_a and A_m refer to the level of TFP in agriculture and manufacturing. The functions F and G are assumed to satisfy the standard neoclassical properties. The output elasticity of a factor is then equal to the respective elasticity of F or G. Accordingly, the equations for RKP and RLP read as

$$RKP = \frac{\frac{\partial F}{\partial K_a}\frac{K_a}{F}}{\frac{\partial G}{\partial K_m}\frac{K_m}{G}} d_k \tag{3.5}$$

$$RLP = \frac{\frac{\partial F}{\partial L_a}\frac{L_a}{F}}{\frac{\partial G}{\partial L_m}\frac{L_m}{G}} d_l. \tag{3.6}$$

The elasticities on the right hand side only depend on the allocation of factors of production, but no longer on unobservable technology levels. This property simplifies the following econometric analysis considerably and enables me (together with additional identifying assumptions that are explained in the next section) to use standard non-linear estimation methods[6]. Intuitively, this approach decomposes the variation of RKP and RLP into a part that is due to the variation of observable factors of production and an unobservable part that is attributed to marginal product differentials. The details will be explained in the following section.

[6] In principle, to the extend that the technology level does affect the output elasticity, variation in the relative technology level in the two sectors between countries could explain part of the RKP and RLP variation across countries and time periods. This means that it is theoretically possible that there is a purely technological explanation for the observed variation of average product ratios even without marginal product differentials. At the same time more general forms of technology could also have the opposite effect and be consistent with *true* marginal product differentials that are larger than the ones identified by this paper.

We can also relate the two equations back to the discussion of Figure 3.1 in the introduction and the example of Cobb-Douglas production functions. For a Cobb-Douglas, the output elasticity of a factor is a constant and equal to the exponent to the factor in the production function.[7] Accordingly, any variation of RKP or RLP across countries or time periods must be attributed to changes of the marginal product differential. One could visualize this by drawing a horizontal line in Figure 3.1 at a level equal to the logarithm of the output elasticity ratio (even though we do not know it yet) and any deviations from the line would then represent the logarithm of the marginal product differential. This illustrates again how restrictive the Cobb-Douglas production function is in this context. Accordingly, this paper allows for more flexible production functions where the output elasticity may depend on the allocation of factors of production.

In contrast to an econometric procedure, a calibration approach could rely on the theoretical prediction that under perfect competition factor prices are equal to marginal value products. This in turn implies that the output elasticity of a factor is equal to the factor income share[8]. The marginal product differentials could then be calculated directly from data on RKP, RLP and sector-specific factor income shares for each country and time period. From this discussion it is clear that both approaches have their benefits. While the calibration approach is independent of a specific functional form of the production functions, restrictions on technological change and further identifying assumptions described in the next section, the econometric approach does not rely on factor prices being equal to marginal value products. Given the controversial discussion in development economics on factor price formation and especially wages in agriculture, the independence of the econometric approach from the way factor prices are related to marginal products is an important advantage. Another complication of the calibration approach is that reported data on labor income shares tends to underestimate the true labor income share due to the fact that labor income of the self-employed is often treated as capital income as argued by Gollin (2002). This measurement problem might be particularly severe in the agricultural sector in poor countries due to the existence of small-scale subsistence farming. Though I do not

[7] Therefore, Cobb-Douglas production functions are also immune to the "purely technological" criticism mentioned in the previous footnote. A Cobb-Douglas can only explain variation in RKP and RLP when the share parameters vary.

[8] For example for labor, if the wage rate w is given by $w = p\frac{\partial Y}{\partial L}$ then the labor income share $\frac{wL}{pY}$ is given by $\frac{wL}{pY} = \frac{\partial Y}{\partial L}\frac{L}{Y}$.

follow the calibration approach here, I am planning to investigate it in more detail in future work.

3.2.2 The Econometric Model

This section discusses the estimation method and further identifying assumptions. This paper follows a standard practice in applied economics by first deriving an economic model and then adding a stochastic error term to the equations. Taking logarithms of equations (3.5) and (3.6) and adding error terms ε_k and ε_l yields

$$\ln(RKP_{it}) = f(X_{it}, \beta) + \delta_{ki} + \varepsilon_{kit} \quad (3.7)$$

$$\ln(RLP_{it}) = g(X_{it}, \beta) + \delta_{li} + \varepsilon_{lit}, \quad (3.8)$$

where

$$f(.) \equiv \ln\left(\frac{\partial F}{\partial K_a}\frac{K_a}{F} / \frac{\partial G}{\partial K_m}\frac{K_m}{G}\right),$$

$$g(.) \equiv \ln\left(\frac{\partial F}{\partial L_a}\frac{L_a}{F} / \frac{\partial G}{\partial L_m}\frac{L_m}{G}\right),$$

$$X_{it} \equiv (K_{ait}, L_{ait}, T_{ait}, K_{mit}, L_{mit}),$$

$$\delta_{ki} \equiv \ln(d_{ki}),$$

$$\delta_{li} \equiv \ln(d_{li}),$$

and β are the parameters of the two production functions, i is a country index and t a time index. The error terms ε_k and ε_l are assumed to have mean zero and are *iid* across countries and time periods and strictly exogenous with respect to the explanatory variables of the model. But the errors of the two equations may be correlated. Accordingly, $(\varepsilon_{kit}, \varepsilon_{lit})'$ may have a general variance-covariance matrix Σ.

The country-specific marginal value product differentials d_k and d_l, and hence the terms δ_{ki} and δ_{li}, are assumed to be constant over time. This assumption is based on the view that institutional features of a country tend to be very persistent over time. Examples of such institutions that could drive a wedge between marginal products include tax rates on wage income or expropriation risk of capital returns that differ between the two sectors, or there could be intersectoral mobility barriers like migration costs or costly education requirements in the urban sector. Though it seems plausible that these institutions are unlikely to change very much from year to year, I also allow the marginal product differentials to vary over time in the sensitivity analysis of section 3.4.2. There I find that the results obtained in the baseline specification with constant wedges are relatively robust. The prime interest of the paper is to identify the terms δ_{ki} and δ_{li} (and consequently d_{ki} and d_{li}) for each country and to check whether they are zero (one).[9]

The choice of the production functions is also an important identifying assumption. This choice could in principle be guided by evidence from past studies, but essentially it is to a certain extent arbitrary. One could address this issue by experimenting with different functional forms and reporting their robustness or by adopting very general production functions that nest different well-known cases. This paper uses to some extend both approaches by employing the flexible Translog production function which can be interpreted as a second order approximation to a general production function[10] and by also reporting results for the Cobb-Douglas as a sensitivity analysis. Specifically, the production function in agriculture is given by

$$\ln F = \alpha_{Ka} \ln K_a + \alpha_{La} \ln L_a + \alpha_{Ta} \ln T_a + \frac{1}{2}\beta_{KKa}(\ln K_a)^2 + \frac{1}{2}\beta_{LLa}(\ln L_a)^2 \\ + \frac{1}{2}\beta_{TTa}(\ln T_a)^2 + \beta_{KLa}(\ln K_a)(\ln L_a) + \beta_{KTa}(\ln K_a)(\ln T_a) + \beta_{LTa}(\ln L_a)(\ln T_a),$$

and the one in manufacturing by

$$\ln G = \alpha_{Km} \ln K_m + \alpha_{Lm} \ln L_m + \frac{1}{2}\beta_{KKm}(\ln K_m)^2 + \frac{1}{2}\beta_{LLm}(\ln L_m)^2 \\ + \beta_{KLm}(\ln K_m)(\ln L_m).$$

The relevant functions for the estimation, the ratios of the output elasticities, f and g, in equations (3.7) and (3.8) then read as

$$\begin{aligned} f(X_{it}, \beta) &= \ln\left(\frac{\alpha_{Ka} + \beta_{KKa} \ln K_{ait} + \beta_{KLa} \ln L_{ait} + \beta_{KTa} \ln T_{ait}}{\alpha_{Km} + \beta_{KKm} \ln K_{mit} + \beta_{KLm} \ln L_{mit}}\right) \\ g(X_{it}, \beta) &= \ln\left(\frac{\alpha_{La} + \beta_{LLa} \ln L_{ait} + \beta_{KLa} \ln K_{ait} + \beta_{LTa} \ln T_{ait}}{\alpha_{Lm} + \beta_{LLm} \ln L_{mit} + \beta_{KLm} \ln K_{mit}}\right). \end{aligned}$$

[9]Note that the specifications in this paper do not include time effects. Exploring the possibility that there are shocks affecting all countries in the same way in a given year could be interesting, but is left for future research.

[10]Greene (2003, pp. 12-13) provides a derivation and Berndt and Christensen (1973) discuss some properties of the Translog.

These two equations reveal two related identification problems associated with production functions of the Translog form. First, one can multiply all parameters in the numerator and denominator of the ratio by the same constant without changing the ratio. Second, numerator and denominator implicitly contain an additive constant term in addition to the country-specific terms δ_{ki} and δ_{li} reflecting the marginal product differentials. Rewriting the two equations of the model in terms of identifiable parameters yields

$$\begin{aligned} \ln(RKP_{it}) &= \ln\left(\frac{\pi_{k1} + \ln K_{ait} + \pi_{k2}\ln L_{ait} + \pi_{k3}\ln T_{ait}}{\pi_{k4} + \ln K_{mit} + \pi_{k5}\ln L_{mit}}\right) + \alpha_{ki} + \varepsilon_{kit} \quad (3.9)\\ \ln(RLP_{it}) &= \ln\left(\frac{1 + \pi_{l1}\ln L_{ait} + \pi_{l2}\ln K_{ait} + \pi_{l3}\ln T_{ait}}{\pi_{l4} + \ln L_{mit} + \pi_{l5}\ln K_{mit}}\right) + \alpha_{li} + \varepsilon_{lit}, \quad (3.10) \end{aligned}$$

where[11]

$$\begin{aligned} \pi_{k0} &= \frac{\beta_{KKa}}{\beta_{KKm}}, \pi_{k1} = \frac{\alpha_{Ka}}{\beta_{KKa}}, \pi_{k2} = \frac{\beta_{KLa}}{\beta_{KKa}}, \pi_{k3} = \frac{\beta_{KTa}}{\beta_{KKa}}, \pi_{k4} = \frac{\alpha_{Km}}{\beta_{KKm}}, \pi_{k5} = \frac{\beta_{KLm}}{\beta_{KKm}},\\ \pi_{l0} &= \frac{\alpha_{La}}{\beta_{LLm}}, \pi_{l1} = \frac{\beta_{LLa}}{\alpha_{La}}, \pi_{l2} = \frac{\beta_{KLa}}{\alpha_{La}}, \pi_{l3} = \frac{\beta_{LTa}}{\alpha_{La}}, \pi_{l4} = \frac{\alpha_{Lm}}{\beta_{LLm}}, \pi_{l5} = \frac{\beta_{KLm}}{\beta_{LLm}},\\ \alpha_{ki} &= \ln\pi_{k0} + \delta_{ki}, \alpha_{li} = \ln\pi_{l0} + \delta_{li}. \end{aligned}$$

Accordingly, one cannot identify the distortion terms δ_{ki} and δ_{li} independently from π_{k0} and π_{l0}, but only the fixed effects $\alpha_{ki} = \ln\pi_{k0} + \delta_{ki}$ and $\alpha_{li} = \ln\pi_{l0} + \delta_{li}$ are identified. Nevertheless, the difference between the fixed effects estimates of two countries is meaningful. If one is able to choose a reference country, say country 1, with perfectly integrated factor markets then $\widehat{\alpha}_{ji} - \widehat{\alpha}_{j1}$ is an unbiased estimate of country i's distortion term δ_{ji} for factor $j = k, l$, and the implied $d_{ji} = \exp(\widehat{\alpha}_{ji} - \widehat{\alpha}_{j1})$ estimates country i's marginal value product differential. Of course, this means that the final estimates of the marginal product differentials depend to a certain extend on the preconceptions of the researcher[12]. This paper will use the United States as the reference country with

[11]This transformation implicitly rests on the assumption that the parameter ratios $\frac{\beta_{KKa}}{\beta_{KKm}}$ and $\frac{\alpha_{La}}{\beta_{LLm}}$ are positive. If they are negative one will need minus signs at the appropriate places inside the log functions. The implementation of the estimation problem allows for this possibility. Also note that no restrictions such as constant returns to scale are imposed on the parameters.

[12]We can again relate this finding back to the discussion of the Cobb-Douglas production function, which is a special case of a Translog. Consider drawing a "no marginal product differential" line in Figure 3.1. It is obvious now that data on output and factor allocations in the two sectors alone do not inform us about the location of this line. Instead one needs additional information or assumptions. Standard calibration exercises based on the Cobb-Douglas production function as the one of Vollrath (2009) typically rely on the assumption that wages are equal to marginal products (in the U.S.) and thus estimate the relevant parameter by factor income shares from the United States.

presumably undistorted factor markets. But the reader is welcome to choose another country. In case the reference country should not have perfectly integrated factor markets, i.e. have true δ's equal to zero, one still estimates the distortion of country i relative to the one of the reference country in an unbiased way.

Equations (3.9) and (3.10) form a system of two nonlinear seemingly unrelated regression equations[13]. The production function parameters π and country fixed-effects α_i are estimated by a two-step feasible generalized least squares procedure. The first step uses the identity matrix as an initial weighting matrix for the errors of the two equations. In the absence of cross-equation parameter restrictions, as is the case here, this is equivalent to equation-by-equation nonlinear least squares (NLLS). From the estimated first step residuals one can compute a consistent estimate of the error covariance matrix Σ, which is then used as the weighting matrix in the second step.

3.3 Data Sources

The data set used in this study is a panel of 49 developed and developing countries observed between 1967 and 1992.[14] The panel is unbalanced and the observed time periods range from 3 to 26 years with the average being approximately 21.5 years. The total number of country-year observations is 1055. The remainder of this section describes the data sources and provides a discussion of how this paper attempts to control for human capital differences between the two sectors.

The ratio of non-agricultural to agricultural output in domestic prices, $\frac{p_m Y_m}{p_a Y_a}$, can be computed as $\frac{1-s}{s}$, where s is the share of agriculture in nominal value added obtained from the World Bank's (2007) World Development Indicators.

[13]Gallant (1987) provides an excellent discussion of non-linear estimation methods including system estimation.

[14]The observed countries and time periods are: Argentina 1967-1992, Australia 1971-1992, Austria 1971-1992, Canada 1971-1992, Chile 1967-1992, Colombia 1967-1992, Costa Rica 1967-1992, Cyprus 1975-1992, Denmark 1971-1992, Dominican Republic 1967-1992, Egypt 1975-1992, El Salvador 1990-1992, Finland 1971-1992, France 1971-1992, Greece 1971-1992, Guatemala 1967-1992, Honduras 1967-1992, India 1967-1992, Indonesia 1967-1992, Iran 1967-1992, Ireland 1971-1992, Italy 1971-1992, Japan 1971-1992, Kenya 1967-1992, Malawi 1967-1992, Malta 1967-1992, Mauritius 1977-1992, Netherlands 1971-1992, New Zealand 1971-1992, Norway 1971-1992, Pakistan 1967-1992, Peru 1970-1979&1986-1992, Philippines 1967-1992, Poland 1985-1992, Portugal 1971-1992, South Africa 1967-1992, South Korea 1967-1992, Sri Lanka 1967-1992, Sweden 1971-1992, Syria 1985-1992, Tanzania 1990-1992, Trinidad & Tobago 1984-1992, Tunisia 1967-1992, Turkey 1968-1992, United Kingdom 1971-1992, United States 1971-1992, Uruguay 1983-1992, Venezuela 1967-1992, Zimbabwe 1967-1992.

The Food and Agriculture Organization of the United Nations (2004) provides data on the number of economically active in agriculture and in total, that is used to determine the number of workers in agriculture, N_a, and non-agriculture, N_m.

Data on physical capital is taken from Crego, Larson, Butzer, and Mundlak (1998), who have constructed a database on agricultural, manufacturing and economy-wide fixed capital. For the capital stock in agriculture, K_a, I use their series on total agricultural capital that contains fixed capital, as well as livestock and tree capital. Non-agricultural capital stocks, K_m, are calculated as economy-wide fixed capital minus agricultural fixed capital.

For agricultural land, T_a, I use data on the amount of arable land in hectares from the World Development Indicators.

Following the literature, this paper determines the amount of labor input in efficiency units as the product between the number of workers and their average human capital stock. Human capital stocks of workers are modeled as a function of years of schooling. Years of education by sector are obtained from Timmer (2002), who provides a panel data set of 65 developing countries on average years of education per person over the age of 25 for the rural and urban population. I apply the rural years of education to all agricultural workers and the urban years to all non-agricultural workers. Since the data set only contains developing countries and only has data at a five year frequency from 1960 to 1985 I follow Vollrath (2009) and rely on an extra- and interpolation technique to construct sectoral years of schooling for other countries and time periods. There exists a regularity between the sectoral education data and data on the level of overall education provided by Barro and Lee (2001). Specifically, one can observe that rural years of education converge towards urban years of education as total years of education increase. This relationship is then used to construct years of schooling for mainly industrialized countries that only have data on the overall level of education as well as for all countries for the time periods 1990 and 1995 that are not covered by the sectoral data. Finally, a linear interpolation between the real and imputed data points at a five year frequency is used to get annual data. Details on the applied procedure are provided in appendix 3.A.

Average years of schooling by sector, s, are translated into sectoral human capital stocks, h, using a Mincerian technique, specifically $h = \exp(\phi(s))$. Hall and Jones (1999) assume that $\phi(s)$ is piecewise linear and that the return to schooling is 13.4%

for the first four years, 10.1% for the next four years and 6.8% for subsequent years[15]. The baseline specification of this paper follows Vollrath (2009) in using this functional form, but applying these rates of return only to the non-agricultural sector and only one half of them to the agricultural sector. Essentially, this choice of a relative return to schooling in agriculture is arbitrary. Accordingly, a sensitivity analysis explores the robustness of the baseline results with respect to the two polar cases of either the relative return being equal to one or zero, i.e. returns to schooling being identical in both sectors and the case where schooling does not increase labor productivity in agriculture. These robustness checks will reveal that the choice of a *relative* return does affect the results, but this effect is limited.[16]

3.4 Results

3.4.1 Baseline Specification

This section presents the results of my baseline specification which is characterized by the use of Translog production functions, constant marginal product wedges and returns to schooling in agriculture that are only one half of those in non-agriculture as described above. The next section then performs a sensitivity analysis concerning changes to production technologies, assumptions on the wedges and the human capital calibration.

Table 3.1 columns (1) and (2) present the marginal value product differentials of labor and capital implied[17] by the fixed effects estimates for the baseline specification. For

[15] Specifically, $\phi(s) = 0.134 \cdot s$ if $s \leq 4$, $\phi(s) = 0.134 \cdot 4 + 0.101 \cdot (s-4)$ if $4 < s \leq 8$, $\phi(s) = 0.134 \cdot 4 + 0.101 \cdot 4 + 0.068 \cdot (s-8)$ if $s > 8$. Bils and Klenow (2000) also use Mincer regression coefficients to construct human capital stocks and provide evidence of diminishing returns to schooling.

[16] Another potential concern with the used numbers for the return to schooling is that they are based on wage regressions (Psacharopoulos 1994) that not always included industry dummies. If wage differentials between sectors exist and the well-educated workers work in the high-wage sector then one would overestimate the pure human capital effect of schooling that makes workers more productive. This might be inconsequential if one is only interested in the individual return to schooling such that the possibility to work in a high-wage sector could be regarded as an important component of this return. But if the primary interest lies in appropriately accounting for the labor input in efficiency units then one would make a mistake. As a robustness check, returns to schooling were scaled down by a factor of 0.75. The (not reported) results of this exercise revealed that the *absolute* level of returns to schooling does not affect the baseline results in a significant way.

[17] As explained above, the reported estimates are based on choosing the United States as the reference country whose factor markets are suspected to be perfectly integrated. In case the reference country

purposes of clarity, figure 3.2 plots the estimated marginal product differentials against the employment share of agriculture, which may be considered as a measure of economic development, and contrasts the findings for capital and labor. In 44 countries the marginal value product of labor is higher in non-agriculture than in agriculture indicating that factor markets might be distorted towards agriculture. For 5 countries the opposite holds. Marginal value products of capital are higher in agriculture than in non-agriculture in 30 countries and lower in 19 countries. For most countries, the deviation of marginal product differentials from one is statistically significant. This finding suggests that labor markets seem to be mostly distorted towards agriculture and capital markets in favor of non-agriculture. In addition to the type of distortions it is also interesting to compare the magnitude of the marginal product differentials. The wedges between marginal products of labor seem to be higher and more common than those of capital. For example, there are 16 countries with marginal product of labor differentials above 2, but only five countries with marginal product of capital differentials below 0.5. Using the agricultural employment share as a measure for development, many industrialized countries seem to be relatively close to marginal product of labor equalization across sectors, while in developing countries the marginal product of labor is higher in non-agriculture. In developing countries, the marginal product of capital tends to be higher in agriculture than in non-agriculture. In contrast, for industrialized countries the results for capital vary more across countries. When considering the combination of distortions to capital and labor markets there are two main groups of countries. One group distorts both labor and capital markets in favor of agriculture, while the other one distorts labor markets in favor of agriculture and capital markets in favor of non-agriculture. Most developing countries in the data set except the least developed ones are members of the latter group. A third smaller group seems to distort both capital and labor markets in favor of non-agriculture. These results support the view that economic dualism, understood as imperfectly integrated factor markets, might be an important characteristic of developing countries.

should not have perfectly integrated factor markets, i.e. have true δ's equal to zero, one still estimates the distortion of country i relative to the one of the reference country in an unbiased way. For convenience, I will base most of the following discussion on the assumption that the United States indeed have perfectly integrated factor markets. But the reader is welcome to insert the caveat at the appropriate places. Table 3.1 provides all necessary information for the reader to compute the implied marginal product differentials choosing another country as the reference country.

Table 3.1: Marginal Value Product Differentials

					Sensitivity Analysis															
	Baseline Specification				**Cobb-Douglas**				**Time Trends**				**Zero Return in Agric.**				**Same Return**			
	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)		(9)		(10)	
Country	$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$		$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$		$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$		$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$		$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$	
Argentina	1.30	(0.07)	1.16	(0.08)	1.68	(0.07)	1.39	(0.11)	1.38	(0.08)	1.18	(0.06)	1.46	(0.07)	1.18	(0.08)	1.17	(0.06)	1.14	(0.07)
Australia	0.77	(0.04)	0.74	(0.04)	0.97	(0.05)	0.89	(0.04)	0.84	(0.05)	0.90	(0.04)	0.75	(0.04)	0.73	(0.04)	0.81	(0.05)	0.76	(0.04)
Austria	1.67	(0.09)	1.11	(0.05)	1.91	(0.06)	1.00	(0.03)	1.79	(0.14)	0.81	(0.05)	1.82	(0.11)	1.17	(0.05)	1.51	(0.08)	1.06	(0.05)
Canada	1.11	(0.05)	1.12	(0.04)	1.24	(0.05)	1.33	(0.04)	1.19	(0.05)	1.24	(0.04)	1.09	(0.04)	1.12	(0.04)	1.15	(0.05)	1.14	(0.04)
Chile	2.19	(0.13)	1.40	(0.07)	3.18	(0.19)	1.55	(0.07)	2.36	(0.19)	1.15	(0.07)	2.52	(0.15)	1.47	(0.08)	1.86	(0.11)	1.34	(0.07)
Colombia	1.54	(0.09)	0.81	(0.04)	2.38	(0.10)	0.70	(0.04)	1.63	(0.14)	0.50	(0.04)	2.02	(0.12)	0.90	(0.05)	1.15	(0.07)	0.72	(0.04)
Costa Rica	1.53	(0.13)	1.02	(0.08)	2.31	(0.11)	1.14	(0.06)	1.80	(0.23)	0.68	(0.06)	1.76	(0.17)	1.17	(0.10)	1.28	(0.11)	0.91	(0.07)
Cyprus	1.50	(0.13)	0.85	(0.07)	2.60	(0.13)	0.82	(0.04)	1.85	(0.24)	0.58	(0.05)	1.46	(0.13)	0.92	(0.08)	1.44	(0.12)	0.82	(0.06)
Denmark	1.34	(0.06)	1.07	(0.05)	1.32	(0.06)	1.16	(0.03)	1.51	(0.09)	0.92	(0.05)	1.34	(0.06)	1.10	(0.05)	1.34	(0.07)	1.06	(0.04)
Dominican Rep.	1.37	(0.09)	0.52	(0.03)	2.59	(0.15)	0.48	(0.02)	1.52	(0.14)	0.38	(0.03)	1.74	(0.12)	0.59	(0.04)	1.04	(0.06)	0.48	(0.03)
Egypt	2.32	(0.16)	0.84	(0.06)	4.33	(0.27)	0.39	(0.01)	2.40	(0.22)	0.36	(0.04)	3.49	(0.27)	1.05	(0.08)	1.49	(0.10)	0.66	(0.04)
El Salvador	1.56	(0.12)	0.56	(0.05)	3.00	(0.16)	0.42	(0.03)	1.73	(0.17)	0.33	(0.03)	2.06	(0.16)	0.67	(0.06)	1.14	(0.09)	0.49	(0.04)
Finland	1.12	(0.06)	0.77	(0.04)	1.50	(0.06)	0.74	(0.03)	1.23	(0.09)	0.65	(0.03)	1.17	(0.06)	0.79	(0.04)	1.06	(0.06)	0.75	(0.03)
France	1.39	(0.05)	0.96	(0.03)	1.64	(0.06)	0.85	(0.02)	1.40	(0.06)	0.81	(0.04)	1.63	(0.07)	1.00	(0.03)	1.20	(0.05)	0.92	(0.03)
Greece	1.45	(0.08)	0.61	(0.03)	2.87	(0.15)	0.47	(0.02)	1.53	(0.12)	0.47	(0.03)	1.67	(0.10)	0.63	(0.03)	1.23	(0.07)	0.59	(0.03)
Guatemala	1.77	(0.11)	0.63	(0.04)	3.83	(0.15)	0.54	(0.03)	1.98	(0.19)	0.46	(0.04)	2.39	(0.16)	0.73	(0.04)	1.27	(0.08)	0.56	(0.03)
Honduras	1.93	(0.14)	0.85	(0.07)	3.91	(0.20)	1.18	(0.11)	2.24	(0.23)	0.83	(0.06)	2.36	(0.17)	0.91	(0.07)	1.53	(0.11)	0.80	(0.06)
India	2.20	(0.12)	0.81	(0.06)	3.59	(0.13)	0.58	(0.02)	2.25	(0.15)	0.66	(0.06)	3.25	(0.18)	0.83	(0.06)	1.46	(0.08)	0.72	(0.04)
Indonesia	2.11	(0.13)	0.54	(0.04)	4.03	(0.23)	0.25	(0.01)	2.20	(0.16)	0.36	(0.04)	3.02	(0.21)	0.54	(0.04)	1.43	(0.09)	0.49	(0.03)
Iran	1.46	(0.15)	0.47	(0.04)	3.72	(0.39)	0.34	(0.03)	1.47	(0.13)	0.45	(0.03)	2.26	(0.24)	0.50	(0.04)	0.96	(0.10)	0.43	(0.04)
Ireland	1.20	(0.08)	0.96	(0.05)	1.48	(0.06)	1.07	(0.04)	1.38	(0.13)	0.77	(0.05)	1.23	(0.08)	1.00	(0.05)	1.15	(0.08)	0.93	(0.05)
Italy	2.04	(0.11)	1.37	(0.06)	2.37	(0.08)	1.13	(0.04)	2.05	(0.13)	1.00	(0.06)	2.58	(0.15)	1.47	(0.06)	1.63	(0.09)	1.26	(0.05)
Japan	2.74	(0.20)	1.45	(0.08)	2.85	(0.10)	0.88	(0.05)	2.59	(0.24)	0.77	(0.06)	3.29	(0.31)	1.55	(0.09)	2.18	(0.14)	1.32	(0.07)
Kenya	4.18	(0.28)	1.21	(0.07)	10.17	(0.50)	0.84	(0.05)	4.62	(0.45)	0.87	(0.08)	5.69	(0.40)	1.31	(0.08)	2.99	(0.20)	1.07	(0.06)
Malawi	3.37	(0.24)	0.83	(0.09)	9.71	(0.55)	0.35	(0.02)	3.91	(0.42)	0.62	(0.08)	4.33	(0.33)	0.89	(0.10)	2.49	(0.18)	0.78	(0.08)
Malta	0.68	(0.07)	0.29	(0.04)	1.10	(0.08)	0.29	(0.01)	0.91	(0.12)	0.19	(0.02)	0.64	(0.08)	0.36	(0.05)	0.61	(0.06)	0.28	(0.03)
Mauritius	0.80	(0.06)	0.33	(0.03)	1.58	(0.07)	0.30	(0.01)	0.95	(0.11)	0.21	(0.02)	0.89	(0.07)	0.39	(0.04)	0.67	(0.05)	0.30	(0.03)

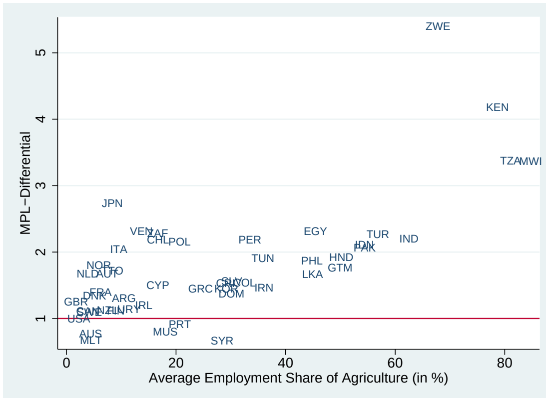
Table 3.1 (Continued)

	Baseline Specification				Sensitivity Analysis: Cobb-Douglas				Time Trends				Zero Return in Agric.				Same Return			
	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)		(9)		(10)	
Country	$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$		$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$		$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$		$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$		$\frac{MPL_m}{MPL_a}$		$\frac{MPK_m}{MPK_a}$	
Netherlands	1.67	(0.13)	1.02	(0.05)	1.20	(0.05)	0.86	(0.03)	1.75	(0.19)	0.64	(0.04)	1.89	(0.19)	1.11	(0.06)	1.43	(0.10)	0.95	(0.05)
New Zealand	1.12	(0.07)	1.11	(0.06)	1.04	(0.05)	1.40	(0.07)	1.33	(0.10)	1.01	(0.06)	1.04	(0.06)	1.12	(0.06)	1.20	(0.08)	1.11	(0.06)
Norway	1.80	(0.11)	1.13	(0.06)	1.82	(0.07)	1.03	(0.03)	2.02	(0.18)	0.84	(0.05)	1.91	(0.14)	1.18	(0.06)	1.66	(0.11)	1.09	(0.05)
Pakistan	2.07	(0.11)	0.76	(0.04)	3.60	(0.12)	0.68	(0.03)	2.19	(0.16)	0.60	(0.04)	3.10	(0.18)	0.84	(0.04)	1.35	(0.07)	0.66	(0.03)
Peru	2.18	(0.15)	0.67	(0.04)	4.58	(0.22)	0.45	(0.02)	2.26	(0.18)	0.49	(0.03)	2.81	(0.18)	0.70	(0.04)	1.65	(0.12)	0.63	(0.04)
Philippines	1.87	(0.11)	0.76	(0.04)	3.09	(0.12)	0.48	(0.03)	1.96	(0.16)	0.33	(0.03)	2.27	(0.15)	0.78	(0.04)	1.48	(0.08)	0.72	(0.04)
Poland	2.15	(0.26)	1.34	(0.16)	2.85	(0.32)	1.16	(0.10)	2.21	(0.18)	1.01	(0.09)	2.36	(0.32)	1.36	(0.17)	1.90	(0.21)	1.30	(0.15)
Portugal	0.92	(0.06)	0.40	(0.02)	1.68	(0.09)	0.34	(0.02)	1.01	(0.08)	0.32	(0.02)	1.21	(0.08)	0.44	(0.03)	0.69	(0.05)	0.37	(0.02)
South Africa	2.28	(0.11)	1.09	(0.04)	3.81	(0.20)	1.01	(0.04)	2.38	(0.15)	1.03	(0.04)	2.93	(0.15)	1.13	(0.04)	1.77	(0.09)	1.04	(0.04)
South Korea	1.45	(0.09)	0.63	(0.05)	2.29	(0.08)	0.31	(0.02)	1.50	(0.13)	0.26	(0.03)	1.73	(0.12)	0.67	(0.05)	1.17	(0.08)	0.57	(0.05)
Sri Lanka	1.66	(0.11)	0.81	(0.08)	2.83	(0.10)	0.41	(0.05)	1.77	(0.17)	0.24	(0.03)	2.03	(0.15)	0.94	(0.09)	1.30	(0.08)	0.68	(0.06)
Sweden	1.10	(0.05)	0.73	(0.04)	1.26	(0.05)	0.70	(0.03)	1.19	(0.06)	0.62	(0.03)	1.15	(0.05)	0.75	(0.04)	1.03	(0.05)	0.72	(0.03)
Syria	0.67	(0.06)	0.30	(0.02)	1.38	(0.14)	0.29	(0.02)	0.71	(0.06)	0.28	(0.02)	0.83	(0.07)	0.31	(0.02)	0.53	(0.05)	0.29	(0.02)
Tanzania	3.38	(0.23)	1.08	(0.08)	6.61	(0.31)	0.85	(0.04)	3.70	(0.35)	0.51	(0.07)	4.80	(0.35)	1.33	(0.12)	2.30	(0.15)	0.82	(0.06)
Trinidad & Tobago	1.72	(0.13)	0.66	(0.06)	4.33	(0.16)	0.40	(0.01)	2.04	(0.25)	0.38	(0.03)	1.78	(0.14)	0.74	(0.07)	1.56	(0.12)	0.63	(0.05)
Tunisia	1.91	(0.13)	0.84	(0.04)	4.14	(0.30)	0.93	(0.04)	2.16	(0.19)	0.87	(0.05)	2.54	(0.17)	0.91	(0.05)	1.43	(0.10)	0.79	(0.04)
Turkey	2.27	(0.13)	0.64	(0.03)	5.10	(0.21)	0.49	(0.02)	2.39	(0.17)	0.66	(0.04)	3.18	(0.18)	0.66	(0.03)	1.61	(0.09)	0.61	(0.03)
United Kingdom	1.26	(0.05)	1.09	(0.04)	1.03	(0.04)	1.05	(0.03)	1.30	(0.07)	0.82	(0.04)	1.36	(0.06)	1.15	(0.04)	1.14	(0.05)	1.05	(0.04)
United States	1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00	
Uruguay	1.14	(0.09)	0.91	(0.08)	1.36	(0.09)	1.38	(0.12)	1.34	(0.11)	0.82	(0.06)	1.21	(0.09)	0.96	(0.09)	1.05	(0.08)	0.89	(0.08)
Venezuela	2.31	(0.21)	1.43	(0.15)	3.41	(0.30)	1.35	(0.16)	2.42	(0.19)	1.10	(0.07)	3.04	(0.29)	1.57	(0.17)	1.77	(0.17)	1.31	(0.14)
Zimbabwe	5.40	(0.42)	1.50	(0.10)	14.77	(0.81)	1.12	(0.06)	6.19	(0.66)	1.30	(0.11)	7.02	(0.54)	1.59	(0.10)	4.04	(0.32)	1.39	(0.09)

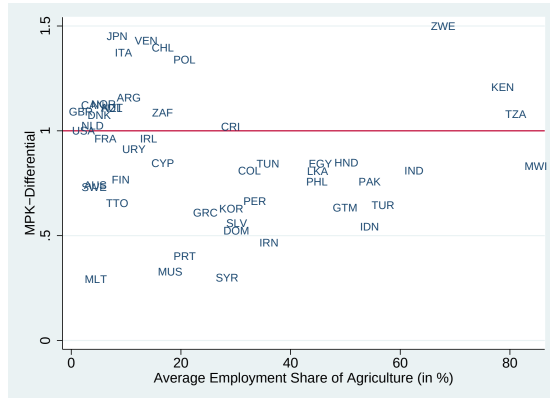
Notes: Marginal value product differentials between non-agriculture and agriculture are calculated as $\widehat{d}_{ji} = \exp\left(\widehat{\alpha}_{ji} - \widehat{\alpha}_{jUS}\right), j = l, k$, where $\widehat{\alpha}_{ji}$ are the country fixed effects from an estimation of equations (3.9) and (3.10). A value above (below) one means that the marginal value product is higher (lower) in non-agriculture than in agriculture. Heteroskedasticity robust standard errors in parentheses.

Figure 3.2: Baseline Results

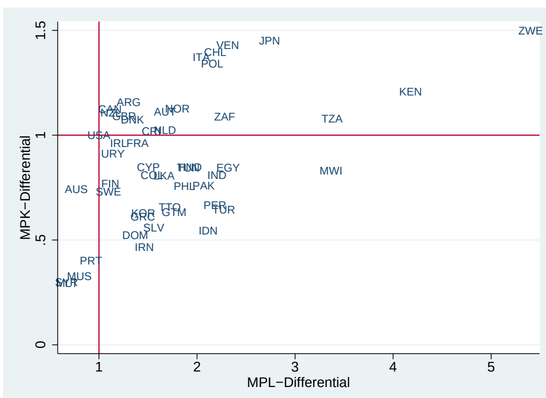
MPL-Differential vs. Agricultural Employment



MPK-Differential vs. Agricultural Employment



MPK- vs. MPL-Differential



Notes: Estimates of marginal value product differentials between non-agriculture and agriculture for baseline specification.

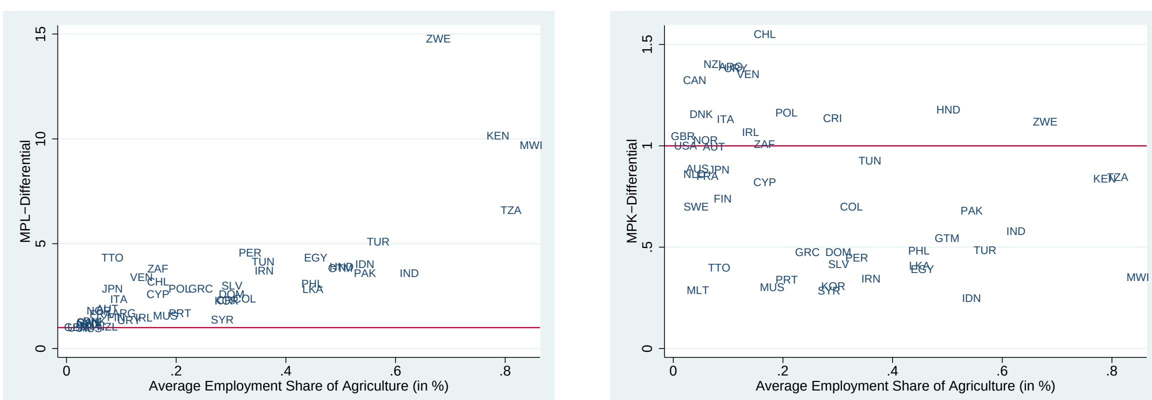
3.4.2 Sensitivity Analysis

The sensitivity analysis investigates the effect of alternative production technologies, assumptions on the wedges and human capital calibrations on estimated marginal product differentials. Four main cases are considered. First, the production technology is changed from Translog to standard Cobb-Douglas production functions. Second, the marginal product differentials are allowed to vary over time. Third, the relative return to schooling in agriculture is set to zero, which means that only non-agricultural workers increase their productivity by schooling and the agricultural schooling data is neglected in the estimation. Fourth, this relative return is set to one implying that returns to schooling are identical in both sectors. Table 3.1 columns (3) to (10) contain the estimated marginal product differentials for all these scenarios.

For Cobb-Douglas technologies, estimated marginal product of labor differentials are much higher for most countries, cf. figure 3.3 Increases by more than 50% are relatively common and are stronger in developing countries. In contrast, changes to marginal product of capital differentials are less pronounced and differ more across countries. In industrialized countries marginal product of capital differentials increase by up to 50% or decrease slightly, while they decrease in developing countries by up to 50%. Accordingly, using Cobb-Douglas technologies one considerably overestimates labor market distortions in favor of agriculture in most countries compared to Translog technologies and overestimates capital market distortions in favor of non-agriculture at least in many developing countries. This finding confirms the role of differing factor combinations for output elasticities that the Translog production function takes into account.

Figure 3.3: Cobb-Douglas

MPL-Differential vs. Agricultural Employment MPK-Differential vs. Agricultural Employment



Notes: Estimates of marginal value product differentials between non-agriculture and agriculture assuming Cobb-Douglas production functions.

One of the main identifying assumptions so far was that the marginal product differentials are constant over time. However it is likely that in the long-run the institutional framework that causes non-equalization of marginal products does change. Here I check how sensitive the results are when the wedges between marginal products are allowed to vary over time. Specifically, I allow for country-specific time trends of the marginal product differential terms. The constant terms d_{ki} and d_{li} are then replaced by

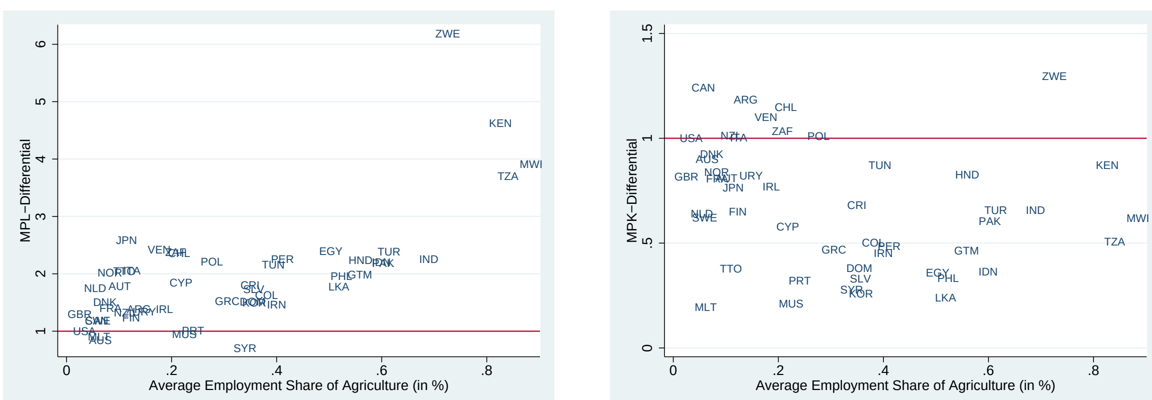
$$d_{kit} = d_{ki0} \times (1 + g_{ki})^t \tag{3.11}$$

$$d_{lit} = d_{li0} \times (1 + g_{li})^t. \tag{3.12}$$

Accordingly, the estimated equations (3.9) and (3.10) now also include $\gamma_{ki} \times t$ and $\gamma_{li} \times t$ terms, where $\gamma_k i \equiv \ln(1 + g_{ki})$ and $\gamma_l i \equiv \ln(1 + g_{li})$. I include these time trends for all countries except for the United States for which the assumption of efficient factor markets is maintained. Figure 3.4 reports the estimated differential for the middle of the time period during which each country is observed in the data set. The estimated labor differentials are similar to the baseline results. But the estimated capital differential decreases somewhat for many countries such that capital markets appear to be more distorted towards non-agriculture. However, overall the results seem to be relatively robust to a change of this assumption.

Figure 3.4: Time Trends

MPL-Differential vs. Agricultural Employment | MPK-Differential vs. Agricultural Employment

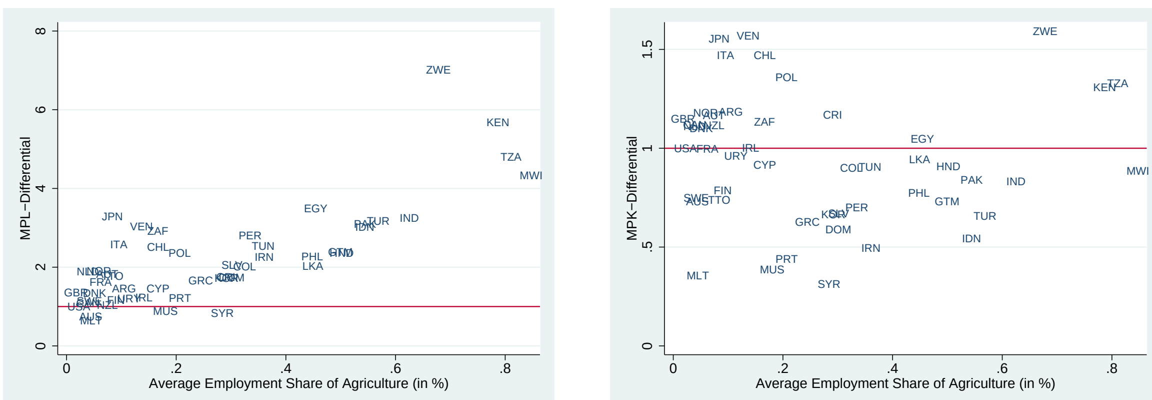


Notes: Estimates of marginal value product differentials between non-agriculture and agriculture allowing for time trends of the marginal product wedges. The figure reports the estimated differential for the middle of the time period during which each country is observed in the data set.

Figure 3.5 shows that imposing a zero return to schooling in agriculture, i.e. neglecting data on agricultural years of schooling, also leads to higher estimated marginal product of labor differentials relative to the baseline scenario. But the magnitude of the effect is on average much lower than in the Cobb-Douglas case such that the maximum increase is below 60%. Again the marginal product differential of developing countries increases the most. The intuition for this finding is the following. Since the agricultural schooling data is effectively neglected in this specification, industrialized countries with their higher absolute schooling levels in non-agriculture (and agriculture) now have the greatest difference in human capital stocks between sectors. This reduces their RLP relative to developing countries and accordingly this greater cross-country variation of RLP is at least partly accounted for by a higher marginal product of labor differential. Marginal product of capital differentials generally increase by up to 25%. Accordingly, distortions go up in countries that were found to distort the capital market in favor of agriculture in the baseline specification and go down in countries that were found to distort it in favor of non-agriculture. This result must come purely from the effect of intersectoral human capital differences on the ratio of output elasticities.

Figure 3.5: Zero Human Capital Return in Agriculture

MPL-Differential MPK-Differential

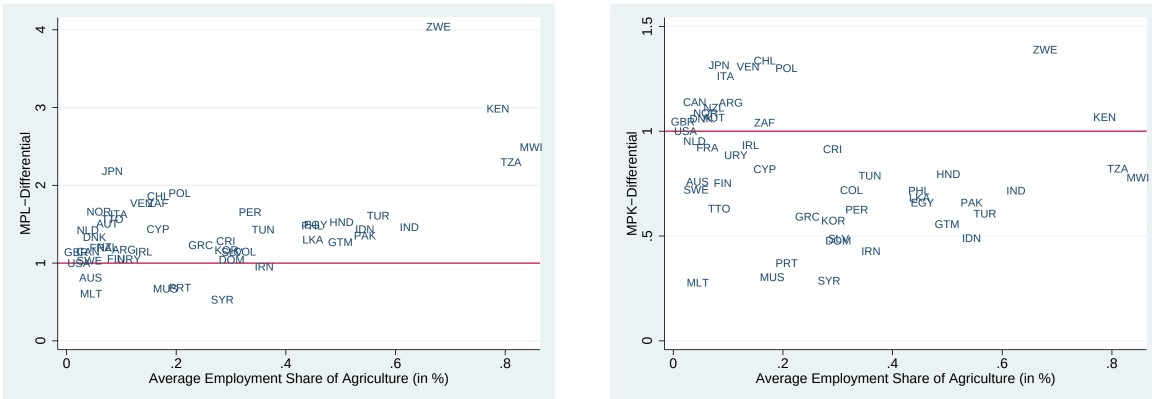


Notes: Estimates of marginal value product differentials between non-agriculture and agriculture assuming a return to schooling of zero in the agricultural sector.

Finally, the relative return to schooling in the two sectors is set to one, i.e. returns to schooling are made identical in both sectors. Marginal product of labor differentials reported in figure 3.6 then decrease by up to 40% and the strongest in developing economies. The intuition follows the same line of reasoning as before, but in the opposite direction. Now developing countries have the greatest intersectoral difference in human capital stocks because of their intersectoral schooling differences. Marginal product of capital differentials decrease as well by up to 25% implying that distortions go up in countries that were found to distort the capital market in favor of non-agriculture and go down in countries that were found to distort it in favor of agriculture.

Figure 3.6: Same Human Capital Return in Both Sectors

MPL-Differential MPK-Differential



Notes: Estimates of marginal value product differentials between non-agriculture and agriculture assuming identical returns to schooling in the two sectors.

The analysis of these four alternative cases have shown that the results are sensitive to certain changes of the assumptions. But the sensitivity of the result is limited and the main results of the paper still hold - in many countries marginal products between the agricultural and non-agricultural sector are not equalized. In particular, for labor the first and third case have shown that the baseline estimates of factor misallocation are relatively modest compared to the ones implied by other assumptions. Adopting the fourth case would diminish the estimates to some extent, but still not imply marginal product of labor equalization. This suggests that from the perspective of identifying factor misallocation, investigations of returns to schooling in agriculture would be an important area for future research. The sensitivity analysis has also shown that the results for capital are less sensitive compared to those of labor.

3.4.3 Discussion

While this paper finds sizeable marginal product differentials, the estimates are lower than the ones found by Vollrath (2009), which is the only other study I am aware of with a similar cross-country focus. In a calibration exercise of Cobb-Douglas production functions Vollrath (2009) found marginal product of labor differentials[18] in excess of two even in many industrialized economies and higher ones up to eight or even twelve

[18]This refers to his marginal product of human capital differentials, which are the ones corresponding to my estimates.

in developing countries. The reasons why the estimates differ are numerous, so I will name some that I suspect to be important. First, I employ the more flexible Translog production function compared to the relatively restrictive Cobb-Douglas. As the sensitivity analysis has shown my methodology would also yield much higher estimates under a Cobb-Douglas assumption. The factor combinations that are in operation in developing countries seem to affect output elasticities and explain a substantial part of the cross-country RLP variation. Second, the production parameters in my paper are estimated from the whole sample and not only from the United States. Third, since Vollrath is ultimately interested in a development accounting question, he uses output evaluated at purchasing power adjusted prices, while this paper uses domestic prices, which seem to be the appropriate choice for the analysis of factor allocation. This difference alone leads to lower levels of RLP in developing countries. Fourth, he only considers a cross-section in 1985, while this paper uses panel data from 1967 to 1992. Overall, the finding of this paper that marginal product differentials are smaller than the ones found by Vollrath (2009) might also diminish how much of the cross-country income distribution and the variation in aggregate efficiency can be explained by this form of factor misallocation. But such a conclusion would have to be based on further analysis.

3.5 Conclusions

This paper has developed a framework for the estimation of the degree of factor misallocation between agriculture and non-agriculture from cross-country panel data. The findings are that there is evidence for sizeable marginal product differentials and factor misallocation. The marginal product of labor is higher in non-agriculture than in agriculture in most countries of the world. In developing countries the marginal product of labor differential is around 1.5 to 5. In contrast, the allocation of capital does not allow such a clear distinction between countries. In most developing countries the marginal product of capital is around 1.2 to 3 times higher in agriculture than in non-agriculture. Industrialized countries tend to be closer to marginal product of capital equalization, but there are also countries where the marginal product is higher in non-agriculture. Thus, this paper provided macroeconomic evidence supporting the view that economic dualism, understood here as non-equalized marginal products, is an important characteristic of developing countries. Even though this study has not investigated specific institutions or sources of market failure, the results indirectly suggest that factor markets for the agricultural and non-agricultural sector might not be well-integrated in developing countries. An obvious question for future research is then, what are the deeper economic reasons for the marginal product differentials in developing countries? Are there policies or institutions that distort factor markets? Alternatively, are at low levels of development migration and reallocation costs so high that factor mobility is effectively very limited? Answering these questions is a key step towards formulating policy implications of the findings.

When thinking about the welfare implications of this paper one needs to keep in mind that it is very likely that people in poor countries do in fact take optimal decisions given the institutional framework they face. The results of this study can thus not be used to argue that deliberately moving people out of agriculture is a good idea. However, to the extent it is possible to identify the institutions that are responsible for misallocation, changing these institutions could be an important area for policy reforms in poor countries.

Factor misallocation could contribute to our understanding of several phenomena of the world economy. If factor misallocation is more severe in developing countries then the associated output losses might partly explain why these countries are poor. Furthermore, the existence of wedges between marginal products of labor would imply that the sectoral reallocation of workers from a low marginal product to a high marginal product sector might be an important source of economic growth during the structural transformation. Finally, the existence of wedges could be part of the explanation why some developing countries are still so dominantly agricultural. I hope that my quantitative results on the degree of factor misallocation in different countries can form a basis for future research on these questions.

Appendix 3.A Construction of Missing Sectoral Schooling Data

Following Vollrath (2009) I exploit a regularity between the sectoral education measures of Timmer (2002) and average years of schooling of persons over the age of 25 for the total population provided by Barro and Lee (2001). This relationship is that rural years of education converge towards urban years of education as total years of education increase. Vollrath uses this relationship until total years of education are greater than 6.6 years because for higher total education levels rural and urban school years become very similar. The empirical justification for this assumption is relatively weak since there are only few observations with available sectoral education data and total years of schooling greater than 6.6. But I also adopt this procedure since it generates human capital stocks that are more unequally distributed between the two sectors in developing countries compared to industrialized countries and so effectively biases the results against finding labor misallocation in developing countries.[19] A simple OLS regression of rural education years, S_R, on urban years, S_U, for all observations with total education smaller than 6.6 years yields

$$\underset{}{S_{Rit}} = \underset{(49.96)}{0.912}S_{Uit} - \underset{(-10.99)}{0.847}, \quad R^2 = 0.87, \quad N = 388, \tag{3.13}$$

where t-statistics are in parentheses. These estimates differ from Vollrath's[20] since I use all observations with available sectoral education data while he only uses the countries that are in his final data set and only the time periods from 1970 to 1985 resulting in 77 observations. Furthermore, average years of schooling of the overall population, S_T are a weighted average of rural and urban years of schooling given by

$$S_{Tit} = l_{ait}S_{Rit} + (1 - l_{ait})S_{Uit}, \tag{3.14}$$

where l_{ait} is the rural share of the population. Equations (3.13) and (3.14) can be solved simultaneously to find expressions for S_{Rit} and S_{Uit} as functions of years of education of the total population and the rural population share. Total years of schooling for

[19]The effect of not using the cutoff of 6.6 years on the results of the baseline specification are negligible (results not reported).

[20]His estimates are

$$S_{Rit} = \underset{(16.18)}{1.071}S_{Uit} - \underset{(5.17)}{1.518}, \quad R^2 = 0.78, \quad N = 77.$$

persons over the age of 25 are taken from Barro and Lee (2001)[21] and the share of economically active in agriculture is used as the rural population share[22]. I construct sectoral education years for the countries that are not in Timmer's data set every five years from 1965-1985 and for all countries for the time periods 1990 and 1995. For the observations with less than 6.6 years of overall education the expressions derived above are used and for the observations with more than 6.6 years the total level of education is used for both rural and urban sectors. Finally, I interpolate linearly between the real and imputed data points that are available at five year frequency to get annual data.

[21]For Tanzania I use average years of schooling of persons over the age of 15 instead since these were the only ones available.

[22]This is exact if as assumed for the human capital calibration the rural population constitutes the agricultural and the urban population the non-agricultural labor force and the participation rates in both areas are identical.

Bibliography

Aguiar, M. and E. Hurst (2009). Deconstructing Lifecycle Expenditure. Working Paper.

Altig, D., A. J. Auerbach, L. J. Kotlikoff, K. A. Smetters, and J. Walliser (2001). Simulating Fundamental Tax Reform in the United States. *American Economic Review 91*(3), 574–594.

Amromin, G. and A. L. Paulson (2009). Comparing Patterns of Default among Prime and Subprime Mortgages. *Federal Reserve Bank of Chicago Economic Perspectives Q2*, 18–37.

Arrazola, M. and J. de Hevia (2004). More on the Estimation of the Human Capital Depreciation Rate. *Applied Economic Letters 11*, 145–148.

Attanasio, O., S. Kitao, and G. L. Violante (2007). Global Demographic Trends and Social Security Reform. *Journal of Monetary Economics 54*(1), 144–198.

Attanasio, O. P. (1999). Consumption. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1b, Chapter 11, pp. 741–812. Amsterdam: Elsevier Science B. V.

Auerbach, A. J. and L. J. Kotlikoff (1987). *Dynamic Fiscal Policy*. Cambridge: Cambridge University Press.

Bajari, P., S. Chu, and M. Park (2010). An Empirical Model of Subprime Mortgage Default from 2000 to 2007. *Working Paper*.

Banerjee, A. and E. Duflo (2005). Growth Theory through the Lens of Development Economics. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of Economic Growth*, pp. 473–552. Amsterdam: Elsevier.

Bar, M. and O. Leukhina (2010). Demographic Transition and Industrial Revolution: A Macroeconomic Investigation. *Review of Economic Dynamics 13*(2), 424–451.

Barro, R. J. and J.-W. Lee (2001). International Data on Educational Attainment: Updates and Implications. *Oxford Economic Papers 53*(3), 541–563.

Becker, G. (1967). Human Capital and the Personal Distribution of Income: An Analytical Approach. Woytinsky Lecture: University of Michigan.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy 75*(4), 352–365.

Berndt, E. R. and L. R. Christensen (1973). The Translog Function and the Substitution of Equipment, Structures, and Labor in U.S. Manufacturing 1929-68. *Journal of Econometrics 1*(1), 81–114.

Bils, M. and P. J. Klenow (2000). Does Schooling Cause Growth? *American Economic Review 90*(5), 1160–1183.

Börsch-Supan, A., A. Ludwig, and J. Winter (2006). Ageing, Pension Reform and Capital Flows: A Multi-Country Simulation Model. *Economica 73*, 625–658.

Boucekkine, R., D. de la Croix, and O. Licandro (2002). Vintage Human Capital, Demographic Trends, and Endogenous Growth. *Journal of Economic Theory 104*, 340–375.

Browning, M., L. P. Hansen, and J. J. Heckman (1999). Micro Data and General Equilibrium Models. In J. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, pp. 543–633. North-Holland.

Bucks, B. K., A. B. Kennickell, and K. B. Moore (2006). Recent Changes in U.S. Family Finances: Evidence from the 2001 and 2004 Survey of Consumer Finances. *Federal Reserve Bulletin 92*, A1–A38.

Calhoun, C. A. (1996). *OFHEO House Price Indexes: HPI Technical Description*. Washington, D.C.: Office of Federal Housing Enterprise Oversight.

Campbell, J. Y. and J. F. Cocco (2003). Household Risk Management and Optimal Mortgage Choice. *Quarterly Journal of Economics 118*, 1449–1494.

Campbell, J. Y. and J. F. Cocco (2011). A Model of Mortgage Default. *Working Paper*.

Carroll, C. D. (1997). Buffer-Stock Saving and the Life-Cycle/Permanent Income Hypothesis. *Quarterly Journal of Economics 112*(1), 1–55.

Caselli, F. (2005). Accounting for Cross-Country Income Differences. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of Economic Growth*, pp. 679–741. Amsterdam: Elsevier.

Caselli, F. and W. J. Coleman (2001). The U.S. Structural Transformation and Regional Convergence: A Reinterpretation. *Journal of Political Economy 109*(3), 584–616.

Chanda, A. and C.-J. Dalgaard (2008). Dual Economies and International Total Factor Productivity Differences: Channeling the Impact from Institutions, Trade and Geography. *Economica 75*, 629–661.

Chatterjee, S. and B. Eyigungor (2011). A Quantitative Analysis of the U.S. Housing and Mortgage Markets and the Foreclosure Crisis. *working paper*.

Corbae, D. and E. Quintin (2011). Mortgage Innovation and the Foreclosure Boom. *working paper*.

Córdoba, J. C. and M. Ripoll (2006). Aggriculture, Aggregation, and Cross-Country Income Differences. *Working Paper*.

Corradin, S. (2009). Household Leverage. *Working Paper*.

Crego, A., D. Larson, R. Butzer, and Y. Mundlak (1998). A New Database on Investment and Capital for Agriculture and Manufacturing. *World Bank Policy Research Working Paper 2013*.

Davis, M. A., A. Lehnert, and R. F. Martin (2008). The Rent-Price Ratio for the Aggregate Stock of Owner-Occupied Housing. *Review of Income and Wealth 54*(2), 279–284.

de la Croix, D. and O. Licandro (1999). Life Expectancy and Endogenous Growth. *Economics Letters 65*, 255–263.

de Melo, J. A. P. (1977). Distortions in the Factor Market: Some General Equilibrium Estimates. *Review of Economics and Statistics 59*(4), 398–405.

De Nardi, M., E. French, and J. Jones (2009). Why Do the Elderly Save? The Role of Medical Expenses. *NBER Working Paper 15149*.

De Nardi, M., S. İmrohoroğlu, and T. J. Sargent (1999). Projected U.S. Demographics and Social Security. *Review of Economic Dynamics 2*(1), 575–615.

Deaton, A. (1991). Saving and Liquidity Constraints. *Econometrica 59*(5), 1221–1248.

Demyanyk, Y. and O. Van Hemert (2011). Understanding the Subprime Mortgage Crisis. *Review of Financial Studies 24*(6), 1848–1880.

Deng, Y., J. M. Quigley, and R. Van Order (2000). Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options. *Econometrica 68*(2), 275–307.

Diamond, P. and J. Gruber (1999). Social Security and Retirement in the United States. In J. Gruber and D. Wise (Eds.), *Social Security and Retirement Around the World*, Chapter 11, pp. 437–474. Chicago: University of Chicago Press.

Domeij, D. and M. Flodén (2006). The Labor-Supply Elasticity and Borrowing Constraints: Why Estimates are Biased. *Review of Economic Dynamics 9*, 242–262.

Dougherty, C. and M. Selowsky (1973). Measuring the Effect of the Misallocation of Labor. *Review of Economics and Statistics 55*(3), 386–390.

Duarte, M. and D. Restuccia (2010). The Role of the Structural Transformation in Aggregate Productivity. *Quarterly Journal of Economics 125*(1). forthcoming.

Dupor, B., L. Lochner, C. Taber, and M. B. Wittekind (1996). Some Effects of Taxes on Schooling and Training. *American Economic Review Papers and Proceedings 86*(2), 340–346.

Echevarria, C. A. and A. Iza (2006). Life Expectancy, Human Capital, Social Security and Growth. *Journal of Public Economics 90*, 2324–2349.

Ehrlich, I. and J. Kim (2007). Social Security and Demographic Trends: Theory and Evidence from the International Experience. *Review of Economic Dynamics 10*(1), 55–77.

Elul, R., N. S. Souleles, S. Chomsisengphet, D. Glennon, and R. Hunt (2010). What 'Triggers' Mortgage Default? *American Economic Review Papers and Proceedings 100*(2), 490–494.

Fernández-Villaverde, J. and D. Krüger (2007). Consumption over the Life Cycle: Facts from Consumer Expenditure Survey Data. *Review of Economics and Statistics 89*(3), 552–565.

Food and Agriculture Organization of the United Nations (2004). *FAOSTAT*. Food and Agriculture Organization of the United Nations. CD-ROM.

Foote, C. F., K. S. Gerardi, L. Goette, and P. S. Willen (2008). Subprime Facts: What (We Think) We Know about the Subprime Crisis and What We Don't. *Federal Reserve Bank of Boston Public Policy Discussion Paper No. 08-2*.

Foote, C. F., K. S. Gerardi, L. Goette, and P. S. Willen (2009). Reducing Foreclosures: No Easy Answers. *NBER Macroeconomics Annual 24*, 89–138.

Foote, C. F., K. S. Gerardi, and P. S. Willen (2008). Negative Equity and Foreclosure: Theory and Evidence. *Journal of Urban Economics 64*(2), 234–245.

Fougère, M. and M. Mérette (1999). Population Ageing and Economic Growth in Seven OECD Countries. *Economic Modelling 16*, 411–427.

Frederick, S., G. Loewenstein, and T. O'Donoghue (2002). Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature 40*(2), 351–401.

Gallant, A. R. (1987). *Nonlinear Statistical Models*. New York: John Wiley & Sons.

Garriga, C. and D. E. Schlagenhauf (2009). Home Equity, Foreclosures, and Bailouts. *working paper*.

Geanakoplos, J. and S. P. Zeldes (2009). Reforming Social Security with Progressive Personal Accounts. *Cowles Foundation Paper 1276*.

Gerardi, K. S., A. Lehnert, S. M. Sherlund, and P. S. Willen (2008). Making Sense of the Subprime Crisis. *Brookings Papers on Economic Activity* (Fall), 69–159.

Gerardi, K. S., A. H. Shapiro, and P. S. Willen (2007). Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures. *Federal Reserve Bank of Boston Working Paper No. 07-15*.

Ghent, A. C. and M. Kudlyak (2010). Recourse and Residential Mortgage Default: Theory and Evidence from U.S. States. *Federal Reserve Bank of Richmond Working Paper No. 09-10R*.

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

Gollin, D., S. L. Parente, and R. Rogerson (2002). The Role of Agriculture in Development. *American Economic Review Paper and Proceedings 92*(2), 160–164.

Gollin, D., S. L. Parente, and R. Rogerson (2004). Farm Work, Home Work and International Productivity Differences. *Review of Economic Dynamics 7*(4), 827–850.

Gollin, D., S. L. Parente, and R. Rogerson (2007). The Food Problem and the Evolution of International Income Levels. *Journal of Monetary Economics 54*(4), 1230–1255.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect Inference. *Journal of Applied Econometrics 8*, S85–S118.

Greene, W. H. (2003). *Econometric Analysis*. Prentice Hall. Fifth Edition.

Guiso, L., P. Sapienza, and L. Zingales (2011). The Determinants of Attitudes towards Strategic Default on Mortgages. *Working Paper*.

Guvenen, F. (2006). Reconciling Conflicting Evidence on the Elasticity of Intertemporal Substitution: A Macroeconomic Perspective. *Journal of Monetary Economics 53*(7), 1451–1472.

Guvenen, F. and B. Kuruscu (2009). A Quantitative Analysis of the Evolution of the U.S. Wage Distribution: 1970-2000. In *NBER Macroeconomics Annual 24*. NBER.

Hall, R. E. and C. I. Jones (1999). Why do some Countries produce so much more Output per Worker than Others? *Quarterly Journal of Economics 114*(1), 83–116.

Hansen, G. D. (1993). The Cyclical and Secular Behaviour of the Labour Input: Comparing Efficiency Units and Hours Worked. *Journal of Applied Econometrics 8*(1), 71–80.

Hansen, G. D. and S. İmrohoroğlu (2008). Consumption over the Life-Cycle: The Role of Annuities. *Review of Economic Dynamics 11*, 566–583.

Hayashi, F. and E. C. Prescott (2008). The Depressing Effect of Agricultural Institutions on the Prewar Japanese Economy. *Journal of Political Economy 116*(4), 573–632.

Heckman, J., L. Lochner, and C. Taber (1998). Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogenous Agents. *Review of Economic Dynamics 1*, 1–58.

Heijdra, B. J. and W. E. Romp (2008). A Life-Cycle Overlapping-Generations Model of the Small Open Economy. *Oxford Economic Papers 60*(1), 88–121.

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

Huang, H., S. İmrohoroğlu, and T. J. Sargent (1997). Two Computations to Fund Social Security. *Macroeconomic Dynamics 1*(1), 7–44.

Huggett, M., G. Ventura, and A. Yaron (2010). Sources of Lifetime Inequality. *American Economic Review*. forthcoming.

Human Mortality Database (2008). *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)*. www.mortality.org.

Imai, S. and M. P. Keane (2004). Intertemporal Labor Supply and Human Capital. *International Economic Review 45*(2), 601–641.

Jagtiani, J. and W. W. Lang (2010). Strategic Default on First and Second Lien Mortgages During the Financial Crisis. *Working Paper*.

Jeske, K., D. Krueger, and K. Mitman (2011). Housing and the Macroeconomy: The Role of Bailout Guarantees for Government Sponsored Enterprises. *working paper*.

Kalemli-Ozcan, S., H. E. Ryder, and D. N. Weil (2000). Mortality Decline, Human Capital Investment, and Economic Growth. *Journal of Development Economics 62*, 1–23.

Kau, J. B., D. C. Keenan, and T. Kim (1994). Default Probabilities for Mortgages. *Journal of Urban Economics 35*, 278–296.

Krüger, D. and A. Ludwig (2007). On the Consequences of Demographic Change for Rates of Returns to Capital, and the Distribution of Wealth and Welfare. *Journal of Monetary Economics 54*(1), 49–87.

Lau, M. I. and P. Poutvaara (2006). Social Security Incentives and Human Capital Investments. *Finnish Economic Papers 19*(1), 16–24.

Lee, R. D. and L. Carter (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association 87*, 659–671.

Lee, R. and A. Mason (2010). Fertility, Human Capital, and Economic Growth over the Demographic Transition. *European Journal of Population 26*(2), 159–182.

Lord, W. (1989). The Transition from Payroll to Consumption Receipts with Endogenous Human Capital. *Journal of Public Economics 38*(1), 53–73.

Ludwig, A. (2007). The Gauss-Seidel-Quasi-Newton Method: A Hybrid Algorithm for Solving Dynamic Economic Models. *Journal of Economic Dynamics and Control 31*(5), 1610–1632.

Ludwig, A., D. Krüger, and A. Börsch-Supan (2007). Demographic Change, Relative Factor Prices, International Capital Flows, and their Differential Effects on the Welfare of Generations. In J. Brown, J. Liebmann, and D. A. Wise (Eds.), *Social Security Policy in a Changing Environment*. University of Chicago Press. forthcoming.

Ludwig, A., T. Schelkle, and E. Vogel (2012). Demographic Change, Human Capital and Welfare. *Review of Economic Dynamics 15*(1), 94–107.

Ludwig, A. and E. Vogel (2009). Mortality, Fertility, Education and Capital Accumulation in a Simple OLG Economy. *Journal of Population Economics 23*(2), 703–735.

Magnac, T. and D. Thesmar (2002). Identifying Dynamic Decision Processes. *Econometrica 70*(2), 801–816.

Mayer, C., K. Pence, and S. M. Sherlund (2009). The Rise in Mortgage Defaults. *Journal of Economic Perspectives 23*(1), 27–50.

McGrattan, E. R. and R. Rogerson (2004). Changes in Hours Worked, 1950-2000. *Federal Reserve Bank of Minneapolis Quarterly Review 28*(1), 14–33.

Mian, A. and A. Sufi (2009). The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. *Quarterly Journal of Economics 124*(4), 1449–1496.

Perroni, C. (1995). Assessing the Dynamic Efficiency Gains of Tax Reform when Human Capital is Endogenous. *International Economic Review 36*(4), 907–925.

Psacharopoulos, G. (1994). Returns to Investment in Education: A Global Update. *World Development 22*(9), 1325–1343.

Quercia, R. G. and M. A. Stegman (1992). Residential Mortgage Default: A Review of the Literature. *Journal of Housing Research 3*(2), 341–379.

Restuccia, D. and R. Rogerson (2008). Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Review of Economic Dynamics 11*(4), 707–720.

Restuccia, D., D. T. Yang, and X. Zhu (2008). Aggriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis. *Journal of Monetary Economics 55*(2), 234–250.

Rogerson, R. and J. Wallenius (2009). Micro and Macro Elasticities in a Life Cycle Model with Taxes. *Journal of Economic Theory 144*(6), 2277–2292.

Rosenzweig, M. R. (1988). Labor Markets in Low-Income Countries. In H. Chenery and T. N. Srinivasan (Eds.), *Handbook of Development Economics*, Volume I, pp. 713–762. Amsterdam: North-Holland.

Sadahiro, A. and M. Shimasawa (2002). The Computable Overlapping Generations Model with Endogenous Growth Mechanism. *Economic Modelling 20*, 1–24.

Smith, A. A. (1993). Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions. *Journal of Applied Econometrics 8*, S63–S84.

Storesletten, K. (2000). Sustaining Fiscal Policy through Immigration. *Journal of Political Economy 108*(2), 300–323.

Temple, J. (2002). The Cost of Dualism. *Department of Economics discussion paper no. 02/532, University of Bristol*.

Temple, J. (2005). Dual Economy Models: A Primer for Growth Economists. *The Manchester School 73*(4), 435–478.

Temple, J. and L. Wößmann (2006). Dualism and Cross-Country Growth Regressions. *Journal of Economic Growth 11*(3), 187–228.

Timmer, P. (2002). Agriculture and Economic Development. In B. L. Gardner and G. C. Rausser (Eds.), *Handbook of Agricultural Economics*, pp. 1487–1546. Amsterdam: Elsevier.

Trostel, P. A. (1993). The Effect of Taxation on Human Capital. *Journal of Political Economy 101*(2), 327–350.

United Nations (2007). *World Population Prospects: The 2002 Revision*. United Nations Population Division. New York.

Vandell, K. D. (1995). How Ruthless is Mortgage Default: A Review and Synthesis of the Evidence. *Journal of Housing Research 6*(2), 245–264.

Vollrath, D. (2009). How Important are Dual Economy Effects for Aggregate Productivity? *Journal of Development Economics 88*(2), 325–334.

Williamson, J. G. (1987). Did English Factor Markets Fail during the Industrial Revolution? *Oxford Economic Papers 39*(4), 641–678.

World Bank (2007, April). *World Development Indicators*. World Bank.

Zeldes, S. (1989). Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence. *Quarterly Journal of Economics 104*(2), 275–298.