Essays in Demographic and Macroeconomic Trends

D I S S E R T A T I O N
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by

Martin Maniera

from

Germany

Approved on the application of

Prof. Guido Cozzi, PhD

and

Prof. Dr. Uwe Sunde
Prof. Dr. Reto Föllmi

Dissertation no. 5002

EduPrint Publishing, Konstanz, 2020

The University of St.Gallen, School of Management, Economics, Law, Social Sciences and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, June 15, 2020

The President:

Prof. Dr. Bernhard Ehrenzeller

Summary

This cumulative PhD thesis consists of three empirical essays about demographic and macroeconomic trends. They all investigate questions which are currently intensely debated in politics, economics and media.

The first essay analyses how migration within the Euro area and across its borders affects the stability of the currency union. In line with the Optimal Currency Area theory I find that migration within the Euro area reduces macroeconomic imbalances between its member states, that is, differences in terms of GDP per capita growth, current account, unemployment rate and central government borrowing. Thus, internal migration increases the stability of the currency union. Yet, migration across its borders reduces stability as the imbalances in central government borrowing increase.

The second essay focuses on the effect of the age distribution on inflation. In many industrial and developing countries the population share of the young cohort is declining while the share of elderly is increasing. Both cohorts increase inflation as they increase aggregate demand relative to aggregate supply. In contrast, a larger share of working age people increase aggregate supply relative to aggregate demand, thus exercising a downward pressure on price levels. Central banks have to take the powerful demographic trends into consideration as they will increase inflation in the long run in countries which find themselves in the late stage of the demographic transition.

The third essay investigates the boundaries of inequality. With higher per capita GDP, income in the most equal countries tends to become less equally distributed. In contrast, income in the most unequal countries becomes more equally distributed. This narrowing corridor can be observed for income before and after redistribution via taxes and social transfers. Four factors are driving the corridor: real GDP per capita growth, schooling, economic complexity and the interaction between schooling and economic complexity. Highly unequal countries have to promote schooling and economic complexity in order to enhance economic development. Highly equal countries, in contrast, are advised to reduce their redistribution since they channel too many resources into schooling and since they lower incentive for capital investments and labour market participation.

Zusammenfassung

Die vorliegende kumulierte Doktorarbeit besteht aus drei empirischen Aufsätzen, die sich mit demographischen und makroökonomischen Trends befassen.

Der erste Aufsatz analysiert, wie Migration innerhalb der Eurozone und über ihre Grenzen hinaus die Stabilität der Währungsunion beeinflusst. Er bestätigt in Übereinstimmung mit der Theorie der Optimalen Währungsräume, dass Migration innerhalb der Eurozone die makroökonomischen Ungleichgewichte reduziert. Die gemessenen Ungleichgewichte bestehen hinsichtlich des BIP-Wachstums pro Kopf, des Leistungsbilanzsaldos, der Arbeitslosenquote und der Neuverschuldung der Zentralregierung. Allerdings dämmt externe Migration über die Grenzen der Währungsunion bzw. der Europäischen Union entstehende Ungleichgewichte nicht ein. Externe Migration baut sogar Ungleichgewichte auf.

Der zweite Aufsatz konzentriert sich auf die Auswirkungen der Alterspyramide auf die Inflation. In vielen Industrie- und Schwellenländern sinkt der Anteil der jungen Alterskohorte, während jener der Älteren steigt. Beide Kohorten erhöhen die aggregierte Nachfrage relativ zum aggregierten Angebot, so dass sie die Inflation steigern. Im Gegensatz dazu steigert ein grösserer Bevölkerungsanteil von Personen im erwerbsfähigen Alter das Angebot stärker als die Nachfrage, so dass die Inflation sinkt. Notenbanken sei empfohlen, diese starken demographischen Entwicklungen zu berücksichtigen, weil sie auf lange Sicht die Inflation erhöhen werden.

Der dritte Aufsatz untersucht die Grenzen der Ungleichheit. Mit wachsendem BIP pro Kopf wird das Einkommen in den gleichsten Staaten tendenziell ungleicher verteilt. Im Vergleich hierzu wird das Einkommen in den ungleichsten Staaten zunehmend gleichmässiger verteilt. Dieser sich verjüngende Korridor kann sowohl vor als auch nach staatlicher Umverteilung beobachtet werden. Vier Faktoren treiben die Verjüngung des Korridors voran: reales BIP-pro-Kopf-Wachstum, Bildung, ökonomische Komplexität und die Interaktion zwischen den letzten beiden Faktoren. Die ungleichsten Länder sollten die Bildung und ökonomische Komplexität ausweiten, um weiter zu wachsen. Die gleichsten Staaten hingegen sollten ihre staatliche Umverteilung reduzieren, weil sie sonst zu viele Ressourcen in die Bildung investieren und die Anreize für Kapitalinvestitionen und Arbeitsmarktpartizipation senken.

List of Tables

List of Figures

1. Does migration mitigate macroeconomic imbalances in the Euro area?

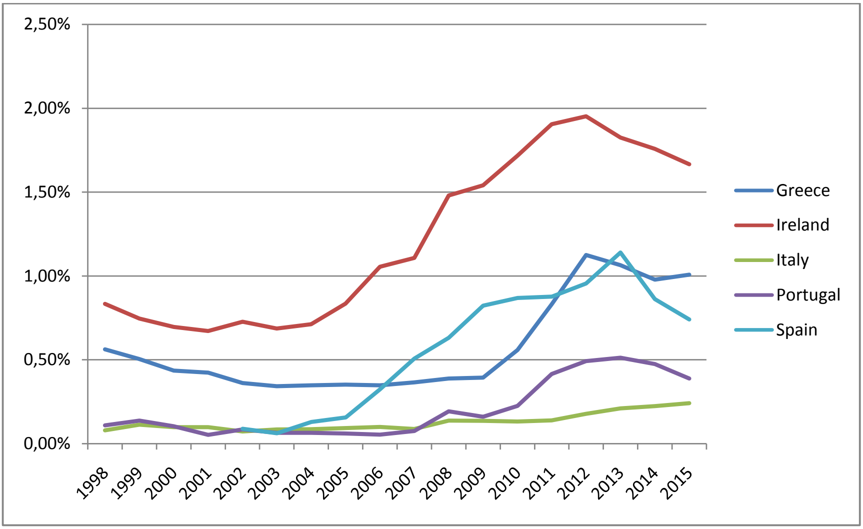
The theory of the Optimal Currency Area suggests that macroeconomic imbalances between the member states render monetary policy powerless. The theory also argues that labour migration reduces the imbalances. However, a critical analysis of the underlying assumptions indicates that in contrast to the predictions of the OCA theory, migration could amplify the imbalances and reduce the stability of the currency union. By using dyadic country data in a Panel VAR set up, this paper confirms empirically that migration within the Euro area reduces the imbalances. Yet, migration across the outer borders of the Euro area reduces the stability of the currency union.

JEL classification: C33, F22, F45

1.1. Introduction

Migration has been dominating the political and public debate in Europe for several years. Great Britain voted to leave the European Union mainly since its citizens want to curb immigration from Eastern Europe. However, within the EU labour has not only migrated from "east to west". In the aftermath of the Global Recession in 2008 many migrants returned home, i.e. they moved back from "west to east". In addition, one could observe migration from "south to north", or to be more precise, from the periphery of the EU to its core. Particularly Greece, Italy, Spain, Portugal and Ireland saw an increase of emigration. Chart 1 shows the share of persons leaving the country relative to its population.

Chart 1: Emigration in percent of the respective country's population between 1998 and 2015. Source: Eurostat



All citizens of the European Union enjoy the Freedom of movement. They are allowed to seek employment in any of the EU member states without discrimination. In contrast to this internal EU migration, the external migration across the outer borders of the EU faces much stricter regulations. It is more difficult for EU citizens to settle down and work outside the union than within. Likewise, third-country citizens face stricter regulation to work in the EU than local citizens. This paper differentiates between internal and external migration in order to analyze if they affect the stability of the Euro area differently.

According to the Optimal Currency Area theory (Mundell, 1961) the member states of a currency union should be as similar as possible with respect to their economic structures. The less similar the members are, the larger the differences between their GDP growth rates, unemployment rates, etc. Those differences are called macroeconomic imbalances. A macroeconomic imbalance is the difference of a given indicator of economic development between two countries.

Large imbalances lower the efficacy of monetary policy. If Germany's economy is booming while France suffers from a recession - which prime interest rate shall the European Central Bank set? Germany's boom requires a high interest rate in order to prevent an overheating and inflation. France, however, needs a low interest rate to stimulate its economy and to keep deflation at bay. Hence, the larger the imbalance economic growth imbalance, the less effective the prime interest rate.

In fact, if an imbalance becomes too large the whole currency union is at the threat of collapse. As monetary policy by itself is not capable of reducing the imbalances, other mitigating instruments are needed. One of these instruments is labour mobility. According to OCA theory, migration should counteract imbalances and raise the stability of the currency union.

However, the OCA theory makes simplified assumptions since it presumes perfect labour mobility as well as homogenous labour. Furthermore, it takes only the internal migration into account but not the external migration from and to non-member countries.

This paper analysis theoretically the weaknesses of these three assumptions and provides extensions of the OCA theory. Furthermore, it investigates empirically if migration mitigates macroeconomic imbalances in the Euro area. It uses two different measures of migration and four different imbalances: unemployment rate, GDP growth, current account and fiscal borrowing. The migration and the imbalance variables are computed as dyadic data, i.e. as country-pairs for all countries in the sample. This approach increases the amount of observations and variation and measures the imbalances accurately. It follows from the OCA theory that migration and all imbalances are endogenous, i.e. they affect each other over time and between member countries. Causality may run into all directions. A Panel VAR set up is an appropriate method to account for those interlinkages.

The paper adds proof that internal migration within the Euro area as well as within the EU reduces the imbalances. The results suggest that the economic structures of the different member countries are similar enough to each other such that migrants relocate efficiently and reduce imbalances. The Freedom of movement for EU citizens may play an important role in this context since it reduces the costs of migration.

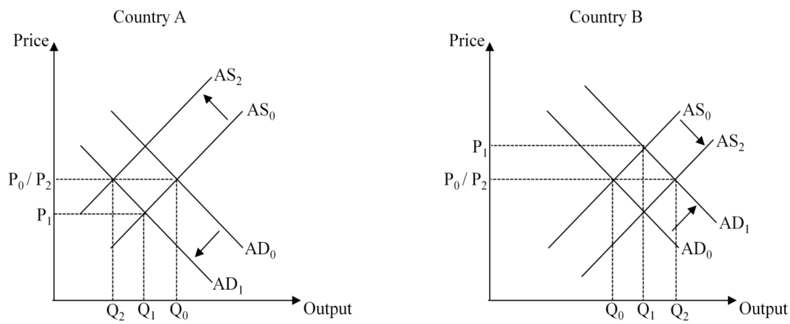
In contrast, the external migration across the borders of the EU does not mitigate the imbalances. The immigration from non-EU countries amplifies the magnitude of the imbalances, makes them more persistent and volatile. One explanation might be that these immigrants do not fit the requirements of the European labour markets, for instance, they might lack the needed education or professional or linguistic skills. Another interpretation could be that the regulations in place are too strict and prevent the external migrants from integrating properly into the EU labour markets.

The contributions of this paper to the literature are threefold. First, it adds a new perspective to the political debate by arguing that internal and external migration might be relevant for the stability of the Euro area. Second, to my knowledge there is no other paper which applies a Panel VAR in the context of migration in the Euro area and the European Union. Third, this paper is pioneering the investigation of the bi-directional causality of migration and imbalances.

1.2. Theory

The theoretical framework for the following empirical analysis is based on Mundell's (1961) standard model of the Optimal Currency Area (OCA) theory. The two countries A and B form a currency union. Each country produces one good respectively. In period t_0, they find themselves in an equilibrium where price levels and outputs are the same, the economies operate at full employment. There are no imbalances between both countries. The situation is illustrated in chart 2.

Chart 2: OCA standard model of aggregate demand and supply in countries A and B



In Period t_0 the quantity and the price of goods are the same in both countries A and B. A shock in consumer preferences at period t_1 shifts aggregate demand AD_1 in country A inwards and in country B outwards. Prices P_1 drop in country A and move upwards in country B. Labour moves in period t_2 from country A to country B as the latter pays higher wages. Aggregate supply AS_2 shifts inwards in country A and outwards in country B. In the new equilibrium, prices are restored to their previous level ($P_2 = P_0$) in both countries. Output Q_2 remains lower in country A than in country B.

The OCA theory describes how an unexpected exogenous asymmetric shock affects both countries differently. The theory also points out how migrants mitigate the impact of the shock.

1.2.1. Unemployment rate

This subchapter focuses on the unemployment rate imbalance. At the beginning, the unemployment rates are equal in both countries: $U_{A0} = U_{B0}$. There are no imbalances in period t_0.

Period t_1 shall begin with a shock in consumer preferences. Consumers shift their preference towards the good produced in country B. The demand curve in country B shifts outwards, leading to a larger output and higher prices. As companies demand more workers for production the unemployment rate u shrinks from period t_0 to period t_1 : $U_{B1} < U_{B0}$.

In contrast, the demand curve in country A moves inwards, reducing output and prices. Companies employ less labour and the unemployment rate rises : $U_{A1} > U_{A0}$.

Eventually, an imbalance has occurred between both countries with respect to their unemployment rates as in period t_1 the unemployment rate is larger in country A than in B : $U_{A1} > U_{B1}$ or $U_{A1} - U_{B1} > 0$.

Monetary policy is the same in both countries as they form a currency union. In fact, monetary policy cannot resolve the difference in price levels or the unemployment imbalance. A restrictive monetary policy, i.e. an increase in the primary interest rate, would be adequate for country B as prices would go down and the unemployment rate up. However, this policy would also reduce the prices in country A by the same amount and its unemployment rates would increase. Eventually, the price differences and the unemployment imbalance prevail.

Analogously, an expansionary monetary policy, i.e. a cut of the primary interest rate, would also fail to resolve the price differences and the unemployment imbalance. To that end, the currency union requires other adjustment mechanisms. One such mechanism is labour migration between countries.

Wages in country B are higher than in country A as a result of the unemployment rate imbalance. The unemployed workers of country A emigrate to country B in order to earn higher wages.

At period t_2, the supply of labour in country A shrinks, shifting the supply curve inwards. The unemployment rate falls, pushing up wages and prices until the previous equilibrium is restored. The unemployment rate drops to its initial level : $U_{A1} > U_{A2} = U_{A0}$.

In country B, the immigrants increase the supply of labour, shifting the supply curve of country B outwards. As a result, companies reduce wages which translates into lower prices. Eventually, prices reach their previous level, just as the unemployment rate : $U_{B1} < U_{B2} = U_{B0}$.

The temporary imbalance of unemployment rates was eliminated by labour migration : $U_{A2} = U_{B2} = U_{A0} = U_{B0}$.

Other mechanisms which could potentially mitigate imbalances are flexible wages and prices. OCA theory argues that fully flexible wages and prices would allow to establish a new equilibrium between both countries A and B. Thus, migration would not be needed to compensate for imbalances. However, in practice, prices and wages are downward rigid, i.e. companies in the Euro area barely reduce salaries (Schmitt-Grohé and Uribe, 2013).

Decressin and Fatás (1994) find that labor market adjustment in the United States takes mainly place via migration whereas in Europe the participation rate is the most important adjustment mechanism. The unemployment rate plays a minor role in both regions. Yet, Pasimeni (2014) points out that the labor market in the Euro area functions as the main adjustment mechanism to counterbalance asymmetric shocks. This leads to sustained and higher unemployment rates in the Euro area compared to the US and other regions worldwide. Dao et al. (2014) confirm all these conclusions.

Cavelaars and Hessel (2007) analyze the direction of causality between the imbalances and migration. They use OLS in order to estimate net migration in percent of the population as a function of relative wages (using income per capita as a proxy) and relative unemployment as well as regional fixed effects due to the strong persistence in regional migration. They find that regional unemployment is a weaker trigger for migration the higher the overall level of unemployment in the whole EU.

In addition to the unemployment imbalance, the standard model allows to derive three other imbalances: GDP per capita, fiscal borrowing and current account.

1.2.2. GDP per capita

In the standard framework of the OCA theory, GDP can be computed by multiplying the output with the price level. In period t_0, the GDP as well as the population size are assumed to be the same for both countries A and B. It follows that the GDP per capita has to be equal for both countries.

The consumer preferences shock in period t_1 disturbs this balance. While country A faces a lower output and therefore lower GDP per capita, country B increases both. At this period, no migration has taken place yet, hence the population sizes have remained constant. An imbalance has appeared with GDP per capita being larger in country B than in A.

GDP per capita can also be interpreted as a proxy for wages. The unemployed workers of country A are attracted by the higher GDP per capita and migrate to country B. Thus, the labour supply in country A shrinks, prices go up and output falls. Since this emigration reduces the population size, per capita GDP returns to its original level.

Country B, however, sees an extension of its labour supply, prices fall and output increases. But the population grows due to the immigration and GDP per capita drops to its previous level. Eventually, the temporary imbalance vanishes as labour migration equalizes the per capita GDP in both countries.

Zimmermann et al. (2014) report that Lithuania has seen a large scale emigration since it became a member of the EU. The reasons for the outmigration are higher salaries and lower unemployment rates in the EU as compared to Lithuania. The migration outflow reduced the excess labor supply in Lithuania, lowered the local unemployment and increased wages.

1.2.3. Fiscal borrowing

Let us assume that each country runs a public unemployment insurance scheme and that the fiscal budgets in both countries are balanced out. The rise of unemployment in country A in period t_1 implies that the government has to provide more unemployment benefits. In order to prevent a further decline of aggregate demand the government prefers not to increase taxes but to borrow money from financial markets. Thus a fiscal budget deficit corresponds to an increase in fiscal borrowing.

While country A runs a fiscal budget deficit, country B benefits from a fiscal budget surplus as less benefits have to be paid out to less unemployed workers. In other words, the consumer preference shock has lead to an imbalance of fiscal budgets and of fiscal borrowing. Emigration in period t_2 relieves the pressure from country A as the unemployed rate drops to its original level and the fiscal budget is balanced out again. Likewise, the extension of the labour supply in country B leads to more unemployment and the previous fiscal surplus disappears. The imbalance between both countries in terms of fiscal borrowing is resolved by labour migration.

The original OCA theory also points out that fiscal transfers between countries A and B could mitigate the unemployment imbalance. Country B would transfer its fiscal surplus to country A in order to lower aggregate demand in the former and to raise aggregate demand in the latter. As such, labour migration would not be required as an imbalance mitigation mechanism. However, in practice the countries of the Euro area do not have a common fiscal policy. The member countries defend their nation-focused policies and resist efforts to create a common fiscal policy. Fiscal policy is a minor instrument to reduce imbalances and does not replace labour migration in the Euro area.

1.2.4. Current account

The fiscal borrowing imbalance is often accompanied by a current account imbalance. Such a situation is referred to as twin deficits. The reason is that the current account mirrors capital flows. The current account can be interpreted as the difference between domestic investments and savings. A government which borrows money is importing capital from abroad which leads to a current account deficit.

The current account also reflects the cross border trade of goods and services. As Kenen (1969) points out the current account can also be inferred implicitly from the standard model. Let us assume that in period t_0 country A produces one good and country B produces another good. Both goods are consumed to equal shares by both countries, i.e. country A exports half of its good to country B, and country B also exports half of its good to country A. Thus, the current accounts of both countries are in equilibrium as exports and imports cancel each other out. The shock of consumption preferences in period t_1, however, creates an imbalance. Demand for the good of country B increases at the expense of the good of country A. Country A will export less of its own good and import more of country B's good. Country A runs now a current account deficit. The situation is opposite for country B, now achieving a current account surplus. The difference between deficit and surplus marks the current account imbalance between both countries. The migration flows in period t_2 restore the previous equilibrium.

A country running a trade deficit borrows capital from the rest of the world to finance the trade deficit. If the current account deficit persists over an extended period, the country accumulates debts which may require at a certain point in time painful price and wage adjustments in order to prevent a default. Eichengreen (2014) indicates that the southern European countries have accumulated large current account deficits over several years until the Great Recession in 2008/09 while the central and northern countries have been running current account surpluses. Persistent current account imbalances have built up within the Euro area. These imbalances led to large debt accumulations and some of the southern countries nearly went bankrupt. Unemployment rates increased. The southern countries implemented adjustment mechanisms to correct the imbalances. Emigration was one of these adjustment mechanisms. Indeed, net migration from the south to the north surged in the aftermath of the Great Recession. The standard model would suggest that migration helped to correct the unemployment and current account imbalances.

Berger and Nitsch (2010) prove that trade imbalances between the Euro member states have become even more persistent after the introduction of the common currency. However, those trade imbalances triggered migration which in turn partially lowered those trade imbalances.

Lange and Gollin (2013) argue that immigrated labour increases the marginal profit of capital as the immigrated workers need to be equipped with capital in order to be productive. A rise in labour immigration can induce capital inflows and thus worsen a current account deficit. On the other hand, many immigrants send money to their relatives who stayed behind in the home country. These remittances improve the current account balance (OECD, 2006).

1.3. Criticism of theory

The standard model of the OCA theory provides insights how labour migration is able to mitigate economic imbalances against which lower the efficacy of monetary policy. However, I would like to analyze three shortcomings of standard model: the assumption of perfect labour mobility, the assumption of homogenous labour and the exclusive focus on union members.

1.3.1. Limited labour mobility

The standard model assumes that labour is perfectly mobile and unemployed workers will relocate to the country with the highest wages. However, this assumption often does not hold in reality.

A common approach to model migration between two countries is the gravity model (Garcia et al., 2015). Applied to the standard model of countries A and B, a worker from country A will emigrate to country B if the following condition is fulfilled, otherwise he will stay in his home country A:

$$E(w_B) - C_{BA} > E(w_A)$$

The condition establishes that a worker will migrate from country A to country B only if the expected wage in country B $E(w_B)$ minus the costs of migration between both countries C_{BA} is larger than the expected wage in country A $E(w_A)$. In empirical research the expected wages are often approximated with the unemployment rate – to estimate the probability of finding a job – and the GDP per capita as a proxy for the average income. In this sense, the standard model of Mundell (1961) already incorporates that migration depends on the difference in expected wages between country A and B. However, the standard model abstracts from the costs of migration, setting them to zero. But the costs are always positive. Migration costs include travel expenses, relocation expenditures, linguistic barriers, social costs, non-transferability of social rights and non-recognition of educational or professional certificates.

In fact, these migration costs can be prohibitively high such that a worker decides not to migrate. The migration costs also depend on the individual worker as well as to which country he wants to emigrate. For instance, a native worker in France might have much smaller costs to emigrate to the French-speaking Quebec in Canada than to its neighboring but English-speaking province of Ontario.

In the context of Mundell's (1961) standard model, in period t_0 no worker will migrate from country A to country B as both countries pay the same wages and with positive migration costs the condition of the gravity model is not fulfilled. In period t_1, the consumer preference shock increases the expected wage in country B and reduces the expected wage in country A.

If this wage imbalance is larger than the migration costs, then in period t_2 the unemployed workers of country A will emigrate to country B. However, in contrast to the standard model the migration costs are not zero but positive. Each single migrant increases the labour supply in country B, thus pushing the local wage down a little bit. Likewise this migrant reduces labour supply A, pushing up the local wage there gradually. The imbalance in expected wages shrinks step by step until it equals the positive migration costs. No more unemployed workers will migrate anymore to country B. In country B, immigration is smaller than compared to the standard model. The labour supply expands less, the unemployment rate ends up at a lower level than in period t_0 while prices are elevated. Analogously, country A ends up having a higher unemployment rate and lower prices. The migration costs prevent the unemployment imbalance to disappear completely. The same applies to the other imbalances with respect to GDP per capita, fiscal balance and current account.

Cavelaars and Hessel (2007) argue that labor mobility is not a mechanism for stabilization in the Euro area. They find that migration rates react just little to regional unemployment differences and even less to wage differences. They conclude that migration is not just a weak instrument to mitigate imbalances but a threat to the stability of the currency union since other mechanisms such as wages and prices have to carry more of the adjustment burden.

In order to understand the important role of the costs of migration, a comparison of labour mobility in the United States and in Europe may yield insights. Literature has traditionally regarded labor as more mobile in the United States than in the EU as costs of migration are lower in the US. First such evidence was brought up by Blanchard and Katz (1992) who showed that labor mobility is a crucial mechanism to adjust to economic shocks in the United States.

According to Feldstein (2008) the Euro area's labor market is characterized by many obstacles. Social factors such as the large variety of languages and cultures decrease labor mobility. Furthermore, trade unions and labor market regulations also hinder inter-industry and cross-border migration. Zimmermann et al. (2014) add that EU-intra mobility is hindered by problematic transferability of social rights, consumer rights and health insurance issues. Similarly, Jauer et al. (2014) refer to migration obstacles such as liquidity constraints, poorly functioning housing and rental market as well as the far from optimal approval of foreign academic and professional qualifications. Meardi (2012) indicates that immigrants in the EU have below average bargaining power compared to their new employers and do not fully integrate into the host labor market.

However, Beyer and Smets (2014) and Molloy et al. (2011) document that the gap in labor mobility between the US and Europe is shrinking. Mobility in the US has declined over the past 30 years while it has risen in Europe as the costs of migration have increased in the United States. Frey (2009) argues that a higher rate of home ownership, the ageing baby-boom generation as well as higher female labor market participation might account partially for the decline in US mobility over the past decades. Especially the Great Recession made it more challenging for potential movers to find employment in other regions of the United States. The crisis raised the difficulty to finance a house at affordable conditions or to find a buyer for the estate.

In contrast, the costs of migration have sunken in Europe. Beine et al. (2019) explain the increase of the mobility in Europe over the last decades with the Schengen agreement, hence the Freedom of movement, as well as the introduction of the currency union.

1.3.2. Non-homogenous labour

The standard model assumes that labour is homogenous. According to Kenen (1969) this assumption does not necessarily hold as workers have different levels of human capital, consisting of formal education, skills and working experience. Furthermore, this assumption covers up important consequences resulting from the heterogeneity of labour.

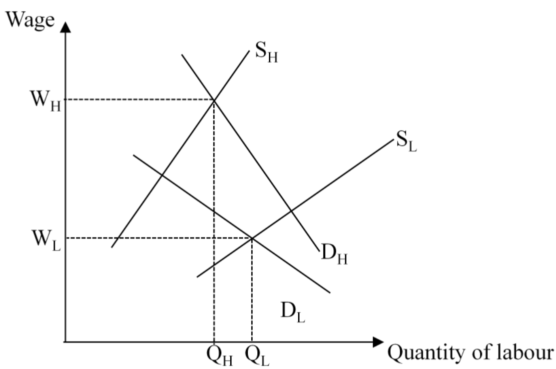
Let us allow for more heterogeneity by distinguishing highly skilled workers and lowly skilled workers. Highly skilled workers are more productive than lowly skilled workers. The demand for highly skilled workers is therefore larger, resulting in lower unemployment rates and in higher wages than for lowly skilled workers.

During the Great Recession did not only the amount of net migrants from southern Europe (namely Portugal, Spain, Italy and Greece) to the core and the north of Europe (Germany, Netherlands, Sweden, Finland and others) increase, but also the share of the highly skilled emigrants at the expense of the share of the lowly skilled emigrants (European Commission, 2018).

The reason behind this self-selection bias is that the expected wages abroad are higher for highly skilled workers than for lowly skilled workers. Thus, migration costs are more binding for lowly skilled than for highly skilled workers.

The home country faces not only an increase of absolute net emigration but in particular an emigration of highly skilled labour. This so-called brain drain effect eventually amplifies the imbalances.

Chart 3: Labour market for highly skilled and lowly skilled workers



S_H and S_L describe the supply curves of highly skilled labour and lowly skilled labour, respectively. D_H and D_L refer to their demand curves. As highly skilled labour is more productive than lowly skilled labour, the demand is larger for the former than for the latter. Yet, as acquiring high skills requires investment costs, the supply of highly skilled labour is more rare and these workers ask for higher wages W_H than lowly skilled labour (W_L). Therefore, the market-clearing quantity of highly skilled labour Q_H is smaller than of lowly skilled labour Q_L. Source: www.economicsonline.co.uk.

Chart 3 illustrates a labour market which distinguishes highly skilled labour demand and supply and lowly skilled labour demand and supply. The demand is stronger for highly skilled than for lowly skilled workers as the former is more productive and generates more output. However, workers need to invest time and money to build up high skills. Therefore, the supply of highly skilled workers is smaller than the supply of lowly skilled workers. To compensate for this investments, highly skilled workers expect higher wages.

Let us use again the standard OCA theory model. Country B attracts migrants from country A. Let us assume that migration costs are initially binding for lowly skilled immigrants such that only highly skilled workers are able to migrate from country A to B.

Expected wages for highly skilled workers in country i are represented by $E(w_i^H)$, for lowly skilled workers by $E(w_i^L)$. The migration costs C_{BA} are identical for all workers irrespective of their skills. In period t_0, wages for highly skilled workers are the same in country A and B. The same applies to lowly skilled workers' wages. The shock in period t_1 lowers wages for both worker groups in country A and raises their wages in country B. It is assumed that migration costs are only too large for lowly skilled workers:

$$E(w_B^H) - E(w_A^H) > C_{BA} > E(w_B^L) - E(w_A^L)$$

Therefore, only highly skilled labour migrates from country A to B. The influx of highly skilled workers shifts their supply curve outwards in country B. Thanks to their high productivity they increase the aggregate output significantly. By definition, the output of a highly skilled worker is larger than the output of a lowly skilled worker. Therefore, any additional highly skilled immigrant increases the per capita GDP in country B.

Another feature of the heterogeneous labour market is that highly skilled workers and lowly skilled workers are complementary. A highly skilled worker could be, for instance, an entrepreneur who founds a new company and hires lowly skilled workers to produce goods. The increase of highly skilled labour supply raises the demand for lowly skilled workers, leading to an outward shift of the demand for lowly skilled workers in country B. Their wages rise.

Thanks to this higher expected wage the migration costs are not binding anymore for the lowly skilled workers in country A. They will also outmigrate to country B, and they add to the supply of lowly skilled workers there and thus increase output even further. Eventually, the unemployment rate shrinks and GDP per capita grows.

The effects for country A, however, have the opposite sign. The initial outflow of the highly skilled workers implies that the most productive labour leaves the country. With each highly skilled worker leaving, GDP per capita of country A falls. Furthermore, due to the complementarity the demand for lowly skilled workers in country A decreases, too, and the unemployment rate rises.

However, the immigration of lowly skilled workers adds to their labour supply in country B, thus pushing down their wages until the migration costs become binding once again. Due to this constraint labour migration is not large enough to fully offset the imbalances in terms of GDP per capita and unemployment rates. In contrast, as in particular highly skilled workers migrate, the GDP per capita is persistently higher in country B than in country A. Likewise, the current account and fiscal borrowing imbalances build up as well and they are not fully cut back by labour mobility.

Borjas (1987) finds that migration depends on self-selection. Particularly young and highly-educated individuals decide to move due to better employment opportunities abroad.

O'Rourke and Taylor (2013) point out that emigration depletes the home country's tax base which is needed to pay for expenditures and pensions, thus raising the fiscal borrowing and debt and, eventually, the probability of the country's bankruptcy.

According to Dhéret et al. (2013) the Great Recession fostered migration in the Euro area from geographically peripheral to central countries as the former group faced a severe economic downturn while the latter achieved higher economic growth rates. In detail, 1.7% of Greek residents left the country in 2011, 0.7% Spain and 0.6% Portugal. Also emigration out of the Euro area has risen though to a smaller degree than inside the currency union.

1.3.3. Non-member countries

According to Mundell (1961) the optimal currency area is a region where labour and capital are fully mobile within but immobile externally. However, this definition does not even apply strictly to the Euro area. Also external labour is mobile since the citizens of the European Union and of the EFTA (European Free Trade Association) outside of the currency area enjoy the same rights of movement as the Euro countries. In addition, the theory has not investigated how external migration across the borders of the currency area affect the internal imbalances within the currency area.

Adding an external country C which is not a member of the currency union can have an indirect effect on the internal imbalances between the two countries A and B which form the currency union.

Referring to the standard model with homogenous labour, once more a shock in consumption preferences in favor of country B at the expense of country A occurs. This leads to an unemployment rate imbalance between countries A and B. The unemployed workers in country A have to decide if they stay at home, if they emigrate to country B or if they emigrate to country C.

Assuming that the expected wages net of migration costs are the same for both destinations and that they are higher than the wage in country A, half of the unemployed will emigrate to country B and the other half to C.

Thus, the labour supply in country A shifts inwards until its previous price level and unemployment rate are re-established. Yet, country B widens its labour supply less than in the standard model since it has received less immigrants as the other unemployed workers of country A migrated to country C. Thus, the price level in country B remains elevated while the unemployment rate remains under its initial level. The imbalance between countries A and B does not fully disappear.

Another scenario is that the exogenous shock shifts aggregate demand in country B outwards and, therefore, also its demand for labour increases. In turn, wages in country B rise which attracts not only migrants from country A but country C as well. Not all immigrants in country B will now stem exclusively from country A but also from C. The immigrants increase the labour supply in country B until its previous equilibrium in terms of prices and unemployment is achieved again. Yet, country A faces an elevated unemployment rate and lower price level since not enough unemployed workers emigrated.

Analogously, introducing labour mobility with a non-member country in the standard model may have not only an impact on the unemployment rate imbalance between member countries but also on the other imbalances, i.e. GDP per capita, fiscal borrowing and current account.

Also empirical literature suggests that migrants inside the currency union should be distinguished from those outside since their impact on the imbalances between Euro countries might be different. Zimmermann et al. (2014) investigate the macroeconomic effects of the EU-8 enlargement. They report a large increase of the migration flows from "east to west" which helped to raise GDP per capita and stabilize inflation rates in the EU.

Zimmermann and Zaiceva (2012) show that especially migrants from the new eastern EU member states returned home from the Euro countries due to the financial crises. These results support the argument that nationals of the Euro area are less mobile while the labor adjustments are shouldered to a significant degree by EU citizens who do not belong to the currency area. Also Jauer et al. (2014) suggest that the labor market adjustment in the Euro area was mainly shouldered by non-Euro area nationals. That includes people from the new EU-member states who gained full labor market access only in 2007 as well as from third countries.

In sum, theoretical and empirical literature suggests that migration and imbalances interact with each other and that causality runs both ways. It is, however, unclear if migration always reduces the imbalances or if migration can also amplify the imbalances, leading to less stability of the currency union. Furthermore, internal migration and external migration should be distinguished from each other since they could react differently to the imbalances and they might affect the imbalances differently themselves. Based on the theoretic argumentation I derive the following hypotheses for the Euro area:

1. Imbalances are mitigated by internal migration.

2. Imbalances are not mitigated by external migration.

3. Internal migration does not cause imbalances.

4. External migration causes imbalances.

1.4. Data set

This subchapter discusses how the migration and the economic variables are computed which will be used in the following Panel VAR model. It also debates the use of dyadic data and describes the data set.

1.4.1. Migration variables

The main variable of interest is migration. A migrant is a person who relocates from one country to another with the intention of a longer stay, that is, for at least one year. There are three options to calculate a migration variable.

The first option is perhaps the most intuitive one: the direct migration flow between two countries. For instance, for the country-pair Belgium - Germany all Belgians, Germans and third country citizens migrating from Belgium to Germany and the other way around are included. I call this variable "direct net migration" and compute it as net migration, that is, immigration minus emigration. In addition, I divide the direct net migration through its respective country's population. This variable requires data on the previous residence of an immigrant, respectively the destination of the emigrant. Unfortunately, data for this option is most scarce as compared to both next options for defining a migration variable.

The second option is based on citizenship or based on country of birth. This data is much more broadly available than data based on previous or next residence. If migration flows are not available then these figures can be calculated by the stock of foreign citizens on any country's territory, for instance via regression or Bayesian estimation techniques which even correct for birth and death rates (compare Abel and Sander, 2014). However, this data acts in best case just as a proxy for data of previous residence for two reasons. First, data sources often pool the criteria citizenship and country of birth. Yet, these two definitions are not always identical. A person born on US territory, for instance, becomes a US citizen. A person born on Swiss soil, however, does not claim automatically Swiss citizenship unless one of her parents is a Swiss citizen. Second, using data on citizenship as a proxy for data on previous residence does not represent migration flows accurately. It assumes that any migrant with a specific citizenship immigrated from her home country. Therefore, this data rests on the assumption that, for instance, an Austrian in Spain must have immigrated directly from Austria, his home country. The data ignores the possibility that the Austrian might have well lived for many years in the Netherlands before relocating to Spain. Furthermore, this second option also neglects any potential home comers, i.e. circular migration. That is, for instance, a Spanish citizen who had left his home country to move to Austria in the past and who now returns to Spain. Due to these disadvantages I do not base my research on this second option.

The third option does not look at the migration flows between two specific countries. Instead, it subtracts total net migration of one country from the total net migration flow of another country. As an example, let us consider the Portugal-France pair. I first compute the total net migration for Portugal by calculating the immigration to Portugal from all other countries around the world minus the emigration from Portugal to the rest of the world. Then, I divide Portugal's total net migration rate through its population. The same steps I repeat for France. Finally, I calculate Portugal's total net migration minus France's total net migration.

One advantage of this option is the broad availability of data. Furthermore, this option is not limited to the citizens of the respective country, but it includes home-coming natives and third country foreigners as well. Moreover, the total net migration variable covers not just direct migration between two countries but also migration flows from and to third countries. It accounts for the indirect effects of those flows on the imbalances between the two respective countries.

Eventually, I use direct net migration and total net migration as variables for migration. I will not use these two variables simultaneously in the PVAR model. Instead, I conduct the PVAR two times, once with direct net migration and another time with total net migration.

Parts of literature focus on a certain kind of migrants based on their labour market status or socio-economic criteria. Depending on the object of research it might make sense to distinguish workers, their relatives, seasonal workers, refugees etc. This paper aims to capture any potential interlinkage between migration and macroeconomic imbalances. Therefore, it is appropriate to not differentiate migrants by their labour market status or socio-economic criteria. However, this paper does distinguish between migrants from the Euro area, the EU and from outside. Furthermore, the data set reaches until the year 2014 in order to prevent any bias due to the unique, large-scale refugee influx in the year 2015 and thereafter.

1.4.2. Imbalance variables

The following imbalances are derived from the OCA theory framework as stated before. All four imbalances affect each other directly. Theory also suggests that they have a causal effect on migration and, in turn, migration on them:

- Annual growth of real GDP per capita: A fast growing economy might attract migrants from abroad as they use the per capita GDP as an indicator for their expected wage. On the other hand, immigrants might increase per capita GDP, especially if they are highly skilled and complementary to the local work force.
- Current account: The current account mirrors international trade and capital flows. A high surplus indicates a highly competitive economy which attracts foreigners. In turn, immigrating workers might raise demand for goods from their home countries, thus increasing imports.
- Unemployment rate: A low unemployment rate signals a high probability of finding employment which will attract immigrants. On the other hand, immigrants can limit upward wage pressure by increasing the labour supply. Furthermore, they can create additional demand for labour in the destination country if they are complementary to the local work force, reducing the local unemployment rate. They also lower the unemployment rate in their home country if they used to be unemployed there.
- Fiscal borrowing: If immigrants can expect a generous welfare system in the destination country, those expenditures will increase the borrowing of the government. In contrast, immigrants often pay more taxes and social contributions than they receive in return which lowers the government's need to borrow money from financial markets. Emigrants who received unemployment benefits in their home country release the constrain on the fiscal budget of their home country which had to borrow capital in order to finance the unemployment benefits.

1.4.3. Description of data set

All data used in this paper is provided by Eurostat[1]. My data set stretches from year 1998 until 2014, i.e. over 17 years. During that period no significant refugee flows occurred within the Euro area and the EU nor across its borders. The large inflow of refugees from Syria, Afghanistan, Iraq and elsewhere started in 2015, hence in the year after my data set ends.

I consider all Euro, EU and EFTA member states. The EU had 15 members in 1998 and consists nowadays of 28 states. EU membership grants its citizens the Freedom of

[1] http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/database

movement which allows all EU citizens to seek employment in any of the member states without facing discrimination. However, this freedom has not been granted immediately to all new east European member states since 2004. The negotiations for all EU enlargements after 2004 allowed the old member countries to shield their labor markets from the new fellow EU citizens for up to seven years. Most old member states did use this option.

EU membership is a prerequisite to become a member of the Euro area. The currency union started in 1999 with 12 countries and enlarged itself to currently 19 members.

There are currently four EFTA countries, namely Switzerland, Norway, Iceland and Liechtenstein. They all have access to the EU single market and share the Freedom of movement. Their laws and regulations are also very close to the EU. I exclude Liechtenstein from the sample as there are too few observations. However, due to the country's small population it should be safe to assume that its migration would not have any significant impact on Euro area imbalances anyway. Hence, in my dataset the EU and EFTA amount to 31 countries.

I calculate all variables as the difference for any available country-pair, e.g. the difference in growth rates between Germany and Belgium, Germany and Spain, Belgium and Spain etc. There are 441 country-pairs in total.

To ensure consistency, I compute the two migration variables and the four imbalance variables in the same direction. For instance, to compute the real GDP per capita growth imbalance for the country-pair Belgium - Germany I calculate the growth rate of Belgium minus Germany's growth. Likewise, I calculate the direct net migration variable as immigration to Belgium from Germany minus emigration from Belgium to Germany.

Thus, with a time dimension of 17 years and a cross-sectional dimension of 441 country-pairs, a maximum of 7497 observations is possible. The data, however, is unbalanced. Table 1 shows the actual amount of observations, mean, standard deviation, minimum and maximum values. All variables are computed as imbalances, that is, as bilateral differences between two countries.

Table 1: Descriptive statistics

	N	Mean	Std. deviation	Min.	Max.
Direct net migration in percent of population	3068	0.00	0.05	- 1.26	1.28
Total net migration in percent of population	6061	-0.03	0.83	- 4.01	4.42
Real GDP per capita growth in percent	6670	0.03	3.96	-17.76	17.05
Current account in percent of GDP	5522	-0.93	8.52	-33.60	28.90
Unemployment rate in percent	7211	0.26	6.06	-22.70	24.10
Fiscal borrowing in percent of GDP	5961	0.37	4.65	-32.00	32.30

Direct net migration - that is the bilateral migration flow between two countries - is the most intuitive and suitable variable to capture migration. Yet, the data base is rather limited with 3068 observations.

In contrast, total net migration offers almost twice as many observations. This variable subtracts net migration for one country (immigration minus emigration from to the rest of the world) from net migration of a second country, relative to the reference country's population. Its magnitude is larger than of direct net migration since total net migration aggregates all migration flows of a country.

The next four variables denote the macroeconomic imbalances derived from the OCA theory. Just as the migration variables also the macroeconomic imbalances are computed as dyadic data. While the average values are comparable to single-country data sets in the literature, the standard deviation as well as the minimum and maximum values indicate the advantage of more variation of dyadic data.

1.4.4. Dyadic data

Macroeconomic literature usually uses total volume, aggregate, country-specific data. In contrast, this paper computes all migration and imbalance variables as dyadic data, i.e. as country-pairs. Dyadic data is commonly used in trade literature (Head and Mayer, 2014) and migration literature (Beine et al., 2016). This paper combines macroeconomic literature and migration literature. Dyadic data is better suited to investigate the stated hypotheses.

The OCA theory suggests that the magnitude of the imbalance between two member countries has an impact on the effectiveness of monetary policy, i.e. the larger the imbalance the harder it is for the central bank to find the appropriate monetary policy. Dyadic data captures the magnitude of the imbalance accurately. In contrast, if the imbalance between two member countries is measured on a country-level then the magnitude is biased as the country-level data reflects the average imbalance between a particular country and the rest of the world. The latter magnitude may be smaller or larger than the dyadic-measured imbalance. The bias also persists if the imbalances are calculated as the difference between a member state and the average value of the currency area as can be found in the literature.

Following the OCA theory, this paper defines a macroeconomic imbalance as the difference between two countries with respect to a certain economic variable. The use of dyadic data is coherent as it reflects bilateral relations between two countries as in contrast to the single-country data which aggregates the relation with the whole rest of the world. Furthermore, migration flows are usually captured as net bilateral flows between countries. To ensure coherence, the imbalances should be calculated as dyadic data as well.

Country-specific data aggregates the migration flows and the economic flows between one country and the rest of the world. It thus summarizes effects of different directions and magnitudes which can cancel each other out and falsely lead to non-significant coefficients. Dyadic data, however, allows to disentangle the different flows and their coefficients can be statistically significant. For instance, Colen et al. (2014) point out that the early literature on Foreign Direct Investments (FDI) which used aggregate country-specific data found that bilateral investment treaties do not have any statistically significant impact on FDI. However, more recent studies applied bilateral data and confirmed that the treaties do stimulate FDI flows.

The gravity model formulates a migrant's decision to emigrate to a particular country. It typically defines the dependent variable as the net migration between two countries. The independent variables such as GDP growth rates and unemployment rates are also defined as the difference between two countries. Therefore, dyadic data is particularly useful in order to estimate the gravity model on which the PVAR model in this paper build upon. The dyadic data reflects the migration decisions for a particular country-pair. In contrast, the aggregated country-specific data only summarizes migration between one country and the whole rest of the world.

The Euro area can be thought of as a network with each country as a node. Any change in one node in terms of migration or imbalance has direct and indirect effects on all other nodes and might fire back on the original node. Due to its network structure, migration between two nodes might regulate indirectly the imbalances between two other nodes. Dyadic data allows to capture these effects. Furthermore, image a worker who wants to emigrate and that he can chose between two destination countries. If the attractiveness of the first country drops, the probability that the worker will migrate to the second country increases automatically. Dyadic data replicates those decisions.

Beine et al. (2016) point out that dyadic data is useful for the analysis of pair-specific institutions such as currency unions and free trade agreements which is appropriate for the context of this paper.

Berger and Nitsch (2010) show that the creation of the Euro area led to an increase in trade between member states and a closer synchronization of their business cycles. Dyadic data captures this integration and synchronization better than country-level data which is a weighted average of that country's relation with all member but also non-member countries around the world.

The European Union and therefore the Euro area have established free trade between their member states, removing tariffs and other trade barriers. Trade agreements foster trade between the connected countries, leading usually to more-intense trade between member countries than with outsiders. As this paper distinguishes between internal and external migration, it also makes sense to differentiate trade between members and outsiders. The dyadic data allows to distinguish these trade flows, which are reflected in the current account imbalance.

Another advantage of dyadic data for regression estimations is that it grants more variation than single-country data. The coefficient of variation allows to compare the dispersion of different samples. It is defined as the ratio between the standard deviation and the average of a time series. However, the coefficient of variation can only be interpreted in a meaningful way if all observations of a sample are strictly positive or strictly negative.

My sample consists of 31 EU/EFTA-countries with observations between the years 1998 and 2014. However, the data set is unbalanced. For each country I first calculate the total net migration rate, i.e. the amount of immigrants minus the amount of emigrants with the rest of the world divided through the country's population. Then, I compute the mean and the standard deviation of the absolute total net migration rates for each country over time. This single-country data yields on average for the 31 EU/EFTA countries a coefficient of variation is 0.59. The smallest coefficient is 0.29, the largest 1.11.

For comparison I look now at country-pairs. Therefore, I calculate for each EU/EFTA country-pair the difference between their respective total net migration rates. These differences are available for 441 country-pairs. This dyadic total net migration data yields on average a coefficient of variation of 0.61. The smallest coefficient is 0.09, the largest 1.43 (all computed for the absolute differences of the total net migration rates). As compared to the single-country data the variation of the observations has increased.

However, these total net migration rates do not show true bilateral flows between two countries. They rather represent the difference in net migration rates between one country with the rest of the world and another country with the rest of the world. My data set also includes the direct bilateral migrant flows between all EU/EFTA countries without any flows related to the rest of the world. However, the dyadic direct migration flows are reported less often then the dyadic total migration rates but this subsample still contains 329 country-pairs. The coefficient of variation for the dyadic direct migration rates is on average 0.74, and at least 0.01 and 2.1 at maximum. Thus, the variation of the dyadic direct migration rates is larger than in the case of the single-country data and of the total migration rates.

Moreover, the difference of total migration rates might be misleading as a country could be a net immigration country as compared to the rest of the world but might be a net emigration country if compared to a specific other country. When for a country-pair both direct and total migration rates are available, the signs of the net total migration and net direct migration are not always the same. In fact, the signs are different in one third of all those observations.

1.5. The Panel VAR Model

The aim of this paper is to shed light on the question whether migration increases the stability of a currency union by countering imbalances. To that end, I take into account the limits of the OCA theory as well as bilateral causalities between migration and imbalances. In particular, I run a Panel VAR.

1.5.1. Drawbacks of traditional VAR and panel data models

There are various approaches to capture migration flows on the micro and the macro level. The most traditional approach is the gravity model (compare Reilly, 1931; Garcia et al., 2015; Bodvarsson et al., 2014; Kennan and Walker, 2011; Beine et al., 2019).

Another popular approach are Vector Autoregressive models (VAR). VAR models are often used to analyze macroeconomic time series since they allow capturing the interdependent dynamic and static relations between different variables. In addition, all variables interact with each other and their respective lags which is advantageous to capture bidirectional causalities.

Blanchard and Katz (1992) were the first to apply a VAR model in order to identify international and regional factors driving migration. Their results suggest that until that time US citizens were more mobile than Europeans as the costs of migration, such as language barriers, were smaller in the US. Beyer and Smets (2014) have replicated this approach more recently. They show that labor mobility within the Euro area has increased over the past years and converged towards the mobility level of the US.

Unfortunately, international migration data is usually available at an annual frequency only and the time dimension does not reach very far. These circumstances reduce the amount of available observations for VAR models.

To compensate for the rather short time dimension cross-sectional data might be used. Therefore, panel data models are of interest. In particular, the "difference" and "system" GMM dynamic panel estimators have found broad application in literature (compare Arellano and Bond, 1991; Blundell and Bond, 1998; Roodman, 2009; Mayda, 2010).

Yet, there are two drawbacks of the panel models. First, they allow only for one direction of causality, either from imbalance to migration or from migration to imbalance. But theory and empirics suggest that causality runs into both directions. The purpose of this paper is to exploit this endogeneity. Second, this paper rests upon a dynamic framework but the calculation of Fixed Effects in panel data models violate the exogeneity assumption of the residuals (Boubtane et al., 2012).

1.5.2. Advantages of the Panel VAR

Panel VARs (PVAR) overcome the three weaknesses of the VAR models and of the panel data models (Holtz-Eakin et al., 1988). First, the short time horizon of VARs is compensated by the cross-sectional dimension of panel data. Mitze (2012), for instance, applies a PVAR to estimate migration between different regions within Germany by combining the short time dimension with a large cross-section dimension. Second, the mono-direction of causality in panel data models is overcome as PVAR models capture the bi-directional causality between all variables. Third, the GMM panel estimator is not consistent as it would be correlated with the lags of the variables. Forward mean differencing in PVAR models (also known as Helmert procedure) provides a solution: "In this approach, all variables are transformed into deviations from forward means, and each observation is weighted to standardize the variance. This transformation preserves orthogonality between transformed variables and lagged regressors, allowing to use the lagged regressors as instruments and estimate the coefficients by the GMM procedure", as Gnimassoun and Mignon (2012, page 5) put it.

The PVAR takes the following form

$$X_{it} = \Theta(L)X_{it} + u_i + e_{it}$$

where the vector X_{it} includes all variables, i.e. migration and imbalances. $\Theta(L)$ denotes the lag operator, u_i refers to the country-pair fixed effects and e_{it} is a vector of idiosyncratic errors. The model includes country-pair fixed effects, thus controlling for the individual heterogeneity of all 441 country-pairs which may elevate or hinder migration such as language, culture, migrant diasporas and similar. The model is estimated with GMM.

Canova and Ciccarelli (2013) give an overview of the theoretical and empirical PVAR literature and explain in detail the characteristics, benefits and disadvantages of PVARs. They identify three main advantages of PVARs:

1. PVARs capture dynamic as well as static interdependencies between countries.
2. PVARs allow for heterogeneous effects.
3. PVARs allow to simulate how an asymmetric shock evolves over different countries.

These characteristics are well-suited to replicate the Euro area with its many several countries as (i) the member states are economically closely linked to each other due to trade and migration, hence they influence each other dynamically, (ii) the countries possess heterogeneous economic characteristics and as (iii) the Great Recession affected the countries asymmetrically.

This PVAR approach is summarized in a Stata modul developed by Love and Zicchino (2006) which I use for my calculations.

1.5.3. Econometric assumptions of the PVAR

Based on their concept, PVAR models can be interpreted as regular VAR models to which a cross-sectional dimension was added. As such, PVAR models have to fulfill the same statistical and econometrical requirements as regular VAR models.

One requirement is stationarity. Stationarity implies that the mean and the variance of a variable do not depend on time. Otherwise, the statistical features of the time series would not remain constant over time and any analysis would not be reliable. There are various unit root tests to check for stationarity. All migration and imbalance variables are first tested in levels for stationarity. If a variable is found to be non-stationary, then it is a common procedure in the literature to take first differences and to test again for stationarity.

To test for stationarity I use the set of four Fisher-type unit root tests as they can be applied to unbalanced data sets. The Phillips-Perron test option used is robust to serial correlation as in contrast to the augmented Dickey Fuller option. Autoregressive parameters are panel specific, panel means are not subtracted, no time trends or lags are included. The unit roots tests share the null hypothesis of non-stationarity.

The tests in table 2 reject the null hypothesis on levels for direct migration, GDP growth and fiscal borrowing. Yet, the unit root is only rejected for the first differences for total net migration, current account and unemployment rate. The results are the same when I apply the unit root test only to Euro countries.

Table 2: Fisher type unit root tests, Phillips-Perron test option

	Direct net migration	First difference total net migration	GDP pc growth	First difference current account	First difference unemployment rate	Fiscal borrowing
Inverse chi-square P	1375.29 (0.000)	3227.28 (0.000)	2539.37 (0.000)	4004.36 (0.000)	2660.74 (0.000)	1052.11 (0.000)
Inverse normal Z	-6.43 (0.000)	-34.73 (0.000)	-26.19 (0.000)	-42.95 (0.000)	-27.03 (0.000)	-8.68 (0.000)
Inverse logit L*	-11.81 (0.000)	-41.51 (0.000)	-29.63 (0.000)	-53.34 (0.000)	-31.74 (0.000)	-9.03 (0.000)
Modified inv. chi-squared Pm	20.64 (0.000)	59.32 (0.000)	39.46 (0.000)	78.34 (0.000)	42.35 (0.000)	8.94 (0.000)

P-values are in parentheses.

As in the case of regular VAR models, the amount of lagged variables is crucial. Their number is selected based on information criteria. The PVAR models in this paper are estimated using GMM. Common information criteria such as the Akaike or Bayesian information criteria are available only for over-identified GMM models, i.e. when the amount of lags of the dependent variable is larger than the order of the PVAR. In this paper, however, the PVAR models are not over-identified but just-identified. Just-identified models estimated with GMM rely therefore on the coefficient of determination (Love and Zicchino, 2006). The coefficient of determination shows how much of the variation of the covariance matrix of the dependent variables is explained by a given regression model. The coefficient of determination is used to select the appropriate number of lags for a Panel VAR model. The coefficients are reported in the following chapter 1.6. It should be noted in the context of PVAR models, additional lags do not necessarily increase the coefficient of determination but they can reduce the coefficient.

Traditional fixed effects panel models eliminate unobservable panel-specific, time-invariant effects by demeaning. This within-transformation subtracts the panel-specific time average of the variables. However, the strict exogeneity assumption is violated which establishes that the error term has to be uncorrelated with any past and future realizations of the dependent variable. The within transformation implies that the lagged error term is correlated with the contemporaneous independent variable. The demeaning is also troublesome for unbalanced panel data sets. If the observation of a panel for a particular year is missing, the demeaned values for the current and the next year cannot be computed. Both years for the panel will be left out, thus widening the gaps in an unbalanced data set (Love and Zicchino, 2006).

This paper applies the Helmert procedure to overcome the difficulties. The procedure eliminates for each year and panel the mean of future observations. In contrast to the traditional demeaning, the Helmert procedure does not create endogeneity. It also does not reduce the time dimensions, which is particularly beneficial for unbalanced panel data sets as the existing gaps are not widened. Furthermore, the Impulse Response Function require an ordering of the variables which implies that the error terms become orthogonalized. The error terms also must not be autocorrelated and be homoscedastic. The Helmert procedure maintains all three features for the transformed error terms (Love and Zicchino, 2006).

The OCA theory suggests that migration and the imbalances should have a long term relation: if an imbalance increases, then migration should grow, too, and nullify the imbalance over time. From a statistical point of view this long term relation can be interpreted as cointegration. This paper, however, investigates if such a theoretical relation exists indeed. A PVAR model is therefore appropriate as it leaves the result open if the time series are cointegrated or not. The usual approach would be to test for cointegration of time series by checking for the stationarity of their residuals and, if cointegration is confirmed, to implement a Vector Error Correction model. First, the gaps in the unbalanced panel data set in this paper are too large to test for potential cointegration of the variables. Second, a Vector Error Correction model would impose a priori a cointegration structure on the estimations which would be contrary to the aim of this paper to investigate whether there exists a long term relation between migration and the imbalances.

An impulse response function (IRF) shows how a variable in its steady state reacts after it was hit by an exogenous shock. In this paper the steady state corresponds to the theoretical initial state in which all imbalances between Euro countries are balanced out. It could be argued that migration has created such a state without imbalances. The shock refers to the sudden emergence of an imbalance or of a difference in migration rates between countries. The IRF computes how migration and the other imbalances react to the shock and if they return to their steady state after some time. According to the OCA theory migration should react to an exogenous imbalance and lead to a return of the imbalance to its steady state.

The annual inflation rates in most Euro countries are between 0% and 2%. They deviated in the past only temporarily from that steady state corridor. The biggest deviation took place after the Great Recession in 2008/09 when the Euro area faced a deflation. But even then the inflation rates of all countries moved into the same direction and to a similar extent.

Based on country-level data, the annual real per capita GDP growth rate in the Euro area is 0.8%. At any given year the growth rate of the majority of the Euro countries has deviated less than 0.5 percentage points from the year's average growth rate. Therefore, per capita GDP growth in the Euro area can be regarded as steady. However, with the inclusion of new member countries from Eastern Europe the steady state stretched slightly since those countries have a smaller per capita GDP level but larger GDP growth rates than the Western European countries.

Existing differences in unemployment rates between country-specific levels and the Euro area average have largely remained stable over time as they developed roughly in parallel to the Euro area average. With the establishment of the currency union the business cycles of the member states have become more synchronized. After the European debt crisis, however, unemployment has increased more strongly in the periphery than in core Europe.

With respect to the current account and fiscal borrowing two groups can be identified in the Euro area. Northern and central European countries have accumulated current account surpluses and reduced their fiscal borrowing. Southern or peripheral countries built up current account deficits and extended their fiscal borrowing. Both developments happened roughly with the same pace. This implies that within each group, the dyadic data is rather steady and that among the northern countries virtually no imbalances are created. Likewise no imbalances appear between the southern countries. When comparing northern and southern countries the imbalances have grown. But as the imbalances grew at a steady pace the current account and the fiscal borrowing imbalances can be regarded as being in their steady state.

1.5.4. Identification strategy

I use the dyadic variables in PVAR models in order to compute an Impulse Response Function (IRF). I use the IRF to simulate a shock to a variable in its steady state and to investigate the reaction of migration and the other imbalances. According to the OCA theory, a shock to an imbalance should vanish over time as migration mitigates the impact. However, due to the shortcomings of the theory other results cannot be ruled out. I postulate that internal migration mitigates an imbalance shock but external migration does not. I use the IRFs to analyze if imbalances become too large or persist too long and threaten to destabilize the currency union.

Neither theory nor literature provide a quantitative indication for how large an imbalance has to be or how long it has to last in order to render the monetary policy powerless or to cause the collapse of the currency union. It is only clear that the larger the imbalance is or the longer it persists, the higher the probability of collapse.

As the OCA theory emphasizes the role of migration, I would expect migration to react stronger to the shock than the macroeconomic variables.

On the one hand, in order to be an effective mitigation tool, migration should react within a short time after the shock as otherwise the imbalance would accumulate and threaten the stability of the currency union. On the other hand, in practice no person takes the decision to emigrate recklessly. After Greece entered the depression, virtually no person decided in the first years to emigrate but to stay in her home country. Only with time, the emigration rates increased as the expected income at home grew smaller relative to the expected income abroad. Given this circumstances, I would expect migration to react and to reach its peak within three years after the shock occurred.

Furthermore, the adjustment mechanisms may take some time to materialize but eventually the imbalance should disappear as it otherwise turns into a structural burden to the monetary policy. It seems reasonable to argue that an imbalance should disappear within ten years. Otherwise, migration would have not prevented the imbalance of becoming a structural, burdensome phenomenon.

There is no clear threshold for the magnitude of the imbalance to trigger the collapse of the currency union or to render its monetary policy powerless. For both to happen the imbalance has to be extraordinarily large which by definition has to be a rare occurrence. I impose therefore two thresholds. First, if an imbalance reaches a magnitude of two standard deviations then this is an event which occurs with just 5% probability. This could be strong enough to cause severe stress to the stability of the currency union. The same result could happen if a large imbalance persists over an extended period of time. The second threshold is, therefore, that an imbalance reaches the size of one standard deviation for at least three consecutive years.

The IRF, however, is based on the Cholesky decomposition of the covariance matrix which in turn requires an ordering of the variables. There is no unambiguous way to define the order. Based on economic theory and political relevance, I will calculates IRFs for two different scenarios each with an own ordering of the variables.

The first scenario is designed to simulate the impact of the Great Recession in 2008: this shock reduced GDP growth in all EU countries yet to a different degree. Also some countries recovered much faster than others. In the aftermath, large migration flows have been observed. Therefore, I order the variables in the following way: GDP growth, migration, unemployment, current account.

I argue that a shock in GDP per capita growth imbalances implies that income increases in the faster growing country relative to the slower growing country. The faster growing country will observe larger net migration since more immigrants will try to maximize their income. With more migrants the unemployment imbalance between both countries becomes smaller as the faster growing country's labor market has to absorb more employees which requires time whereas the slower growing country should observe an outflow of unemployed. Eventually, with some delay also trade, i.e. the current account, will be affected as the migrants will import more goods from their home countries.

The second scenario reflects the recent political circumstances: many migrants entered the EU from outside due to political reasons. These immigrants are to a large extent refugees seeking shelter. Hence, the EU countries will increase their fiscal spending which should raise the budget deficit. Furthermore, due to certain restrictions such as language barriers or labor market regulations refugees usually do not enter the labor market immediately but the earliest after one year. Afterwards they should become productive and their income increase, hence GDP growth accelerate and eventually current account will be affected, too. The second order is therefore: migration, fiscal borrowing, unemployment rate, GDP growth, current account.

Each of the two scenarios contains four different specifications of the PVAR model. The specifications vary due to the choice of the migration variable - direct net migration or total net migration - and due to the sample selection - Euro area member states only or all EU countries.

The first specification focuses on the direct net migration within the Euro area. The specification concentrates solely on the migrant flows and imbalances between pairs of countries who are exclusively members of the Euro area. It captures the fact that the currency area was enlarged over time and previous non-member countries introduced the Euro.

The second specification refers to dyadic migration flows between all EU members, that is, between members and non-members of the currency area in Europe. In comparison to the first specification this implies an extension of the sample with a few Northern European and several Eastern European countries which are members of the EU but not of the Euro area. The economies are still strongly interwoven and costs of migration are small since the Freedom of movement applies to the whole EU and not just to the Euro area.

In contrast, the third and fourth specifications do not use direct net migration but total net migration as the migration variable. Total net migration refers to the migration an Euro or EU country exchanges with the rest of the world, hence also with outside of Europe. Specifications three and four can be interpreted as global migration flows which the Euro area and the European Union, respectively, exchange.

1.6. Results

In this section I present the results for both PVAR scenarios, each constiting of four different specifications.

1.6.1. First scenario: Great Recession

This scenario imitates the impact of the Great Recession. Therefore, I order the variables as GDP growth, migration, unemployment, and current account.

The lag order tests based on the coefficient of determination suggests to use two lags as this lag selection explains most of the variance of the dependent variables, irrespective of whether direct net migration or total net migration are used.

Table 3: Lag order selection tests for Panel VAR, based on whole sample

	Coefficient of determination	
Lags	Direct net migration	Total net migration
1	0.9772	0.8640
2	0.9980	0.9468
3	0.9882	0.9312
4	0.9805	0.9338
Sample years	2000 - 2013	2000 - 2013
Number of observations	475	512
Number of panels	128	132
Average number of T	3.7	3.9

Therefore, my model in the first scenario has two lags. The next step is to ensure that this model fulfills the requirement of stationarity. To that end, the eigenvalues in the following unit root tests have to be smaller than one. If an eigenvalue lies outside the unit circle, then the PVAR is considered unstable. There are four unit root tests since the current scenario distinguishes the four specifications of direct net migration / total net migration as well as Euro states / whole sample. For the purpose of clarity the eigenvalues in each column are sorted in a descending order and they are limited to the three largest eigenvalues.

Table 4: Unit root tests for Panel VAR based on the modulus of eigenvalues, three largest eigenvalues in descending order

Specification	(i)	(ii)	(iii)	(iv)
Net migration variable	direct	direct	total	total
Sample selection	Euro states	whole sample	Euro states	whole sample
1st largest eigenvalue	0.7590	0.7830	0.5077	0.5559
2nd largest eigenvalue	0.5859	0.5189	0.4454	0.5559
3rd largest eigenvalue	0.5859	0.5189	0.4454	0.4151

Table 4 confirms all four model specifications are stationary since all eigenvalues lie within the unit circle.

The following table 5 tests all four model specifications for Granger-causality. The null hypothesis is that the excluded variable does not Granger-cause the equation variable. If the null hypothesis is rejected, then the excluded variable does Granger-cause the equation variable. The values displayed refer to the chi square probabilities.

Table 5: Granger-causality tests for the four specifications of the first scenario, P-values shown based on chi square

Specification		(i)	(ii)	(iii)	(iv)
Net migration variable		direct	direct	total	total
Sample selection		Euro states	whole sample	Euro states	whole sample
Equation variable	**Excluded variable**				
GDP growth	Migration	0.528	0.716	0.007***	0.002***
GDP growth	Unemployment rate	0.001***	0.005***	0.000***	0.000***
GDP growth	Current account	0.005***	0.000***	0.030**	0.000***
GDP growth	All	0.000***	0.000***	0.000***	0.000***
Migration	GDP growth	0.345	0.089*	0.000***	0.000***
Migration	Unemployment rate	0.873	0.193	0.000***	0.000***
Migration	Current account	0.521	0.343	0.008***	0.000***
Migration	All	0.528	0.163	0.000***	0.000***
Unemployment rate	GDP growth	0.000***	0.000***	0.000***	0.000***
Unemployment rate	Migration	0.140	0.844	0.000***	0.000***
Unemployment rate	Current account	0.000***	0.000***	0.000***	0.000***
Unemployment rate	All	0.000***	0.000***	0.000***	0.000***
Current account	GDP growth	0.125	0.000***	0.005***	0.000***
Current account	Migration	0.293	0.008***	0.000***	0.000***
Current account	Unemployment rate	0.000***	0.018**	0.000***	0.016**
Current account	All	0.000***	0.000***	0.000***	0.000***

* p<0.1, ** p<0.05, *** p<0.01. Null hypothesis: excluded variable does not Granger-cause the equation variable.

In most cases, the null hypothesis is rejected and it is confirmed that Granger-causality runs into both directions between migration and any imbalance. Furthermore, mostly the imbalances do Granger-cause each other.

The first specification with direct net migration among Euro countries shows that direct net migration does not Granger-cause the imbalances, and neither do they Granger-cause direct net migration.

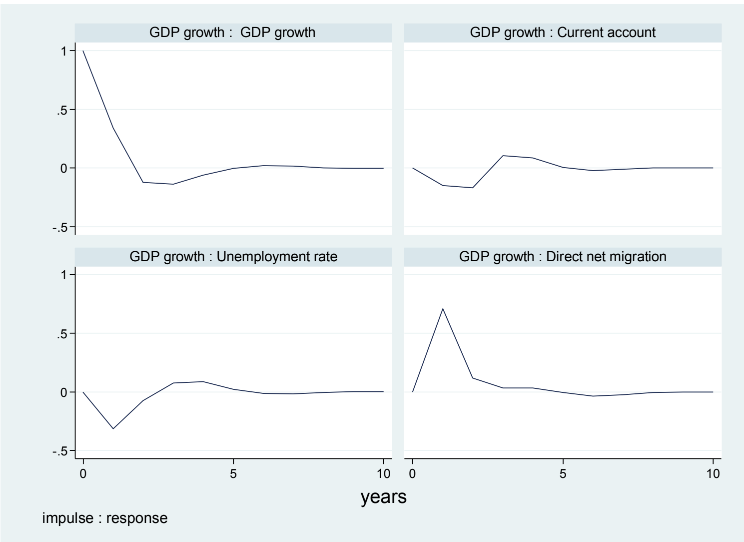
The second specification indicates that direct net migration does Granger-cause the current account imbalances but it is not significant for GDP growth and unemployment even for the whole sample. Direct net migration does not react significantly to the imbalances, except to GDP growth.

The third and fourth specifications apply total net migration to the Euro countries and the whole sample, respectively. Total net migration does Granger-cause all macroeconomic imbalances and, in turn, migration also reacts to them.

1.6.1.1. Direct net migration, Euro states only

The first specification uses the direct net migration variable and limits the sample to the Euro states. Shocking the GDP growth reveals that direct net migration reacts fast and intensely. After roughly five years migration returns to its stable level. The other imbalances do not react much. In other words, intra Euro area migration does mitigate a GDP growth shock, that is, the widening of growth rates between member countries, such as OCA theory predicts.

Chart 4: IRF with GDP shock and direct net migration for Euro countries

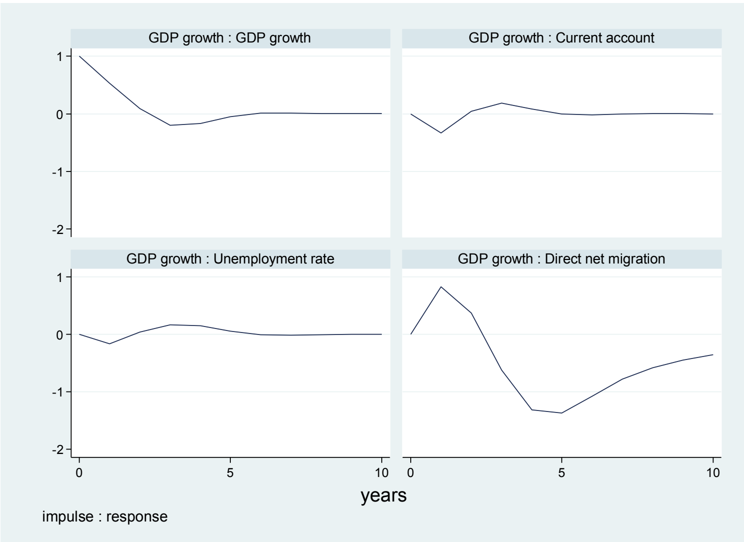


If a positive shock in terms of GDP growth occurs to an Euro country relative to the other member states then it will attract more migrants from there. Furthermore, its unemployment rate will fall as compared to the other member states. Its current account is likely to fall into deficit for a short while as with its economy outperforming in growth it is likely to import more goods and services.

1.6.1.2. Direct net migration, whole sample

Repeating the exercise with all EU countries (instead of Euro states only) and using direct net migration again yields similar results as the first specification. The direct net migration in chart 5 reacts to a GDP shock even more pronounced than in chart 4. Again, after an initial increase in direct net migration, i.e. the faster growing economy receives more net immigration, a stark slump can be observed to recover only after twenty years. As the other imbalances do not react virtually, I conclude that migration does again act as a buffer and allows to keep imbalances stable.

Chart 5: IRF with GDP shock and direct net migration for whole sample



In the second specification migration reacts much stronger than in the first specification. The second specification combines direct net migration and the whole sample, that is, all EU countries and intra-EU-migration which thus proves to be an effective buffer to a GDP growth imbalance shock. The difference to the first specification lies in the fact that the second specification contains especially the new

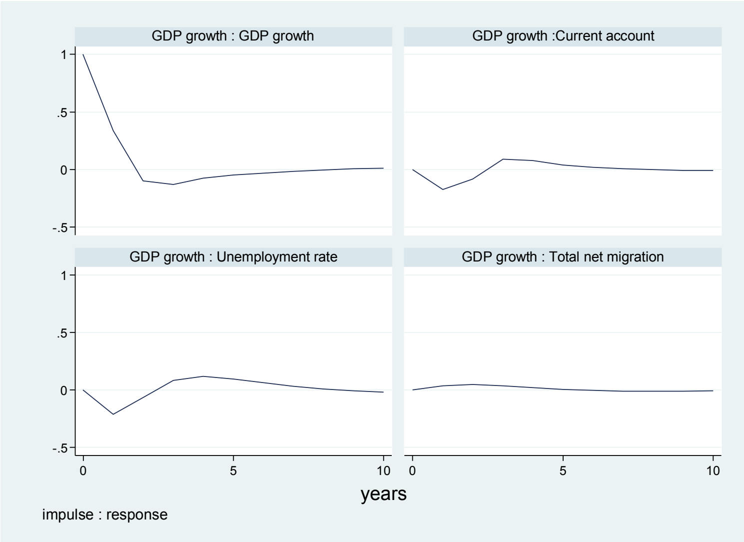
Eastern European countries which have not become members of the currency area. In other words, migrant from Eastern Europe are an effective instrument to mitigate imbalance shocks. This is in line with the findings of Zimmermann and Zaiceva (2012).

1.6.1.3. Total net migration, Euro states only

Now the third specification changes from direct net migration to total net migration. The difference between these two variables lies in the fact that the former considers only migration flows between two countries whereas the former refers to the difference in total migration rates between two nations.

It can be observed in chart 6 that total net migration reacts to a GDP growth shock to a much smaller degree than direct net migration. The other imbalances are now much more affected, though. The unemployment rate and the current account shrink imbalances turn negative relative to the other Euro area member states. However, both imbalances react more strongly than before as total net migration absorbs less of the GDP shock than direct net migration. In other words, while migration between Euro countries is an effective buffer for an imbalance shock, migration across the borders of the currency area is less effective in mitigating such shocks.

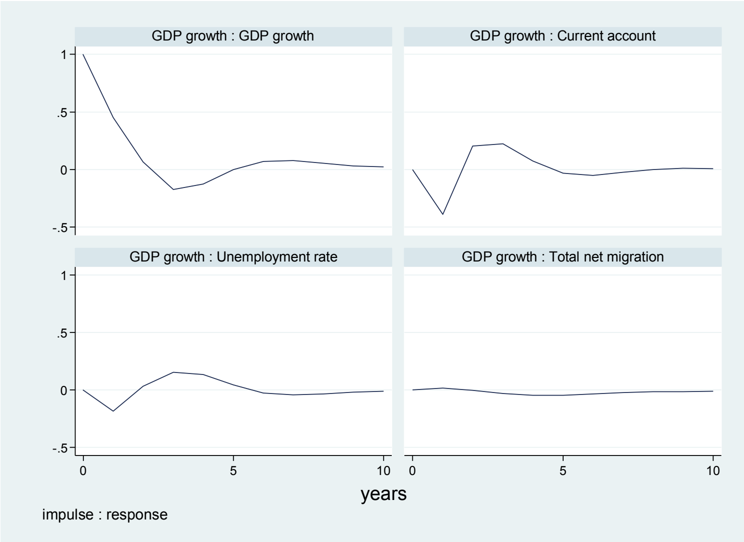
Chart 6: IRF with GDP shock and total net migration for Euro countries



1.6.1.4. Total net migration, whole sample

The first two specifications showed that direct net migration between the member states mitigates an imbalance shock even more effectively when the sample is enlarged from the Euro to all EU countries. Yet, the results are opposite the third and fourth specifications which use total net migration. The fourth specification takes into consideration all migration flows the EU exchanges with the rest of the world. As such, total net migration barely reacts to a shock in GDP growth. Consequently the unemployment rate and the current account imbalances grow even larger in order to absorb the GDP shock.

Chart 7: IRF with GDP shock and total net migration for whole sample



Therefore, the currency area can trust intra-Euro-area-migration to absorb a macroeconomic imbalance shock. The Euro area can also rely on migration with other EU countries which are not members of the currency area to absorb an imbalance shock.

However, migration flows exchanged with the rest of the world across the borders of the Euro area or the EU, respectively, do not mitigate a shock in GDP growth. Therefore, the current account and unemployment rate build up imbalances to absorb the GDP shock.

One explanation is that Euro area and EU countries are more similar to each other than to the rest of the world in terms of economic structure and labour force. Furthermore, the European countries trade intensely with each other and migration costs are low thanks to the institutional solutions such as the Freedom of Movement. It is therefore not surprising that migration within the Euro area or EU reacts much stronger to an imbalance shock than external migration with the rest of the world. However, it is surprising how well internal migration absorbs such shocks. As predicted by the OCA theory the direct net migration is an effective tool to mitigate economic shocks and to shift the adjustment burden away from other factors such as unemployment or current account.

Nevertheless, no imbalance in any specification has reached a significant magnitude of two standard deviations immediately nor of one standard deviation over three consecutive years. In this scenario the currency area remained stable in all four specifications, that is, irrespective of whether direct net migration or total net migration has been applied and irrespective of whether the focus was on the Euro area or the whole sample.

1.6.2. Second scenario: Migration crisis

The second scenario simulates a shock of migration. Both direct and total net migration are computed in such a way that the shock simulates immigration of one standard deviation. This I can investigate how the sudden movement of migrants within the Euro area and an influx of migrants from outside the EU might affect the stability of the currency area. I order the variables as migration, fiscal borrowing, unemployment rate, GDP growth and current account.

First, the number of lags has to be selected using the coefficient of determination.

Table 6: Lag order selection tests for Panel VAR, based on whole sample

	Coefficient of determination	
Lags	Direct net migration	Total net migration
1	0.9843	0.9979
2	0.9928	0.9995
3	0.9955	0.9997
4	0.9947	0.9994
Sample years	2000 - 2013	2000 - 2013
Number of observations	442	1566
Number of panels	111	297
Average number of T	4.0	5.3

The coefficient of determination in table 6 suggests to use three lags for the direct and the total net migration specifications since they explain most of the variance of the dependent variables. Yet, the lag order does not guarantee the stability of a PVAR system. In fact, when I did use three lags, two of the four specifications were unstable since some of their eigenvalues were larger than one. No reliable conclusions can be drawn from an unstable PVAR.

Yet, table 6 indicates that the variation explained by two lags is very similar to the variance explained by three lags. The choice of two lags also increases the amount of observations which can be used for the regressions and it allows to compare the second with the first scenario.

Using two lags for the unit root tests suggest that all four specifications are stable given that all their eigenvalues lie inside the unit circle (see table 7).

Table 7: Unit root tests based on the modulus of eigenvalues, three largest eigenvalues in descending order

Specification	(i)	(ii)	(iii)	(iv)
Net migration variable	direct	direct	total	total
Sample selection	Euro states	whole sample	Euro states	whole sample
1st largest eigenvalue	0.7988	0.5741	0.7811	0.8921
2nd largest eigenvalue	0.6228	0.5741	0.7811	0.5556
3rd largest eigenvalue	0.6228	0.5673	0.5636	0.5556

As in the case of the first scenario, Granger-causality between migration and imbalances can be detected less often when direct net migration is used instead of total net migration. Using Euro states only or the whole sample does not yield any significant differences in terms of causality.

The first two specifications in table 8 relate to direct net migration and to Euro states or the whole sample, respectively. No imbalances do Granger-cause direct net migration. On the other hand, direct net migration has statistically significant causal power only on the current account imbalance. Most imbalances affect each other in terms of Granger-causality. The third and fourth specification, though, suggest that all relations between total net migration and the imbalances as well as among the imbalances themselves are statistically significant for Granger-causality.

Table 8: Granger-causality tests for the four specifications of the second scenario, P-values shown based on chi square

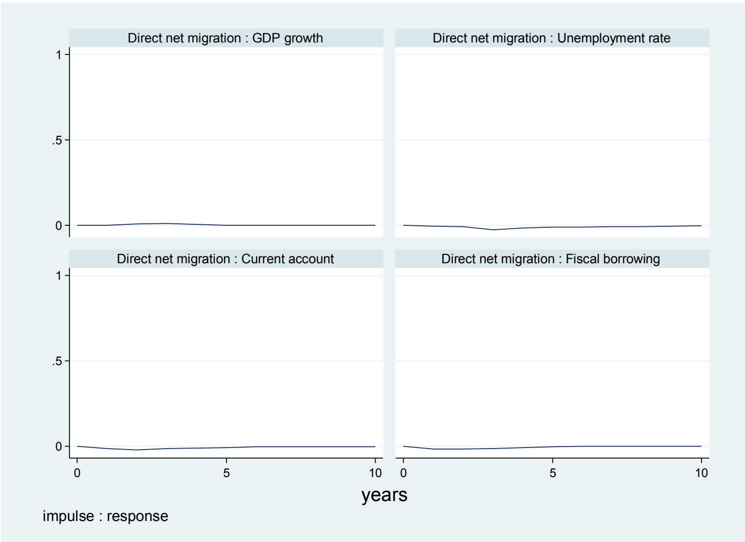
Specification		(i)	(ii)	(iii)	(iv)
Net migration variable		direct	direct	total	total
Sample selection		Euro states	whole sample	Euro states	whole sample
Equation variable	**Excluded variable**				
Migration	Fiscal borrowing	0.150	0.657	0.000***	0.000***
Migration	Unemployment rate	0.364	0.778	0.000***	0.000***
Migration	Current account	0.295	0.409	0.000***	0.000***
Migration	GDP growth	0.128	0.680	0.000***	0.000***
Migration	All	0.557	0.526	0.000***	0.000***
Fiscal borrowing	Migration	0.499	0.703	0.000***	0.000***
Fiscal borrowing	Unemployment rate	0.140	0.000***	0.001***	0.000***
Fiscal borrowing	Current account	0.000***	0.000***	0.000***	0.000***
Fiscal borrowing	GDP growth	0.005***	0.001***	0.000***	0.000***
Fiscal borrowing	All	0.005***	0.000***	0.000***	0.000***
Unemployment rate	Migration	0.139	0.178	0.000***	0.000***
Unemployment rate	Fiscal borrowing	0.000***	0.000***	0.000***	0.000***
Unemployment rate	Current account	0.002***	0.000***	0.058*	0.000***
Unemployment rate	GDP growth	0.000***	0.000***	0.000***	0.000***
Unemployment rate	All	0.000***	0.000***	0.000***	0.000***
Current account	Migration	0.555	0.033**	0.005***	0.013**
Current account	Fiscal borrowing	0.000***	0.000***	0.000***	0.000***
Current account	Unemployment rate	0.000***	0.000***	0.000***	0.000***
Current account	GDP growth	0.274	0.000***	0.034**	0.000***
Current account	All	0.000***	0.000***	0.000***	0.000***
GDP growth	Migration	0.219	0.547	0.036***	0.001***
GDP growth	Fiscal borrowing	0.000***	0.000***	0.000***	0.000***
GDP growth	Unemployment rate	0.037**	0.060*	0.000***	0.000***
GDP growth	Current account	0.009***	0.000***	0.093*	0.000***
GDP growth	All	0.000***	0.000***	0.000***	0.000***

* p<0.1, ** p<0.05, *** p<0.01. Null hypothesis: excluded variable does not Granger-cause the equation variable.

1.6.2.1. Direct net migration, Euro states only

As expected by the OCA theory, a shock in direct net migration does not cause turbulences in the imbalances between the Euro area member states as no imbalance reacts in a meaningful way, that is, no imbalance reacts with even 10% of a standard deviation.

Chart 8: IRF with direct net migration shock for Euro countries



1.6.2.2. Direct net migration, whole sample

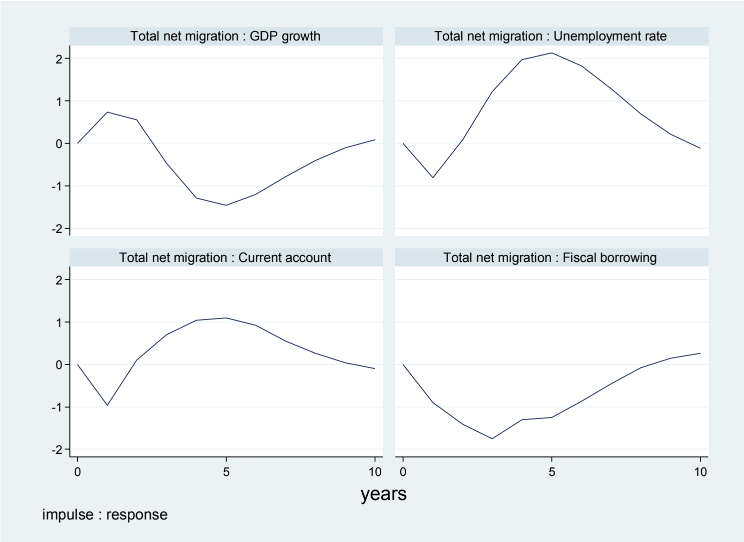
The picture remains unchanged when the sample is extended from Euro countries to the whole European Union. Direct net migration still does not cause imbalances to build up as shown in chart 9.

Chart 9: IRF with direct net migration shock for whole sample

1.6.2.3. Total net migration, Euro states only

This third specification simulates a shock of one standard deviation to total net migration. The model is computed as dyadic data, meaning that this shock equals an increase of net immigration (measured in total net migration rates) in an Euro country relative to the other Euro area member states. In contrast to the OCA theory, however, a shock in total net migration causes manifold instabilities to the Euro area.

Chart 10: IRF with total net migration shock for Euro countries



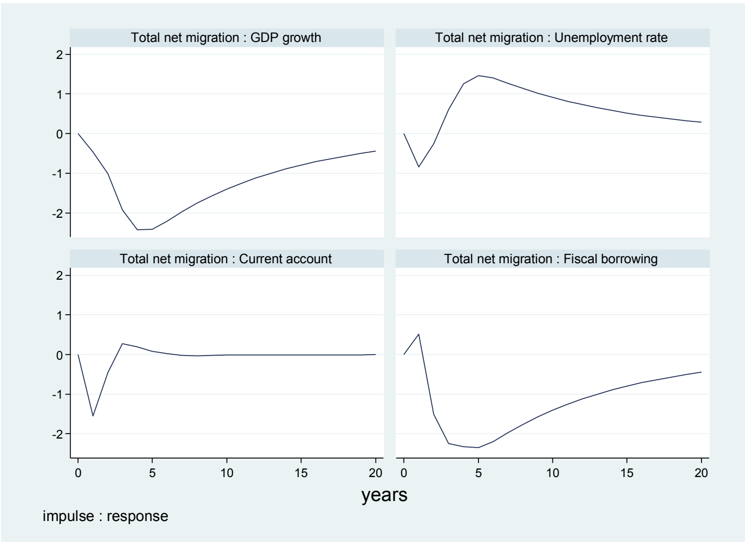
In the first two years, GDP growth rises as the influx of migrants increases aggregate demand. To serve the demand, more labour is hired which reduces the unemployment rate. However, as both the current account and fiscal borrowing imbalances turn negative, it seems that the demand is being supplied with imports and fiscal spending. This is not sustainable as the GDP growth changes its sign quickly, turning from positive to negative for an extended period of time. The large immigrant inflow cannot be absorbed by the labour market either, leading to a rise in the unemployment rates. Instead the immigrants are sustained by the public hand which has to continue borrowing capital. The current account imbalance turns positive as a weaker economic growth goes hand in hand with a reduction of imports.

The immigration shock led to an instability of the currency area as defined by the previous rules. GDP growth shrunk by more than one standard deviation for more than three consecutive years. The unemployment rate even crossed the threshold of two standard deviations implying a strong deviation from the other Euro area member states and putting significant stress on the effectiveness of the monetary policy. That deviation is not just a short term phenomenon as the unemployment rate prevails for several consecutive years by more than one standard deviation each year. Fiscal borrowing also exceeds one standard deviation for more than three years. Only the current account appears to be more robust as it reacts less strongly and returns quickly to its long term state. The stability rules as defined before are thus broken by three imbalances caused by the total net migration shock within the Euro area.

1.6.2.4. Total net migration, whole sample

The patterns of the third specification remain broadly the same when the sample is extended from the Euro to all EU countries. This fourth specification also simulates a total net migration shock, that is, immigration within Europe as well as across its borders with the rest of the world. Again, the three imbalances GDP growth, unemployment rates and fiscal borrowing break the stability rules.

Chart 11: IRF with total net migration shock for whole sample



GDP growth does not turn positive in the first years after the shock but the imbalance turns constantly negative and even reaches a magnitude of two standard deviations, implying a significant burden on the monetary policy. The fiscal borrowing imbalance reaches a similar level. The unemployment rate imbalance increases by more than one standard deviation for several years. It is remarkable that the imbalances in the EU need about 20 years to return to their long term steady state. This is twice as long as in the Euro area. Please note that for this reason even the abscissa had to be doubled in chart 11 as compared to the preceding charts. One explanation could be that immigrants to EU-countries outside the currency area fit the needs of the labour markets less well than immigrants from the rest of the world to the Euro states. An argument would be that the Eastern European countries are less developed than the western Euro countries hence the requirements towards external immigrants are less demanding. Unskilled labour from the rest of the world has a better chance to enter the EU via its non-Euro countries. Yet, as they do not match the requirements especially of the labour markets in the Euro countries a large scale immigration causes the buildup of significant imbalances within the Euro area and puts the stability of the currency union in danger.

The first and the second specifications showed that internal Euro area and EU migration do not cause turbulences to the macroeconomic imbalances GDP growth, current account, unemployment or fiscal borrowing. In line with the OCA theory, internal migration does not put the currency area under stress. Yet, the third and fourth specifications simulate an immigration shock from the rest of the world to Europe. This immigration is not taken into account by the OCA theory. Yet, my theoretical arguments specified before that external migration might not match the needs of the European labour markets and economies have been confirmed empirically as the migrant influx increases the instability of the currency area.

1.7. Robustness checks

This section introduces a series of robustness checks to test for the validity of the results.

1.7.1. Extension of the first scenario

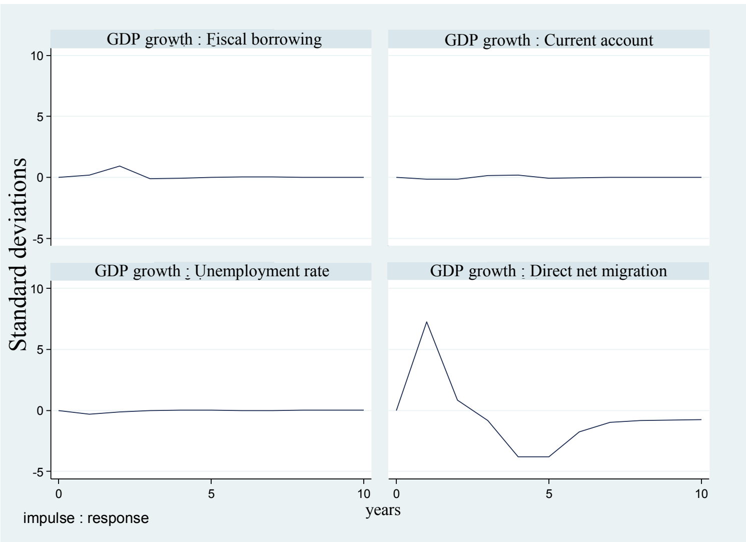
The first scenario simulates the Great Recession and it includes all variables except fiscal borrowing. One could argue that at the beginning of the Great Recession the economic growth nosedived. Governments reacted swiftly and increased fiscal spending to restore GDP growth. However, those measurements were usually financed by an increase in government borrowing. As a robustness check, therefore, fiscal borrowing should be put on the second place in the variable ordering of the first scenario. The modified order reads GDP growth, fiscal borrowing, migration, unemployment and current account. All four specifications are stable with two lags again.

The results from all four previous specifications are very similar. The first two specifications use direct net migration. As before, only migration mitigates the GDP growth shock, fiscal borrowing while other imbalances barely react. This result holds for both the Euro countries subsample and the whole sample. In the third and fourth specification, total net migration reacts much less strongly than direct net migration. All imbalances including fiscal borrowing react more strongly as compared to the first and second specifications which use direct net migration.

The following two charts replicate the first and the third specifications of the first scenario. For reasons of comparability, the plots in the top left corner in both charts picture the fiscal borrowing imbalance and thus replace the GDP growth plots. The replications of the second and of the fourth specifications are not shown here as they do not yield additional insights.

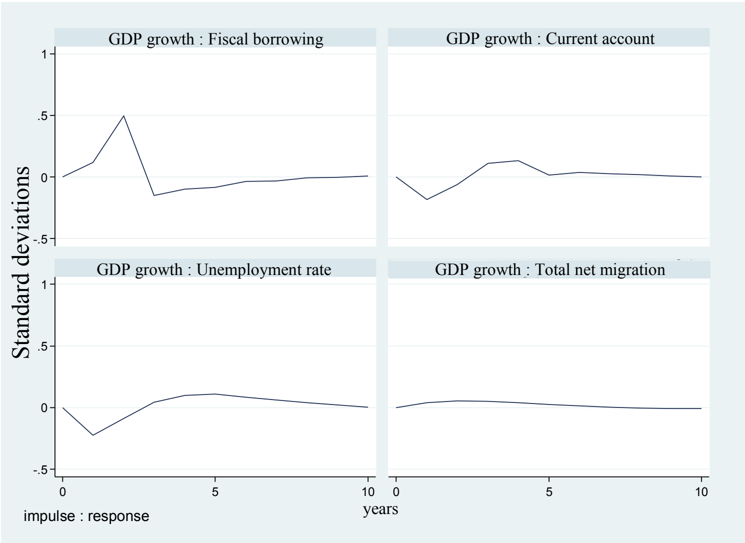
Chart 12 replicates the first specification of the first scenario, that is, using direct net migration within the Euro countries. Chart 12 is comparable to chart 4.

Chart 12: IRF with GDP shock and direct net migration for Euro countries with fiscal borrowing as additional variable



The next chart 13 uses total net migration within the Euro area and is comparable to chart 6.

Chart 13: IRF with GDP shock and total net migration for Euro countries with fiscal borrowing as additional variable

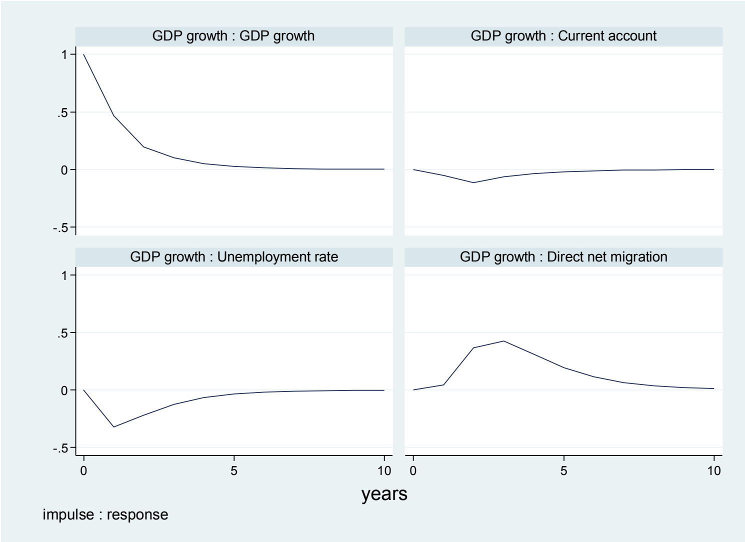


1.7.2. Time split

The data base reaches from year 1998 until 2014. It therefore includes the Great Recession which began in 2008. It is interesting to investigate if the role of migration in mitigating economic shocks has changed due to the Great Recession. I split the sample into two subsamples, one lasting from 1998 until 2007 and the other from 2008 until 2014, and I repeat the first scenario. In particular, I replicate the first and the third specification, that is, direct net migration and total net migration in the Euro countries. As in the benchmark model, the time split models use two lags and they are stable.

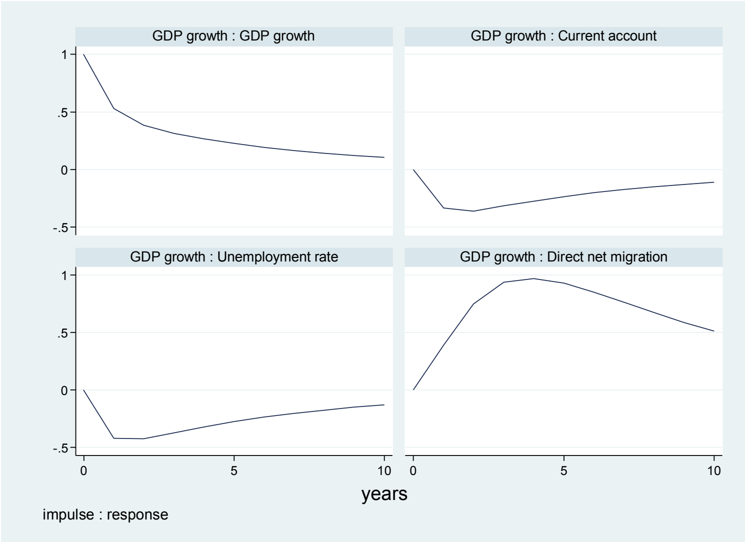
Charts 14 and 15 replicate the first specification with direct net migration in the Euro area. The former refers to the time span from 1998 until 2007, the latter to the time from 2008 until 2014. They can be compared to chart 4.

Chart 14: IRF with GDP shock and direct net migration in Euro countries 1998 - 2007



As in the benchmark model which spans over the whole time period, migration reacts strongest to a GDP growth shock. Yet, the magnitude of the direct net migration is smaller before the Great Recession. Both the current account and the unemployment imbalances turn negative once again.

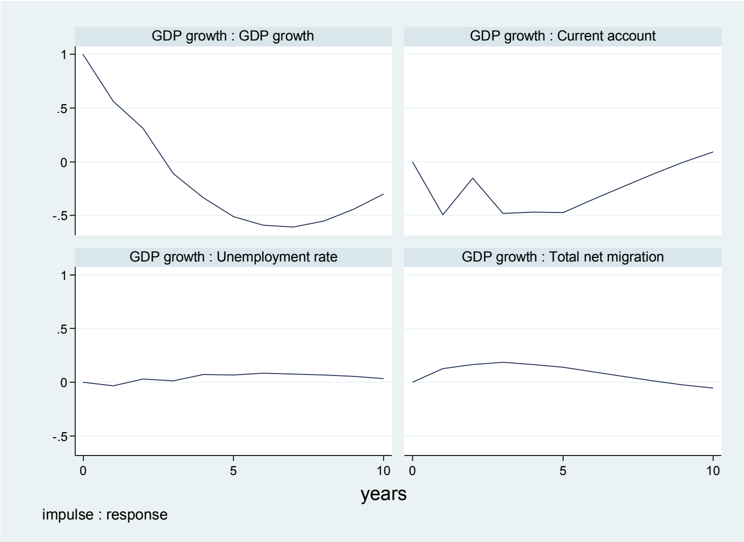
Chart 15: IRF with GDP shock and direct net migration in Euro countries 2008-2014



The outburst of the Great Recession in 2008 amplified the reaction of migration and the imbalances to a GDP growth shock as shown in chart 15. In particular, the direct net migration has doubled its reaction from 0.5 to nearly a full standard deviation. Yet, the unemployment rate and the current account imbalances need more time to return to their steady state than in the benchmark model in chart 4.

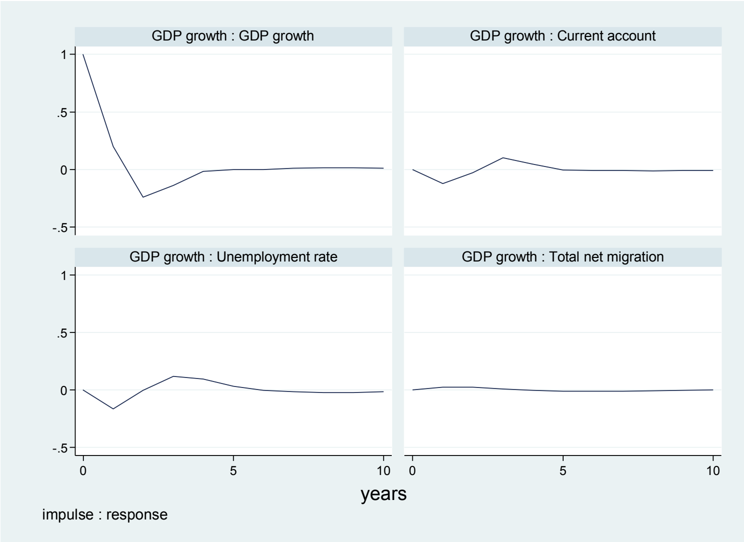
The two following charts 16 and 167refer to total net migration in the Euro area, i.e. they consider migration flows to and from the rest of the world. They can be compared to chart 6. Already in this benchmark model, total net migration reacted much less strongly to a GDP growth shock than direct net migration. This pattern repeats itself in chart 16 which refers to the time span between 1998 and 2007.

Chart 16: IRF with GDP shock and total net migration in Euro countries 1998 - 2007



The Great Recession did not alter this finding either since the reaction of total net migration in the Euro area remains weak even after 2008 (compare chart 17).

Chart 17: IRF with GDP shock and total net migration in Euro countries 2008 - 2014



1.7.3. Dynamic Panel data model

According to the OCA theory an imbalance should be mitigated by migration, i.e. in the context of dyadic data the imbalance should react in the opposite direction of migration. A positive imbalance implies a negative coefficient for migration whereas a negative imbalance suggests a positive migration coefficient.

Hence, in a regression framework with the imbalance (e.g. GDP growth) being the dependent variable and migration the independent variable, their signs should oppose each other. For instance, if the GDP growth difference is positive for a country-pair, then the coefficient of migration should be negative. Likewise, a negative GDP growth difference implies a positive migration coefficient. This robustness check investigates if causality runs from migration to the imbalances, i.e. if an increase in net migration reduces the imbalance and vice versa. Yet, this robustness check does not test if migration reacts to an imbalance.

I conduct an Arenallo-Bond style dynamic panel data model, also known as System GMM. The dependent variable is real GDP per capita growth. I split the dependent variable into two baskets: one for positive GDP growth rates and the other for negative

GDP growth rates. I regress the dependent variable on its own first lag, migration and the other imbalances. I run the System GMM for each basket separately and I also distinguish between direct net migration and total net migration. I use all country-pairs of the data set.

In this setting I assume that causality runs only from migration to the imbalance. Due to this abstraction the magnitude of the coefficients cannot be interpreted reliably. Yet, the null hypothesis should be still valid, that is: in case of a positive imbalance the migration coefficient should be negative whereas the coefficient should be positive for negative imbalances.

Table 9 shows the result. The first two columns refer to the direct net migration. Its sign is positive for positive GDP growth imbalances and positive but not significant for negative GDP growth imbalances. These results match only partially the expectations.

Columns three and four use the total net migration. In contrast to the expectations of the OCA theory, its sign is coherent with the signs of the GDP growth imbalances. An increase in the total net migration raises positive GDP growth imbalances and lowers negative GDP growth imbalances. These findings confirm that the total net migration is not neutral towards the imbalances but even causes them to build-up.

Table 9: Robustness checks with System GMM

Dependent variable: GDP growth (in %)	(i)	(ii)	(iii)	(iv)
GDP growth imbalance	Positive	Negative	Positive	Negative
Direct net migration	0.001***	0.000		
	(0.000)	(0.000)		
Total net migration			0.402**	-0.057
			(0.177)	(0.259)
1st lag GDP growth	-0.051	0.156***	0.087***	0.052*
	(0.337)	(0.045)	(0.033)	(0.028)
Current account	-0.147***	-0.259***	-0.187***	-0.126***
	(0.039)	(0.000)	(0.026)	(0.025)
Unemployment rate	-1.025***	-0.684***	-0.753***	-0.834***
	(0.080)	(0.000)	(0.060)	(0.060)
Fiscal borrowing	-0.137	0.073***	0.015	0.030
	(0.024)	(0.073)	(0.025)	(0.019)
Constant	2.015***	-2.19***	-2.255***	1.857***
	(0.126)	(0.000)	(0.100)	(0.000)
Country-pair fixed effects	Yes	Yes	Yes	Yes
Wald chi^2	320.14	215.63	554.94	587.97
	(0.000)	(0.000)	(0.000)	(0.000)
N	811	800	1577	1560
N of countries	206	207	323	314
N of instruments	139	140	140	140

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors in parentheses. System GMM estimation with country-pair fixed effects. First two columns use direct net migration, the last two columns total net migration. First and third column use only positive values of annual real per capita GDP growth imbalance. Second and fourth column use only negative values of annual real per capita GDP growth imbalance. Tests confirm the absence of autocorrelation, exogeneity and validity of instruments. Migration variables are measured in percent of population, economic variables in percent (of GDP).

1.8. Conclusion

The Optimal Currency Area theory is the theoretical foundation of the Euro area. The theory suggests that macroeconomic imbalances might built up between the member states. The imbalances could impair the effectiveness of monetary policy and even the stability of the currency union. Yet, labour migration is supposed to be an effective instrument to mitigate the imbalances.

This paper argues that the theory has three important shortcomings: the assumptions of perfect labour mobility and of homogenous labour as well as the exclusive focus on member states. These shortcomings could theoretically affect the ability of migration to mitigate the imbalances. Furthermore, the analysis of the three shortcomings implies that two types of migration should be distinguished from each other. On the one hand, migration within the currency area. On the other hand, migration across the outer borders of the currency area.

This paper combines dyadic data with Panel VAR in the context of the Euro area and the European Union. It defines two scenarios - the Great Recession and an immigration crisis – in order to investigate the theoretical hypothesis empirically.

The hypotheses postulate that the internal migration effectively mitigates imbalances and that they do not cause imbalances. In contrast, the external migration does not mitigate imbalances but it is rather a cause of imbalances itself. These hypotheses are confirmed empirically.

If the GDP growth of an Euro country suddenly increases or drops much stronger than in the other member states, then this imbalance can be a burden for the monetary policy. However, this imbalance disappears within a few years as it is absorbed by the internal migration within the Euro area and within the European Union. The strong reaction of the internal migration maintains the stability of the currency union as other economic variables such as current account, fiscal borrowing and unemployment rates do not turn into balances. In addition, sudden large migration flows within the Euro area and the European Union do not generate imbalances between the member states with respect to any economic variable.

A GDP growth shock is not mitigated by migrants who cross the outer border of the Euro area and the European Union. This external migration does not react in a meaningful extent such that the economic growth shock prevails for a longer time period as compared to the case of internal migration. A shock in economic growth has to be absorbed by current account, fiscal borrowing and unemployment. They build up large and persistent imbalances within the Euro area which severely affect the monetary policy and which could even threaten the stability of the currency union.

Furthermore, a large scale immigration from the rest to the world to Europe generates large imbalances to GDP growth, current account, fiscal borrowing and unemployment rates in the Euro area. In this sense, external migration can be regarded as a source of instability in the Euro area.

There are two potential reasons explaining the important differences between the effects generated by the internal migration and by the external migration. Firstly, the Euro and EU countries are similar with respect to their economic structures, business cycles and labour forces. Therefore, a migrant willing to migrate from one EU country to another has a high probability to find employment as his skills are likely to fulfill the requirements of the destination labour market. Secondly, the Freedom of movement removes institutional barriers to a large degree, thus lowering the costs of migration.

In comparison, migrants from outside the European Union might not possess the required skills in order to fit the European labour markets. In addition, their relatively high costs of migration suggests that their decision to move to the EU does not depend on the development of the internal imbalances.

EU policy makers and peoples have expressed their will to retain the Euro area even in spite of large economic and fiscal costs. If their target is to maximize the probability of survival of the Euro area, EU policy makers face different options with respect to internal and external migration.

One such option is to lower the costs of migration for internal migrants. Enhancing language skills, cultural exchange programs, the recognition of foreign education and professional certificates are measurements which could reduce the costs of migration in order to increase the effectiveness of internal migration in mitigating economic imbalances.

EU policy makers, however, face a potential dilemma regarding external migration. Immigrants from the rest of the world to Europe could become more effective in mitigating imbalances if their costs of migration were lower and if they were able to blend in the EU labour markets more efficiently. EU policy makers could therefore seek to harmonize and lower the requirements of immigration in all EU countries. Yet, the empirical analysis of this paper suggests that a large scale of immigration from outside the EU might trigger economic imbalances. If the reason for this causality are political institutions, then EU policy makers could remove those obstacles and the immigration would be less likely to cause imbalances. If on the other hand the reason is that these immigrants lack the skills needed to succeed in the EU labour markets, then the options for EU policy makers are limited. The training of skills for immigration would be cost-intensive and would require time. In the meanwhile, the immigrants could cause the build-up of economic imbalances.

1.9. References

Abel, G. J. (2013): Estimating global migration flow tables using place of birth data. Demographic Research: Vol. 28. No. 18. pp. 505-546.

Abel, G. J. / Sander, N. (2014): Quantifying Global International Migration Flows. Science. Vol. 343. No. 6178. pp. 1520-522.

Arellano, M. / Bond, S. (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. Review of Economic Studies. Vol. 58. No.2. pp. 277–297.

Arellano, M. / Bover, O. (1995): Another look at the instrumental variable estimation of error-components model. Journal of Econometrics. Vol. 68. No. 1. pp. 29-51.

Beine, M., / Bourgeon, P., / Bricongne, J. C. (2019). Aggregate Fluctuations and International Migration. The Scandinavian Journal of Economics. Vol. 121. No. 1. pp. 117-152.

Beine, M. / Bertoli, S. / Huertas Moraga, J.F. (2016): A Practitioners' Guide to Gravity Models of International Migration. The World Economy. Vol. 39. No. 4. pp. 496-512.

Berger, H. / Nitsch, V. (2012): The Euro's Effect on Trade Imbalances. IMF Working Paper WP/10/226.

Beyer, R.C.M. / Smets, F. (2014): Has mobility decreased? Reassessing regional labor market adjustments in Europe and the US. Presented at the 60th Panel meeting of Economic Policy in October 2014.

Blanchard, O. J. / Katz, L.F. (1992): Regional Evolutions. Brookings Papers on Economic Activity. Vol. 1 (1992). pp. 1-75.

Blundell, R. / Bond, S. (1998): Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics. Vol. 87. No. 1. pp. 115-143

Bodvarsson, O. B. / Hou, J.W. / Shen, K. (2014): Aging and Migration in a Transition Economy: The Case of China. IZA Discussion Paper No. 8351.

Borjas, G. J. (1987): Self-selection and earnings of immigrants. American Economic Review. Vol. 77. No. 4. pp. 531-553.

Boubtane, E. / Coulibaly, D. / Rault, C. (2012): Immigration, Growth and Unemployment: Panel VAR Evidence from OECD Countries. IZA Discussion Paper No. 6966.

Canova, F. /Ciccarelli, M.(2013): Panel Vector Autoregressive Models: A Survey. ECB Working Paper Series No. 1507.

Cavelaars, P. / Hessel, J. (2007): Regional labor mobility in the European Union: adjustment mechanism or disturbance? DNB working paper No.137/2007.

Colen, L. / Persyn, D. / Guariso, A. (2014): What type of FDI is attracted by bilateral investment treaties? LICOS Discussion Papers 34614. LICOS - Centre for Institutions and Economic Performance, KU Leuven.

Dao, M. / Furceri, D. / Loungani, P. (2014): Regional Labor Market Adjustments in the United States and Europe. IMF Working Paper WP/14/26.

Decressin, J. / Fatás, A. (1994): Regional Labor Market Dynamics in Europe. CEPR Discussion Papers No. 1085.

Dhéret, C. / Lazarowicz, A. / Nicoli, F. / Pascouau, Y. / Zuleeg, F. (2013): Making progress towards the completion of the Single European Labor Market. EPC Issue Paper No. 75.

Doz, C. / Giannone, D. / Reichlin, L. (2012): A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models. Review of Economics and Statistics. Vol. 94. No. 4. pp. 1014-1024.

Dumitrescu, E. / Hurlin, C. (2012): Testing for Granger Non-causality in Heterogeneous Panels. HAL Working Papers halshs-00224434.

Eichengreen, B. (2014) : The Eurozone Crisis: The Theory of Optimum Currency Areas Bites Back. Notenstein Academy White Paper.

Esteves, R. P. / Khoudour-Castéras, D. (2009): A Fantastic Rain of Gold: European Migrants' Remittances and Balance of Payments Adjustment during the Gold Standard Period. The Journal of Economic History. Vol. 69. No. 4. pp. 951-985.

European Central Bank (2012): Competitiveness and external imbalances within the Euro Area. Occasional Paper series No. 139.

European Commission (2018): Study on the movement of skilled labour. Final Report. EUROPEAN COMMISSION. Directorate-General for Employment, Social Affairs and Inclusion.

Feldstein, M. (2008): Optimal Currency Areas. Speech at the ECB's Fifth Central Banking Conference on November 14, 2008 in Frankfurt.

Frey, W. (2009): The great American migration slowdown. Brookings Institution, Washington, DC.

Garcia, A. J. / Pindolia, D. K. / Lopiano, K. K. / Tatem, A. J. (2015): Modelling internal migration flows in sub-Saharan Africa using census microdata. Migration Studies. Vol. 3. No. 1. pp. 89-110.

Gnimassoun, B. / Mignon, V.(2013):How macroeconomic imbalances interact? Evidence from a panel VAR analysis. CEPII Working Paper No. 2013-42.

Gori, F. (2014): Monetary Policy and Real Cost Imbalances in Currency Unions. Graduate Institute of International and Development Studies Working Paper No. 12/2014.

Hayakawa, K. (2011): An Improved GMM Estimation of Panel VAR Models with Applications to Granger Causality Test and Impulse Response Analysis. European Economic Association & Econometric Society 2011 Parallel Meetings.

Head, K. / Mayer, T. (2014): Gravity Equations: Workhorse, Toolkit, and Cookbook. Handbook of International Economics. Elsevier.

Holtz-Eakin, D. / Newey, W. / Rosen, H.S. (1988): Estimating Vector Autoregressions with Panel Data. Econometrica. Vol. 56. No. 6. pp. 1371-1395.

Hurlin, C. / Dumitrescu, E. (2012): Testing for Granger Non-causality in Heterogeneous Panels. Economic Modelling. Vol. 29, No. 4. pp. 1450-1460.

Jauer, J. / Liebig, T. / Martin, J.P. / Puhani, P. (2014): Migration as an adjustment mechanism in the crisis? A comparison of Europe and the United States. OECD Social, Employment and Migration Working Papers No 155.

Kenen, P. (1969): The theory of optimum currency areas: an eclectic view. In R.A. Mundell and A.J. Swoboda (Eds.): Monetary Problems of the International Economy. Chicago University Press. pp. 41-60.

Kennan, J. / Walker, J.R. (2011): The Effect of Expected Income on Individual Migration Decisions. Econometrica. Vol. 79. No. 1. pp. 211-251.

Lange, F. / Gollin, D. (2013): Equipping immigrants: migration flows and capital movements in small open economies. Review of World Economics. Vol. 149. No. 4. pp: 749–777.

Love, I. / Zicchino, L. (2006): Financial development and dynamic investment behavior: Evidence from panel VAR. The Quarterly Review of Economics and Finance. Vol. 46. No. 2. pp. 190-210.

Mayda, A.M. (2010) : International migration: a panel data analysis of the determinants of bilateral flows. Journal of Population Economics. Vol. 23. No. 4. pp. 1249-1274.

Meardi, G. (2012): Social failures of EU Enlargement. In Meardi, G.: A Case of Workers Voting with their Feet. New York: Routledge. pp. 87-105.

Mitze, T. (2012): A Panel VAR Approach for Internal Migration Modelling and Regional Labor Market Dynamics in Germany. In Mitze, T.: Empirical Modelling in Regional Science. Towards a Global Time–Space–Structural Analysis. Springer-Verlag Berlin Heidelberg. pp. 19-52.

Molloy, R. / Smith, C. L. / Wozniak, A. K. (2011): Internal migration in the United States. Technical report. National Bureau of Economic Research.

Mundell, R.A. (1961): A Theory of Optimum Currency Areas. American Economic Review. Vol. 51. No. 4. pp. 657-665.

OECD (2006): International Migration Outlook. Sopemi 200 Edition. Vol. 2006. No. 5.pp. 139 – 161.

O'Rourke, K. H. / Taylor, A.M. (2013):Cross of Euros. Journal of Economic Perspectives. Vol. 27. No. 3. pp. 167-192.

Pasimeni, P. (2014): An Optimum Currency Crisis. The European Journal of Comparative Economics. Vol. 11. No. 2. pp. 173-204.

Pedroni, P. (2013): Structural Panel VARs. Econometrics. Vol. 1. No. 2. pp. 180-206.

Pesaran, M.H. (2007): A simple panel unit root test in the presence of cross-section dependence. Journal of Applied Econometrics. Vol. 22. No. 2. pp. 265-312.

Reilly, J.W. (1931): The Law of Retail Gravitation. New York Knickerbocker Press.

Roodman, D. (2009): How to do xtabond2: An introduction to difference and system GMM in Stata. The Stata Journal. Vol. 9. No. 1. pp. 86-136.

Schmitt-Grohé, S. / Uribe, M. (2013): Downward Nominal Wage Rigidity and the Case for Temporary Inflation in the Eurozone. Journal of Economic Perspectives. Vol. 27. No. 3. pp. 193-212.

Westerlund, J. (2007): Testing for error correction in panel data. Oxford Bulletin of Economics and Statistics. Vol. 69. No. 6. pp. 709-748.

Zimmermann, K. F. / Krause, A. / Rinne, U. (2014): How Far Away Is a Single European Labor Market? IZA Discussion Paper No. 8456.

Zimmermann, K. F. / Pytliková, M. / Kahanec, M. (2014): The Free Movement of Workers in an Enlarged European Union: Institutional Underpinnings of Economic Adjustment. IZA Discussion Paper No. 8456.

Zimmermann, K. F. / Zaiceva, A. (2012): Returning Home at Times of Trouble? Return Migration of EU Enlargement Migrants during the Crisis.IZA Discussion Paper No. 7111.

2. How demographics affect inflation - an NTA approach

In many countries the population share of young people is declining while the share of elderly is increasing. Yet, neither theory nor empirical evidence provide clear guidance whether the change of population shares will increase or reduce inflation. Empirical literature has so far relied solely on population data without integrating age-specific consumption and production patterns. I fill the gap using the National Transfer Accounts Project data. I run Fixed Effects regressions with dependency ratios and population distribution polynomials. My results suggest that both young and old increase inflation since their consumption exceeds production. These findings are robust to various robustness checks, instrumental variable estimations and country-specific regressions.

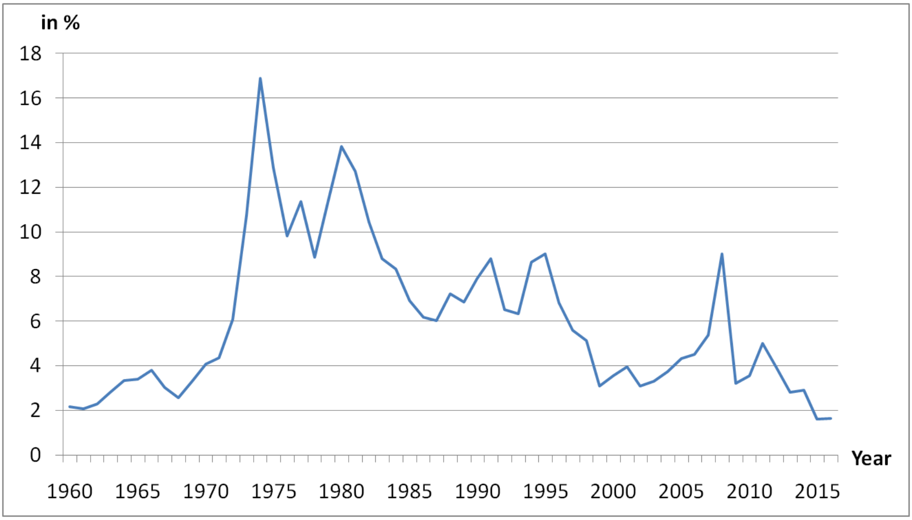
JEL Classification: D15, E31, J11, J21

2.1. Introduction

The term 'demographics' refers to any change in the size and structure of a population. Nowadays most advanced economies as well as several developing economies are characterized by falling fertility rates and increasing life expectancy. Thus, the population share of young declines, the population share of working age people stagnates roughly while the population share of old people virtually explodes.

Demographics has recently captured the attention of economists since it might contribute to the so-called Secular Stagnation (Aksoy et al., 2018; Eichengreen, 2015). Demographics might contribute to low GDP growth rates, low interest rates and puzzling low inflation.

Chart 18: Global median inflation rate in percent 1960 - 2015. Source: World Bank



Both population size and population structure matter for economic development and eventually inflation. To explain this, I first describe the setup of the AS-AD model. Afterwards, the life cycle theory explains the effects on the level of an individual person. The demographic transition theory explains the aggregate economic effects on country-level.

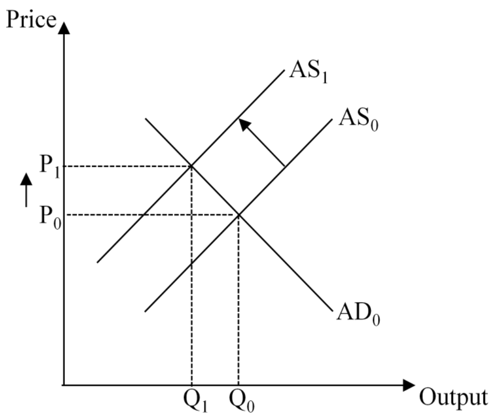
2.1.1. Aggregate supply and demand

Consider a model of aggregate supply and aggregate demand. Aggregate supply includes labour supply. Aggregate demand consists of private consumption besides other factors. At the beginning, the economy shall find itself in an equilibrium. Demographic changes, i.e. population size growth or changes of the population distribution, can affect both supply and demand.

When the population size grows but the distribution remains unchanged, both the aggregate supply and the aggregate demand curves are shifted outwards. The aggregate supply curve is shifted outwards since the labour force grows. Private consumption increases, too, and shifts the aggregate demand curve outbound by the same amount. Eventually, output increases in real terms and depending on whether demand or supply is affected stronger, also the price level might adapt.

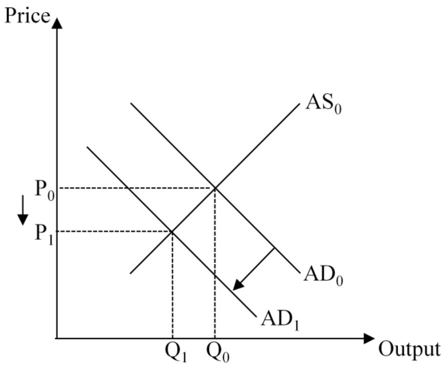
When the population size is constant but the age distribution changes, i.e. the population shares of age cohorts vary over time, an increase of the non-working young and old people at the expense of the working-age population will affect the supply and demand curves differently. The aggregate supply curve moves inwards since young and old do not engage in the labour market (compare chart 19). As the labour force shrinks, this production factor becomes more expensive and drives up prices. This view is represented by the BIS strand in literature.

Chart 19: Higher share of young or old reduces aggregate supply and increases prices



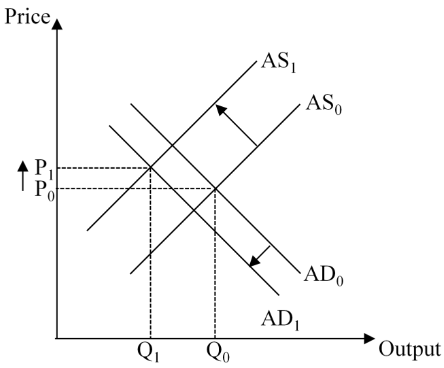
Whether the demand curve reacts depends upon the consumption of the young and old relative to the consumption of working age population. Young people have less income available than working cohort and therefore consume less. Old people, however, have an income comparable to working age people and it is therefore unknown a priori if the former consume as much or less than the latter. Eventually, the aggregate demand curve might remain constant or shift inwards, depending on just how much less the young and old consume less than the working age cohort. The literature represented by the IMF usually follows this line of argumentation and suggests that a higher population share of young or of old are deflationary (chart 20).

Chart 20: Higher share of young or old reduces aggregate demand and reduces prices



The impact on inflation therefore depends on which curve shifts stronger inbound. If the demographic change moves the aggregate supply stronger than the aggregate demand, then prices go up. Otherwise, if the aggregate demand curve is shifted stronger inwards, then prices will shrink. The hypothesis of this paper is that the supply curve shifts more strongly inwards than the demand curve and therefore inflation increases (chart 21).

Chart 21: Higher share of young or old could reduce supply more strongly than demand



2.1.2. Life cycle theory

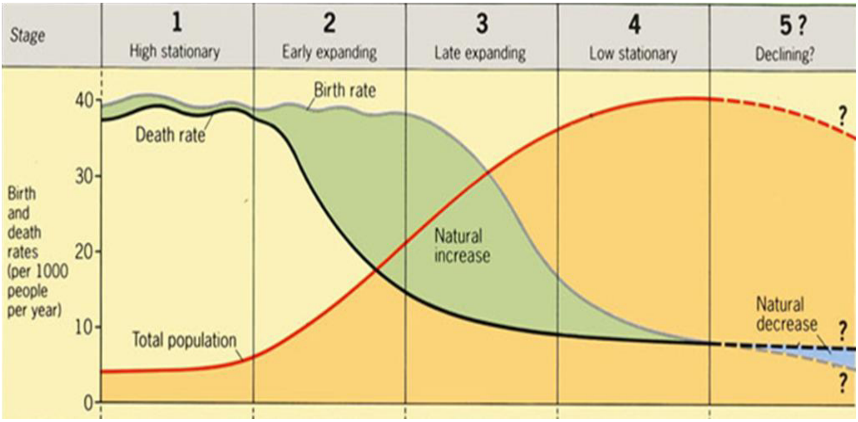
According to the life cycle theory each person wants to smooth consumption over her life time (Ando and Modigliani, 1963). Consumption increases over childhood and adolescence and remains very constant afterwards over working age and at retirement. Therefore, working and saving patterns vary with age.

At a young age, almost no person works and hence does not earn income from labour. Instead, children rely on transfers from their parents, family and the public. After schooling, people start to work, they gain work experience which yields increasing labour income. Income peaks at the age of 45 to 50 years. Afterwards, people tend to reduce their workload since they prefer to enjoy more leisure time or due to health problems. Furthermore, work experience loses value because of technological progress and income declines. At working age, people earn more than they spend on consumption and thus accumulate savings. When people reach the age of 60 to 65 years, labour income drops significantly since the majority of the people do not work anymore due to health problems, public pension schemes or private transfers from family. At this final life stage people live off their savings.

2.1.3. Demographic transition theory

The life cycle theory predicts age-specific consumption and labour supply of the individual person. The demographic transition theory builds upon those predictions and introduces demographic moments. The demographic transition theory describes how a country moves from high stationary levels of fertility and mortality to low or even declining stationary levels. This transition has important effects on economic development (Issa, 2005; Bloom and Canning, 2003; Lee and Mason, 2007; Prskawetz et al., 2000).

Chart 22: Model of demographic transition. Source: www.revisealevel.co.uk



Poor countries are characterized by high rates of mortality and fertility which implies a stable population size and structure. However, improvements in nutrition, hygiene and medical care cause a decline in mortality. In particular infants survive more often.

Hence the population size grows and the population share of the young cohort increases. The young are getting educated but they are not available for the labour market yet. Hence the young consume resources which could otherwise be invested in physical or human capital to increase economic productivity. Economic growth slows down.

Two decades later the children have finished schooling and have become adults. They reach the working age, enter the labour market and begin to earn their own labour income which exceeds their own consumption. The increase of the population share of working age people boosts economic growth.

The country finds itself now in the midst of the demographic transition which has a significant impact on education and, in turn, on fertility. Three effects are at play. The first effect is that labour income rises as the economy grows. An even higher income can be obtained with better education.

Before the demographic transition kicks in, people require many children to ensure the survival of the whole family. With declining mortality and rising income, people can have fewer children and they still survive. Having children implies an opportunity cost in the sense of foregone income. This is particularly true for women who have entered more and more the labour market over the past decades. Adults decide therefore to have fewer children and fertility begins to decline.

Lower mortality implies a higher life expectancy. Hence, people can work over a longer life time and earn more income. With economic growth also economic complexity and technological progress move forward and both reward education with higher income. This is a strong motivation for parents to invest into the education of their children, either since they are altruistic or since they hope that their children will take care of them once they retire. Eventually, the amount of years of schooling increases as well as the quality of education.

The simultaneous decline in fertility and rise in education is labeled the "quality-quantity trade off" (Becker and Lewis, 1973). According to this trade off, parents have only limited resources. With high mortality such as at the beginning of the demographic transition, parents prefer to have many children yet little educated to ensure survival. With economic and technological progress, however, education yields higher income. Parents now decide to have fewer children but to invest into their education. Thus, a country's human capital reservoir grows constantly and fuels economic and technological progress further. This, in turn, yields higher labour income and provides an even stronger motivation for investments into education. Education continues its rise while fertility keeps declining.

In other words, the initial fall of mortality triggers a decline in fertility - but only with a time lag. During this time lag the population size increases. The period of increasing population growth rate is called early expanding. The later expanding period refers to a declining but still positive population growth rate. During both expanding periods a country can take advantage of the so-called first and second demographic dividend in order to boost its economic growth.

The first demographic dividend is a simple identity. GDP per capita is calculated by dividing the GDP ($Y(t)$) by the population $N(t)$. This can be extended by introducing labour: the amount of workers $L(t)$ is divided by the total population - which corresponds to the population share of working age people - and at the same time the GDP is divided by the amount of workers:

$$\frac{Y(t)}{N(t)} = \frac{L(t)}{N(t)} \times \frac{Y(t)}{L(t)}$$

Taking logs and deriving by time yields the following equation:

$$\dot{y}^{n}(t) = \dot{L}(t) - \dot{N}(t) + \dot{y}(t)$$

Thus, growth of GDP per capita, $\dot{y}^{n}(t)$, becomes a linear function. If labour $\dot{L}(t)$ grows faster than the population $\dot{N}(t)$, then GDP per capita increases. This is the first demographic dividend. It implies that the population structure matters for economic growth because an increase of the population share of working age people raises per capita GDP.

The first demographic dividend contributed one third of the "growth miracle" observed in the East Asian countries between 1965 and 1990, i.e. it added between 1.4% to 1.9% to the annual GDP per capita growth (Bloom and Williamson (1998); Bloom et al. (2003)).

According to the life cycle theory, the labour income of people at working exceeds their consumption. While the first demographic dividend endures, working age people accumulate savings. They invest their savings in physical capital or in human capital, i.e. the education of their children. Physical capital and human capital enhance growth of productivity, which is denoted with $\dot{y}(t)$ in the preceding equation. A higher productivity increases GDP per capita growth. $\dot{y}(t)$ is called the second demographic dividend.

The second demographic dividend is a self-fueling mechanism since savings are invested into education which, in turn, generates a higher income and therefore more savings. The second demographic dividend has the potential to last forever - in contrast to the first dividend. The first dividend solely depends on the population share of working age people. At some point in time the working age people will retire, leading to the decline of population share of working age people in favor of the old aged. This implies a low stable or even slightly increasing mortality rate. From now on, the population size remains stable or even declines slightly. The country finds itself in the low stationary or even declining phase. The demographic transition has been completed.

The length of the demographic transition and in particular of both expanding periods and of the first demographic dividend usually lasts for several decades and can take even more than one hundred years. The duration varies between countries and depends on the availability of contraceptive methods, technological progress, policies, migration and other factors. China, for instance, went through the demographic transition at a very high pace within only three decades due to its one child policy. This opportunity window contributed considerably to China's skyrocketing wealth accumulation. But since 2015 the population share of workers has been declining and China's first demographic dividend has faded and could even constrain GDP per capita growth in the future. However, China has accumulated large savings which serve as the basis for the second demographic dividend and future economic growth.

The life cycle theory and the demographic transition theory teach us that not only population growth but also the population structure affects economic development. A larger share of workers implies more labour supply, as well as investments into education and physical capital. In contrast, population shares of young or old do not contribute to labour nor capital supply. Instead, they bind resources for consumption.

Based on these thoughts I argue that the demographic structure has an important impact not only on economic growth but also on inflation. To see this, imagine a simple aggregate supply and aggregate demand framework.

Children do not work nor possess any savings, they consume exclusively. Their income stems from public or private (family) transfers. In other words, a higher population share of young increases demand but does not affect supply.

Working age people are the only cohort whose labour income exceeds their consumption. A higher population share of working age people increases aggregate supply stronger than aggregate demand since they contribute with their labour supply and with capital created with their savings.

Eventually, people retire and leave the labour force. They do not earn any labour income anymore and reduce labour supply. Furthermore, they reduce their savings in order to finance their consumption. Their dissaving reduces capital. In the end, a higher population share of old increases demand and lowers supply.

While the economy rests in equilibrium, an outward shift in aggregate demand causes an increase in inflation. Likewise a decline in aggregate supply feeds inflation. Both shifts can be the result of a change in the population structure.

Demographic transition is a very long term development which takes decades to materialize. Economic agents might not be foresighted enough over such a long time horizon. Furthermore, the transition moves slowly. In this context, economic agents might overlook or underestimate the effect of the demographic transition on economic development. Correspondingly, changes in demand occur in parallel with the change in the population structure. However, it can take more time to adapt supply, for instance when it comes to using more capital to replace retired employees. Since supply adjustments lag behind the changes in demand, inflation arises.

Eventually, I want to investigate the following hypotheses that an increase in the population share of...

1. young increases demand but does not affect supply and therefore inflation rises.
2. workers increases supply stronger than demand and therefore reduces inflation.
3. old increases demand but reduces supply and therefore inflation rises.

This paper confirms all three hypothesis empirically. It supports the strand of the literature which argues that the old age cohort increases inflation instead of reducing it.

The related literature usually uses only demographic but not age-specific consumption and labour supply data. This paper, however, combines population data with age-specific consumption and labour supply data for each single age as provided by the National Transfer Accounts project (NTA).

NTA shows the consumption and labour income of the average person at any age. In other words, I observe consumption and production over the whole life cycle of a person. The difference between labour income and consumption is called the life cycle deficit. The deficit is positive when consumption exceeds labour income. When the population share of a cohort with a positive deficit rises, aggregate demand increases relative to aggregate supply which in turn raises inflation. According to NTA data, young and old people consume more than they produce, i.e. they have a positive life cycle deficit and therefore increase inflation if their population shares rise.

In between the young and the old cohort, the deficit turns negative as people earn more from labour than they consume. This working age cohort reduces inflation.

On global average, the age border for the young cohort is 26 years, and for the old cohort 59 years. Yet, as this paper shows the exact age borders vary between countries and over time. This is in contrast to most of the related literature which defines the age borders in an arbitrary way, i.e. 14 years for the young cohort and 65 years for the old cohort.

Furthermore, literature assumes implicitly that the impact on inflation in terms of magnitude is the same for any age, only that the sign is opposite for the working age as compared to the young and old cohorts. This paper proves that the size of the effect varies strongly with age. Indeed, even within the same cohort the impact depends on the specific age.

2.2. Literature review

Two important strands of the existing theoretical and empirical literature can be summarized by the view of the International Monetary Fund (IMF, 2014) and of the Bank for International Settlements (BIS, 2015). IMF argues that a higher population share of both young and of old reduces inflation. This is the rather common view in literature (Gajewski, 2015; Lindh 2004). In contrast, BIS suggests that both shares increase inflation. I first describe and evaluate these two strands of literature before turning to other channels identified by literature.

2.2.1. IMF strand of literature

IMF builds on the assumption that the young and old cohorts only consume but do not work. Young and old are so-called dependents since they need transfers from the working age people. All working age people are assumed to work full time. Since they produce and earn more than they consume, they are so-called supporters with their labour income partially being transferred directly or indirectly (via taxes) to the young and old cohorts.

Most of the literature including the IMF defines the young cohort as persons of an age from 0 to14 years, the working age cohort as 15 – 64 years and the old cohort as 65 years and over. Based on these definitions, IMF calculates two so-called dependency ratios. The young dependency ratio divides the amount of young by the amount of working age people. The old dependency ratio is derived analogously by dividing the amount of old by the amount of working age persons. IMF finds that both dependency ratios reduce inflation.

Yet, the assumptions and definitions are biased in three respects. First, the age definitions of young, working age and old are ambiguous and do not reflect reality: in many countries worldwide young people enter the labour market only at the age of 20 or later. In contrast, some old do not retire but still work at ages beyond 65. Second, only because people are at working age does not automatically mean that they are actually employed or that they work fulltime. They can be unemployed for educational purposes, taking care of children or other family members or due to health problems. They might also work part time instead of full time. Furthermore, some workers might earn less labour income than they spend for own consumption, a pattern particularly often observed at the age of 14 years and more, as well as at ages well before 65 years. Third, some old people consume more, not less than persons of working age, at least when public medical expenditures are taken into account.

The view represented by the BIS tries to overcome the biased assumptions.

2.2.2. BIS strand of literature

BIS does not use the broad, fixed age definitions nor dependency ratios as the IMF. Instead, BIS wants to take account of the whole population structure, that is, the share of each single age cohort of the total population for a given country in a given year. The population structure is approximated with a polynomial function of 4th degree by age to fit the population structure for various countries worldwide since 1950. The four polynomials are used as independent variables in a Fixed Effects regression which is run on inflation as the dependent variable. Afterwards, BIS reverses the calculation to analyze how each 5-year age interval affects inflation. BIS finds that young people until an age of roughly 30 years as well as old people above 60 years increase inflation.

These findings mark a strong contrast to IMF's view. The IMF focus solely on the demand side, that is, young and old cohorts consume less than working age people and if their population shares rise, inflation will go down eventually. BIS points out that also the supply side has to be taken into account. Young and old do not only consume but also work less than the working age cohort. They also consume more than they produce themselves. If the population share of young and old increases, then aggregate demand rises stronger relative to aggregate supply which gives rise to inflation.

However, BIS bases its analysis purely on demographic data but does not include any empirical data about age-specific consumption and working behavior. Furthermore, BIS uses 5-year age intervals instead of single age years, hence still defines ambiguously age intervals and aggregates data over age though to a smaller degree than IMF with its fixed broad age definitions.

2.2.3. Other literature

In addition to the literature strands represented by the IMF and BIS, further channels on how demographics might affect inflation have been investigated. A good starting point is Japan. Japan is the most rapid ageing country in the world, with an already large yet still growing share of old. Therefore, the country has been subject to various research papers.

Anderson et al. (2014) use a global DSGE model developed by the IMF and conclude that Japan's large old population share reduces GDP growth and thus causes deflationary pressures. Furthermore, old people run down their savings. In the case of Japan many people have purchased assets abroad which they repatriate. Thus, Japan's currency Yen appreciates, imports become cheaper and as a consequence they add to the deflationary pressures.

Liu and Westelius (2016) investigate the effect of ageing on productivity and inflation in all Japanese prefectures. They use 10-year age cohorts as well as population growth and the old dependency ratio as the main independent variables in a Fixed Effects regression. They find that the age-group of 40-49 years is the most productive whereas much younger and old cohorts are significantly less productive. A higher average age as well as a declining population size go hand in hand with lower inflation

On a more detailed level Liu and Westelius analyze which elements of the consumer basket underlying the price index are affected. They show that an increase in the old dependency ratio raises prices of reading, recreation and of furniture. In contrast, prices for food and beverages as well for medical care go down. The latter comes as a surprise since NTA data and literature show that medical expenditures increase with old age which implies higher prices of medical care.

Liu and Westelius (2016) quote the former governor of the Bank of Japan, Masaaki Shirakawa, with reference to a theoretical expectations channel: "Specifically, he postulates that as agents gradually realize that demographic headwinds will lower future growth, and thus their expected permanent income, they cut back on current consumption and investment which leads to deflationary pressures."

Demographics could have an indirect impact on inflation via its link to productivity, such as a reduction of innovative young, a high share of non-innovative old and lower investments. Furthermore, old have a different consumption pattern than young and population of working age, they prefer labour intense services (such as health care, travel and leisure) over more productive industrial goods or services. All those channels lower productivity and therefore increase inflation. In contrast, their working experience could boost productivity. Young working age people, however, are also likely to be very productive due to their application of modern technologies and higher entrepreneurial spirit (compare Liu and Westelius, 2016; Anderson et al., 2014; Feyrer, 2007; Jones, 2010).

Demographics might be partially responsible for the current break down of the Phillips curve in the USA and many European countries. This question is addressed by Nickel et al. (2017). They find that the old dependency ratio and secular inflation trends are cointegrated. Thus the authors claim that a higher population share of old reduces inflation.

Aksoy et al. (2018) suggest that larger population shares of young and old drive up inflation since these age-groups consume but do not produce. They generate excess demand and make labour scarcer, hence more expensive. Production costs are transferred to consumers via higher prices.

Ikeda and Saito (2012) find that asset prices fall since a higher life expectancy and a larger population share of old increase dissaving.

Likewise Lindh and Malmberg (1998 and 2000) argue that a higher population share of net savers (i.e. workers) reduces inflation. In contrast, a higher share of old promotes inflation as they use their pension inflows.

Bullard et al. (2012) describe a political economy channel. They argue that since old people do not work but live off their assets and public pension schemes they prefer low inflation. With old people gaining a larger population share they also increase their power on political elections. Thus, a higher population share of old leads via a political process to less inflation. However, it can be argued that political election processes do not influence directly monetary policy but fiscal policy. As such, old people would vote in favor of more redistribution, i.e. higher public pension expenditures. Thus, resources for productive investments - which lower inflation - would become scarcer and thus, inflation could rise.

Konishi and Ueda (2013) study another political economy channel by inserting fiscal policy into an overlapping-generations model. Ageing reduces inflationary pressures if life expectancy rises but stimulates inflation when fertility drops.

Eventually, also the other strands of literature besides the IMF and BIS provide no clear guidance whether young and old cohorts increase or reduce inflation.

2.3. Data

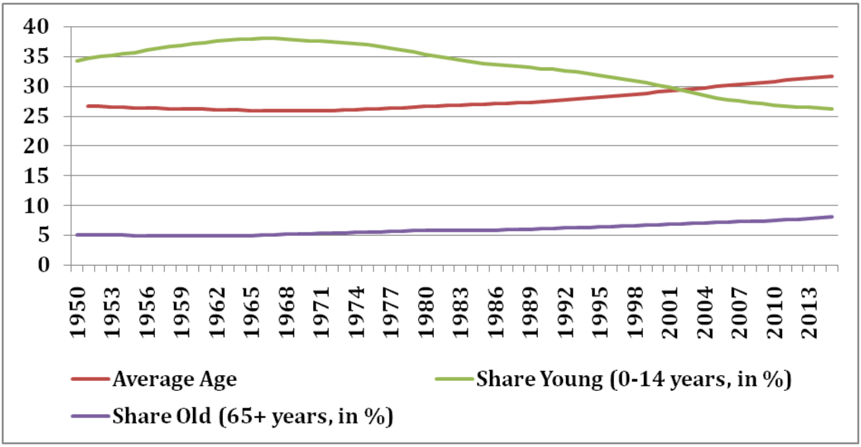
Most of this reviews literature as well as the IMF and BIS papers use demographic data provided by the UN. I use the same data source.

2.3.1. Description of demographic data

The United Nations World Population Prospects are the most common source for demographic data. The data base provides the amount of persons for any age in about 200 countries in the world, spanning from the year 1950 until 2015. For each age, I divide the amount of persons of this age through the total population to calculate the age-specific population share. Whenever I speak about shares I refer to population data in order to distinguish it from NTA data.

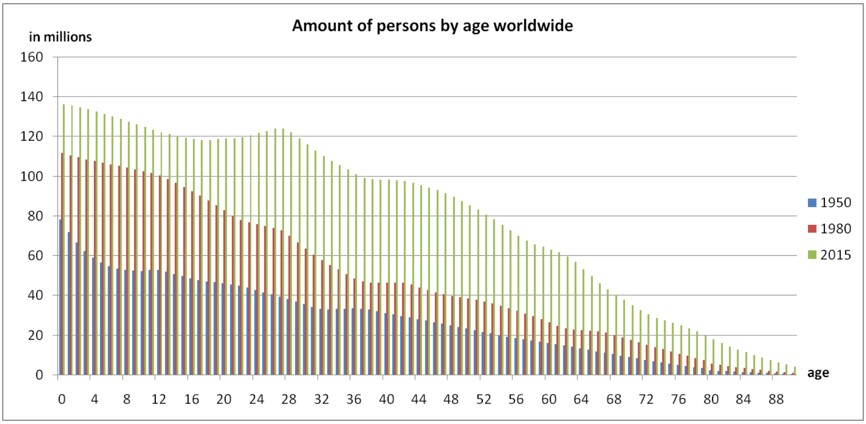
Global population has tripled from 2.5 billion people in 1950 to 7.5 billion in 2015. Not only has the size changed but also the population structure, the share of old (65 years and more) has been growing constantly and accelerated in the last few years while the share of the young 0 - 14 aged peaked in the 1960s and has been falling since (see chart 23).

Chart 23: World average age and population shares of young and old, from year 1950 until 2015. Source: UN



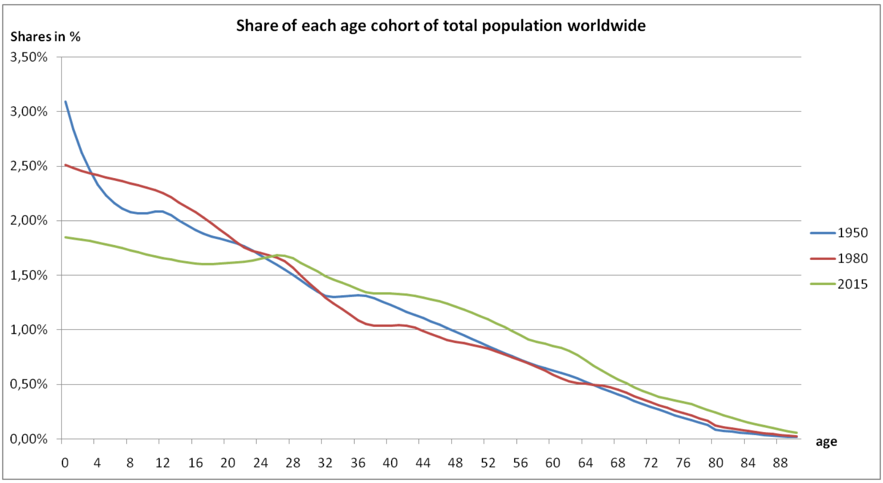
Looking at the absolute amount of persons by age we see that the size of each age cohort has at least doubled since 1950. While in 1950 particularly the young dominated, nowadays we observe a bulk of middle ages and a strong increase in the old ages (chart 24).

Chart 24: Global amount of persons in millions by age cohort. Source: UN



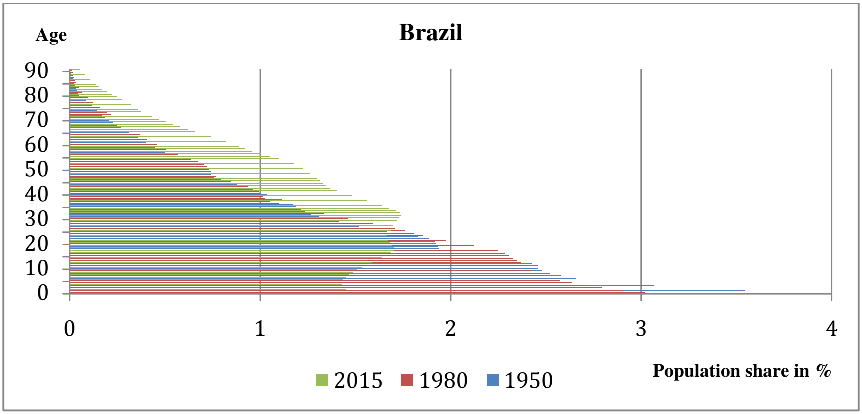
Thus, out of total global population, the share of young age cohorts declined from what was more than 3% to less than 2%. The share of the working age population increased slightly as well as the population share of the older people. Age distribution has become more equal (chart 25).

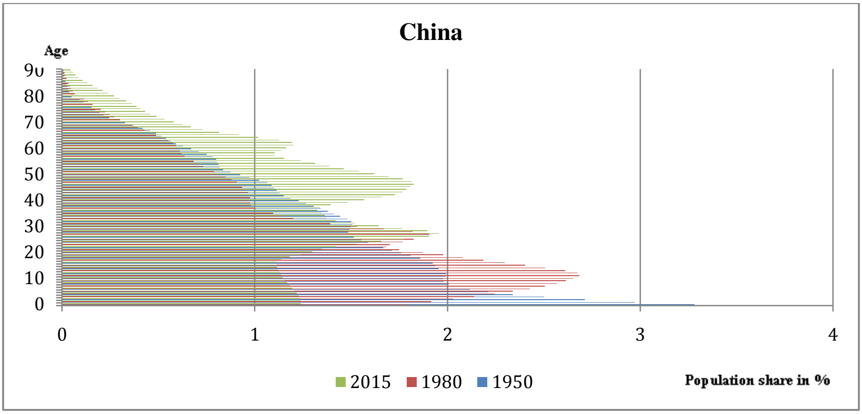
Chart 25: Global population share by age. Source: UN

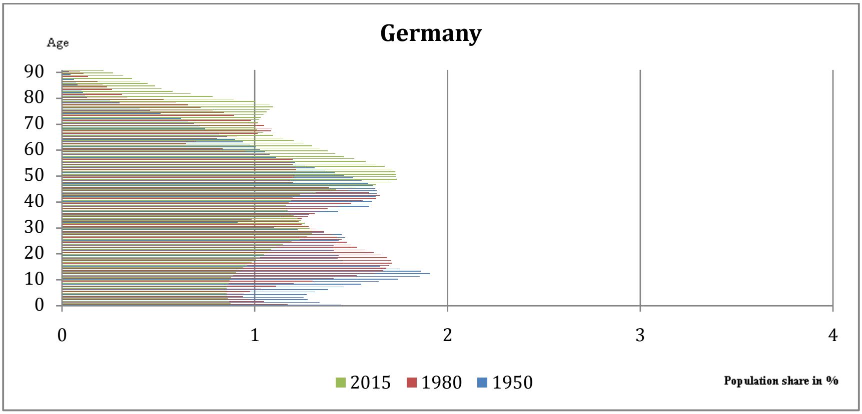


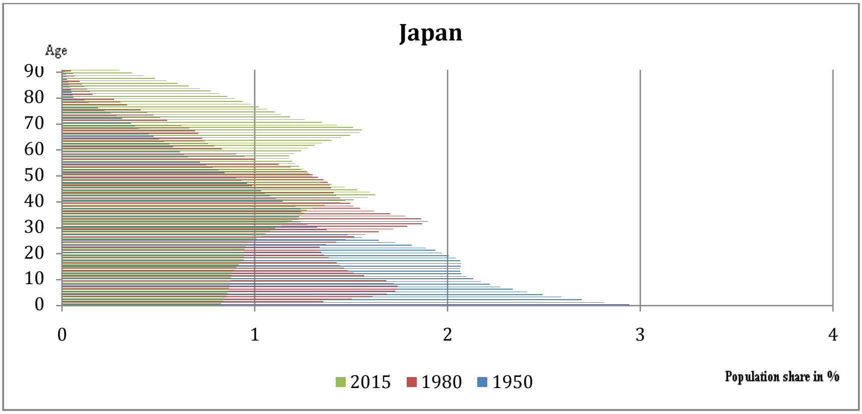
The ageing of societies is not observed exclusively in the vast majority of the advanced economies but also in several developing economies (chart 26).

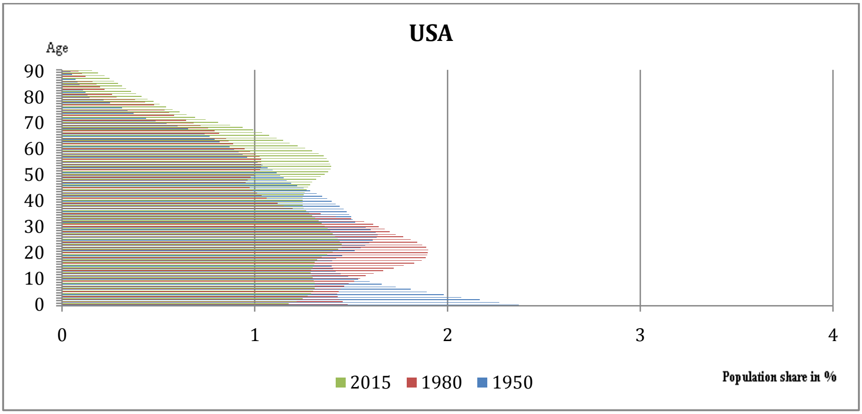
Chart 26: Population shares by age in percent in 1950, 1980 and 2015 for selected countries. Source: UN






Germany
Age
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
2015
1980
1950
Population share in %


Japan
Age
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
2015
1980
1950
Population share in %


USA
Age
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
2015
1980
1950
Population share in %

2.3.2. Description of NTA data

According to the life cycle theory, consumption and labour supply vary with age. The National Transfer Accounts project (NTA) gathers corresponding empirical data. Hereinafter I summarize the NTA manual book (2013).

The NTA data is based on the idea that any financial outflow has to be financed by some sort of financial inflow. Let us consider the following equation:

$$YL(x) + t^{+}(x) + Y^{k}(x) + Y^{p+}(x) = C(x) + t^{-}(x) + Y^{p-}(x) + S(x)$$

The left hand side of the equation represents inflows:

- Labour income YL(x) is the most important inflow from the perspective of this paper. It consists of gross salaries, wages, fringe benefits and social contributions paid by employers on behalf of the employees. Labour income also consists of self-employed earnings which is calculated as two thirds of national gross mixed income.
- Transfer inflows $t^{+}(x)$ refer to private, family-based inflows (e.g. elderly who receive financial support from their working-age children, remittances, charitable contributions) as well as public inflows, i.e. by the government mediated inflows such as publicly financed education and health services or pension and other social benefits.
- Capital income $Y^{K}(x)$ is defined as capital which is invested into production processes in order to generate a future income stream. Private capital income summarizes housing, consumer durables, and similar. Also one third of national gross income is attributed to capital income. Public capital income, however, is rather negligible.
- Property income inflows $Y^{p+}(x)$ refers to private financial assets, interest rates, royalties on natural resources and similar. Furthermore, public property such as public debt and sovereign wealth funds are included as well. Any property income inflow is financed by a counterparty.

On the right hand the equation defines the outflows:

- Consumption C(x) is the most important outflow. It consists of both private consumption and public consumption. Education, health and other items are consumed both privately and publicly.

- $t^-(x)$ means transfer outflows and measures in particular taxes paid to the government.
- Property income outflows are assigned as $Y^{p-}(x)$, mirroring the property income inflows.
- S(x) represents savings. This variable balances the equation out. If inflows exceed outflows then the savings are positive.

This paper focuses on the labour supply and consumption in an AS-AD-framework. Therefore, I rearrange the preceding equation:

$$C(x) - YL(x) = t^+(x) + t^-(x) + Y^k(x) + Y^{p+}(x) + Y^{p-}(x) + S(x)$$

Now, the left hand side of the equation defines the life cycle deficit which is calculated as the difference between consumption C(x) and labour income YL(x), i.e. the individual's person consumption and labour supply. If consumption exceeds labour income, the life cycle deficit is positive and it needs to be financed by any other source of income. Those other sources can be net transfers ($t^+(x)$ - $t^-(x)$), capital income, net property flows ($Y^{p+}(x)$ - Y^p-(x)) or savings / debts. It is assumed that the savings (= wealth) of a new born, i.e. at birth, are equal to zero.

Likewise it is assumed that wealth equals zero at the end of the life time, i.e. upon death. If a person has nonzero savings, these are handed down to the younger generations. In other words, bequests are allowed for. Bequests are computed as part of private transfers.

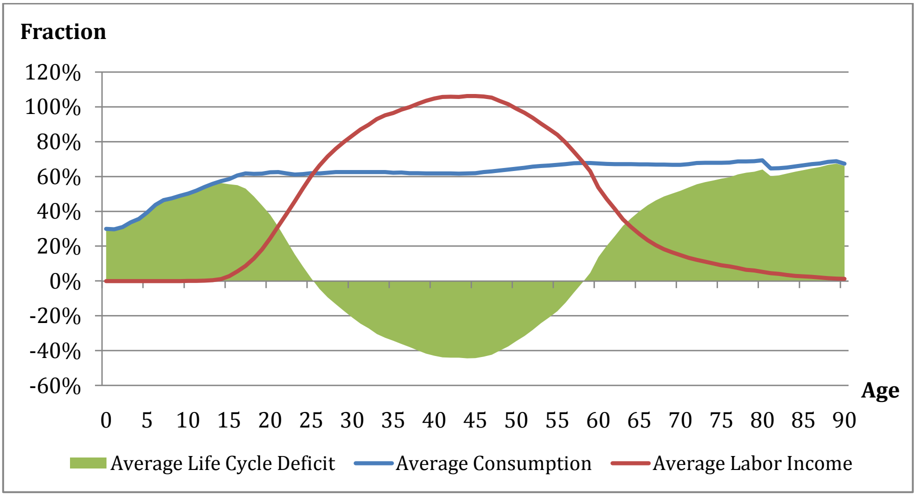
Related to this, it should be noted that the total life cycle deficit accumulated over life time must not equal zero at the end of the life time. The reason is that in this paper the life cycle deficit is defined as the difference between consumption and labour income at the given age. However, capital income, transfers and dissaving are other forms of income used for consumption. It is assumed that upon death consumption is completely financed. When total income exceeds consumption and other outflows by the point of death, it is assumed that is wealth is transferred to younger generations as bequests.

The NTA computes all flows for any singular age between 0 and 90 years. The NTA measures inflows and outflows in local currencies. For the purpose of comparability between countries, NTA calculates all variables as a fraction of the simple average of labour income received between ages 30 to 49 years.[2] I use the term fraction to distinguish NTA data from the UN population data.

The NTA data is an accounting system based on the System of National accounts. The Appendix contains more information about how NTA data is constructed.

The following chart 27 shows the global average of consumption, labour income and the life cycle deficit fraction for each age.

Chart 27: Global average consumption, labour income and life cycle deficit by age, expressed as fractions of simple average labour income at age 30 - 49. Source: NTA, own calculations



The empirical data confirms the life cycle theory that consumption is smoothed over life time. Consumption doubles from 30% at birth to 60% at 15 years, measured as a fraction of the average labour income earned between 30 to 49 years. Working and saving patters vary with age. Labor income peaks between 45 to 50 years, reflecting the most productive age of workers. Depending on the age, the life cycle deficit ranges from +60% to -40%. A positive deficit implies that a person cannot finance her consumption with her own labour income at a given age and therefore requires private or public income inflows. A negative deficit means that a person's labour income

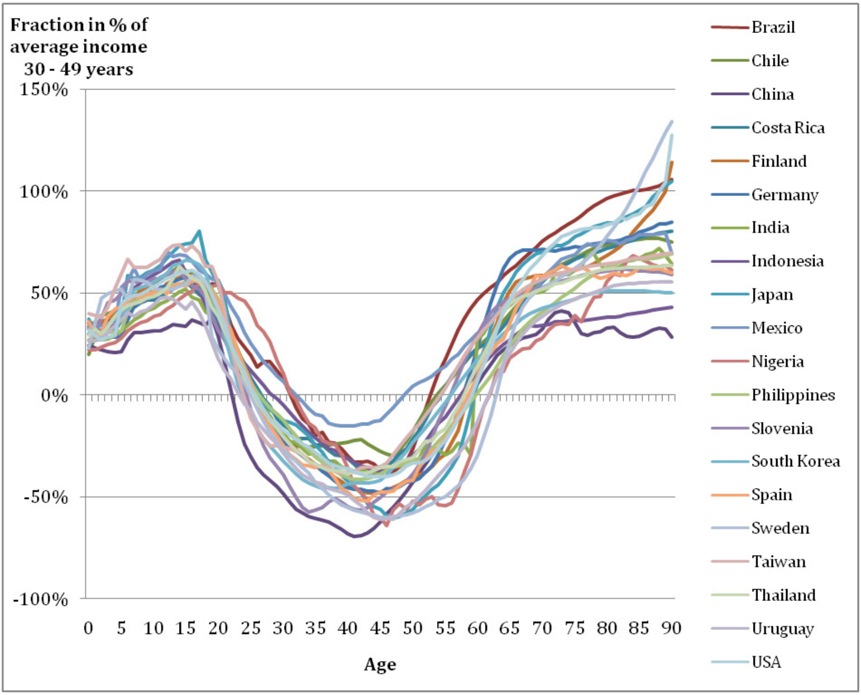
[2] For an overview of published NTA papers, please see: http://www.ntaccounts.org/web/nta/show/Published%20NTA%20Papers%20Public

exceeds her consumption. Yet, the deficit is not equal to savings as the other inflows and outflows have not been taken into account yet.

The life cycle deficit shows a common pattern among different countries, that is, a positive deficit for young people until the age of roughly 25 and for old persons above 60. In between, the deficit turns negative indicating the working life of the average person.

However, a closer look on the NTA data shows that the age when a person enters and leaves the labour market varies between countries. For instance, in the random subsample of countries shown in chart 28, people enter the labour market the earliest at an age of 18 years and at the latest at 34 years. On the other hand, they exit the labour market at ages between 48 and 62 years. These large variations reflect the institutional differences between countries with respect to schooling system, pension scheme, retirement age, private family transfers as well as GDP per capita, fertility rate and life expectancy.

Chart 28: Life cycle deficits by age for a random selection of NTA countries expressed as fractions of simple average labour income earned at age 30 to 49. Source: NTA, own calculations



NTA provides data for 42 developed and developing countries over all continents. The time horizon reaches from year 1993 until 2015. However, data is not available for all countries each year. Chart 29 indicates for which year data is available for the countries in the sample. There are 60 country-year pairs in total. For each of these 60 pairs the life cycle deficit fraction at every age from 0 years until 90 years is known.

The NTA project is very resource-intense. It breaks down macro country-level data of the System of National Accounts to the micro-level of the individual person for any given age. This requires combining a large amount of different macro level sources (government reports, public pension funds, national and international statistic agencies, etc) and micro level sources (tax data, surveys, etc) and is therefore highly computational-intensive. This explains why NTA data is not available for all countries in all years.

Chart 29: Availability of original NTA data by country and year. Source: NTA

Country / Year	1993	1994	1995	1996	1997	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Argentina					X																	
Australia											X						X					
Austria			X				X					X					X					
Brazil				X					X													
Cambodia																X						
Canada																		X				
Chile					X																	
China									X													
Colombia															X							
Costa Rica											X									X		
El Salvador																	X					
Ethiopia												X										
Finland											X		X									
France								X				X						X				
Germany										X												
Ghana												X										
Hungary												X										
India											X											
Indonesia												X										
Italy															X							
Jamaica									X													
Japan											X											
Kenya		X										X										
Mexico											X											
Moldova																					X	
Mozambique															X							
Nigeria											X					X						
Peru														X								
Philippines						X																
Senegal												X						X				
Slovenia											X						X					
South Africa												X										
South Korea							X										X	X	X			
Spain							X															
Sweden										X												
Taiwan	X																					X
Thailand											X											
Turkey													X									
United Kingdom																			X			
United States										X								X				
Uruguay													X									
Vietnam															X							

2.3.3. Extension of the NTA data set

I would like to extend the NTA dataset in order to obtain a larger data base which in turn increases the validity of my estimation results. According to the demographic transition theory age-specific consumption and labour supply are heavily influenced by fertility, life expectancy and per capita GDP (Carroll and Summers (1991), Lee and Mason (2010), Cervellati and Sunde (2011)). These three variables are the most crucial though not only relevant factors. Those three variables are provided by the World Development Indicators and are available for a large number of countries and for several decades, thus allowing to extend the NTA data set manifold.

At the beginning of its demographic transition and economic development, each country is characterized by a high fertility rate, low life expectancy and low per capita GDP.

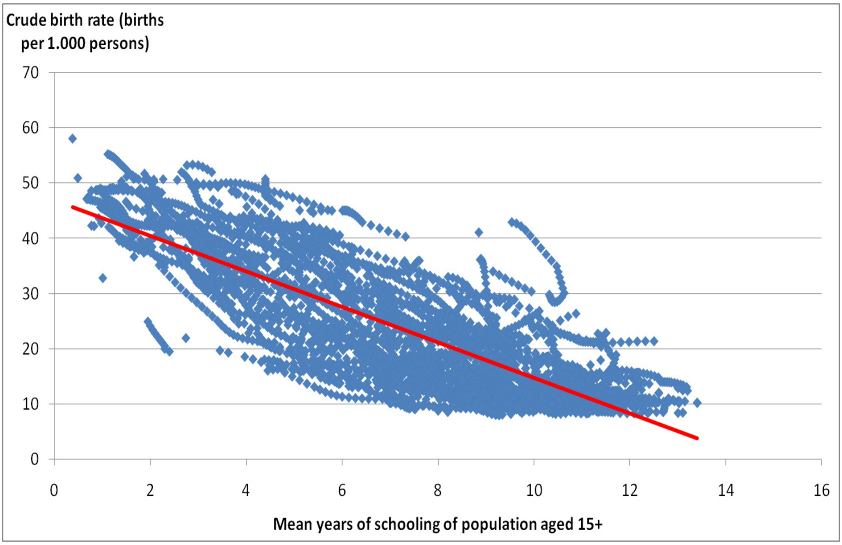
Under such circumstances, parents have many children to make them work and thus to secure the survival of the family. Technological progress of agricultural production methods increases food yield. Thus, less children are needed to secure survival and fertility declines. Better nutrition also implies that people live longer, i.e. life expectancy increases.

Following the technological progress, more complex products are created which in turn require more sophisticated human capital. Returns on education increase which give the incentive to invest into schooling. Returns are supported additionally by higher life expectancy since higher income is earned over a longer life span.

Parents, therefore, decide to spend more money on the education of their children - either due to an altruistic or an insurance motive i.e. parents hope for a better care and higher income transfers once they retire and their children will take care of them. This implies higher education consumption at young age. As the financial resources of parents are limited they tend to have less children but to invest more capital into each child. Eventually, having to care for fewer children has allowed women to engage in the labour market. Their participation rate has increased and they earn their own income. The opportunity costs of bearing children have therefore increased and women tend to have less children.

This situation leads to a decline of fertility and the rise of education, the so-called Quantity-Quality trade off: parents can channel their resources on fewer children, thus increasing the spending per child. Chart 30 suggests a linear negative correlation between fertility (measured as crude birth rate, i.e. amount of births per 1000 persons) and mean years of schooling. In terms of NTA data, lower fertility increases consumption of education particularly at a young age.

Chart 30: The Quality-Quantity-Trade Off illustrated by crude birth rate and mean years of schooling. Sources: World Bank, Barro-Lee



Education fosters human capital which in turn increases GDP per capita. With higher GDP per capita a country can build up a welfare system, in particular to finance education for young and health care for old people. Thus, a higher GDP per capita tends to increase consumption of the young (mainly in the form of expenditures for public schooling) and of the old (mainly in the form of expenditures for public health care and public pension schemes).

Moreover, with higher GDP per capita overall consumption increases though marginal propensity to consume is declining. This implies, that consumption measured as a fraction of the labour income increases slowly with GDP per capita.

A rising life expectancy means that people live longer. With each additional year, their total consumption increases. At old age in particular the consumption of health care increases.

The effects of fertility, life expectancy and GDP per capita are not limited to consumption but also stretch to labour income.

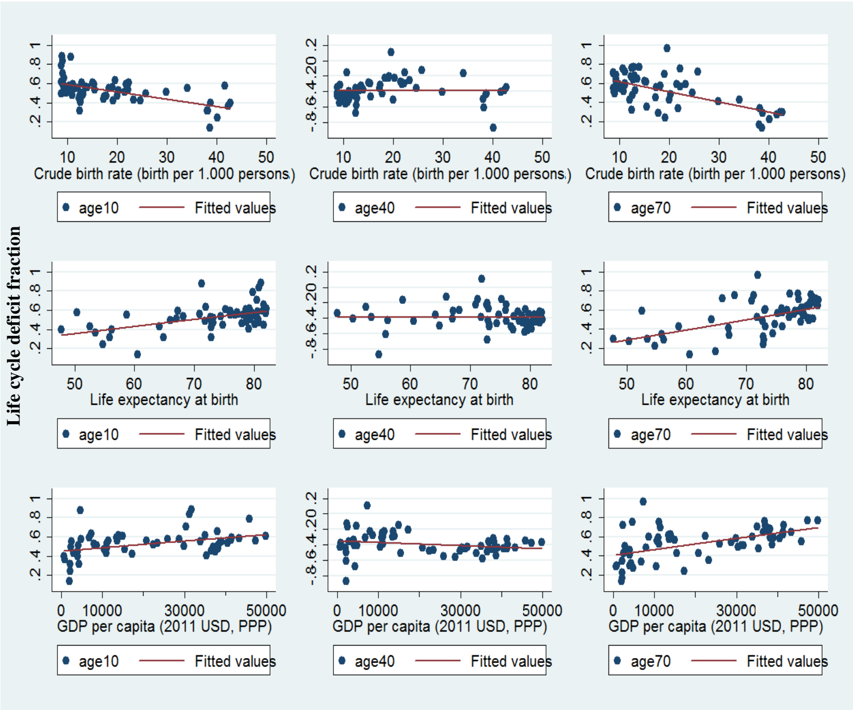
Thinking of GDP per capita as a budget constraint it proves to be a crucial determinant on how much resources can be spent on schooling. A larger per capita GDP allows for longer and more quality schooling.

A higher GDP per capita implies also higher salaries which are usually obtained by schooling. In other words, the returns on education increase and people are likelier to dedicate more time to schooling. Eventually, with higher life expectancy the returns for education prevail longer, therefore increasing additionally the time spend on schooling.

Spending more time for schooling implies a delay in entering the labour market. Yet, higher education yields a higher labour income once a person is working. Better educated persons also tend to longer in order to collect the returns on their human capital over more time. As a consequence, the leave of the labour market is delayed, too. Due to their longer life expectancy people are aware that they well need to finance more consumption at pension age, hence they reduce their consumption during their working age.

Chart 31 shows the linear relation between each of the three independent variables crude birth rate, life expectancy and per capita GDP and the life cycle deficit fraction at ages 10, 40 and 70 years. Those ages were chosen to represent the young, working age and old cohort, respectively.

Chart 31: Life cycle deficits at ages 10, 40 and 70 and crude birth rate, life expectancy and per capita GDP. Sources: NTA, World Bank, own calculations



Each scatter plot consists of 42 countries at different years between 1995 and 2015, 60 country-year observations in total. Life cycle deficits are shown for ages 10, 40 and 70 years to represent the young, working age and old cohorts. The life cycle deficit is calculated as the fraction of average labour income earned between 30 and 49 years.

Each point in the scatter plots marks a country-year observation. For most observations, the life cycle deficit is positive at ages 10 and 70, but negative at age 40. This corresponds to the life cycle theory.

The life cycle deficit as well as the independent variables are moving very slowly over time, thus allowing for the linear OLS approximation. At the age of 10 years, for instance, the deficit increases by 10 percentage points only when the birth rate drops by 10 births per 1000 persons, or when the life expectancy increases by 15 years or when per capita GDP rises by 25.000 USD.

The correlation between the deficit and an independent variable depends upon both, the variable itself and the age cohort. At the age of 70, the life cycle deficit correlates positively with life expectancy and per capita GDP, but negatively with the crude birth rate. In turn, per capita GDP correlates positively with the deficit at the ages 10 and 70 years but negatively at age of 40 years.

Finally, the scatter plots suggest a linear relation between all independent variables and the life cycle deficit for each age cohort.

In order to extend the NTA data set, I regress the life cycle deficit for each single age between 0 and 90 years for 60 country-year pairs, i.e. for 42 countries in 20 years. The deficit is measured as the fraction of average labour income earned at ages 30 till 49 years. The deficit data is provided by NTA. The deficit is linearly regressed on the three variables crude birth rate (amount of births per 1000 persons), life expectancy at birth and GDP per capita measured in international 2011 USD at purchasing power parity. This data is provided by the World Bank. In other words, I first pool all the original NTA data and then I make out-of-sample predictions country by country. Regressing the deficit on all three independent variables for all ages separately in OLS regressions yields an adjusted R^2 of up to 80%.

These OLS estimations allows to extend the NTA data set in the cross-section and the time dimension as fertility, life expectancy and GDP per capita are available for many more countries and years. Eventually, I estimate the life cycle deficit for each single age for 184 countries between 1960 and 2015. The data set is unbalanced, meaning that I have between 8 and 56 year observations depending on the country. I observe between 101 to 184 countries for a given year. I end up with about 8400 observations.

I am not controlling for heterogeneous behaviors between countries with respect to changes in income, life expectancy and fertility rate. This seems to be a rather strong assumption given that cultures and corresponding economical behaviors vary strongly between countries. However, the NTA data set is based on countries from different continents and years and the linear extrapolation is the average behavior. Furthermore, the incentives (quality-quantity trade off, marginal propensity to save with rising income, change of age when entering and leaving labour market due to an increase in life expectancy) are the same among all countries and people react similarly.

Given that according to theory, the decline of mortality is at the beginning of the demographic transition, I could have used mortality as an independent variable in the linear regressions instead of fertility and life expectancy. However, I do not use mortality as an independent variable in the regressions as I want to distinguish two effects which are relevant to the age-specific economic behavior.

First, according to the demographic transition theory, the decline of (child) mortality precedes the reduction of fertility with a delay of about two decades. Second, any change in mortality affects directly life expectancy, i.e. the less people die the larger the life expectancy. Furthermore, life expectancy does not only rise due to less deaths at a young age but also due to medical and technological progresses extending the life time for old persons as well.

2.3.4. Description of the extended NTA data set

In order to provide more intuition to the extended NTA data set, the three following charts show the minimum, first quartile, medium, third quartile and maximum values of the life cycle deficit fraction, population shares and their interactive term. The latter will become an independent variable in the regressions in the following subchapters. The reason is that I want to investigate how inflation reacts if the population share of a particular age increases by one percent. To answer this question, I need to know the age-specific economic behavior which I obtain by multiplying the deficit fraction with the corresponding population share for any given age.

Chart 32: Life cycle deficit by age as fraction of simple average of labour income between 30 to 49 years based on extended NTA data set

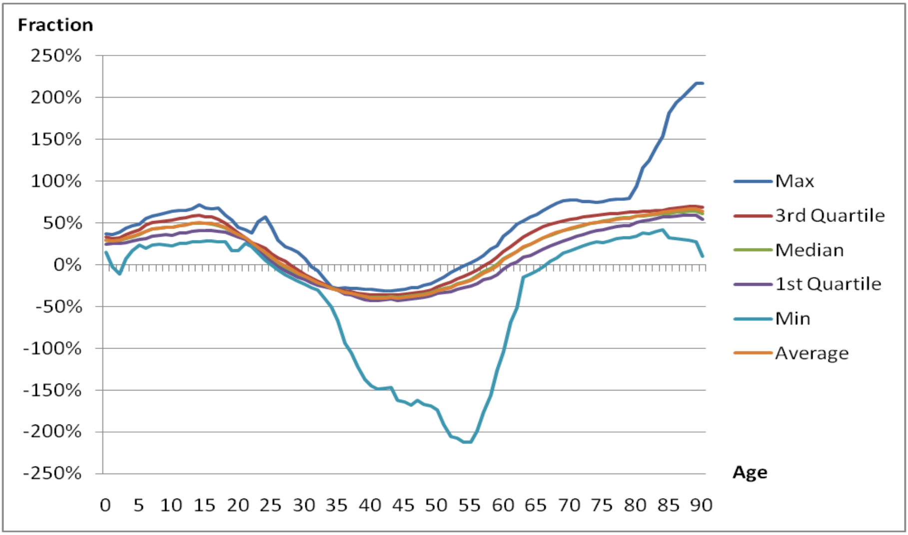


Chart 33: Population shares by age based on extended NTA data set

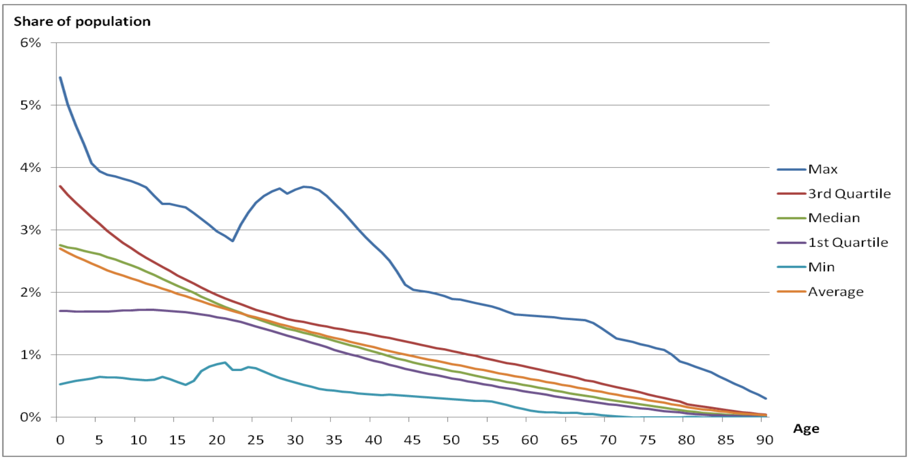
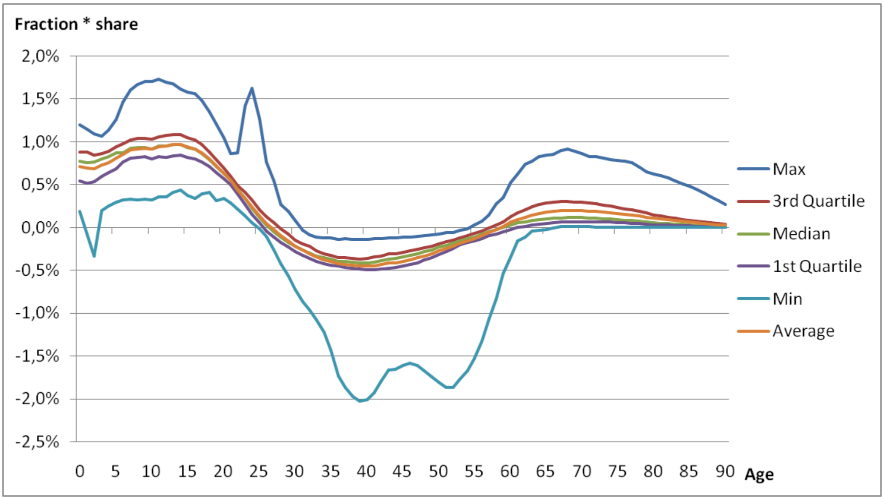


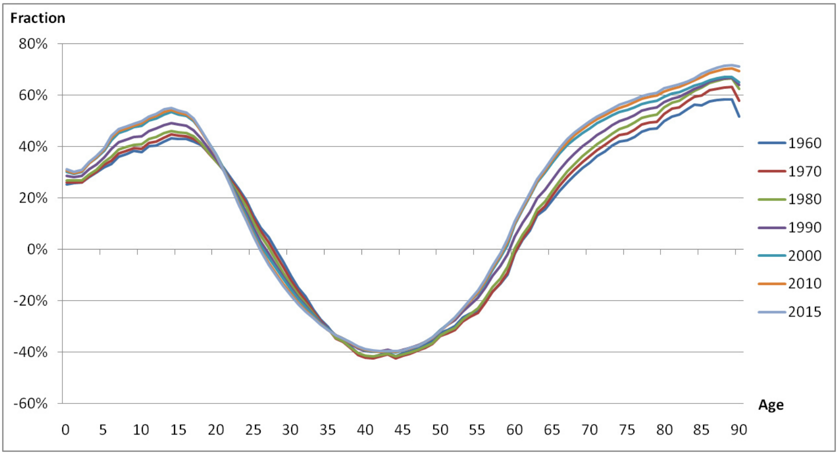
Chart 34: Labour cycle deficit fractions multiplied with population shares by age based on extended NTA data set



Charts 32 until 34 suggest a large variation between countries with respect to the labour cycle deficit, population shares and their respective interaction term. Yet, the extreme minimum and maximum observations do not bias the average values since they stay close to the median observations. That is an indication that the following Fixed Effects regressions are not driven by outliers.

The following chart 35 shows how the average life cycle deficit by age has developed over time, for each decade since 1960.

Chart 35: Average life cycle deficit by age per decade since 1960



Between 1960 and 2015, the average life cycle deficit has deepened for young and old dependents. At the ages of 15 and 75, for instance, the deficit fractions have widen from 40% to almost 60%.

The working age duration as measured between the age when the deficit turns from negative to positive and back seems to have narrowed over time. People enter the labour market at a later age and leave it earlier, although its magnitude has remained fairly stable, peaking at -40% between the age of 40 to 45 years. This implies that the life cycle deficit aggregated over life time has increased since 1960. It has become more unlikely that the average person is able to cover all her consumption with her own labour income over her life time. This does not have to be necessarily problematic as consumption can also be financed by capital income, private transfers and public transfers. Eventually, it is assumed that wealth at the end of the life time is zero.

Table 10 shows descriptive statistics for the whole data set. My dependent variable in all Fixed Effects regressions is the inflation rate, measured as the annual change of any country's Consumer Price Index. Like the other macroeconomic data it is provided by World Bank. Population data is derived from the UN.

The adjusted young dependency ratio and the adjusted old dependency ratio are the independent variables of the IMF approach regressions in the following subchapter. It corresponds to the traditional dependency ratios which are adjusted with NTA data.

The next four polynomials are the independent variables of the following BIS approach regressions. The polynomials approximate the distribution of the life cycle deficit fraction multiplied with the population share by age.

As standard control variables I use annual broad money (M2 aggregate) growth as a proxy for monetary policy, annual growth of population and annual real GDP growth. All three control variables are expressed in per cent.

In order to control for other potential factors driving inflation rates all over the world, I will add dummy variables for the oil price shocks in the years 1974 and 1980. Furthermore, I control for the collapse of the Bretton Woods system by including the annual gold price per ounce in US-Dollar as well as the real effective exchange rate of the US-Dollar.

Eventually, I want to ensure that endogeneity does not bias by regression estimations. To that end, I use contraceptive prevalence and an armed conflict dummy variable as instruments.

As mentioned before, the crude birth rate, life expectancy and GDP per capita are used in OLS regressions to extend the NTA data set.

Extreme outliers are excluded from the data set. The data set is unbalanced.

Table 10: Descriptive statistics of the data set. Sources: World Bank, NTA, UN

	N	Mean	Std. dev.	Min	Max
Inflation (in %)	7231	9.37	13.44	-11.69	109.68
Adjusted Young Dependency Ratio	8387	2.32	0.81	0.30	5.13
Adjusted Old Dependency Ratio	8387	0.39	0.30	0.02	1.47
1st polynomial	8387	-5.86	3.51	-24.51	1.20
2nd polynomial (x 10^2)	8387	5.49	1.77	2.12	14.73
3rd polynomial (x 10^3)	8387	-1.09	0.32	-2.63	-0.50
4th polynomial (x 10^4)	8387	0.06	0.02	0.03	0.14
Broad money growth (M2, annual, in %)	6932	18.19	19.78	-49.68	174.43

Real GDP growth (annual %)	8638	3.84	5.43	-29.59	35.63
Population growth (%)	10740	1.87	1.57	-3.95	19.36
Oil price in year 1974 (dummy variable)	12539	0.02	0.13	0.00	1.00
Oil price in 1980 (dummy variable)	12539	0.02	0.13	0.00	1.00
Gold price in US-Dollar per ounce (annual average)	12535	438.24	421.54	35.15	1668.00
Real Effective Exchange Rate Index for the US-Dollar	12318	70.42	9.74	54.09	89.95
Contraceptive prevalence (any method used by woman or partner, in percentage of women aged 15-49)	7536	42.78	24.59	0.20	87.60
Armed conflict on the country's territory (dummy variable)	12539	0.11	0.31	0.00	1.00
Crude birth rate (per 1.000 persons)	10761	29.46	13.17	7.37	58.23
Life Expectancy (at birth in years)	10761	63.59	11.44	18.91	83.80
GDP per capita (2011 US-Dollar, PPP)	8989	13079.55	18696.55	246.12	245077.80

2.4. Model

I constructed this data set to improve the estimation techniques used in the literature as represented by the IMF and the BIS. Those techniques apply Fixed Effects regressions, using the inflation rate as the dependent variable and relying solely on population data without including age-specific economical behavior data. They differ, however, with respect to their demographic independent variables: the IMF uses the young and the old dependency ratios, whereas the BIS calculates polynomials for the population structure.

In general, they run the following Fixed Effects regression:

$$INF_{it} = \alpha + \beta_1 D_{it} + \beta_2 X_{it} + c_t + \varepsilon_{it}$$

INF_{it} represents the dependent variable inflation, i.e. the annual change of a country's consumer price index measured in percent. α, c_i and ε_{it} refer to the constant, time fixed effects and the random error, respectively. Country fixed effects are not included since they were always omitted in my regressions and since the F-tests for joint statistical significance indicated that they are not relevant.

D_{it} denotes the demographic independent variables of interest. In the case of IMF, the young and the old dependency ratios. In contrast, BIS uses four polynomials of the population distribution structure. I replicate both approaches but will adjust their demographic variables with NTA data.

X_{it} is a vector of control variables, namely real GDP growth, money aggregate growth and population growth. Inflation is linked to real economic activity. A high real GDP growth implies an outward shift of both demand and supply, hence it could raise or reduce inflationary pressures. Furthermore, I control for monetary policy by including annual growth of the broad money aggregate M2 (in percent). From a theoretical point of view not only the population structure but also the population size could influence inflation since size has a direct impact on total demand and supply.

The following two subchapters explain how the demographic variables D_{it} of the IMF and BIS are adjusted with the help of the NTA data. Chapter 2.5 presents the results.

2.4.1. The IMF method: dependency ratios

IMF finds that both the young and the old dependency ratio reduce inflation. IMF defines young as persons aged 0 to 14 years and old at age 65 and above. Correspondingly, working age is defined at the years from 15 until 64. Then IMF calculates the young dependency ratio by dividing the amount of young people through the amount of working age people. The old dependency ratio is calculated analogously. The traditional dependency ratios indicate how many young or old persons are supported by a single working age person with direct or indirect monetary flows.

Yet, the traditional dependency ratios calculated by IMF do not take actual consumption and working behavior into account. The traditional dependency ratios assume implicitly that consumption is constant over the course of the whole life. It is also assumed that working age people earn a constant labour income over their work life while young and old do not have any labour income.

These assumptions, however, do not hold as the empirical NTA data shows. In fact, with each single age year, consumption and labour income and therefore the life cycle deficit vary. Those assumptions can be left aside the purely demographic data can be adjusted with the NTA data.

The consumption expenditure can even vary within the same cohort. Within the young cohort, for instance, an one year old baby consumes less than a ten year old child.

While some young people start working at age 15 they usually do not earn enough to cover their consumption expenditures, i.e. their life cycle deficit is positive, and they are therefore still dependents. On global average, young people are dependent until an age of roughly 25 years. Thus, the age boundaries of 0 to 14 years of the young dependency ratio are not accurate, the upper boundary needs to be shifted upwards. However, the exact age boundary varies between countries and over time.

The traditional old dependency ratio begins at age 65 by definition. This boundary reflects the official pension age in many countries. However, NTA data reveals that the old age dependency - again, defined as a positive life cycle deficit - actually often begins much earlier or later than the official retirement age, depending on the country.

The false assumptions of the traditional dependency ratios do not only focus on the demand side but also stretch to the supply side. It is assumed that working age persons earn the same amount of labour income all work life long. In a closer detail, this assumption implies that a person is never unemployed or on an extend sick leave, works full time, and all years from age 15 to 64 years. These assumptions are proven wrong by the NTA data. On average, most persons start earning a labour income already at an age of about 15 years, but this income is far from covering the own consumption entirely. This indicates that persons often start with part-time jobs. Labour income increases and peaks at an age of 45 to 50 years. Afterwards, labour income gradually drops as people reduce their workload, they become more often unemployed or cannot work do not health problems. These patterns are reflected in the NTA labour income data which thus mirrors effective labour supply. Yet, all those options are not reflected by the traditional dependency ratio.

Therefore, I adjust the traditional dependency ratios with NTA data following Hammer et al. (2014) who also provide a summary of the advantages and disadvantages of dependency ratios as well as guidance on how to adjust dependency ratios using NTA data.

In particular, I use the life cycle deficit for each age to define if the person is a dependent or a supporter. As mentioned before, the deficit is positive when a person cannot finance her consumption with her own labour income and thus depends on direct or indirect financial transfers from other people. Those other people are the so-called supporters.

I consider any young person as dependent until the age when her life cycle deficit turns from positive to negative. This threshold age varies between countries and years. Analogously, a person is considered a supporter as long as her life cycle deficit is negative. When at a later age her deficit turns back positive, I refer to her as an old dependent.

With this definition, young people are not only dependent until the age of 14 but 27 years. On the other hand, old people do not become dependent exclusively at the age of 65 years but a bit earlier, at the age of 61 years.

In order to adjust the young dependency ratio, I sum up the amount of all young people with a positive deficit and divide them by the sum of all people who have a negative deficit. Each age is weighted with its respective size of the life cycle deficit fraction. Thus, I obtain the adjusted young dependency ratio (aYDR). I proceed analogously to calculate the adjusted old dependency ratio (aODR). The respective formulas are:

$$aYDR = \frac{\sum_{i=0}^{i=L}(C_i - YL_i)}{\sum_{i=L+1}^{i=O-1}(YL_i - C_i)}$$

$$aODR = \frac{\sum_{i=O}^{i=90}(C_i - YL_i)}{\sum_{i=L+1}^{i=O-1}(YL_i - C_i)}$$

Index L refers to the age until when young people are still dependent and have a positive life cycle deficit. Correspondingly, O implies the old age when people turn from supporters back to dependents, i.e. when their life cycle deficit turns from negative to positive. Thus, $L+1$ and $O-1$ mark the life time span when a person is engaged in the labour market earning enough to cover her consumption with her own labour income. YL_i and C_i denote age-specific labour income and consumption, respectively. Whereas the regular dependency ratios are usually interpreted as the amount of dependents per worker, the adjusted dependency ratios can be referred to as the amount of effective consumers per effective producer (charts 36 and 37).

Chart 36: Adjusted Young Dependency Ratio and Inflation (in percent)

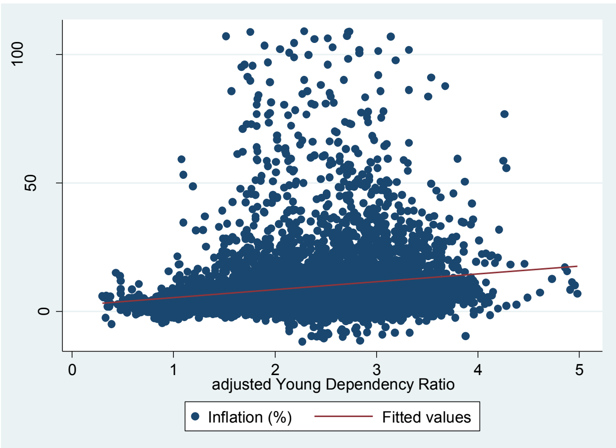
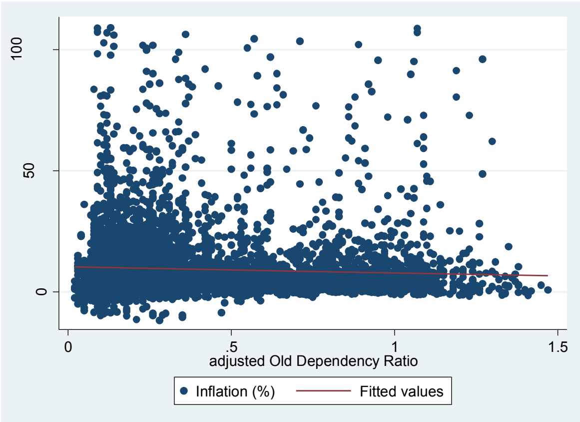


Chart 37: Adjusted Old Dependency Ratio and Inflation (in percent)



I use the adjusted dependency ratios of the young and of the old as main independent variables in the first part of my regressions in chapter 2.5 following the IMF-approach.

2.4.2. The BIS method: polynomials

The main research question is how the change of a population share of a given age affects inflation. To answer this question I need to combine population with economic behavior data. In particular, I multiply the population share for each single age year with its corresponding life cycle deficit fraction. This multiplication is done for each single age from 0 years to 90 years.

I cannot use the each single age year as independent variables in a regression framework as this would not only imply a loss of efficiency but would also lead to multilinearity since any given age is strongly correlated with its neighboring ages.

The majority of literature uses 5- or 10-year age intervals such as 0-4 years, 5-9 years, 0-10 years instead. This method reduces the amount of independent variables drastically, but even those age intervals are still strongly correlated with their neighboring intervals. Furthermore, there is the danger that in a regression the age interval coefficients could change signs in an economic puzzling way. For instance, 0-4 years and 10-15 years could have a positive sign while the middle interval from 5-9 years shows a negative sign. Eventually, the choice of the age interval is ambiguous, e.g. whether it should consist of five or ten years.

For these reasons BIS (2015) calculates a polynomial function to approximate the population structure for each country in any year separately. A polynomial function takes in general the following form:

$$f(x) = \sum_{k=0}^{n} a_k x^k$$

In the case of BIS the *f(x)* refers to the population shares by age and *x* to age. *k* refers to the power of the polynomial function.

BIS reports that a polynomial function of 4th degree fits their population data best, hence they use the following equation:

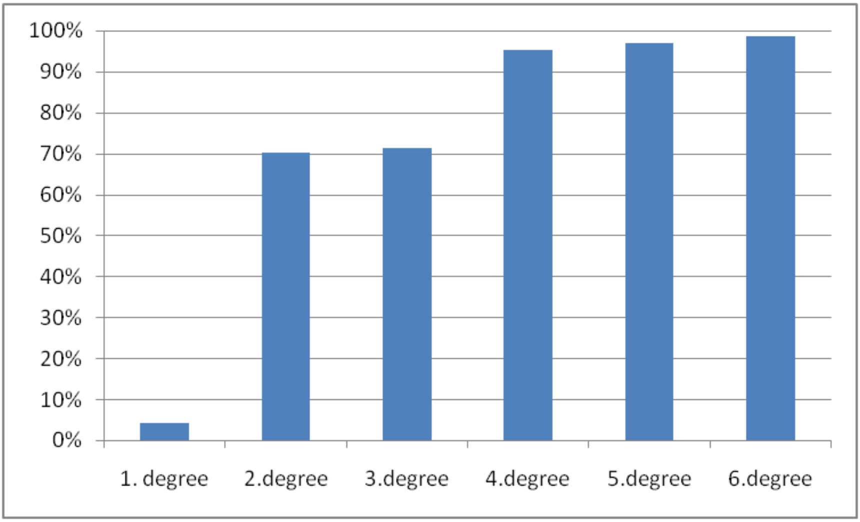
$$f(x) = a_0 + a_1 x^1 + a_2 x^2 + a_3 x^3 + a_4 x^4$$

For defining the independent variable *x*, BIS uses not single ages but 5-year age intervals (0-4, years, 5-9 years, etc.). After approximating the population distribution with this polynomial function for each country in any given year, BIS regresses inflation on the four polynomials a_1 until a_4 (without a_0) as independent variables. BIS avoids multilinearity, excludes the danger of sharply changing coefficient signs and does not choose ambiguously age intervals.

I adjust this approach. In contrast to BIS my data is not based on 5-year age intervals but on single age years from 0 to 90 years. Thus, I avoid defining arbitrary age intervals. Furthermore, I do not want to approximate solely the population structure as this would be purely demographic data but instead I want to approximate the population share multiplied with the age-specific life cycle deficit fraction by using the polynomial function. In other words, I weight each population share by age with their respective life cycle deficit.

Hence, I need to define the polynomial function, i.e. its degree of power. For each country in any year I calculate six polynomial functions, from the 1st till the 6th power. For each of those functions I calculate the adjusted R^2. Chart 38 shows the average adjusted R^2 for all six polynomial functions.

Chart 38: Average adjusted R^2 for polynomial functions from 1st till 6th degree for the product of life cycle deficit fractions times population shares



The functions from 1st to 3rd degree fit the product of deficit fractions times population shares rather poorly as the adjusted R^2 values of 70% and less indicate. Only at the 4th power and higher, the adjusted R^2 reaches at least 90%.

Therefore, I have run the Fixed Effects regression with the polynomial of the 4th, 5th and 6th degree separately. While the polynomials until 4th degree were statistically significant, the 5th and 6th polynomial were omitted due to collinearity. As a consequence I select the polynomials of 4^{th} degree to be the main independent variables in the following regressions of the BIS-approach.

2.5. Results

My hypotheses are that the young and old cohorts increase inflation due to their positive life cycle deficits while the working age cohort is deflationary due to its negative deficit. I am investigating those hypotheses by adjusting the IMF and BIS methods. It should be noted that both methods describe changes of a steady state, i.e. how the long-term inflation rate will change due to demographic developments. They do not refer to transition paths.

2.5.1. Results with IMF method

The following results replicate the IMF approach, using dependency ratios as independent variables. Table 11 reports the main results.

Table 11: Main results of IMF approach

Dependent variable: inflation (in %)	(i)	(ii)	(iii)
Traditional young dependency ratio	-2.870 (2.071)		
Traditional old dependency ratio	-59.25*** (9.857)		
Adjusted young dependency ratio		1.152*** (0.368)	2.470** (1.086)
Adjusted old dependency ratio		-6.411*** (2.307)	-1.920 (3.242)
Money growth (in %)	0.333*** (0.009)	0.335*** (0.009)	0.427*** (0.0695)
Real GDP growth (in %)	-0.382*** (0.031)	-0.381*** (0.031)	-0.482*** (0.107)
Population growth (in %)	-0.532** (0.243)	-0.749*** (0.243)	-1.484* (0.766)
Constant	9.250*** (2.417)	1.894 (2.113)	-0.603 (2.551)
Time fixed effects	Yes	Yes	Yes
Population weighted	No	No	Yes
N	5120	5120	5120
N of countries	153	153	153
R^2 within	0.363	0.360	0.426
R^2 between	0.460	0.579	0.673
R^2 overall	0.383	0.403	0.400

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Clustered standard errors in parentheses. Adjusted dependency ratios are traditional dependency ratios adjusted by life cycle deficit.

The first column uses the traditional definitions of the dependency ratio. It suggests in line with the IMF results that both dependency ratios have a deflationary impact.

However, the young dependency ratio is not statistically significant and the old dependency ratio is implausibly high.

Instead of the traditional dependency ratios, the second column uses the adjusted dependency ratios. Thus, young have a positive significant effect on inflation. The adjusted old dependency ratio's coefficient remains negative yet loses in magnitude.

The first two columns treated each country in the data set equally, irrespective of their population size. The third column weights each country by its population size. The coefficient of young remains positive and significant and rises in magnitude. In contrast, the old dependency ratio is not significant anymore and its coefficient shrinks further. Table 12 reports various robustness checks for the IMF approach.

Table 12: Robustness checks for IMF approach

Dependent variable: inflation (in %)	(i)	(ii)	(iii)	(iv)
Adjusted Young dependency ratio	2.624** (1.236)	3.721*** (1.035)	4.277*** (1.081)	3.620*** (1.03)
Adjusted Old dependency ratio	-0.748 (5.129)	1.732 (2.403)	-2.777 (4.433)	-2.871 (3.112)
Money growth (in %)		0.440*** (0.072)	0.447*** (0.069)	0.441*** (0.071)
Real GDP growth (in %)		-0.490*** (0.134)	-0.506*** (0.121)	-0.506*** (0.122)
Population growth (in %)	-1.365 (0.963)	-0.14 (0.904)	0.156 (0.812)	0.284 (0.851)
2nd lag money growth (in %)	0.236*** (0.038)			
3rd lag money growth (in %)	0.129*** (0.0296)			
4th lag money growth (in %)	0.048*** (0.017)			
2nd lag real GDP growth (in %)	-0.107 (0.092)			
3rd lag real GDP growth (in %)	-0.085 (0.065)			
4th lag real GDP growth	0.005			

(in %)	(0.076)			
Oil price shock 1974 (dummy variable)		11.64*** (4.436)		
Oil price shock 1980 (dummy variable)		3.382** (1.408)		
Gold price (in US-Dollar per ounce)			0.002** (0.001)	
Real effective exchange rate of USD (index)				-0.094 (0.060)
Constant	-0.532 (3.796)	-5.500* (3.032)	-6.520** (3.072)	2.251 (4.294)
Time Fixed Effects	Yes	No	No	No
Population weighted	Yes	Yes	Yes	Yes
N	4666	5120	5120	5120
N of countries	152	153	153	153
R^2 within	0.354	0.344	0.330	0.330
R^2 between	0.536	0.642	0.486	0.510
R^2 overall	0.325	0.371	0.336	0.346

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Clustered standard errors in parentheses. Adjusted dependency ratios are traditional dependency ratios adjusted by life cycle deficit.

One potential difficulty regarding the estimations so far could be caused by endogeneity, meaning that causality does not run from the explaining and control variables to the dependent variable inflation, but the other way round.

The endogeneity bias can be excluded for the variables of interest, namely the dependency ratios, as there is no theoretical reason how inflation could affect demographics. Similarly, there is no link between inflation and the control variable population growth.

Yet, the other two control variables - money growth and real GDP growth - could be endogenous to inflation.

For instance, a central bank can change money supply as a reaction to contemporary inflation or to inflation expected one period ahead. In this case of reversed causality one would expect a negative sign of monetary growth since the central bank would reduce money supply in case of an overshooting inflation. Yet, my results show a positive sign which is in line with the theory that money growth increases inflation, thus being a hint against endogeneity.

Furthermore, inflation could distort investment and consumption decisions and by doing so they could affect real GDP growth. However, the distortion is plausible in the short run but in the long run the "neutrality of money" should prevail. My data set stretches over 55 years, i.e. from the year 1960 until 2015. On average, there are 33 years observed by country. Under these conditions, it seems unlikely that inflation affects real GDP growth in my set up.

In any case, inflation should affect money growth and real GDP growth only contemporarily or in the closest lagged or lead year. The first column in table 12, therefore, does not use contemporaneous money growth and real GDP growth but their second, third and fourth lags. In this way, endogeneity is much less likely to occur. The adjusted young dependency ratio remains positive, significant and its coefficient barely changes its magnitude. Also, the adjusted old dependency ratio remains negative and not significant. It follows that the results are not biased by endogeneity.

Global inflation rates increased strongly in the 1960s and peaked in 1974. Inflation began to decline afterwards yet reached another local peak in 1980. Both peaks were caused by oil price shocks. Oil exporting countries from Near East cut their oil supply for political reasons, thus driving up the oil price and consumption prices all over the world. Therefore, column 2 applies the adjusted dependency ratios and the contemporaneous control variables and adds two dummy variables for the years 1974 and 1980. Since I am interested particularly in these two years, time fixed effects are now excluded. Both dummies are statistically significant and have a positive effect on inflation. The young dependency ratio remains positive and statistically significant, while the old remains not significant. However, its coefficient changes from negative to positive.

Another event which could potentially explain the global inflation pattern that has occurred since the 1960s is the collapse of the Bretton Woods system. After World War II, the international community established a system of fixed exchange rates for their currencies. The US-Dollar was at the core of the system. The other currencies were pegged against the US-Dollar. In turn, the USA guaranteed the convertibility to gold, with the gold price fixed at roughly 35 US-Dollar per ounce. However, in 1971 the USA announced surprisingly that the convertibility would be suspended temporarily. In 1973, the Bretton Woods system collapsed eventually and the fixed exchange rates between the currencies to the US-Dollar and to the gold price were lifted.

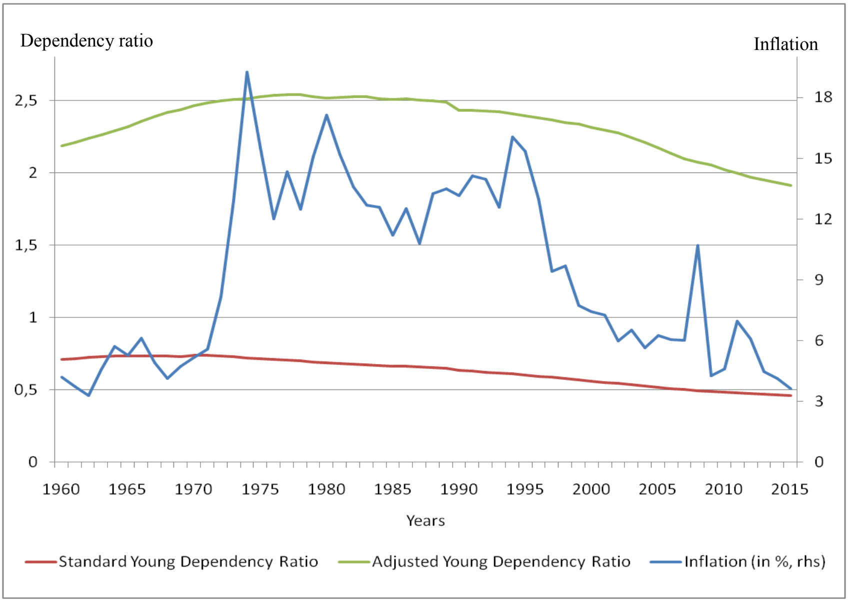
Columns 3 and 4 control for the collapse by including the average annual gold price measured in US-Dollar per ounce and the annual real effective exchange rate of the US-Dollar, respectively. Since both control variables reflect global developments, year fixed effects are excluded. Both estimations do not change materially the results: the young significantly increase inflation while the effect of the old is deflationary but not significant.

Based on the IMF approach, my results suggest that the adjusted young dependency ratio increases inflation in a statistically highly significant manner. By definition, the young consume more than they produce while working age people produce more than they consume. Therefore, the young increase demand as relative to supply. Working age people, in contrast, reduce demand as relative to supply. The dependency ratio puts young people relative to the working age cohort. The former outweighs the latter and inflation increases.

That finding contrasts with the results obtained by the IMF. This can be explained by the different definitions and calculations of the adjusted and the traditional young dependency ratio. The following chart 39 pictures the differences in the global adjusted and traditional young dependency ratios. The different levels of magnitude are striking. The traditional young dependency ratio ranges only between 0.5 to 0.75 points during the years 1960 until 2015. This means that a worker has to support 0.5 to 0.75 young persons. In contrast, according to the adjusted dependency ratio the effective producer has to support 1.9 to 2.5 effective consumers. This large difference occurs as the traditional dependency ratio assumes that the young cohorts ends at the age of 14 years and that the people can finance themselves thereafter and that within the young cohort all persons consume the same.

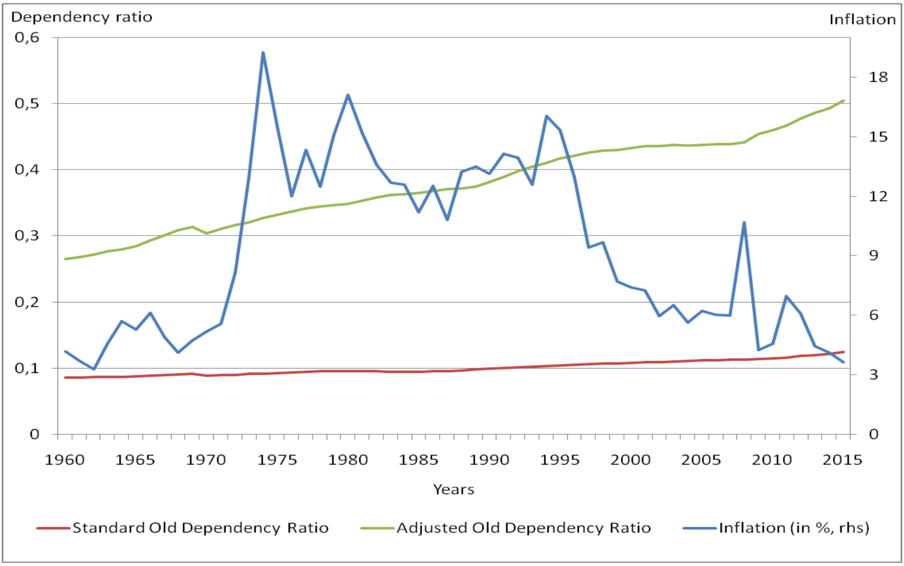
In addition, the adjusted young ratio develops more flexible than the traditional ratio. The adjusted ratio climbs much stronger in the 1960s than its traditional counterpart. The adjusted ratio reaches its climax around 1975, ten years later than the traditional ratio. The adjusted ratio also declines much stronger thereafter. In essence, the adjusted ratio develops much more similar to the global average inflation rate than the traditional ratio. As the coefficient of the adjusted young dependency ratio is positive, it explains partially the parallel decline of inflation.

Chart 39: Global adjusted and traditional young dependency ratios and global average inflation (in percent, right hand side)



The old dependency ratio, however, shows a different pattern (chart 40). On a global scale, the traditional old dependency ratio has increased only slightly over the past six decades. The adjusted old dependency ratio has almost doubled instead. Yet, since both old dependency ratios are increasing they contrast with the inflation which has been declining since the 1970s. This would explain the negative coefficient. However, the coefficient has not been statistically significant.

Chart 40: Global traditional and adjusted old dependency ratios and global average inflation (in percent, right hand side)



The adjusted dependency ratios are more realistic than their traditional counterparts since they take actual age-specific economic behavior into consideration. Yet, irrespective of whether the dependency ratios are defined in a traditional or the adjusted way, the age cohorts are very broadly defined. Furthermore, it remains unclear which way each age cohort contributes to the effect on inflation as the dependency ratio always consists of two age groups simultaneously.

2.5.2. Results with BIS method

Those disadvantages can be solved by using age-specific polynomials. They allow to identify the effect on inflation of each single age. In this section I apply the BIS method of polynomials. BIS found that young and old increase inflation while working age people reduce inflation.

For each country and year, I multiply the life cycle deficit fraction with the population share by age. Subsequently I approximate the population structure with a polynomial function of 4th degree. These four polynomials are the independent variables in my replication of the BIS approach. The main results are reported in table 13.

Table 13: Main results of BIS method

Dependent variable: inflation (in %)	(i)	(ii)
1st polynomial	8.012***	6.310**
	(1.387)	(2.816)
2nd polynomial (x10^2)	56.783***	51.602**
	(14.813)	(23.934)
3rd polynomial (x10^3)	401.821***	389.924*
	(149.5)	(230.7)
4th polynomial (x10^4)	3073.978**	2952.931
	(1451.241)	(2247.824)
Money growth (in %)	0.331***	0.428***
	(0.009)	(0.0697)
Real GDP growth (in %)	-0.384***	-0.480***
	(0.031)	(0.106)
Population growth (in %)	-0.254	-1.234
	(0.252)	(0.782)
Constant	-17.6***	-3.978
	(4.203)	(9.539)
Time fixed effects	Yes	Yes
Population weighted	No	Yes
N	5120	5120
N of countries	153	153
R^2 within	0.365	0.425
R^2 between	0.579	0.714
R^2 overall	0.401	0.404

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Clustered standard errors in parentheses.

All four polynomials have a positive impact on inflation and are statistically significant. This underlines the relevant impact of demographics on inflation. Moreover, money growth has the expected positive sign and real GDP growth the assumed negative coefficient on inflation. This is in line with the findings of the IMF method. However, population growth is not significant.

The first column refers to the unweighted sample, while the results in the second column are weighted by country's population size. The four polynomials lose slightly on statistical significance but the first two polynomials are still significant at the 5%-level, the third polynomial on the 10%-level. All four polynomials keep their positive sign and the coefficients do not change materially in magnitude.

The results suggest that demographics do have a significant effect on inflation. In order to estimate how the population share of a given age affects inflation, I calculate the coefficients and polynomials from the first column back into the polynomial function.

Chart 41: Effect on inflation by age in percentage points based on table 13, column 1

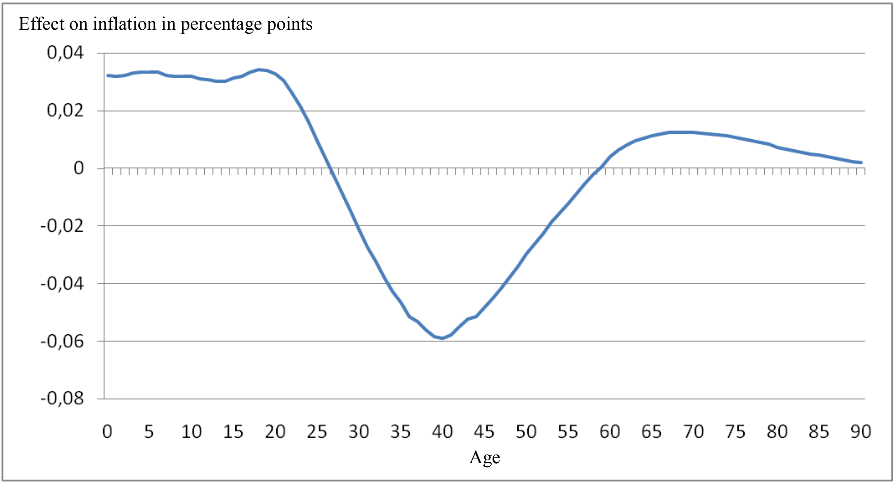


Chart 41 shows the coefficients by any age on inflation. Young people until the age of 26 increase inflation. The same holds true for old people aged 59 years and more. The coefficients in the young cohort are mostly double as large as the coefficients of the old cohort. Working age people prove to be deflationary as expected.

Young and old people with a positive life cycle deficit consume more than they earn from their labour. They create more demand than supply and therefore inflation rises. In contrast, working age people have a negative deficit, i.e. their labour income surpasses their consumption. As labour income is a proxy for production, working age people increase supply relatively to demand, causing deflation.

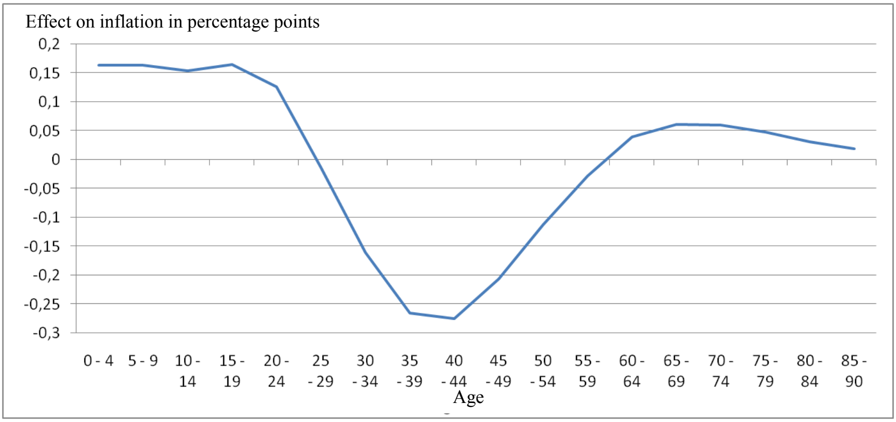
Thus, my three hypothesis from the introduction subchapter are confirmed.

Perhaps surprisingly, the effect on the inflation has a U-shape pattern although the coefficients of all four polynomials are positive. Keep in mind, that already the life cycle deficit has a U-shape pattern as, depending on the age, the deficit is positive or negative. These deficits are basically weighed by the respective population shares by age. Thus, the U-shape of the polynomials of life cycle deficits multiplied with the population shares emerges.

The effect of a single age is not very large, reaching from 0.04% to -0.06%. For instance, if the population share of people aged 0 years increases by 1 percent, then inflation rises by 0.035%. It can be observed that young people until the age of 20 drive up inflation stronger than elderly people. There might by two explanations. First, young people until the age of 20 years have only very little or no labour income whereas the elder only gradually leave the labor market and are thus able to cover their consumption with their own income to a larger degree than young people. Second, young people usually do not have any savings of significant amount. Old people, however, have spent their whole work life accumulating savings. At retirement age, they draw a steady capital income from those financial assets which reduces supply gradually, thus muting the impact of the old on inflation.

BIS does not use single age years but 5-year age cohorts. In order to compare my results with BIS, I aggregate the coefficients for 5-year age cohorts (chart 42).

Chart 42: Effect on inflation by 5-year age intervals in percentage points based on table 13, column 1



The pattern is similar to the one pointed out by the BIS results: young people are inflationary until the age interval of 25 - 29 years. Working age people cause deflation. However, the inflationary old begins at the age interval of 59 - 64 years.

My results suggest, however, that the effect of young and old on inflation is smaller than estimated by BIS. For instance, according to BIS young people increase inflation by up to 0.5%. However, the BIS estimates are based solely on demographic data. Yet, by taking into consideration actual economic behavior as measured by the life cycle deficit I show that young age cohorts drive up inflation only by 0.16%, a third of BIS' results.

As the polynomial method reveals the impact of each population share by age on inflation, it allows to take a closer at the dependency ratios of the IMF method.

The traditional dependency ratio defines young and old people as dependents, working age people as supporters. This way the young and the old are assumed to have an identical impact on inflation in terms of direction and magnitude. While it is true that both young and old increase inflation due to their consumption exceeding their labour supply the magnitude varies with age. On average, young people increase inflation more strongly than old people.

Similarly, the traditional young dependency ratio assumes that all persons belonging to the young cohort have the same effect on inflation. Likewise, the traditional old dependency ratio rests upon the assumption that all old persons affect inflation in the same way. However, this paper has shown that even within the same cohort the specific age is crucial for determining the effect on inflation. Within the young cohort, the effect on inflation is stable between age 0 and 17 years but it declines consequently thereafter. Within the old cohort, the inflationary impact increases from age 59 until 67 years and declines afterwards.

Furthermore, the traditional dependency ratio assumes that the dependents and the supporters have an equal impact on inflation in terms of magnitude just with opposite signs. Yet, also this assumption is false as the deflationary impact of working age persons depends on the specific age. Furthermore, the effect of people aged 35 till 50 years is up to twice as large as of the youngest cohort aged 0 to 17 years. Due to its assumptions, the traditional dependency ratio cloaks the U-shape of the relation between age and inflation.

The life cycle deficit is based on the idea that when consumption is not sufficiently covered by labour income, a person runs a positive deficit. This deficit has to be financed by some other income source such as private or public transfers. This definition makes sense as most people in most countries rely primarily on their labour income in order to survive. Yet, during their working age people have a negative deficit, i.e. their labour income exceeds their income. If this excess is not used for sponsoring private or public transfers, savings can be accumulated. These savings will generate financial income in the future. Hence the definition of the life cycle deficit could be augmented by mirroring consumption not just with labour income but also with financial income. The augmented definition of the deficit would be that the deficit becomes only positive once consumption surpasses the sum of labour income and financial income. The augmentation would make the aggregate supply and aggregate demand framework even more complete as it also takes the capital supply and demand into consideration. The augmentation of the definition also would change the age borders defining the young, working age, and old cohorts. The young cohort is unlikely to be affected as young people usually have no savings which could provide financial income. The border marking the shift from young age to working age would not change. Yet, the second border between working age and old age will be shifted upwards as working age persons begin to use their financial income for consumption purposes, thus delaying the point of time when their deficit turns from negative to positive. The population share of the young would be stable, but it would increase for the working age cohort and decline for the old cohort. Table 14 introduces the same robustness checks as used for the IMF method.

Table 14: Robustness checks for BIS method

Dependent variable: inflation (in %)	(i)	(ii)	(iii)	(iv)
1st polynomial	6.142***	10.433***	10.092***	12.378***
	(1.495)	(1.325)	(1.441)	(1.351)
2nd polynomial (x10^2)	34.581**	90.912***	89.684***	111.457***
	(15.751)	(14.520)	(15.394)	(14.732)
3rd polynomial (x10^3)	157.876	726.323***	716.614***	857.528***
	(157.637)	(151.562)	(156.487)	(152.112)
4th polynomial (x10^4)	569.026	5718.198***	5593.624***	6277.413***
	(1521.804)	(1494.507)	(1521.209)	(1494.225)
Money growth (in %)		0.363***	0.366***	0.364***
		(0.009)	(0.009)	(0.009)
Real GDP growth (in %)		-0.458***	-0.459***	-0.450***
		(0.031)	(0.031)	(0.031)
Population growth (in %)	-0.429	-0.328	-0.309	-0.269
	(0.265)	(0.262)	(0.263)	(0.262)
2nd lag money growth (in %)	0.141*** (0.001)			
3rd lag money growth (in %)	0.095*** (0.010)			
4th lag money growth (in %)	0.061*** (0.009)			
2nd lag real GDP growth (in %)	-0.142*** (0.033)			
3rd lag real GDP growth (in %)	-0.080** (0.033)			
4th lag real GDP growth (in %)	-0.048 (0.031)			
Oil price shock 1974		5.924***		

(dummy variable)		(1.352)		
Oil price shock 1980 (dummy variable)		5.464*** (1.266)		
Gold price (in US-Dollar per ounce)			-0.001 (0.000)	
Real Effective Exchange Rate of USD (index)				-0.150*** (0.021)
Constant	-13.461*** (4.382)	-0.961 (3.614)	1.217 (3.617)	15.521*** (4.139)
Time fixed effects	Yes	No	No	No
Population weighted	Yes	Yes	Yes	Yes
N	4666	5120	5120	5120
N of countries	152	153	153	153
R^2 within	0.267	0.288	0.283	0.290
R^2 between	0.525	0.440	0.435	0.280
R^2 overall	0.336	0.336	0.330	0.274

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Standard errors in parentheses.

In column 1 is not the first lag included but the second, third and fourth lags of money and GDP growth are included in order to rule out endogeneity. All four polynomials remain positive. The first two polynomials are highly statistically significant. The coefficient of the first polynomials remains almost the same, the coefficients of the others polynomials lose in magnitude.

In columns 2 to 4 all four polynomials are highly statistically significant and retain their positive signs. Column 2 includes two dummy variables to address the oil price shocks which occurred in 1974 and 1980, column 3 and 4 control for the collapse of the Bretton Woods system with the gold price and the real effective exchange rate of the US-Dollar, respectively.

2.6. Test for generated regressor problem

I have modified the original BIS approach of a polynomial function. I have regressed the age-specific life cycle deficit on GDP per capita, fertility and life expectancy. This generated life cycle deficit I used as a regressor in the baseline regression with inflation as the dependent variable. Therefore, my baseline regression could suffer from the generated regressor problem. The generated regressor has a variance on its own which has to be taken into account as otherwise the standard errors could be underestimated and the statistical significance of the polynomials could be overestimated.

In order to investigate if the generated regressor problem biases the results, I compute a two-step instrument fixed effects regression. Crucially, in both steps the standard errors are computed by using 100 bootstrap repetitions. That is a standard approach to correct for potential biases due to a generated regressor problem (Pagan, 1984).

In particular, I repeat the baseline scenario in which I regress inflation on the four polynomials of the population shares times deficit fractions plus the three usual control variables. These polynomials, however, are themselves regressed in a first step with a number of instruments: the four polynomials of the population shares, the three variables used before to compute the life cycle deficit fractions via OLS - namely fertility, life expectancy and per capita GDP -, as well as the usual control variables monetary growth, real GDP per capita growth, population growth and time fixed effects.

Table 15 shows that the all four polynomials retain their high statistical significance, their signs continue to be positive and their coefficients vary only slightly in magnitude. The generated regressor problem is ruled out and the previous results are not biased.

Table 15: Second-stage results of two-step instrument fixed effects regression

Dependent variable: inflation (in %)	
1st polynomial	8.350***
	(2.670)
2nd polynomial (x10²)	63.412***
	(26.792)
3rd polynomial ($x10^3$)	531.072***
	(253.228)
4th polynomial ($x10^4$)	4891.296***
	(2464.041)
Money growth (in %)	0.329***
	(0.036)
Real GDP growth (in %)	-0.382***
	(0.0515)
Population growth (in %)	0.0171
	(0.347)
Constant	-24.692***
	(8.682)
Time fixed effects	Yes
Population weighted	Yes
N	5168
N of countries	155
R^2 within	0.363
R^2 between	0.414
R^2 overall	0.374

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Standard errors in parentheses. Standard errors are computed with 100 bootstrap repetitions on both stages. The four polynomials of population shares multiplied with life cycle deficit fraction are instrumented by using four polynomials representing the population shares, life expectancy, fertility, GDP per capita and the three control variables.

2.7. IV regression to test for endogeneity

In the previous robustness checks I used remote lags of money growth and real GDP growth to exclude endogeneity. Another popular approach are instrumental variables. By using instruments I want to show that the demographic variables are exogenous. In order to use them, four requirements have to be fulfilled. First, the instrument and the endogenous independent variable are strongly correlated to ensure the relevance of the instrument. Second, the instrument should affect the dependent variable via the endogenous independent variables only and not via other channels. Third, reversed causality must be excluded, i.e. the dependent variable must not have an impact on the instrument. Fourth, the amount of instruments has to be at least as large as the amount of endogenous independent variables.

I will further use two instrumental variables: the prevalence rate of contraceptives and a dummy variable for armed conflicts.

The first instrument to be used is the prevalence rate of contraceptives. The data source are the UN. The instrument is defined as the self-reported prevalence rate of women or their partner who apply any form of contraception. It is measured for women aged 15 to 49 years and who are married or living in an union. A high rate of contraception lowers the amount of children and is therefore expected to correlate strongly negatively with the population share of the young cohort. In contrast, the instrument should increase the population share of the working age and old cohorts.

Given that contraceptives play only a minor role in any household's consumption basket used to measure the inflation rate and its low price as compared to other consumption products, contraceptives should not have any direct impact on inflation. It also seems implausible that contraceptives would affect inflation via any other channel besides demographics. Furthermore, reversed causality triggered by inflation on contraceptives seems very unlikely.

The second instrument are armed conflicts. It is a dummy variable for military conflicts between countries as well as conflicts within a country where weapons are used, such as in civil wars. Whenever such a conflict happens it is assigned to the country on whose territory the conflict takes place. It does not assign the conflict to all countries involved. The reason is that an armed conflict is very likely to inflict lethal damages to the local population of the country on whose surface the conflict takes place. In contrast, the damage to the populations of other countries involved is relatively small. This is because only the soldiers from the countries involved are primarily in danger, with a much smaller probability for their civil populations. If a war between two countries takes place on both countries' territories, then both countries will have the dummy variable assigned.

The relevance of this instrument is underlined by the fact that 15% of all country-year observations reported an armed conflict.

Armed conflicts are likely to reduce the population size of the targeted country as well as to affect its population structure. However, it is hard to gauge a priori which age cohort will be affected the strongest and thus to estimate the changes in population structure due to armed conflicts. Under war-like circumstances adults are unlikely to bear children, hence the population share of young is set to decline. Furthermore, in particular men aged 20 to 30 years are mostly recruited as soldiers during an armed conflict and thus to perish. In terms of the adjusted young dependency ratio, exactly this age interval defines the moment when a person moves from a positive to a negative life cycle deficit. Therefore, it is difficult to estimate whether the young or rather the working age cohort will be affected by an armed conflict. Moreover, working age people are in a better physical condition than children and old such that they can escape from the conflict region easier.

Besides demographics armed conflicts can influence inflation via changing the demand and supply composition. The supply of agricultural and luxury products might be reduced while the supply of weapon goods increases. Likewise, the demand for luxury and other consumption goods is likely to decline in favor of food staples and war goods. These changes should be reflected in real GDP growth which I include as an exogenous control variable.

Reversed causality from inflation to armed conflict seems rather negligible in my data set. On the one hand, it seems implausible that high inflation rates would cause war between countries. On the other hand, only very high inflation rates (in particular of staple food goods) are likely to cause civil unrest and thus an armed conflict within a country. Yet, I have already excluded such outliers of high inflation rates from the data set.

Table 16: Correlation matrix between instruments and independent variables

	Contraceptive prevalence	Armed conflict
Contraceptive prevalence	1.000	-0.198
Armed conflict	-0.198	1.000
Adjusted young dependency ratio	-0.495	0.187
Adjusted old dependency ratio	0.721	-0.192
1st polynomial	-0.688	0.193
2nd polynomial	0.702	-0.196
Inflation	-0.100	0.121
Money growth	-0.040	0.080
Real GDP growth	-0.048	-0.015

The correlation matrix in table 16 suggests a rather strong correlation between contraceptive prevalence and all demographic variables and with the expected signs. Armed conflict shows a correlation of 20% with the four demographic variables which might indicate a rather weak instrument. Its positive correlation with the adjusted young dependency ratio implies that working-age people die more often due to armed conflict than children, thus driving up the young dependency ratio. In contrast, the correlation is negative for the old dependency ratio which suggests that elder people die more often in armed conflicts compared with working-age people.

I will repeat the Fixed Effects regressions from the previous subchapters. Since I have two instruments I can use only on a maximum of two endogenous variables.. In the case of the IMF approach, I will therefore use the instruments on the adjusted young dependency ratio and the adjusted old dependency ratio. In the case of the BIS approach, I will use the first two polynomials of the life cycle deficit multiplied with population share by age. Throughout the previous results the first and second polynomials have been more statically significant than the third and fourth polynomial.

As control variables I include money growth and real GDP growth and I assume that they are exogenous. I exclude, however, population growth as a control variable. Population growth is another demographic variable and is therefore likely to be affected by the two instruments. Nevertheless, since I only have two instruments I can include only up to two endogenous variables. As population growth is not my main variable of interest but the other demographic variables representing the population structure, I choose to leave it aside. Furthermore, population growth has often not been statistically significant in the regression results so far. The following IV regressions are unweighted and include year fixed effects.

Tables 17 and 18 show the first and the second step results of the IV estimation following the IMF approach, i.e. using dependency ratios.

Table 17: First-stage results of IV fixed effects regression of IMF method

	(i)	(ii)
Endogenous variables	Adjusted Young dependency ratio	Adjusted Old dependency ratio
Contraception	0.007*** (0.001)	0.002*** (0.000)
Armed conflicts	0.111*** (0.053)	-0.004 (0.004)
Money growth (in %)	0.001*** (0.000)	-0.0002** (0.0001)
Real GDP growth (in %)	0.001 (0.001)	-0.0002 (0.0002)
Constant	2.413*** (0.525)	0.206*** (0.009)
Time fixed effects	Yes	Yes
Population weighted	No	No
N	4159	4159
N of countries	130	130
R^2 within	0.2289	0.3473
R^2 between	0.1107	0.5166
R^2 overall	0.0018	0.3581

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Standard errors in parentheses.

As expected the prevalence of contraceptive methods is statistically significant for both adjusted dependency ratios. Armed conflicts, however, are only relevant for the young but not the old dependency ratio. As argued above, armed conflicts could reduce the population share of working age people as they are used as soldiers. At the same time, particularly old people are incapable to flee conflict zones and therefore their population share is also likely to decrease. Both effects might cancel each other out, leading to an insignificant coefficient of armed conflicts for the old dependency ratio. The high F-values indicate that the chosen instruments are not weak.

Table 18: Second-stage results of IV fixed effects regression of IMF method

Dependent variable: inflation (in %)	
Adjusted Young dependency ratio	22.913*** (6.522)
Adjusted Old dependency ratio	-60.584** (26.728)
Money growth (in %)	0.303*** (0.017)
Real GDP growth (in %)	-0.414*** (0.046)
Constant	-41.609*** (12.351)
N	4159
N of counties	130
Time Fixed Effects	Yes
F-Test value	3.95
F-Test probability	0.000
Population weighted	No
R^2 between	0.0146
R^2 overall	0.0378
Wald chi2(49)	3971.15
Prob > chi2	0.000

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Standard errors in parentheses. Both dependency ratios are instrumented on the variables of the first step.

Using the two instrumental variables yields similar results for the IMF method as stated in the previous subchapters. Both dependency ratios are statistically significant. The adjusted young age dependency ratio increases inflation while the old dependency ratio reduces inflation. These findings are in line with the previous results where the young drove up inflation in a statistically significant way while the old had either a negative coefficient or were not significant. Furthermore, both control variables have the expected signs, i.e. money growth enhances inflation while real GDP growth reduces inflation. Now I turn to the BIS method based on polynomials. The baseline specification so far has used four polynomials as this specification yielded the highest adjusted R^2 value (above 90%). Yet, for the sake of comparability I need to modify the baseline specification twofold. Since there are only two instrumental variables available, I repeat the baseline specification with two polynomials only. The corresponding adjusted R^2 yields 70%. Furthermore, I exclude population growth as a control variable just as the following IV estimation does.

Table 19: Baseline specification using first two polynomials of BIS method

Dependent variable: Inflation (in %)	
1st polynomial	3.208*** (1.050)
2nd polynomial (x10²)	7.516*** (2.210)
Money growth (in %)	0.334*** (0.036)
Real GDP growth (in %)	-0.373*** (0.054)
Constant	-19.061*** (6.076)
Time fixed effects	Yes
Population weighted	No
N	5177
N of countries	155
R^2 within	0.3621
R^2 between	0.4232
R^2 overall	0.1287
F(58,152)	20.06
Prob > F	0.000

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Standard errors in parentheses.

Both polynomials in table 19 are highly significant and positive, just as in the baseline specification. Yet, especially the coefficient of the second polynomial has a smaller multitude than before. Both control variables are significant and have the expected signs.

Table 20: First-stage results of IV fixed effects regression of BIS method

	(i)	(ii)
Endogenous variable	1st polynomial	2nd polynomial
Contraception	-0.037***	0.018***
	(0.003)	(0.001)
Armed conflicts	0.082	-0.0291
	(0.0745)	(0.0355)
Money growth (in %)	-0.004***	0.002***
	(0.001)	(0.001)
Real GDP growth (in %)	0.007	-0.004*
	(0.004)	(0.002)
Constant	-3.145***	4.097***
	(0.182)	(0.086)
Time fixed effects	Yes	Yes
Population weighted	No	No
N	4159	4159
N of countries	130	130
R^2 within	0.4093	0.4604
R^2 between	0.5822	0.5663
R^2 overall	0.4336	0.4379

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Standard errors in parentheses.

The results of the first stage of the IV regression in table 20 suggest that the prevalence of contraceptives is highly significant for both polynomials. Armed conflicts, however, are not significant. Nonetheless, the F-test suggests that the instruments as a whole are not weak.

Table 21: Second-stage results of IV fixed effects regression of BIS method

Dependent variable: inflation (in %)	
1st polynomial	119.007*
	(66.677)
2nd polynomial (x10²)	241.875*
	(134.707)
Money growth (in %)	0.273***
	(0.045)
Real GDP growth (in %)	-0.337***
	(0.0957)
Constant	-615.509*
	(342.707)
Time fixed effects	Yes
Population weighted	No
N	4159
N of countries	130
R^2 within	-
R^2 between	0.0185
R^2 overall	0.0057

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Standard errors in parentheses. Both polynomials are instrumented on the variables of the first step.

The second stage results indicate that both polynomials are significant on a 10%-level and they maintain their positive sign (table 21). This quadratic function implies that any age cohort with a positive life cycle deficit increases inflation. The statistical significance of the two polynomials is lower as compared to the previous estimations. However, the previous estimations have used four polynomials which create a better fit to the distribution of the life cycle deficit multiplied with population shares by age than two polynomials do. Yet, it is not possible to include more than the first two polynomials in the IV estimation since there are no more than two instruments available in the data set.

Money growth and real per capital GDP growth have their expected signs. These results confirm that the baseline scenario is not biased by endogeneity. Overall, it seems the prevalence of contraceptives, measuring the share of women using any contraceptive method, is a strong instrument for the chosen demographic variables. The dummy variable armed conflicts, indicating whether a country was involved in an armed conflict at a given year, seems to be a somewhat weaker instrument in particular for the polynomials.

2.8. Assumptions

Both the IMF method and the BIS method rely on a number of theoretical and econometric assumptions. The focus of this paper lies on the impact on inflation. Other macroeconomic variables are of less interest and therefore not explicitly modeled here. Yet, literature has shown that changes in the population shares by age may alter interest rates and productivity (compare chapter 2.2). Furthermore, direct price effects of imported goods and services have not been taken into account yet.

All three channels - interest rates, productivity and import prices - can have a direct impact on inflation. Therefore, I add in table 22 corresponding control variables to the baseline polynomial estimation in order to exclude any omitted-variable bias.

Table 22: Tests of assumptions

Dependent variable: inflation (in %)	(i)	(ii)	(iii)
1st polynomial	3.555	6.099**	7.072**
	(3.004)	(2.833)	(3.134)
2nd polynomial (x10^2)	84.589**	49.558**	54.403**
	(34.228)	(23.875)	(26.267)
3rd polynomial (x10^3)	981.811**	369.814	396.716*
	(382.790)	(228.003)	(244.223)
4th polynomial (x10^4)	9571.327**	2752.329	2956.496
	(3983.972)	(2209.692)	(2310.388)
Deposit interest rate (in %)	-1.003***		
	(0.0405)		
Total factor productivity annual growth (in %)		-0.084***	
		(0.032)	
Import price index annual growth (in %)			0.119**
			(0.055)
Money growth (in %)		0.432***	0.439***
		(0.069)	(0.072)
Real GDP growth (in %)	-0.224***	-0.434***	-0.478***
	(0.081)	(0.106)	(0.109)
Population growth (in %)	-1.358	-1.274	-1.364*
	(0.933)	(0.787)	(0.749)
Constant	-12.684***	-3.374	-7.673
	(12.050)	-(9.865)	(10.474)
Time fixed effects	Yes	Yes	Yes
Population weighted	Yes	Yes	Yes
N	4203	4867	4501
N of countries	167	141	125
R^2 within	0.603	0.426	0.438
R^2 between	0.322	0.728	0.737
R^2 overall	0.421	0.408	0.408

*** = 99% significance-level, ** = 95%-level, * = 90%-level. Standard errors in parentheses.

Column 1 in table 22 replaces the control variable for monetary policy. While the money aggregate was the preferred monetary policy instrument for most central banks until the 1990s or even 2000s, it has lost its practical significance in the more recent time. Central banks have gradually replaced the money growth rate by short term key interest rates. The World Development Indicators database does not provide key interest rates set by central banks. However, the data base does include deposit interest rates which can serve as proxy since they are also short term in their nature and they react strongly to key interest rates.

I replace the annual money growth rate with the deposit interest rate in percent. The deposit interest rate is available for fewer years than the money growth rate which leads to a reduction in observations. As expected the deposit interest rate shows a negative, statistically significant and strong coefficient. The other two control variables retain their respective signs and significances. All polynomials keep their positive signs. While the first polynomial loses its significance, the other polynomials remain highly significant and their coefficients become larger in magnitude. The results are very similar when not the contemporary deposit interest rate is used but its first lag. I conclude that also when changing the control variable for monetary policy, the results remain broadly the same.

Neither supply of nor demand for capital are modeled explicitly in this paper. A positive / negative life cycle deficit should not be mistaken for private savings / debts as the deficit also incorporates private and public transfers. Correspondingly, interest rates are not modeled either. The change of the population structure can affect the supply and demand of capital in manifold ways.

If the population share of the young increases, the supply of capital is reduced as more resources are required for consumption. Given that consumption of the young cohort is financed almost completely by the transfers of the working age and old age cohorts, resources are just shifted across generations but demand for capital remains unchanged.

A rising population share of the working age implies a larger reservoir of savings, hence the capital supply is likely to increase (as long as private and public transfers remain unchanged). Capital demand could also increase as the increase of labour supply requires a larger capital stock to prevent productivity from declining.

A larger population share of the old cohort corresponds to a shrinking capital supply as that cohort liquidates its financial assets in order to finance its consumption. The demand for capital should not be altered as the old cohort consumes at least as much as the working age cohort.

The global decline of fertility fuels the quality-quantity trade off and human capital increases. Physical capital is complementary to human capital, i.e. investments into physical capital should increase when the stock of human capital grows. In the future, the stock of human capital will usually grow thanks to an increase in population size and, even more relevant, due to better education. The older generations are less well educated than the young generations due to the quality-quantity trade-off. As the old generations pass away over time, the average quality of education increases. The younger generations have also spent more years on schooling. Thus, the larger stock of human capital increases the returns on physical capital and therefore investments should rise, too. Finally, the demand for capital will increase.

The higher life expectancy provides a strong incentive to the working age cohort to increase its savings, hence capital supply, in order to generate sufficient capital income to finance its future consumption at pension age. All those channels could have a strong impact on the supply and demand of capital and interest rates accordingly. Introducing the deposit interest rate shall control for those impact.

Demographics may also affect productivity. For instance, the quality-quantity trade off leads to better educated cohorts over time. As the life cycle deficit implies, persons from age 30 until 50 years increase their labour income which should in part reflect an increase in their productivity. The World Penn Tables 9.1 estimate the total factor productivity, i.e. labour and capital productivity together. I compute the annual growth rate of the total factor productivity.

In column 2, the coefficient has the expected negative sign and is highly statistically significant. A higher productivity increases the real aggregate supply at a constant price level. In other words, a higher productivity would lower inflation. All four polynomials retain their positive coefficients but lose somewhat in significance and magnitude.

World Penn Tables 9.1 also provide a price index for imports. I compute its annual growth rate in order to control for the impact of imports on consumption prices. Imports can have a major impact on the price level on the importing country's price level, especially the higher the share of imported goods and services in the consumer price index.

The import prices are statistically significant as shown by column 3. As expected, an increase in import prices raises inflation. The polynomials as well as the other control variables retain their significance and signs to a high degree. Thus, also when explicitly accounting for the open economy the results remain broadly the same.

I want to point out that although the baseline regression does not account for direct price effects due to trade, it does not reflect a fully closed economy either. Rather, the baseline regression assumes a partially closed economy. The medium scenario of the UN World Population Prospects serves as the source for the population data. This source integrates data and simulations of cross-border migration, i.e. an open economy with respect to labour and population movements. Furthermore, labour income is based on observed data implying that income generated by international trade of goods and services is included implicitly as well.

From an econometric point of views, it should be mentioned that slowly moving variables such as demographic trends are considered to have a lack of statistical power. In order to increase the power of my data set, I use a large time horizon of more than 50 years between 1960 and 2015. During this time span, all countries of the data set find themselves in the demographic transition which has changed the population shares dramatically. In the United States the share of 0 to 14 years old children has declined from 30% in 1960 to 20% in 2015. During the same period, the share of the young cohort in China has halved from 40% to 20% while the old cohort aged 65 years and more has increased its share 2.5-fold to 9.6%. In Germany, the average age has increased from 35 years in 1970 to 45 Years in 2015. In other words, the time dimension is large enough to ensure that the demographic changes within countries are remarkable, increasing the statistical power of demographic changes. Secondly, the statistical power of my data set is boosted by using a large cross section of more than 170 countries. The variation is large as the countries find themselves at different stages of the demographic transition at a given year.

Given the long time dimension of about six decades of annual data, I have to ensure stationarity. Out of all unit root tests, I can use only the Fisher-type test as all other tests require a strongly balanced panel data set. Using the Phillips-Perron option confirmed for the dependent variables, for all explaining demographic variables as well as for all control variables that they are stationary on a 99%-significance level. For these tests the autocorrelation parameters were set as panel-specific, panel means were included, time trends and lags not.

Error terms between countries must not be correlated. If they were, not fixed effects but random effects would be an appropriate estimation technique. The Hausman specifies which of both techniques should be used. I have conducted the Hausman test the baselines specifications of the IMF approach and the BIS approach. The tests suggest to reject the random effects in favor of the fixed effects technique. The Hausman test thus confirms indirectly that the error terms between countries are not correlated.

Further requirements for reliable estimates are no autocorrelation and homoscedasticity. the econometric software Stata, which was used for the regressions, offers an option to ensure robust standard errors which are corrected for autocorrelation and heteroskadsticity. To test if both requirements are met, I plotted the residuals against the estimated dependent variable as well as residuals against their own first and second lag. In all plots, the residuals were distributed randomly, no pattern was observed. Similarly, when the residuals were regressed on their own lags, the coefficient was statistically not significant, just as expected. Thus, autocorrelation and heteroskedasticity can be ruled out.

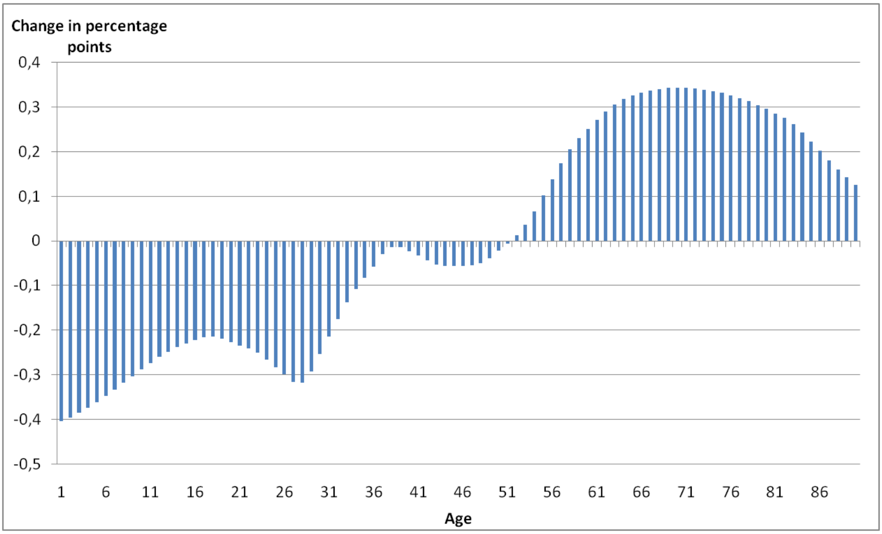
2.9. Forecasts of future inflation

In the past between 1960 and 2015 the global population share of the young cojort has declined. Young persons are inflationary but as their population share has fallen, they have contributed to a lower global inflation. Working age persons are deflationary and as their population share has increased they also kept inflation at low levels. Due to their inflationary impact old persons exercised an up-driving force on inflation on a global scale. Yet, their impact was muted as the population share of the old cohort was mostly small and stable in the past but has risen only in the last few decades.

How will the change of population shares affect the future inflation?

Based on UN population forecasts, chart 43 shows how global population shares by age will change from year 2015 to 2050.

Chart 43: Change of global population shares by age in percentage points between 2015 to 2050. Source: UN, own calculations



The population shares of all ages until 52 years will drop until the year 2050. The shares of persons aged 53 or more will increase until 2050.

Recall that until the age of 26 people increase inflation, yet their population share will drop in the future. Furthermore, also the deflationary persons aged between 27 and 52 years will face smaller population shares. People aged 53 till 59 are deflationary and will gain in population shares. Even older people are inflationary and their population share will also increase.

In order to forecast how inflation will develop over time, I apply the BIS polynomial approach on the future population trends. I use the coefficients of the polynomials as calculated in table 3, column II. I assume that the life cycle deficits in the year 2015 remain constant for all countries until 2050 for two reasons: First, I want to investigate the sole impact of the population structures on inflation and therefore I do not want it to be overlaid by changes of the life cycle deficits. Second, while the UN population database also forecasts future fertility and life expectancy, per capita GDP is not available and thus I am lacking a crucial variable in order to extend the NTA data base into the future. Furthermore, I do not control for country-fixed effects nor do I forecast other control variables such as money growth, real per capita GDP growth or population growth.

Chart 44: Effect on global average inflation by age in percentage points between 2015 and 2050

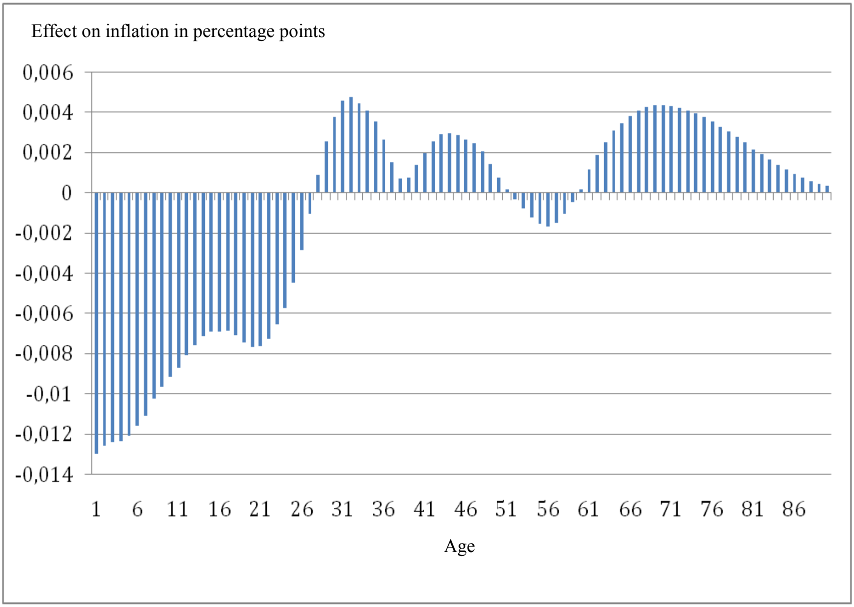


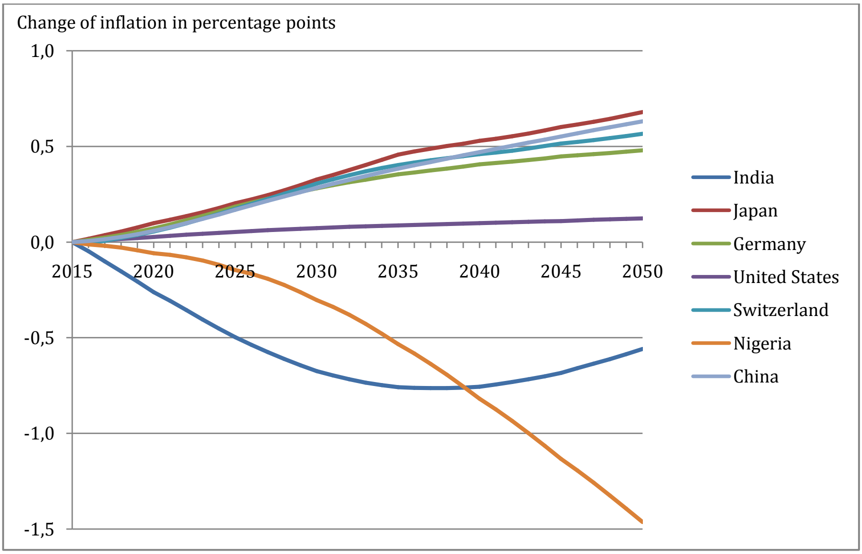
Chart 44 indicates the change of inflation in 2050 as compared to 2015 due to changes in the population shares. The drop of the inflationary young person will have particularly strong effects, it will drag down global average inflation. Yet, as there will be relatively less deflationary people aged 27 till 52 as compared to 2015, inflation will rise. Persons aged 53 till 58 are deflationary and they will gain in population shares. But their impact on inflation is small and so is their relative increase in population. In particular the population shares of inflationary persons aged 59 and more will increase and drive up inflation. In sum, the demographic changes have opposite effects on inflation. Indeed, by summing up the effects over all ages it becomes clear that by 2050 the average global inflation will drop by only 0.1 percentage points as compared to 2015.

As investors and central banks, however, are rather interested into country-specific developments, it is interesting to analyze how inflation will develop in certain countries. I have chosen a number of countries which find themselves currently at different stages of the demographic transition. Their demographic and economic trends vary accordingly.

For each country separately, I replicate the BIS method and regress inflation on the four polynomials with OLS. Just as in the previous estimations I include money growth, real GDP growth and population growth but I do not control for time fixed effects.

In order to forecast inflation in each country separately, I hold the life cycle deficit fractions of the year 2015 constant, and I multiply the changes of the population shares for each age and year between 2015 and 2050. Chart 45 shows for various countries how their inflation rates will develop 2015 due to changes in the population structure relative to 2015. 7

Chart 45: Change of inflation per year until 2050 due to changes in population structure relative to 2015



Japan, China, Switzerland and Germany will see an increase on inflation by between 0.5 and 0.7 percentage points in 2050 as compared to 2015. These countries are defined by a decline of inflationary young, a drop of deflationary working age and a strong increase of inflationary old cohorts. Consumption will grow stronger in those countries than labour supply, leading to an excess demand and an increase in prices. Most of literature expects that Japan, for example, will maintain its deflationary trend due to its changes in its demographic structure. Yet, my results point to the opposite direction with a modest increase in inflation.

In the United States inflation will grow at a much more moderate pace, by a mere 0.12 percentage points until 2050. As compared to 2015, the population share of the working age cohort will fall only little until 2050. In contrast, the drop of the inflationary young cohort share will be absorbed by the increase of the old cohort share, thus both groups' effects on inflation cancel each other out to a large degree.

In contrast, India will benefit from a rising share of the working age cohort. This cohort is defined as persons whose life cycle deficit is negative, i.e. their labour income exceeds their consumption. In the case of India, this is the case for persons aged 26 until 56 years. They exercise a downward pressure on inflation. Similarly, the share of the young aged will continue to drop and thus also lower inflation. The bottom will be reached in year 2036 with inflation being -0.76 percentage points lower as compared to 2015. Until 2036, the share of working age persons will grow. But after this turning point the cohorts share will stagnate and even drop slightly. The share of young will find itself in constant decline while the old persons will capture even larger population shares.

In Nigeria the inflation will take a strong opposite direction as compared to the other countries. The life cycle deficit in Nigeria is positive for young people until the age 28 years. The old age cohort begins at 61 years when the deficit turns from negative to positive again. At any age in between this interval the deficit is positive, defining the working age. For many decades Nigeria found itself at the first stage of the demographic transition. The population share of the young cohort peaked in the 2000s at almost 70%. Having to provide resources to such a large young cohort which consumes more than it earns with its own labour, it cannot surprise that no resources were saved and no meaningful economic progress has been achieved. However, Nigeria has entered by now the second stage of the demographic transition. The share of the young has started to decline and will gain pace in the 2030s. By 2050, its population share will fall to 61%. In parallel, the share of the working age has begun to increase. It will increase from 25% in the 2000s to 33% in 2050. In the same period, the share of the old cohort will grow only from 4% to 6%, it will remain negligible.

The combined deflationary power of a smaller young cohort and a larger working age cohort will cut inflation by 1.5% in 2050 as compared to 2015.

In sum, while the global average effect of demographics on inflation will be small in the future, the picture looks much more heterogonous on the country level. Central banks and investors have to take into account those trends for their monetary policies and their investment strategies.

2.10. Conclusion

I have investigated the impact of young, working age and old people on inflation. I have not only used demographic data but also empirical data about age-specific consumption and labour supply. The difference between consumption and labour supply is the so-called life cycle deficit. Persons with a positive deficit consume more than they earn with their own labour income. Up-scaling from the individual level to an aggregate level, this logic implies that an increase of persons with a positive deficit raises aggregate demand relative to aggregate supply and therefore boosts inflation.

The life cycle deficit varies with age. On global average, young persons until the age of 26 years and old persons aged 59 years or more show a positive deficit. However, persons aged 27 until 58 years have negative deficit, they earn more with their labour income than they consume. The negative deficit of the working age persons finances the positive deficits of the young and old via direct family transfers or indirect public transfers, i.e. taxes and social payments.

Furthermore, this paper has shown that the magnitude of the life cycle deficit varies with age, i.e. that even within a cohort the magnitude of the deficit depends on the specific age. On global average, the impact is strongest for persons aged 0 to 20 years, they increase inflation the most. The effect declines constantly thereafter, turns deflationary at an age of 27 and reaches the bottom at the age of 40 years. Afterwards, the impact tends back towards zero again, turns positive once more and reaches its local climax at an age of 69 years. Eventually, the effect becomes smaller again but stays positive until the age of 90 years.

Based on those insights my papers provides several improvements as compared to literature so far.

First, most literature relied solely on demographic data and has thus to draw assumptions about age-specific economic behavior. In contrast, this paper uses empirical data and estimations about age-specific economic behavior. It has quantified how consumption and labour supply varies with age, over time and between countries.

Second, traditional demographic cohort variables such as dependency ratios define the age cohorts in an arbitrary way. The maximum age border for the young cohort is usually defined as 14 years as the young are assumed to the labour market afterwards. Yet, in fact there is a large heterogeneity within countries as some pupils leave school and enter an apprenticeship while others continue and study until a much later age. Similarly, the age border for the old cohort is often 65 years, irrespective of the broad evidence that many employees enter retirement already at an earlier age. Furthermore, those traditional arbitrary definition impose the same age borders for all countries which is also very unrealistic.

Instead, the definition of my age cohorts rest upon the direction of the life cycle deficit - positive for young and old, negative for working age. On global average, people aged until 26 years belong to the young cohort and anyone above 59 years to the old cohort according to my definition. Those age borders deviate strongly from the traditional definitions. My method also allows for flexible age borders between countries and over time. Thus, my adjusted dependency ratios replicate reality closer than their traditional counterparts.

Moreover, the polynomial method does not even need to define age cohorts as it replicates the whole population distribution. The polynomial approach can draw from all information provided by the age distribution and does not have to make implicit assumptions about direction and magnitude of the effect on inflation of each cohort.

Third, literature assumes a constant magnitude of the effect on inflation within an age cohort. As a result, it is assumed that as soon as a person changes the cohort, e.g. from young to working age or from working age to old, she will also change her behavior drastically. This is a rather strong theoretic assumption. This paper, however, observes a gradual change in behavior, also when a person changes her cohort. As such it is also capable of capturing the heterogeneity within an age cohort better.

Fourth, central banks have a mandate to ensure price stability in the midterm. They compute inflation forecasts to gauge the impact of any potential monetary policy intervention. Those forecast models, however, are incomplete if they do not include the significant demographic trends. By taking into account demographic trends, central banks can improve the quality of their inflation forecasts and increase the effectiveness of their monetary policies.

Fifth, inflation is an important factor for real yields of financial investments. Investors can update their expectations and adjust their investment strategy accordingly. Long-term investors should require a higher premium on bond coupons as a compensation for the expected increase in inflation. They also should shift their portfolio allocation towards more inflation-prone assets such as real estate, equity shares or inflation-linked bonds.

Eventually, this paper strengthens the strand of literature which suggests that young and old cohorts increase inflation instead of reducing it.

Demographic trends develop slowly. This paper has show that demographics contribute in explaining low frequency inflation, such as annual inflation. However, these trends cannot explain inflation movements at higher pace such as monthly frequencies.

The methods presented in this paper, especially the polynomial approach combined with NTA data, can be extended in a couple of ways. On the one hand, the methods are suitable to investigate various macroeconomic dependent variables. In particular, economic growth as well as interest rates could be of interest as they would shed light on the secular stagnation hypothesis. On the other hand, the implicit mechanisms could be modeled explicitly, especially the demographic effects on productivity and financial markets could be researched in depth.

2.11. References

Aksoy, Y. / Basso, H. S. / Grasl, T. / Smith, R.P. (2018): Demographic Structure and Macroeconomic Trends. American Economic Journal: Macroeconomics. Vol. 11. No. 1. pp. 193-222.

Anderson, D. / Botman, D. / Hunt, B. (2014): Is Japan’s Population Aging Deflationary? IMF Working Paper 14/138.

Ando, A. / Modigliani, F. (1963): The ‘life cycle’ hypothesis of saving: aggregate implications and tests. American Economic Review, Vol 53. No. 1. pp. 55-84.

Arnott, R.B. / Chaves, D.B. (2012): Demographic Changes, Financial Markets, and the Economy. Financial Analysts Journal. Vol. 68. No. 1. pp. 23-46.

Bank for International Settlements (BIS), Juselis, M. / Takats, E. (2015): Can demography affect inflation and monetary policy? BIS Working Papers No. 485.

Becker, G., / Lewis, H. (1973). On the Interaction between the Quantity and Quality of Children. Journal of Political Economy. Vol. 81. No. 2. pp. 279-288.

Bloom, D.E. / Canning, D. (2003): Contraception and the Celtic Tiger. The Economic and Social Review. Vol. 34. No.3. pp. 229-247.

Bloom, D.E. / Canning, D. / Sevilla, J. (2003): Economic growth and the demographic transition. NBER Working Paper 8685.

Bloom, D.E. / Williamson, J.G. (1998): Demographic Transitions and Economic Miracles in Emerging Asia. World Bank Economic Review. Vol. 12. No. 3. pp. 419-455.

Börsch-Supan, A. / Härtl, K. / Ludwig, A. (2014): Aging in Europe: Reforms, International Diversification, and Behavioral Reactions. American Economic Review. Vol. 104. No. 5. pp. 224-229.

Bullard, J. / Garriga, C. / Waller, C. (2012): Demographics, Redistribution, and Inflation. Federal Reserve Bank of St. Louis Review. Vol 94. No. 6. pp. 419-439.

Carroll, C.D. / Summers, L.H (1991): Consumption growth parallels income growth: some new evidence. National saving and economic performance. B. Douglas Bernheim and John B. Shoven. Chicago University Press. pp. 305–343.

Cervellati, M. / Sunde, U. (2011): Life expectancy and economic growth: the role of the demographic transition. Journal of Economic Growth. Vol. 16. No. 2. pp. 99-133.

Eichengreen, B. (2015): Secular stagnation: the long view. The American Economic Review. Vol. 10. No. 5. pp. 66-70.

Feyrer, J. (2007): Demographics and Productivity. The Review of Economics and Statistics. Vol. 89. No. 1. pp. 100-109.

Gajewski, P. (2015): Is ageing deflationary? Some evidence from OECD countries. Applied Economics Letters. Vol. 22. No. 11. pp. 916-919.

Ikeda, D. / Saito, M. (2012): The Effects of Demographic Changes on the Real Interest Rate in Japan. Bank of Japan Working Paper Series. No. 12-E-3.

International Monetary Fund (IMF), Jong-Won Yoon, J.-W. / Kim, J. / Lee, J. (2014): Impact of Demographic Changes on Inflation and the Macroeconomy. IMF Working Paper 14/210.

Issa, H. (2005). "Human capital, Demographic Transition and Economic Growth.Journal of Economic Development. Vol. 30. No. 2. pp. 49-65.

Jones, B. F. (2010): Age and Great Invention. Review of Economics and Statistics. Vol. 92. No. 1. pp. 1-14.

Konishi, H. / Ueda, K. (2013): Aging and Deflation from a Fiscal perspective. IMES Discussion Paper Series. No. 2013-E-13.

Lee, R. / Mason, A. (2010): Fertility, Human Capital, and Economic Growth over the Demographic Transition. European Journal of Population. Vol. 26. No. 2. pp. 159-182.

Lee, S.-H. / Mason, A. (2007): Who Gains from the Demographic Dividend? Forecasting Income by Age. International Journal of Forecasting. Vol. 23. No. 4. pp. 603–619.

Lindh, T. (2004): Medium-term forecasts of potential gdp and inflation using age structure information. Journal of Forecasting. Vol. 23. No. 1. pp. 19-49.

Lindh, T. / Malmberg, B. (1998): Age Structure and Inflation—A Wicksellian Interpretation of the OECD Data. Journal of Economic Behavior and Organization. Vol. 36. No. 1. pp. 19-37.

Lindh, T. / Malmberg, B. (2000): Can Age Structure Forecast Inflation Trends?. Journal of Economics and Business, Vol. 52. No. 1-2. pp. 31-49.

Liu, Y. / Westelius, N. (2016): The Impact of Demographics on Productivity and Inflation in Japan. IMF Working Paper 16/237.

Nickel, C. / Bobeicay, E. / Lis, E. / Sun, Y. (2017): Demographics and inflation. ECB Working Paper 2006.

NTA Manual (2013): National Transfer Accounts Manual: Measuring and Analysing the Generational Economy. United Nations Population Division, Department of Economic and Social Affairs. New York.

Pagan, A. (1984): Econometric Issues in the Analysis of Regressions with Generated Regressors. International Economic Review, Vol. 25, No. 1, pp. 221-247.

Prskawetz, A. / Steinmann, G / Feichtinger, G. (2000): Human capital, technological progress and the demographic transition. Mathematical Population Studies. Vol. 7. No. 4. pp. 343-363.

2.12. Appendix

Introduction to the construction of NTA data

The National Transfer Account project (NTA) combines various data sources. It begins with aggregated economic data on the country level and breaks it down to the individual person level by age with the help of microeconomic and population data.

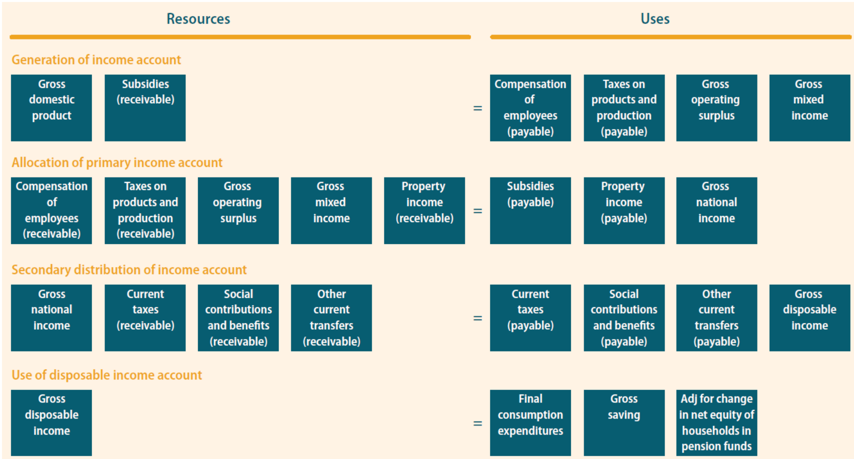
I am focusing on labour income and consumption and will shortly explain how both NTA variables are derived. The NTA manual[3] contains more details, also for constructing all other NTA variables.

NTA data is based in first line on the System of National Accounts (SNA). The SNA method is used primarily to construct the Gross Domestic Product which itself builds upon different accounts. The SNA is an internationally established method applied by the United Nations, OECD, national statistic agencies and others.

The first step is to transform SNA accounts into NTA accounts. They are similar yet not identical since they were constructed for different purposes: while SNA wants to aggregate economic data on country level, the NTA aims on attributing economic data to the average individual person by age. The following schematic flow describes how SNA accounts based on resources are used for various purposes:

[3]The NTA manual can be downloaded here: http://www.ntaccounts.org/web/nta/show/Methodology

Chart 46: Schematic flows in the SNA. Source: NTA Manual



Primary income in NSA is allocated to five resource accounts: compensation of employees, taxes & subsidies on products, operating surplus, gross mixed income and property income.

In contrast, NTA primary income consists only of three accounts: labour income, capital income and property income. To transform primary income from NSA to NTA definition, two adjustments are needed.

First, gross mixed income in SNA merges returns to labour and to capital. Based on empirical estimations, in NTA two thirds of the Gross mixed income are assigned to labour income and is labeled self-employment labour income, and one third to capital income. Second, labour income and capital income are adjusted for taxes and subsidies on products.

Eventually, labour income aggregates on the one hand compensation of employees which in turn consists of wages and salaries and social contributions paid by the employer. On the other hand, self-employment labour income is included too, i.e. two thirds of gross mixed income. Labour income is adjusted for taxes and subsidies, such that labour income is stated tax free.

Consumption in NTA is almost the same as in SNA, as defined in the disposable income account. NTA, however, distinguishes private and public consumption. Private consumption aggregates values from corporations and households, public consumption refers to government expenditures and savings. Consumption is adjusted for taxes and subsidies. When necessary, some expenditures are reclassified between public and private consumption.

The described methods allow to calculate NTA accounts for labour income and consumption on an aggregated macro level. The next step is to break down the data to the average individual person by every single age year. This requires data sources on the micro-level. These are in general household consumption surveys, labour market surveys, government reports, tax data and administrative records. The specific source varies from country to country, often different sources are combined.

Surveys provide detailed socio-economic data of individuals or households on private consumption. However, consumption is not identical with expenditures, especially in the sense of investments that are excluded such as the purchase of equity stocks or bonds or of durable goods like cars. Instead, consumption means particularly private expenditures on education, health and non-durable goods. Expenditures on education include for instance tuition feeds and schooling books, health expenditures refer to costs such as health care insurance premium. Private education and health consumption expenditures are often not reported by person but by household. Based on the self-reports of the surveyed persons the data can distinguish if a person is enrolled at a schooling program and if a person is being treated medically. Using a linear regression and distinguishing the enrolled in schooling or treated medically, the household education and health expenditures are assigned to the respective person and therefore age.

Furthermore, surveys also report different income sources, mainly labour earnings, self-employer income, capital income and social transfers. Since surveys usually ask for the age of head and all members of the households, private consumption and income data is available by age.

Surveys are complemented by government reports and administrative accounts. They help to estimate public education and health consumption. Government and administrative data usually distinguish costs and enrollment rates by age for primary, secondary and tertiary education levels. Thus, one can calculate the public costs of education for the average person by age. Similarly, NTA data on public health expenditures rests upon government reports for health care spending by age.

Based on the micro-level data one obtains NTA data on the individual person level by age. Multiplying this data with population size of each age should yield the macroeconomic NTA accounts derived from NSA accounts. Population statistics are obtained from the UN population division or national statistics agencies.

However, most likely micro data and macro data will not fit neatly, for instance due to reporting and measurement errors in the surveys. Therefore, the micro data is rescaled such that it matches the macro NTA accounts.

3. Boundaries of income inequality

On average, income inequality declines with higher GDP per capita. Yet, with higher income, the maximum cross-country observed income inequality declines while minimum cross-country observed inequality increases. Using a System GMM model, I show that four drivers are creating this narrowing corridor: real GDP per capita growth, schooling, economic complexity and training of employees for economic complexity. As dependent variables I use net inequality and redistribution, each measured by the Gini index as well as households' shares of total income.

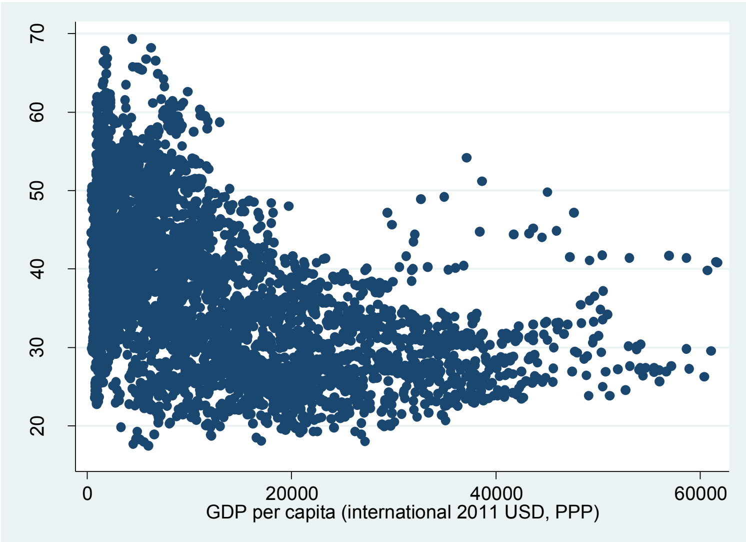
JEL classification: I24, O15

3.1. Introduction

Thomas Piketty's book *The Capital* has relaunched the public debate how much income inequality is desirable. Besides other aspects such as social justice, the debate has focused mainly on the relation between income inequality and the level of income or income growth.

Inequality is usually measured by the Gini coefficient which ranges from 0 points (absolute equality) to 100 points (absolute inequality). Chart 47 refers to net inequality, that is, income after redistribution via taxes and social transfers.

Chart 47: The distribution of net income (after taxes and transfers) and income level for 170 countries between 1960 and 2013. Sources: SWIID, Penntables



On the horizontal axis, income is measured as real GDP per capita in terms of international US-Dollars. Each point represents a country-year observation. To be precise, the data set covers roughly 170 countries over the past 50 years. There are more than 4600 country-year pairs in Chart 47.

I observe a large range between minimum and maximum inequality for low income levels, the net Gini spreads from 20 to 70 points. With rising income, however, the range between observed minimum and maximum inequality narrows down. *Maximum* inequality declines and stabilizes around 40 Gini points at an income of 20.000 USD and more. The only exceptions are Hong Kong and Singapore which both have a higher income and whose Gini values are somewhat above 40 points but which tend towards this threshold nonetheless. On the other hand, the minimum level of inequality rises slightly from 15 to 25 points.

In sum, inequality follows a narrowing corridor as income increases. I refer to the corridor borders as the boundaries of inequality. The upper boundary, i.e. maximum inequality, declines from 70 to 40 Gini points. The lower boundary, i.e. minimum inequality, increases from 15 to 25 points. The distance between maximum and minimum inequality is largest for low income and narrows for high income.

How does theory explain those observations? According to the Unified Growth Theory (Galor and Moav, 2004), the accumulation of physical capital is the most important driver of economic growth at the stage of Industrialization. Income inequality amplifies this mechanism as it channels resources towards rich households which have a higher marginal propensity to save than poor households. With more savings, more physical capital is accumulated.

Yet, with economic development the importance of human capital rises. Physical capital becomes more complex as machines, production processes and products become more sophisticated. In order to be operated, physical capital requires better educated workers. In other words, physical capital and human capital are complementary (Acemoglu, 1998). Education needs financial investment. Financial constraints are particularly binding in countries with a highly unequal distribution of income since poor households cannot afford schooling. In contrast, there are no binding financial constraints in highly equal countries in terms of income distribution.

In sum, high income inequality is not an obstacle to economic growth at an early stage of development. But at later stages of development high income inequality needs to be reduced as it otherwise prevents an economy to grow further.

From Unified Growth Theory it follows that the complexity of production increases with later stages of development. The increase of complexity in turn has manifold effects on income inequality as suggested by the theory of skill-biased technological change.

The production of complex products is difficult to imitate by other companies and this low degree of competition allows for high profits which translate into higher income along the whole value chain, ranging from all tasks at the bottom end over the middle range to the top end of the income distribution. Furthermore, there is a triple down effect. Employees at the top end of the income distribution prefer leisure time such that they increase their demand for time-saving services which are usually provided by employees at the bottom end (Cozzi and Impullitti, 2017). Hence, the income of the latter increases additionally and overall income inequality shrinks (Autor et al., 2006).

The reduction takes place in countries with equal and in countries with unequal income distributions.

Yet, since complex production requires high levels of human capital, the demand for highly educated employees is strong while the supply is relatively scarce. Therefore, complex production pays a high education premium on income. The education premium increases income inequality since particularly the highly educated employees find themselves at the top end of the income distribution.

At this point it is necessary to distinguish between market inequality and net inequality. Market inequality refers to income before redistribution via taxes and social transfers. Net inequality means income after state redistribution.

Production complexity in interaction with education increases market inequality in unequal and equal countries alike.

However, most states have imposed a progressive income tax system which implies that income at the top end is taxed heavier in order to subsidize income at the bottom end. Thanks to this redistribution, production complexity in interaction with education reduces net inequality in both unequal and equal countries.

Redistribution is not only used to channel income directly from the top end to the bottom end of the income distribution. Redistribution also serves to finance education (i.e. schooling). The state faces a so-called equity-efficiency trade off (Okun, 1975). Redistribution can enhance households to become more productive and therefore stimulate economic development. But too much redistribution can distort incentives for labour market participation and for capital investments and, eventually, hinder economic development.

Redistribution is particularly powerful in countries where income is highly unequally distributed. The state subsidizes schooling and ensures that financial constraints are not binding and education also becomes affordable to poor households. Finally, inequality is reduced both in terms of market inequality and net inequality.

However, in countries with a rather equal distribution financial constraints are not binding. Even poor households are educated such that human capital and physical capital are balanced productively. Yet, with even more redistribution states increase human capital beyond the optimal level. Furthermore, too many resources are channeled away from growth enhancing investments. Consequently, economies with a highly equal net income distribution must allow for more inequality if they want to develop further.

Based on those theories I derive the following five hypotheses:

- Hypothesis 1: Schooling reduces inequality in unequal countries but not equal countries since only in the former financial constraints are binding and prevent poor households from getting better educated.
- Hypothesis 2: Economic complexity reduces income market and net inequality in unequal and equal countries due to spill-over effects from rich to poor households.
- Hypothesis 3: Economic complexity in interaction with schooling increases market inequality in unequal and equal countries as the education premium increases income at the top end of the distribution.
- Hypothesis 4: Economic complexity in interaction with schooling reduces net inequality in unequal and equal countries as progressive tax systems redistribute income from rich to poor households.
- Hypothesis 5: Schooling increases net inequality in equal countries since only there too much redistribution prevents growth-enhancing investments.

I confirm all these hypothesis in this paper empirically using a System GMM. I divide the data set into bottom and top subsamples, these are, the most equal and the most unequal countries, respectively. Furthermore, I distinguish between market income (before redistribution) and net income (after redistribution). In order to add robustness to my results, I use three methods to calculate the bottom and top subsamples. I also use annual and five-year-averages data, and I measure inequality with the Gini index and households' shares of income.

3.2. Literature review

This paper focuses on the relation between human capital, economic complexity and income inequality. Yet, literature is still inconclusive about their relation.

According to the Unified Growth Theory neither high nor low inequality prevents poor countries from growing. However, as countries become richer, too much inequality becomes an obstacle to further economic expansion as poor households cannot afford education. Maximum inequality therefore has to shrink. This can be achieved by three means. The government redistributes income from rich to poor households, banks provide student loans, or by triple down effects, that is rich households consume time-saving services provided by poor households and eventually their income increases.

However, too much equality would imply an excessive redistribution of resources towards poor households in the sense that resources are being channeled away from more efficient, growth enhancing investments. Redistribution implies progressive taxes which put a larger burden on rich households. Those households, however, are mostly highly educated themselves. Too much redistribution corresponds thus to a tax on education which in turn lowers the incentive for schooling. Higher redistribution also implies more social benefits to unemployed people who thus participate less in the labour market. Both effects are negative for income growth (Kierzenkowski and Koske, 2012; Anderson, 2015).

Human capital is an important driver of income. It is also relevant for inequality but its correlation - whether positive or negative - is unknown a priori. On the one hand, if poorer households get access to higher education they will earn higher income and thus human capital reduces inequality. On the other hand, education can also increase inequality in combination with technology: in the past, typical middle income jobs (such as secretaries and accountants) required an average duration of schooling. But since these jobs are repetitive they are nowadays being replaced by computers. The so-called labour market polarization implies that jobs at the middle income level are vanishing. However, the salaries both at top and lower end of the income distribution have grown over the last few decades. Jobs at the top end require a high level of human capital, for instance studies in engineering or IT. Jobs at the bottom end require rather little human capital. Curiously, they also pay higher salaries nowadays than in the past. The reason is that top income earners demand more time-saving services provided by employees at the bottom end (Dabla-Norris et al, 2015; Autor et al., 2006; Kearny, 2006; Goos et al., 2009).

Chambers and Krause (2010) analyze how the relative scarcity of human capital to labour capital affects inequality and growth. In lowly educated countries inequality hurts economic growth whereas this relation cannot be observed in highly educated countries. The reason is that as assumed by Unified Growth Theory human capital is particularly important for growth at high income levels. In such spheres income, however, is high enough to overcome the credit market imperfections, i.e. the inability of poor households to borrow money to invest into their education due to financial constraints. Thus, inequality becomes irrelevant for growth at high income levels.

De Gregorio and Lee (2002) argue that the effect of human capital on income inequality is ambiguous, that is, human capital could potentially raise or lower inequality. They argue for instance that with a larger population share of highly educated labour, the return on education declines and inequality decreases, too.

Morrison and Mutrin (2012) find within countries as well as on a global level an inverted U-shape curve for returns to human capital over time. Whereas in the past the return to human capital has declined there is evidence that the returns have begun to increase in the more recent past.

Castelló-Climent (2010) uses a System GMM model and finds that human capital inequality has usually a negative impact on economic growth, particularly in low income counties. The effect loses power in wealthy countries.

According to Güngör (2006), higher inequality of human capital affects economic growth negatively if poor households cannot effort education since they are unable to use new technologies to produce complex goods. Hence, poor households cannot take advantage of their human capital potential and the economy wastes resources. This effect is accentuated the longer the poor households are financially constrained as less human capital is accumulated over time than it could have been otherwise.

Frank (2009) analysis the relationship between the top income decile, economic growth and years of schooling. He runs a VAR on US states from 1920 until 2000 and shows that higher inequality reduces growth whereas more schooling raises growth. The full impact of schooling needs 15 years to accumulate whereas inequality peaks already after 4 years and declines steadily afterwards. Yet, schooling reduces inequality to a small degree. Eventually schooling does not react to growth nor inequality.

Neves et al. (2016) conduct a meta-analysis on papers published over the last two decades which investigate the effect of inequality on growth. In general, inequality reduces growth more in poor than in rich economies. Yet, this negative effect is muted significantly by the inclusion of regional dummies. In particular, Latin America is characterized by a very high inequality.

Barro (2000) shows that below an average income of 2.070 USD (at 1985 US dollars) inequality reduces per capita GDP growth. At low income levels the credit constraint of poor households mutes growth whereas at high income levels the higher saving propensity of rich households feeds investments and thus growth.

Da Silva (2017) applies a Panel VAR to Brazilian states. He shows that inequality stimulates economic growth in low income countries and that growth reduces inequality. These findings are robust to including a variable for human capital accumulation.

Using an ARDL model, Li et al. (2015) find for a panel of Chinese provinces a positive relation between inequality and growth. While private capital investments raise growth in the long run, public capital investments and human capital are not statistically significant.

Grigoli and Robles (2017) employ a dynamic panel model which suggests that the relation between inequality and growth is non-linear, in fact quadratic or even cubic. They show that low inequality enhances GDP growth but the effect turns negative if net Gini exceeds 27%. The decelerating effect is magnified by financial inclusion but reduced by female labor market participation.

Atems and Jones (2015) apply an Panel VAR to US states to investigate the relation between income inequality and the level of income. They show that higher inequality lowers per capita GDP in the long run and vice versa. Yet, the magnitude of this relation varies over time.

Malinen (2016) shows that the top 1% share of income has a positive impact on the credit to GDP ratio. In other words, inequality may leverage credit cycles which in the end could trigger a financial crisis. The author, however, does not observe the reversed causality.

Bordo and Meissner (2012) investigate if a rise in the top income share leads to a credit boom and thus eventually to a recession. By using a panel of 14 advanced countries from 1920 until 2000 they reject this link.

Woo (2011) provides evidence that an initial high level of inequality increases volatility in fiscal policy, that is, more discretionary and procyclical fiscal spending which eventually affects long term growth negatively.

Kolev and Niehues (2016) investigate the effect of inequality on economic growth. They compare OLS, FE, System GMM and Dynamic GMM models. They conclude that in literature the effect of inequality on growth is rather overestimated. Their results suggest that inequality is negative for growth if GDP per capita is below 9000 USD and positive otherwise. The authors add that the relation is a reversed U-shape. Their results also suggest that redistribution might have a negative impact on growth.

Shin (2012) uses a stochastic optimal growth model to argue that high inequality reduces growth at low income levels but increases it at high income levels. Redistribution via highly progressive tax rates does not reduce but increase inequality at low income levels. Therefore Shin concludes that only low tax rates at low income level can maintain both low inequality and high growth. Furthermore, at high income levels redistribution is able to reduce inequality but only at the expense of growth.

Anderson et al (2017) perform a meta-analysis on the relationship between fiscal redistribution and inequality and the relevant instruments of fiscal policy. They point out that theoretically causality could run into both directions, i.e. that redistribution lowers inequality or that a high level of inequality sparks public outcry and could increase redistribution. They find empirical evidence that causality runs from fiscal redistribution to inequality.

Besides the triangle relation between human capital, income level and inequality, literature has identified other channels which might affect inequality. To those channels belong constraints to physical capital accumulation, socio-political instability, changes in institutions, international trade, migration and incentives to innovation. Consult Neves et al. (2016) as well as Halter et al. (2014) for more details.

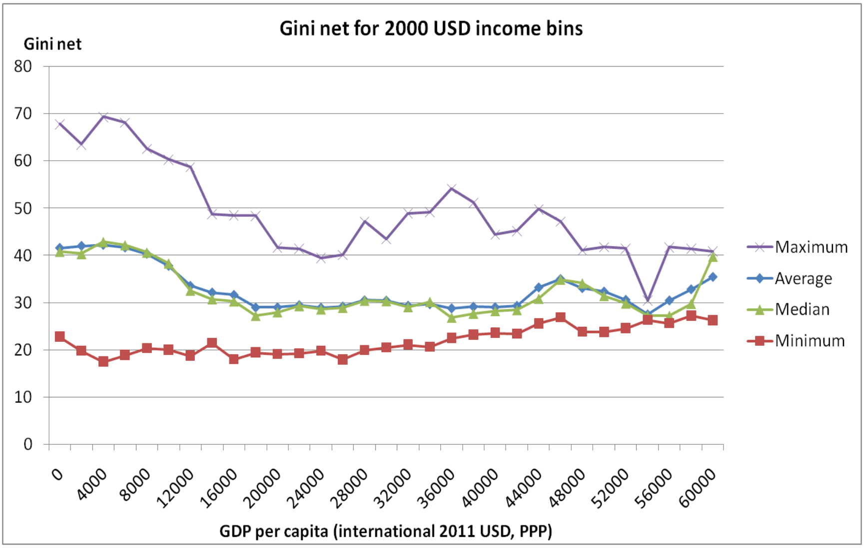
Hartmann et al. (2017) find that the Economic Complexity Index reduces income inequality. The index is based on the number and complexity of products which a country exports. The index reflects available human capital and premium skills.

3.3. Defining boundaries of inequality

As mentioned in the introduction, maximum inequality declines and minimum inequality increases. This is illustrated by chart 48. Based on the same data set as in chart 47, I now put all country-year observations into income bins, each of a width of

2000 USD. I display the maximum and minimum inequality level as well the average and median for each income bin.

Chart 48: Average, median, maximum and minimum net inequality values for each 2000-USD-income bin. Sources: SWIID, Penntables



The pattern of a downsizing inequality corridor is not just a modern age phenomenon. Instead, it has existed for decades. The data set stretches from the year 1960 until 2013. The following panels divide the data set into subsamples of country-year observations by decade, i.e. from 1960 until 1969, 1970 until 1979, etc. The last panel displays the years 2000 until 2013. All panels in chart 49 confirm that the pattern has been stable over time.

Chart 49: Net inequality and GDP per capita by decade. Sources: SWIID, World Bank

A famous representation of the relation between economic growth and income inequality is the Kuznet curve, an inverted U-shape (Kuznet, 1955). In poor countries most labour is working in the agricultural sector leading to a very equal income distribution. With upcoming industrialization and better paid factory jobs these countries become wealthier but also more unequal since workers move from the primary to the secondary sector. At some point inequality will reach its zenith, i.e. the climax of the inverted U-shape curve. As even more people work in factories, income becomes more evenly distributed.

In order to investigate how my database relates to the Kuznet-curve I have selected 20 countries from different continents. Each country provides observations over at least 45 years.

Chart 50: Net inequality and GDP per capita for 20 selected countries. Sources: SWIID, World Bank

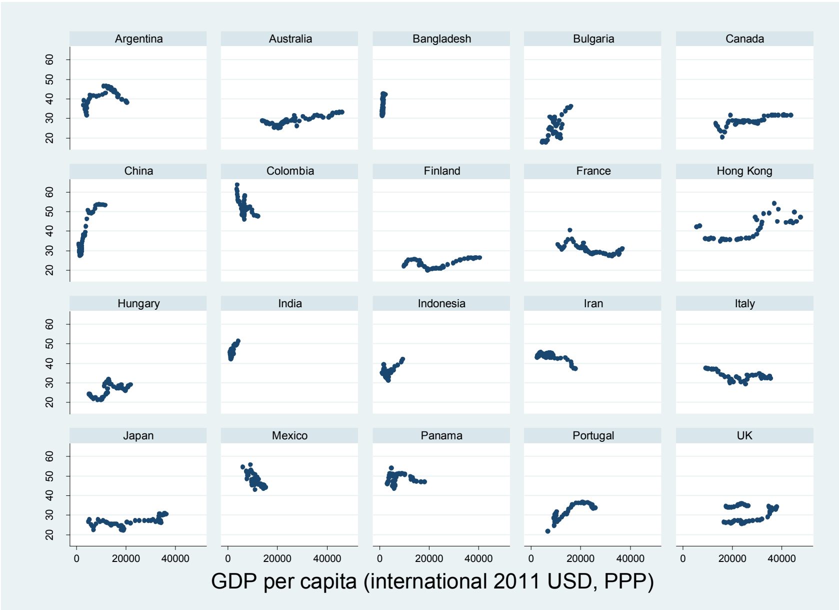


Chart 50 suggests that Kuznets theoretical prediction of a cubic relation between income and inequality can be observed in many countries. However, the pattern also clearly indicates that the exact shape of the Kuznet curve varies.

Firstly, the Kuznet curves of Argentina and China both begin roughly at the same level of income and inequality, i.e. at 30 Gini points. Yet, as income rises the zenith of Argentina's inverse U-shaped curve rests well below 50 Gini points while China's zenith surpasses 50 points.

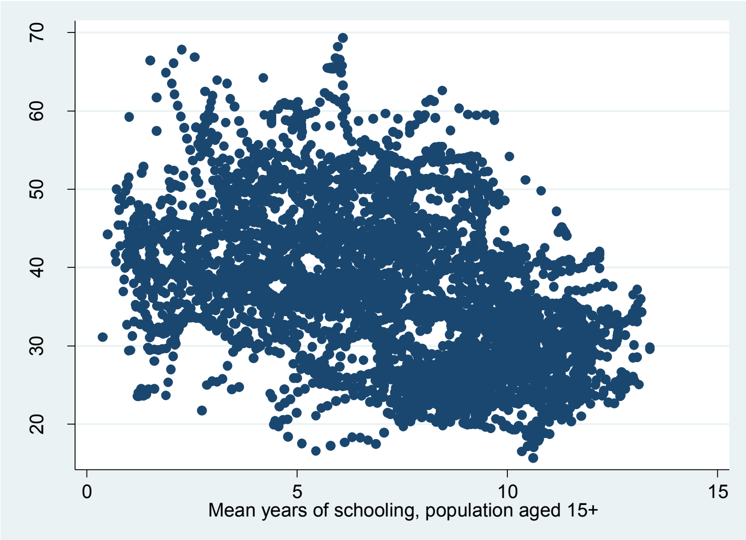
Secondly, the income level at which a country reaches its inequality zenith differs, too. Mexico has reached its zenith at an income of about 10.000 USD per head. In contrast, Canada has reached its maximum inequality at 40.000 USD, and inequality might be still growing.

Thirdly, the steepness of the curve's arches varies between countries. In other words, the distance between the starting and ending point of the curve is different among countries. Portugal has a much more graduate increase of inequality than Bangladesh, for instance.

Given all those differences between the Kuznet curves, one can think of Chart 47 as overlapping Kuznet curves for a high number of countries. With growing income, however, the possible maximum zenith of the Kuznet curve declines. Simultaneously, the minimum level of the curve's right end is increasing.

Out of the various factors which have been identified by literature to correlate with inequality, I focus on human capital. When I display net inequality against human capital - the average years of schooling of the population aged 15 and above, an often used metric for human capital - the data suggests a negative relation (chart 51).

Chart 51: Net inequality and years of schooling. Sources: SWIID, Barro-Lee



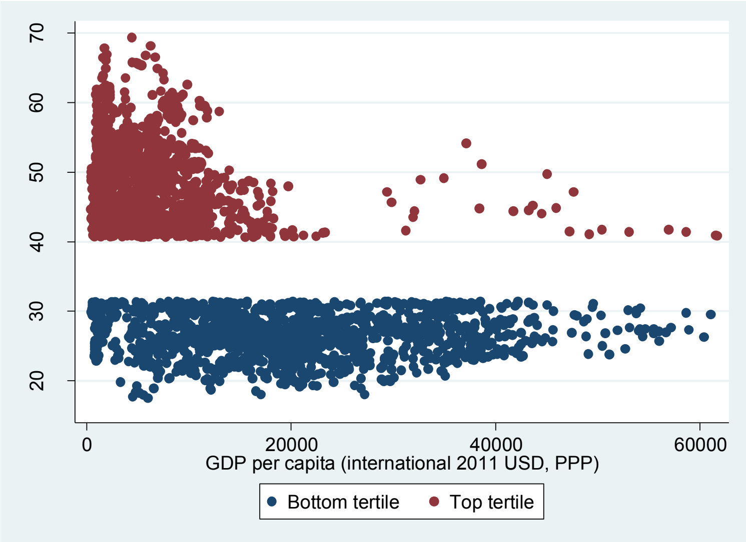
Please recall chart 47 which showed that the maximum observed net inequality declines with income while minimum inequality increases.

I am not interested in the whole data set as such but only in the values closest to the observed maximum and minimum inequality. These observations are the boundaries of inequality. In other words, I do not apply the estimation technique to the whole data sample but just to the subsamples consisting of the observations at the top and bottom of inequality. Nevertheless, I face a trade-off: the smaller each subsample the closer I am to the boundaries. Yet, the subsample must not be too small as otherwise I do not have sufficient observations to draw reliable conclusions from my estimations.

The perhaps most intuitive method is to split the whole sample into equally large subsamples, e.g. to split the whole sample into tertiles. I would apply the estimation technique to the bottom and the top tertile, separately, while I would leave out the observations in the middle tertile.

In case of net inequality, the bottom subsample would consist of country-year observations with 31.4 Gini points or less. The top subsample would only include observations with 40.6 Gini points or more.

Chart 52: Observations of top and bottom tertile of net inequality



While this definition is acceptable for the bottom tertile in the sense that is covers observations from poor to rich countries in a representative manner, the top tertile is underrepresented. The reason is that there are just a few observations left above 40.6 Gini once GDP per capita crosses 20.000 USD. Thus, this top tertile is skewed towards poor countries. Hence, I need better balanced subsamples.

I first divide the whole sample into bins with respect to per capita GDP. Each bin has a width of 2000 USD (first bin: 0 to 2000 USD, second bin: 2001 to 4000 USD, etc.). Secondly, for each bin I define the bottom and top tertile of inequality separately. Thus, there are no fixed overall thresholds (such as 31.4 or 40.6 Gini points as in the previous example). Instead, the thresholds vary from bin to bin. This is illustrated by charts 53 and 54 which show the bottom and top tertile and quartile thresholds of net inequality as well as the corresponding observations.

Chart 53: Thresholds for bottom and top tertiles and quartiles of net inequality for each 2000-USD per capita GDP bin. Source: SWIID

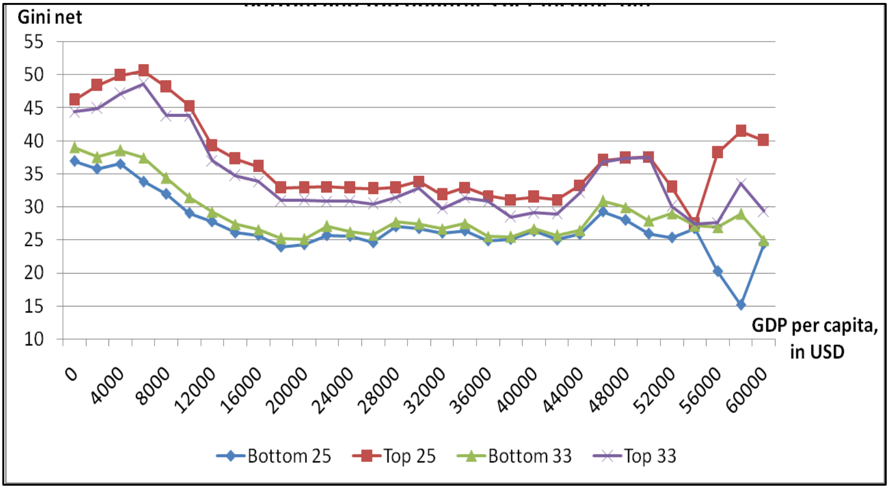
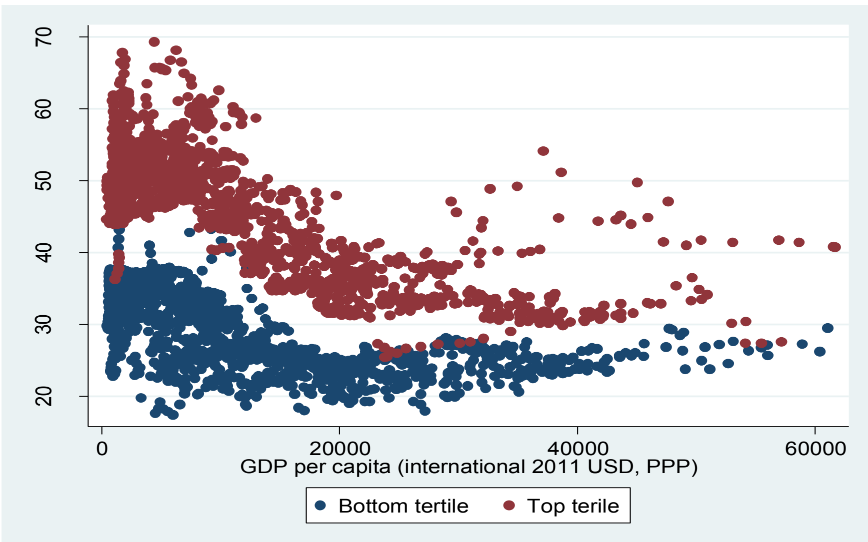


Chart 54: Observations for bottom and top tertiles of net inequality for each 2000-USD per capita GDP bin



The top and bottom tertiles are now more balanced and thus represent also richer countries more adequately.

Regarding market inequality, I define analogously the top and bottom tertiles for each 2000 USD income bin.

The innovation of this paper lies in the way how I split the sample: I do not use the overall sample but extremal data, i.e. the top and bottom tertile subsamples. Literature so far has investigated the average effects of various factors on inequality but it has not distinguished whether those effects are the same in very unequal respectively very equal countries. My paper fills this gap.

This distinction, for instance, I already motivated by the Unified Growth Theory which argues that in very unequal countries poor households are financially constrained from investing into education. This constraint, however, does not hold in very equal countries since there also poor households can afford schooling thanks to income redistribution.

Another motivation for the distinction is the equity-efficiency trade-off: in the very equal countries governments redistribute income from rich to poor households such that the latter is better endowed. However, this redistribution could cause inefficiencies as poor households have a smaller propensity to save, hence less money is used for growth-enhancing investments. In contrast, the equity-efficiency trade-off plays no role in very unequal countries as they barely redistribute little income.

The definition of my threshold does not restrict any country from leaving or entering the sample between income bins or between top and bottom samples over time. For instance, a country in the bottom sample can be part of the first income bin and may or not may not be part of the second income bin. This, however, solely depends upon its development of inequality but not on my definition of the threshold.

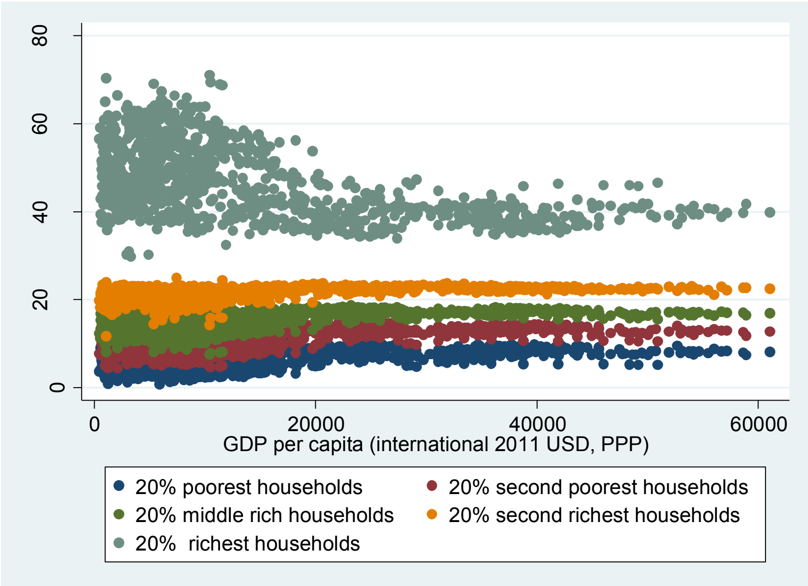
A very equal but poor would at first belong to the bottom sample for an early point in time. But with its income development this country might well become more unequal such that it drops out of the bottom sample for a later year. Indeed, for an even later year this country could become part of the top sample.

The Gini index is by definition fixed between 0 and 100 points. All what I do is to narrow down this range further with thresholds. However, I do not use ambiguous, rigid definitions for the thresholds. Instead, the thresholds vary with income bins which counteracts any rigid definition.

Papers in literature investigating the relation between income (measured by GDP per capita) and inequality have also often split their data set into subsamples. They just did not differentiate with respect to inequality but with respect to GDP per capita, i.e. those papers have split the data set into poor and rich countries instead of into equal and unequal.

The Gini index is the most often used and available metric of inequality. Another popular measurement are household income shares, i.e. how much of total income is owned by the poorest 20% of all households, the second poorest 20%, etc. up to the richest 20% of all households. This measurement, however, is available for less countries and years. Nonetheless it reveals in chart 55 a similar pattern to net inequality, that is, the share of the poorest households in total income increases with GDP per capita while the share of the richest households declines. I will apply my estimation techniques also to those households shares to provide robustness to my results.

Chart 55: Household shares of total income after redistribution based on 1057 country-year observations. Source: World Bank



3.4. Data

The panel data set is unbalanced and taking first differences of GDP per capita and of the Gini index (to ensure stationarity) reduces the amount of observations further. When I consider only countries for which the data set reports values for all variables at a given year, the set amounts to 3067 country-year observations. The cross-sectional dimension covers 105 countries, spread over all continents. The time dimension reaches from the year 1961 to 2013. The amount of total year observations by country are reported in the Appendix.

Table 23: Overview of variables and sources

Variable	Description	Source
Gini net	Measures inequality after state redistribution, includes labor income and capital income and social transfers	SWIID 5.1
Gini market	Measures inequality before state redistribution via taxes and social transfers.	SWIID 5.1
Redistribution	Difference between Gini net and Gini market.	SWIID 5.1
Household shares	Household shares of total income after redistribution.	World Bank
GDP per capita	Real GDP per capita in 2011 international US-Dollar, Purchasing Power Parity. Capped at 65'000 USD to exclude outliers (8 observations of Norway and Switzerland).	World Bank
Mean years of schooling	Mean years of schooling of population aged 15 and older. Available at 5-year intervals. Here interpolated.	Barro-Lee
Economic Complexity Index	Index for the complexity of a country's exported goods. Proxy for labour market polarization.	Hausmann and Hidalgo
Schooling multiplied with Economic Complexity Index	An interaction term of means years of schooling multiplied with the Economic Complexity Index. Proxy for the relation between schooling and labour market polarization.	Barro-Lee, Hausmann and Hidalgo

Gini coefficients are provided by SWIID, the Standardized World Income Inequality Database. The Gini index is available at net and market values, i.e. before and after redistribution via taxes and social transfers, respectively. The difference between net and market Gini is called "redistribution". A positive value indicates that market inequality is larger than net inequality and therefore the government redistributes from richer to poorer households.

GDP per capita is provided by World Bank's World Development Indicators. It is measured as in international 2010 US-Dollar, based on Purchasing Power Parities.

Schooling measures the mean years of schooling of the whole population aged 15 or older and stems from the Barro-Lee data set. Data is available for 5-year intervals only but since the mean years of schooling grows slowly and almost linearly, the data was interpolated.

To capture the labour market polarization I include the Economic Complexity Index (ECI) developed by Hausmann and Hidalgo (2009). The ECI is calculated for each country separately. The ECI is based on the diversity of a country's exports and their ubiquity, i.e. on the amount of countries producing the same good.

Exported products are exposed to international competition by their nature. A country can only maintain a large diversity of different exports if these products are based on special knowledge and know-how such that they are difficult to imitate or compete with. This is particular true for complex and complicated products.

Hence, a country will achieve a high index score if it exports a large diversity of products and if those products are produced by just a few other countries.

The index is based on empirical international trade data which distinguishes more than 1200 products ranging from agricultural goods and crude materials to chemical and manufactured products, e.g. electronic microcircuits, passenger motor cars, wood-based panels, water gas, etc.

Negative values of the ECI imply that a country has either little diversity of exports or that its exports are also manufactured in many other countries. The latter leads to the interpretation that the needed skills to produce those exports are available in many countries, i.e. in countries with low levels of education. In contrast, high values of ECI suggest little ubiquity, i.e. that only few countries are capable of producing such exports, hence they have a well educated workforce.

ECI data is available for all countries in the SWIID data set and from year 1960 until 2013. The index is constructed in such a way, that negative values imply very little export complexity whereas positive values mark high complexity. The ECI ranges from -2.8 points to +2.6 points between countries.

The economic complexity of export goods is shaped by a country's capability to use its resources, institutions, regulations, property rights, labour skills, production technologies, infrastructure and similar. But inequality has no direct causality on ECI, hence the ECI can be regarded as exogenous.

Human capital is an important factor in determining a country's capability to produce complex exports. I would expect a strong positive correlation. However, human capital is just one among other factors needed for the production of exports. Very equal countries face the danger to redistribute too much income to promote human capital accumulation at the expense of the other factors. In the end, large shares of the population would be highly educated but the complexity of exports lags behind what could be expected from such a large stock of human capital. To account for this possibility, I multiply the schooling variable with the ECI.

The interaction term between schooling and ECI can also be interpreted as a proxy for the labour market polarization. Countries with high levels of export product complexity have a high demand for a highly educated labour force. In parallel, those employees seek services to save time which raises demand for lowly educated employees. Thus, demand for middle educated employees is relatively weak. Furthermore, companies producing in countries with high ECI values seek means to replace expensive labour, i.e. they automate routine jobs with computers or outsource those tasks to countries with cheap labour. Countries producing complex goods and services should promote schooling in order to prepare employees for complex job tasks. I expect therefore the interactive term to reduce inequality, irrespective of the already existing level of inequality. In fact, the negative effect should be stronger in highly unequal countries as there poor households have only limited access to education. The descriptive statistics are shown in table 24.

Table 24: Descriptive statistics of full data set. Sources: SWIID, World Bank, Barro Lee, Hausmann and Hidalgo

	Mean	S.D.	Min.	Max.	N
Gini net (after redistribution)	36.76	9.85	15.68	69.35	4626
Gini market (before redistribution)	43.72	8.41	17.34	72.85	4624
Redistribution (Gini market - Gini net)	6.96	6.61	-4.48	26.96	4624
20% poorest households	6.28	2.26	0.80	13.40	1057
20% second poorest households	10.75	2.55	4.30	16.60	1057
20% middle rich households	15.04	2.31	7.50	18.80	1057
20% second richest households	21.42	1.55	11.70	24.90	1057
20% richest households	46.52	8.28	29.70	71.00	1057
GDP per capita (international 2010	11900	11500	383.43	61700	4407

USD, PPP)					
Mean years of schooling (population aged 15+)	7.22	2.98	0.38	13.40	4384
Economic Complexity Index	0.27	0.99	-2.76	2.62	3335
Interactive term between schooling and Economic Complexity Index	3.75	8.00	-14.37	28.95	3286

As mentioned before I am going to split the total into subsamples. The following table 25 shows the means and standard deviations in parentheses of the variables for the top tertile and bottom tertile, each calculated for the income bins.

The bottom subsample characterizes a rather equal income distribution, the top subsample unequal. Little surprisingly the bottom is associated with a stronger redistribution than the top subsample, on average by 3 Gini points more.

Mean real GDP per capita is virtually the same for the top and bottom subsamples. Also the minimum and maximum observed values for both subsamples (not shown here) are very close to each other. The real GDP per capita growth rate is slightly higher in the bottom subsample. This may be due to the higher level of education as mean years of schooling is higher by almost one full year in the bottom than in the top sample.

Furthermore, the ECI is larger in the equal countries indicating that these nations are capable of manufacturing more complex products. One precondition for complex production is the availability of highly skilled labour.

The interactive term between schooling and the ECI is larger in the bottom subsample. This comes almost natural since each, schooling and ECI, are by themselves already larger in the equal countries. Therefore one has to be careful about the interpretation. Yet, the interactive term could suggest that schooling responds more to an increase in ECI as a more complex economic structure implies higher salaries which are only obtainable with high education. The causality might also run the other way round, such that the ECI is more responsive to schooling as a larger pool of highly skilled labour might be a prerequisite for producing complex goods.

Table 25: Mean values and standard deviation in parentheses

	Top 33% in income bins	Bottom 33% in income bins
Redistribution (Gini market - Gini net, in points)	5.425 (5.312)	8.491 (7.691)
GDP per capita (international 2010 USD, PPP)	11953.62 (11559.48)	11490.66 (11517.90)
GDP per capita growth (annual, in percent)	2.521 (6.873)	2.891 (6.778)
Mean years of schooling (population aged 15+)	6.873 (2.869)	7.659 (2.877)
Economic Complexity Index	0.123 (0.821)	0.541 (1.012)
Interactive term between schooling and Economic Complexity Index	2.132 (6.821)	5.894 (8.288)

3.5. Estimation technique

I run the following estimation regression

$$Y_{it} = \beta_o + \beta_1 H_{it} + \beta_2 X_{it} + u_t + e_{it}$$

and I repeat the regression with four different dependent variables Y_{it}

- net inequality measured by the Gini index
- market inequality measured by the Gini index
- each 20% household's share of total income
- redistribution measured as the difference between net and market inequality.

The right hand side of the equation remains the same for each dependent variable.

Human capital is denoted as H_{it}. It is measured as the average of years of schooling for the population aged 15 and older. The vector X_{it} includes per capita GDP, the Economic Complexity Index (ECI) and its interaction with schooling. u_t captures time fixed effects and e_{it} are random errors.

Inequality as well as GDP per capita are highly persistent variables. They change little over time and keep an upward or downward trend over a large time period. Hence, they might suffer from a unit root. Such non-stationary series could bias the estimation results and cause spurious regression. Therefore, I use the first differences of both variables in the estimations, i.e. the annual GDP per capita growth rate in percent and the growth in percentage points of net Gini, which are stationary according to the Fisher-type unit root test (Phillips-Perron specification, no lags, no time trend, panel-specific AR parameters, Panel means included). Other unit root tests are not available as the panel data set is not strongly balanced.

Table 26: Unit root tests

Dependent variable		GDP per capita real annual growth	Gini net, first difference	Gini market, first difference
Inverse chi-squared	P	2306.4288 (0.0000)	2111.1166 (0.0000)	2344.9675 (0.0000)
Inverse normal	Z	-36.3701 (0.0000)	-28.3015 (0.0000)	-31.0894 (0.0000)
Inverse logit	L*	-51.1206 (0.0000)	-44.2909 (0.0000)	-49.9602 (0.0000)
Modified inv. chi-squared	Pm	80.1785 (0.0000)	71.1017 (0.0000)	80.3745 (0.0000)
Number of panels		159	164	164
Average number of periods		25.48	25.84	26.05

p-values in parentheses. Null hypothesis: panels contain unit root

I do not add no further control variables as they would broaden the gaps in the unbalanced panel, thus reducing the observations and the amount of groups even further. Additional control variables would also drive up the instrument count and therefore overfit the model. Eventually, as any other control variable should have an impact on a country's exports complexity, the ECI reflects other control variables.

Traditional Fixed effects and Random effects estimation techniques are not suited for the dynamic panel model, nor to deal with endogeneity or persistent dependent variables (Neves et al., 2016).

Castelló-Climent (2010) argues that the cross-sectional dimension contains a lot of information particularly if variables are highly persistent such as in the case of inequality. First differences would eliminate this information. Hence the Difference GMM method is not appropriate, as in contrast to the System GMM. In line with this argument, Kolev and Niehues (2016) point out that lagged levels might be relatively weak instruments. Thus, one would not prefer the Dynamic GMM but the System since the latter uses additionally lagged differences of variables as instruments.

I estimate the regression equation with System GMM which uses the independent variables as instruments in levels and differences. Literature suggests that schooling, GDP per capital as well as the interaction term between schooling and ECI might be endogenously related to inequality. ECI and time fixed effects, however, are treated as exogenous.

Most papers in the literature average GDP growth rates and inequality variables over periods of 3 to 5 years to rule out business cycles. This technique, however, does not guarantee that the business cycles are cut off properly since their length differs over time and countries. Another disadvantage is the elimination of observations. Thirdly, as Herzer and Vollmer (2011) argue, averaging observations can create spurious correlation.

According to Roodman (2009), any endogenous variable should be lagged and the Arellano-Bond test should indicate no autocorrelation from the second lag onwards, if used. The test indicates for my data set that autocorrelation can usually only be excluded for the 4th and 5th lags.

A too large number of instruments could overfit the endogenous variables. To avoid this problem I use the "collapse"-option of Stata's "xtabond2"- command. This command implies that each instrument used in the System GMM does not have its own column but that the instruments are collapsed to fewer, shared columns. As reported in the following results, the Sargan and Hansen tests suggest that the instruments are exogenous in level and in difference and that they do not overfit the model. Two-step calculation is implemented and standard errors are Windmeijer-corrected cluster–robust errors.

3.6. Results

This subchapter presents the results based on the introduced estimation equation. There are four dependent variables: net inequality and market inequality (for each I distinguish the bottom and top subsamples), households' shares of total income and redistribution. I report the results for each dependent variable in a separate section.

3.6.1. Net inequality, bottom subsample

Net inequality measures the distribution of household income after redistribution via taxes and social transfers. It is measured by the Gini index, 0 points mark absolute equality and 100 points absolute inequality.

This section focuses on net inequality in the bottom sample, that is, on the most equal countries. As a reminder, some poor countries are very egalitarian, as the minimum observed net inequality ranges between 15 and 20 Gini points. With rising income these countries become more unequal.

Table 27 reports four specifications for the bottom subsample. The first three columns use observations in the bottom tertile within each 2000 US-Dollar bin of GDP per capita. Thus, I obtain a more representative subsample as poor and rich countries are balanced adequately. In order to add robustness to my results, column 4 draws from the overall bottom tertile of the whole sample without dividing it into GDP per capita bins.

Table 27: Results for bottom subsample of net inequality

System GMM	(i)	(ii)	(iii)	(iv)
Dependent variable: Change of net inequality	Bottom 33% in income bins	Bottom 33% in income bins	Bottom 33% in income bins	Bottom 33% of whole sample
Real GDP per capita growth (in %)		3.723** (0.453)	2.960 (6.203)	8.408* (4.686)
Schooling (mean of years)	0.030*** (0.007)	0.648** (0.009)	0.505 (0.366)	1.178 (0.750)
Economic Complexity Index (points)			-2.908** (1.036)	1.137 (4.283)
Schooling * ECI			-0.219 (0.146)	-0.358 (0.459)
Constant	0.275 (0.055)	0.627*** (0.032)	27.42*** (2.736)	18.49** (7.133)
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	1315	1281	982	1063
Number of countries	91	90	61	49
Number of instruments	56	59	59	61
F-Value	19.15	64.28	29.71	3.988
F-Test probability	0.000	0.000	0.000	0.007
AR (1)	0.000	0.000	0.005	0.069
AR (2)	0.003	0.002	0.002	0.041
AR (3)	0.000	0.000	0.281	0.017
AR (4)	0.728	0.450	0.160	0.407
AR (5)	0.274	0.253	0.120	0.142
Hansen test of over-identifying restrictions	0.313	0.591	0.444	0.894
Hansen test excluding group	0.285	0.586	0.589	0.846
Difference (null H = exogenous)	0.678	0.399	0.058	1.000
Hansen test excluding group	0.300	0.567	0.660	0.843
Difference (null H =	0.430	0.509	0.660	0.777

exogenous)				
Hansen test excluding group	0.746	0.267	0.735	0.481
Difference (null H = exogenous)	0.265	0.641	0.391	0.887

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors in parentheses.

The first column includes schooling as the only independent variable. In contrast to most of the literature, schooling is here associated with a positive coefficient and not negative. That means, an increase in education goes hand in hand with a rise in inequality. In all the other regression specifications the coefficient remains positive but suffers in statistical significance. The reason for the difference is that most of literature takes average effects into consideration, whereas I focus in table 28 only on the bottom subsample, i.e. countries with a relatively equal net income distribution. In this bottom subsample even poor households can afford schooling, they are not financially constrained as the government redistributes income to them. Redistribution is financed by households with high income and therefore with many years of education. Households are only willing to obtain an additional year of schooling, however, if the government allows for more inequality as otherwise the tax burden on higher income is too high and the financial incentive for additional schooling is too small.

In the most equal countries, GDP per capita growth increases net inequality as according to all regression specifications. In this subsample I am looking on the most equal countries, for any level of GDP per capita. Such low levels of inequality usually require a strong redistribution which in turn relies on a progressive tax system on labour income or on corporate profits which distorts investments needed for growth. Those countries need to cut back redistribution and to allow for more inequality in order to generate growth.

Both economic complexity (ECI) and the interaction term between schooling and ECI show negative coefficients, i.e. they reduce net inequality. It seems that a more complex economy implies a broader variety of jobs such that also previously unemployed households can find jobs. Perhaps more plausible, however, is that ECI has a triple down effect: the most educated employees earn high salaries in complex economies and they demand time-saving services provided by poor households which thus increase their income. With a more complex ECI poor and middle rich households have two strong incentives for schooling, i.e. as they want to reach higher income spheres or since they are afraid to lose their jobs due to globalization and automatisation. The latter is particularly true for middle income employees. This argumentation is underlined by the negative coefficient of the interaction term between schooling and ECI which suggests that middle income households obtain targeted schooling to cope with a more complex economy and therefore more complex job tasks. Thus, the interaction term can be thought of as targeted training enabling employees to produce more complex production. At a given level of economic complexity an increase in years of schooling usually refers to tertiary education or specialized job training. As poor and middle rich households increase their income, net inequality drops.

In sum, as the minimum observed inequality rises with the level of income, the up-driving forces seem to be GDP per capita growth and general schooling. In other words, too much redistribution and too much equality hinder countries from reaching higher income levels. In contrast, the economic complexity and the targeted training for more complex job tasks relieve the pressure for redistribution as poor and middle rich households are able to earn themselves higher income.

The Arellano-Bond tests confirm for all regression specifications that the first three lags usually suffer from autocorrelation and therefore should not be used as instruments. In contrast, the tests reject the null hypothesis of no autocorrelation for the fourth and fifth lags.

The instrument count is smaller than the amount of countries especially in column (3) and is therefore the most reliable specification not suffering from over-identification. Yet, the Hansen tests confirm for all specifications that the null hypothesis that the chosen overall and subsets of instruments are exogenous cannot be rejected which hints at a correct specification of the regressions.

3.6.2. Net inequality, top subsample

The previous section was focusing on the bottom subsample, hence the most equal countries worldwide. This section now looks at the most unequal countries in terms of income distribution. Once more, net inequality, i.e. inequality after redistribution, is the dependent variable. I apply the estimation equation to four specifications for the top tertile of GDP per capita bins and the top tertile of the whole sample.

Table 28: Results for top subsample of net inequality

System GMM	(i)	(ii)	(iii)	(iv)
Dependent variable: Change of net inequality	Top 33% in income bins	Top 33% in income bins	Top 33% in income bins	Top 33% whole sample
Real GDP per capita growth (in %)		-0.620*** (0.122)	-2.725 (4.254)	-2.657 (6.138)
Schooling (mean of years)	-0.014*** (0.001)	-0.888*** (0.003)	-0.962** (0.408)	-0.240 (0.340)
Economic Complexity Index (points)			-1.180 (2.353)	5.961** (2.200)
Schooling * ECI			-0.504* (0.281)	-0.957** (0.289)
Constant	0.0857*** (0.005)	0.066*** (0.016)	53.32*** (3.087)	49.92*** (2.516)
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	1314	1311	1066	992
Number of countries	88	87	60	53
Number of instruments	56	59	58	58
F-Value	93.82	24.36	31.29	4.428
F-Test probability	0.000	0.000	0.000	0.004
AR (1)	0.000	0.000	0.013	0.002
AR (2)	0.040	0.038	0.000	0.001
AR (3)	0.005	0.005	0.329	0.514
AR (4)	0.569	0.580	0.358	0.056
AR (5)	0.433	0.419	0.792	0.254
Hansen test of over-identifying restrictions	0.829	0.808	0.374	0.594
Hansen test excluding group	0.814	0.853	0.308	0.561
Difference (null H = exogenous)	0.513	0.178	0.918	0.565
Hansen test excluding group	0.872	0.886	0.426	0.472

Difference (null H = exogenous)	0.227	0.213	0.273	0.816
Hansen test excluding group	0.082	0.091	0.935	0.703
Difference (null H = exogenous)	0.906	0.915	0.307	0.543

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors in parentheses.

In the most unequal countries seemingly all four independent variables drive down inequality. Starting with per capita GDP growth, the coefficient is negative in all four specifications. The reading could be that higher growth creates especially jobs for poor households and thus lowers inequality.

Schooling is associated with a negative and significant impact in all columns. Industrialization plays a major role for poor countries reaching mid income levels. This development often causes an increase in inequality since, for example, some previous uneducated farmers gain schooling in order to handle more demanding production machines. Once reaching a tipping point, with more schooling for the broad population more people are hired in producing industrialized goods. Thus, the income of more households rises and inequality declines. With more schooling, people also obtain the skills needed for producing complex exports to reach even higher income levels.

A higher ECI reduces inequality in the top sample, though the effect is smaller than in the bottom sample. In contrast, the interaction term between schooling and ECI lowers inequality stronger in the top than in the bottom sample. The top sample presents the most unequal countries, the bottom sample the most equal countries. I argue therefore that in the most unequal countries poor households often have not the necessary training in order to produce complex goods as their financial constraints do not allow them to acquire the necessary skills. A targeted training for complex job tasks is more fruitful in the most unequal countries where poor households are stronger bounded by financial constraints than in equal countries.

As in the previous section, the autocorrelation and instrument tests suggest an adequate specification of the regressions.

3.6.3. Market inequality, bottom subsample

The results so far referred to net inequality, that is, income after redistribution via taxes and social transfers. This section is investigating market inequality, i.e. income before redistribution. It begins with the bottom subsample in table 29.

Table 29: Results for bottom subsample of market inequality

System GMM	(i)	(ii)	(iii)	(iv)
Dependent variable: Change of market inequality	Bottom 33% in income bins	Bottom 33% in income bins	Bottom 33% in income bins	Bottom 33% whole sample
Real GDP per capita growth (in %)		-0.554*** (0.037)	-0.8408*** (0.146)	-1.059*** (0.144)
Schooling (mean of years)	-0.052*** (0.011)	-0.020* (0.012)	-0.046*** (0.004)	-0.057*** (0.005)
Economic Complexity Index (points)			-0.503*** (0.040)	-0.272*** (0.033)
Schooling * ECI			0.061*** (0.005)	0.028*** (0.005)
Constant	-0.351*** (0.083)	0.086 (0.077)	-0.391*** (0.027)	-0.407*** (0.029)
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	1316	912	906	912
Number of countries	99	65	65	65
Number of instruments	56	59	63	63
F-Value	21.87	61.27	74.89	19.52
F-Test probability	0.000	0.000	0.000	0.000
AR (1)	0.000	0.000	0.000	0.000
AR (2)	0.035	0.009	0.007	0.008
AR (3)	0.000	0.000	0.000	0.000
AR (4)	0.010	0.110	0.092	0.095
AR (5)	0.508	0.296	0.260	0.267
Hansen test of over-identifying restrictions	0.286	0.405	0.379	0.366

Hansen test excluding group	0.260	0.336	0.348	0.297
Difference (null H = exogenous)	0.681	0.961	0.531	0.804
Hansen test excluding group	0.314	0.260	0.249	0.250
Difference (null H = exogenous)	0.260	0.928	0.797	0.794
Hansen test excluding group	0.131	0.023	0.084	0.082
Difference (null H = exogenous)	0.351	0.656	0.553	0.541

* $p<0.1$, ** $p<0.05$, *** $p<0.01$. Clustered standard errors in parentheses.

In contrast to the bottom net inequality subsample, GDP growth and schooling reduce market inequality. The coefficients are in all regression specifications negative and statistically significant. This is a strong indication that the most equal countries suffer from too much redistribution which puts a burden on economic growth. Economic complexity reduces market inequality as well as net inequality. This suggests that new jobs are created along the whole value chain and that salaries rise in particular at the bottom end of the income distribution even without state redistribution. Moreover, thanks to their education premium employees at the top end can increase the demand for time-saving services provided by employees at the bottom end, whose salaries rise over proportionally. ECI in interaction with schooling has a positive and significant coefficient in the more reliable specifications, which indicates the existence of an education premium of which particularly employees at the top end of the income distribution benefit.

3.6.4. Market inequality, top subsample

Turning to the most unequal counties in terms of market inequality (table 30), economic growth has the expected negative coefficient. This means that growth benefits in particularly poor and middle class households, for instance by generating particularly jobs for less and middle educated people.

Table 30: Results for top subsample of market inequality

System GMM	(i)	(ii)	(iii)	(iv)
Dependent variable: Change of market inequality	Top 33% in income bins	Top 33% in income bins	Top 33% in income bins	Top 33% whole sample
Real GDP per capita growth (in %)		-1.879 (1.170)	-2.370* (1.375)	-2.703*** (0.200)
Schooling (mean of years)	-0.052* (0.030)	-0.051* (0.027)	-0.057 (0.049)	-0.020 (0.012)
Economic Complexity Index (points)			-1.618*** (0.581)	-1.131*** (0.167)
Schooling * ECI			0.232*** (0.080)	0.162*** (0.023)
Constant	-0.297 (0.233)	-0.266 (0.220)	0.282 (0.276)	0.057 (0.163)
Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	1325	1321	1075	1174
Number of countries	92	91	66	67
Number of instruments	54	57	60	60
F-Value	2.98	3.20	11.79	11.59
F-Test probability	0.088	0.045	0.000	0.000
AR (1)	0.000	0.001	0.000	0.001
AR (2)	0.007	0.005	0.006	0.005
AR (3)	0.002	0.002	0.003	0.003
AR (4)	0.641	0.988	0.767	0.860
AR (5)	0.929	0.739	0.810	0.886
Hansen test of over-identifying restrictions	0.600	0.321	0.480	0.511
Hansen test excluding group	0.562	0.335	0.392	0.485
Difference (null H = exogenous)	0.851	0.285	0.871	0.505
Hansen test excluding group	0.652	0.532	0.324	0.479
Difference (null H = exogenous)	0.315	0.068	0.831	0.512

Hansen test excluding group	0.001	0.001	0.007	0.002
Difference (null H = exogenous)	0.942	0.842	0.870	0.944

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors in parentheses.

While economic complexity reduces inequality, ECI in combination with schooling increases market inequality. Countries with a highly unequal distribution of market income have to lower the inequality in order to be able to reach higher levels of GDP per capita. They should do so by promoting the complexity of its economic structure since it has a strong downsizing effect on market inequality. It is true that the education premium (ECI times schooling) increases market inequality but its coefficient is much smaller than the ECI effect. Overall, the increase of complexity reduces market inequality. Furthermore, recall that most states have imposed a progressive income tax system. Thanks to redistribution the education premium effectively reduces net inequality.

3.7. Robustness checks

This subchapters adds robustness to the results, by using 5-year averages instead of annual data, by using household shares of income as the definition of inequality instead of the Gini index, and by using redistribution as a dependent variable.

3.7.1. Robustness check: 5-year average

The System GMM with several lags on an annual base is a common approach in literature to cope with endogeneity via reversed causality, i.e. to ensure that the impact of economic growth on inequality is investigated and not the other way round. Another common approach is to take 5-year averages of all variables and to regress inequality on a one-period lag of economic growth. Besides reducing the endogeneity bias, gaps in unbalanced data sets are reduced and business cycle effects are averaged out.

Hence, I repeat my previous estimations with 5-year averages (1960-1964, 1655-1969), etc. The last period covers 2010 until 2013. Thus, I end up with 11 time periods. Reducing the amount of time periods drives down the instrument count significantly which increases the effectiveness of the regression. Again, I control for time fixed effects. This approach has the disadvantage of reducing observations severely, from more than 4.600 observations in the overall annual sample to 1187 for the 5-year intervals.

So far I defined the inequality boundaries by using income bins with a width of 2000 USD. Yet, now with the amount of observations being reduced, I need to widen the income bins to 5000 USD. For comparison with my previous results, I use now only one lag in differences and levels as instruments which corresponds to the preceding 5-year interval (table 31).

Table 31: Robustness check of 5-year average for bottom and top subsamples of net inequality

System GMM	(i)	(ii)
Dependent variable: Change of net inequality	Bottom 33% in income bins	Top 33% in income bins
Real GDP per capita growth (in %)	-0.824* (0.436)	-1.005* (0.547)
Schooling (mean of years)	0.410** (0.161)	-0.110* (0.056)
Economic Complexity Index (points)	-3.764** (1.745)	-0.262 (1.115)
Schooling * ECI	-0.574** (0.232)	-0.027 (0.146)
Constant	-1.627 (1.114)	-0.606*** (0.211)
Year Fixed Effects	Yes	Yes
Observations	243	255
Number of countries	51	59
Number of instruments	21	21
F-Value	3.98	3.05
F-Test probability	0.007	0.024
Arellano-Bond test for autocorrelation		
AR (1)	0.128	0.351
Hansen test of over-identifying restrictions	0.903	0.829
Difference-in-Hansen tests of exogeneity of instrument subsets		
GMM instruments for levels		

Hansen test excluding group	0.754	0.381
Difference (null H = exogenous)	0.612	0.438
Endogenous variables		
Hansen test excluding group	0.236	0.349
Difference (null H = exogenous)	0.418	0.542
Exogenous variables		
Hansen test excluding group	0.840	0.716
Difference (null H = exogenous)	0.599	0.628

* $p<0.1$, ** $p<0.05$, *** $p<0.01$. Clustered standard errors in parentheses.

Real GDP per capita growth reduces inequality strongly in unequal countries which corresponds to the previous regressions on an annual base. Yet, here economic growth reduces inequality in equal countries.

Schooling reduces inequality in unequal countries but increases it in equal countries. This finding corresponds to the annual regressions, just as the fact that ECI always lowers inequality. Once again, ECI in interaction with schooling levels income in both unequal and equal countries. I conclude that the previous results largely hold also for the 5-year averages.

3.7.2. Households' shares of total income

The Gini index is the most common metric to measure income inequality in literature. Yet, it suffers from several drawbacks. For instance, the Gini index does not show which households own how much of total income. In order to overcome this shortcoming, I use another metric for inequality. In particular, I use households' shares of total income. This metric shows how much of total income after redistribution is owned by the poorest 20% of all households, the second poorest 20%, etc. up to the richest 20% of all households. This metric allows to analyze whether the independent variables affect poor households differently than rich households. The dependent variable of this section is the household's share of a country's total income. I run the estimation regression for each 20% of the households separately (table 32).

Table 32: Results for households' shares of total income

System GMM	(i)	(ii)	(iii)	(iv)	(v)
Dependent variable: percentage of country's income held by households	20% poorest households	20% second poorest households	20% middle rich households	20% second richest households	20% richest households
Real GDP per capita growth (in %)	3.645** (1.420)	2.969** (1.353)	1.160 (1.517)	-0.109 (0.895)	-7.129 (4.980)
Schooling (mean of years)	0.241 (0.203)	0.420*** (0.105)	0.391*** (0.0839)	0.280*** (0.0813)	-1.272*** (0.297)
Economic Complexity Index (ECI in index points)	-2.470 (1.523)	-3.473** (1.295)	-2.647*** (0.606)	-0.860* (0.497)	9.629*** (2.276)
Schooling * ECI	0.364** (0.134)	0.462*** (0.126)	0.360*** (0.0743)	0.111** (0.0495)	-1.339*** (0.260)
Constant	3.167 (1.962)	6.036*** (1.057)	10.87*** (0.542)	18.93*** (0.703)	60.67*** (2.170)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	827	827	827	827	827
Number of countries	92	92	92	92	92
Number of instruments	46	46	46	46	46
F-Value	16.83	29.96	31.49	13.26	33.55
F-Test probability	0.000	0.000	0.000	0.000	0.000
Arellano-Bond test for autocorrelation					
AR (1)	0.033	0.024	0.005	0.000	0.004
AR (2)	0.371	0.550	0.677	0.141	0.827
AR (3)	0.692	0.754	0.880	0.692	0.991

AR (4)	0.243	0.174	0.664	0.576	0.513
AR (5)	0.081	0.031	0.122	0.388	0.037
Hansen test of over-identifying restrictions	0.330	0.251	0.379	0.474	0.377
Difference-in-Hansen tests of exogeneity of instrument subsets					
GMM instruments for levels					
Hansen test excluding group	0.269	0.268	0.329	0.429	0.313
Difference (null H = exogenous)	0.801	0.267	0.665	0.606	0.790
Endogenous variables					
Hansen test excluding group	0.215	0.282	0.322	0.614	0.301
Difference (null H = exogenous)	0.795	0.292	0.563	0.177	0.626
Exogenous variables					
Hansen test excluding group	0.721	0.588	0.525	0.875	0.683
Difference (null H = exogenous)	0.277	0.217	0.348	0.397	0.325

* p<0.1, ** p<0.05, *** p<0.01. Clustered standard errors in parentheses.

Strong real GDP per capita growth favors the poorest households most, their share of total income increases by 3.6 percentage points if per capita GDP grows by 1% . Also the second poorest 20% of households benefit, their share rises by 2.9%. The three richer household groups are not affected in a statistically significant manner although both richest groups show a negative coefficient.

More years of schooling do not affect the poorest 20% households, the coefficient is positive but not significant. In contrast, the richest 20% households lose 1.3% of their income share with an additional year of schooling. All the remaining 60% households between poorest and richest claim a larger share of total income as schooling rises. It should be noted that the effect is largest for the second poorest 20% households and declines in power thereafter.

The ECI has a negative coefficient for the poorest 80% of all households. A higher ECI implies that a country is more open towards international trade, hence competition. A higher ECI also suggests an increased extent of outsourcing and automation, i.e. that in particular routine middle skilled jobs are being outsourced to cheaper labour countries or being replaced by machines. This is in line with the finding that the negative ECI coefficient is strongest for the second poorest 20% and middle 20% households. The poorest 20% are not affected in a statistically significant way, hence their jobs are not affected by international trade, outsourcing or automation. These jobs are local, non-tradable services. Those results suggest a labour market polarization.

The ECI coefficient is only positive for the richest 20% households. These households are often the owners of companies, either via being entrepreneurs or holding stocks. A high ECI value implies that companies are competitive on a global scale. Those competitive companies earn high profits which benefit their owners directly or indirectly via stock value, i.e. the richest 20% households. An additional interpretation is that the richest households are highly educated. Complex exports pay a premium on education such that the rich households benefit in first line.

The interaction term between schooling and ECI is statistically significant for all household percentiles. All the poorest 80% households increase their total income share if schooling and ECI rise hand in hand. The increase of income of those income groups naturally is at the expense of the richest 20% households. With the economy becoming more complex (as measured with a growing ECI) and their jobs being lost, middle skilled households have two options left: they either accept lower paying jobs or obtain more education. The former option is implied by the non-significant ECI coefficient for the poorest 20% households. The latter option is suggested by the positive coefficient of the interaction term between schooling and ECI. The interaction term can be interpreted as a targeted training in order to enable middle skilled households to cope with more complex jobs, thus also to increase their income share. This targeted training benefits the second poorest, the middle and the second richest 20% households.

Yet, if the ECI increases by one point for any given amount of schooling years, the negative coefficient of the ECI outweighs the positive coefficient of the interaction term, stressing the need for even more schooling if those households want at least to maintain their total income shares. This effect should be particularly strong in wealthy and very equal countries as they already have a high level of schooling. This interplay could explain why the most equal countries become more unequal as they become wealthier.

The estimations are robust. For instance, the number of instruments is just half as big as the amount of countries, hence the estimations are not over-identified. This is confirmed by all Hansen-tests. I have used the fourth and fifth lags of the level and first difference instruments in order to provide comparability to the results in the other sections. I could also have used the second and third lags instead as the Arellano-Bond tests reject any autocorrelation for those lags.

3.7.3. Redistribution

Eventually, I want to analyze how schooling and export complexity affect redistribution of income. Redistribution is calculated as the difference between market and net inequality.

Intuitively, one might think that higher market inequality should be associated with higher redistribution as governments want to prevent social unrest and enable poorer households to go to school. Yet, this intuition holds only partially.

Chart 56: Market inequality and redistribution. Source: SWIID

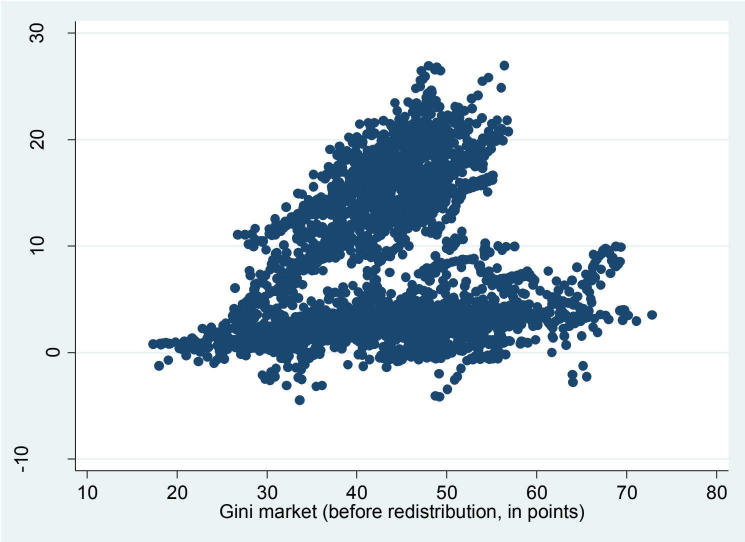
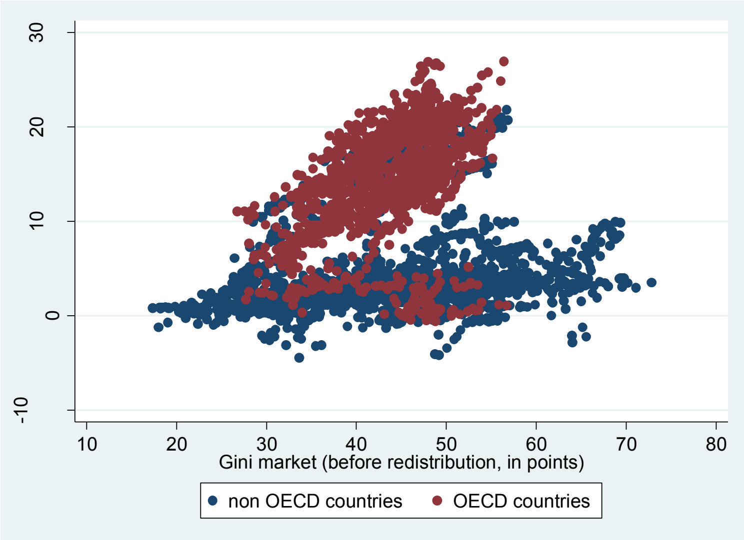


Chart 56 rather suggests that there are two groups worldwide. In the first group, redistribution is rather small with values between 0 to 10 Gini points irrespective of the extent of market inequality. In the second group the positive correlation between market inequality and redistribution is much stronger, it redistributes 10 and 30 Gini points.

Parts of the literature argue that poor countries do not have enough resources for redistribution and hence represent the first group. Rich nations, i.e. OECD countries, have a sufficiently high GDP per capita level such that they can afford redistribution in first place. Secondly, as most OECD countries are democracies they have a lower tolerance for inequality.

Chart 57: Market inequality and redistribution for OECD and non-OECD countries. Sources: SWIID, OECD



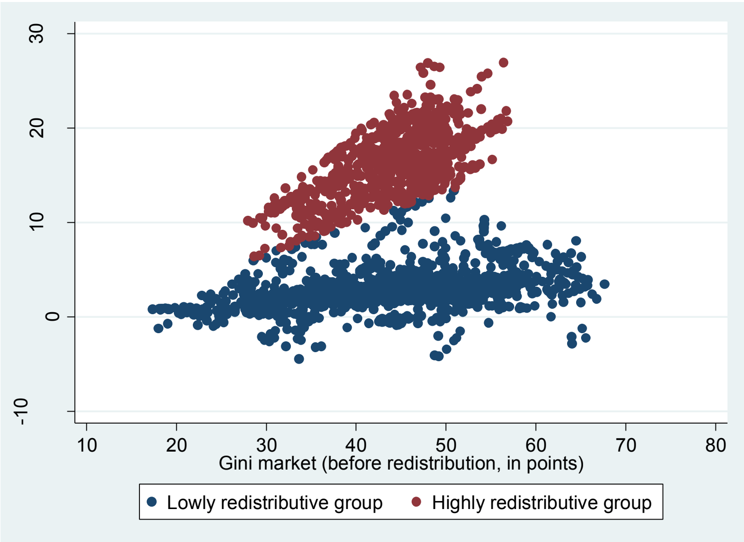
There is a large heterogeneity within both groups in chart 57, however. On the one hand, some non-OECD countries redistribute income as much as the majority of the OECD-countries. On the other hand, some OECD countries barely redistribute income irrespective of the extent of market inequality.

Hence, OECD membership alone does not account for the observation why some countries redistribute heavily or little at a higher level of market inequality. Furthermore, OECD membership is just a mere label but we would still not know the underlying factor actually driving a country's desire for redistribution: it could be any of the treats shared by OECD countries, for instance wealth, democracy, culture etc.

My approach instead is to distinguish both groups by graphical alignment, i.e. I draw a linear function between both groups, and I run the estimation regression on each group separately.

In particular, the linear function takes the form: y = 15/70 * market inequality, with y representing redistribution. Any country-year observation below this linear function is assigned to the lowly redistributive group, and to the highly redistributive group. otherwise (chart 58).

Chart 58: Assignment to lowly or highly redistributive group



Though my definition of the linear function is arbitrary, it captures the optical division which I want to explain between both groups quite precisely. Any more sophisticated approach would be less intuitive and might yet yield the same results.

Using System GMM to estimate the regression equation is not an option as it would produce roughly 60 instruments but the highly redistributive group consists only of 31 countries. This over-identification would bias the results. Hence, I choose to run a Fixed Effects regression on the equation presented in chapter 3.5. Furthermore, I add market inequality as another independent variable and control for fixed country and year effects. The results are reported in table 33.

Table 33: Results for redistribution

Fixed Effects	(i)	(ii)
Dependent variable: redistribution	Lowly redistributive group	Highly redistributive group
Market inequality (in Gini points)	0.083*** (0.007)	0.530*** (0.012)
Real GDP per capita growth (in %)	-0.510 (0.332)	-1.518 (1.051)
Schooling (mean of years)	0.161*** (0.046)	0.085 (0.086)
Economic Complexity Index (points)	-0.318* (0.165)	-2.181*** (0.518)
Schooling * ECI	0.059*** (0.023)	0.265*** (0.051)
N of observations	1920	1166
N of countries	85	31
R^2 within	0.1187	0.7983
R^2 between	0.2439	0.2049
R^2 overall	0.2145	0.4023
Country Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes

* $p<0.1$, ** $p<0.05$, *** $p<0.01$. Clustered standard errors in parentheses.

At this point I would like to remind that the Gini index here reaches from 0 to 100 points. Market inequality is highly statistically significant and positive for both groups. If market inequality increases by one point, then redistribution rises only by 0.08 points. The coefficient of the highly redistributive group, however, shows a magnitude 0.53 points implying that the latter group reacts five time stronger than the former group.

Annual growth of the real GDP per capita does not affect redistribution in any group.

Schooling shows a positive coefficient for both groups. However, it is only significant for the lowly but not for the highly redistributive group.

Exports complexity reduces redistribution though the effect is much stronger and significant in the highly redistributive group.

The interactive term between schooling and ECI reveals a positive and highly significant coefficient. The effect is stronger in the highly redistributive group.

Schooling is positive in both groups, but only significant for the lowly redistributive group. The countries constituting this group have rather a lower average level of schooling which suggests that ECI barely affects middle skilled jobs. This explains also why ECI has a much smaller impact on redistribution and why this group does not emphasize target schooling in order to cope with a more complex economy.

The highly redistributive group redistributes income strongly as a response to rising market inequality. This could reflect the concern that too much inequality constrains poor households financially and thus prevents them for obtaining education. This argument is supported by the result that schooling is not statistically significant by itself but only in interaction with ECI. In other words, an increase in ECI eliminates middle skilled jobs which the countries try to overcome with the targeted schooling to prepare employees for a more complex economy, which requires an increase in redistribution. Once employees are better educated, a higher ECI lowers the need for redistribution.

3.8. Conclusion

While literature investigated factors driving average inequality, I have focused on the observed maximum and minimum inequality. As GDP per capita rises, education is reducing maximum net inequality but also increasing minimum net inequality. I argue that high levels of inequality prevent poor households from investing into education. Without these resources, however, a country is likely not to reach higher income levels. Therefore, a country redistributes income to grant poorer households access to education. Yet, in order to become wealthier, countries have to develop their production capabilities and their goods and services have to become more complex in order to retain their competitiveness. The more households have sufficient skills, the higher their income and the more equal are their countries. Yet, increasing schooling in general does not help households to acquire those particularly needed skills. Rather, the most unequal countries tend to redistribute too much income on general schooling and thus waste resources needed to enhance growth. In contrast, targeted training teaches the skills needed for producing more complex goods and services lowers inequality. Table 34 summarizes the respective signs of the coefficients.

Table 34: Summary of signs of coefficients

Subsample	**Top**		**Bottom**	
Income distribution	**Market**	**Net**	**Market**	**Net**
Real GDP per capita growth	-	-	-	+
Schooling	-	-	-	+
ECI	-	-	-	-
ECI*schooling	+	-	+	-

Each country strives to become wealthy, that is, to achieve a high level of per capita GDP. This paper has shown that too much as well as too little inequality might become an obstacle to this goal.

The most unequal countries of the top subsample should enhance GDP growth rates, schooling and economic complexity in order to reduce market and net inequality. The education premium, represented by the interaction term of ECI and schooling, raises the income of the highly educated employees at the top end of the income distribution. Thus, market inequality grows but the government can counter this increase with a progressive tax system which grants a reduction in net inequality.

At the bottom, i.e. the most equal countries have to allow for a higher level of net inequality as they redistribute too many resources. Physical capital and human capital are complementary. In the most equal countries also poor households can afford education. Governments take too many resources away from growth-enhancing investments in favor of human capital even beyond the optimal level of complementary. Economic complexity tends to reduce market and net inequality in the bottom subsample such that the progressive tax system also takes advantage of the education premium to reduce net inequality.

3.9. References

Acemoglu. D. (1998): Why do new technologies complement skills? Directed technical change and wage inequality. Quarterly Journal of Economics. Vol. 113. No.4. pp. 1055-1089.

Almås, I. / Havnes, T. / Mogstad, M. (2011): Baby booming inequality? Demographic change and earnings inequality in Norway, 1967–2000. The Journal of Economic Inequality. Vol. 9. No. 4. pp. 629–650.

Andersen, T. M. (2015): Human Capital, Inequality and Growth. European Economy Discussion Papers. Vol. 007.2015.

Anderson, E. / D'Orey, M. J. A. / Muvendack, D. / Esposito, L. (2017): Does Government spending affect income inequality? A Meta-Regression analysis. Journal of Economic Surveys. Vol. 31. No. 4. pp. 961–987.

Atems, B. / Jones, J. (2015): Income inequality and economic growth: a panel VAR approach. Journal of Empirical Economics. Vol. 48. No. 4. pp. 1541–1561.

Autor, D. H. / Katz, L. F. / Kearney, M.S. (2006): The Polarization of the U.S. Labor Market. American Economic Review. Vol. 96. No. 2. pp. 189–194.

Baghi, S. / Svejnar, J. (2015): Does wealth inequality matter for growth? The effect of billionaire wealth, income distribution, and poverty. Journal of Comparative Economics. Vol. 43. No. 3. pp. 505–530.

Bircan, C. / Brück, T. / Vothknecht, M. (2010): Violent Conflict and Inequality. IZA Discussion Paper No. 4990.

Bordo, M. D. / Christopher, A. M. (2012): Does inequality lead to a financial crisis? Journal of International Money and Finance. Vol. 31. No. 8. pp. 2147–2161

Castelló-Climent, A. (2010): Inequality and growth in advanced economies: an empirical investigation. The Journal of Economic Inequality. Vol. 8. No. 3. pp. 293-321.

Chamber, D. / Krause, A. (2010): Is the relationship between inequality and growth affected by physical and human capital accumulation? The Journal of Economic Inequality. Vol. 8. No. 2. pp. 153-172.

Cozzi, G. / Impullitti, G. (2016): Globalization and Wage Polarization. The Review of Economics and Statistics. Vol. 98. No. 5. pp. 984-1000.

Dabla-Norris, E. / Kochhar, K. / Ricka, F. / Suphaphiphat, N. / Tsounta, E. (2015): Causes and Consequences of Income Inequality: A global perspective. IMF Staff Discussion Note SDN/15/13.

Da Silva, M. E. A. (2017): Does inequality benefit growth? New evidence using a Panel VAR approach. Working Paper Universidade Federal de Pernambuco, Brazil.

De Gregorio, J. / Lee, J.-H. (2002): Education and income inequality: new evidence from cross-country data. Review of Income and Wealth. Vol. 48. No. 3. pp. 395-416.

De la Croix, D. / Doepke, M. (2003): Inequality and growth: why differential fertility matters. American Economic Review. Vol. 93. No. 4. pp. 1091–1113.

Frank, M.W. (2009): Income Inequality, Human Capital, and Income Growth: Evidence from a State-Level VAR Analysis. Atlantic Economic Journal. Vol. 37. No. 2. pp. 173-185.

Galor, O. / Moav, O. (2004): From Physical to Human Capital Accumulation: Inequality and the Process of Development. Review of Economic Studies. Vol. 71. No. 4. pp. 1001–1126.

Goos, M. / Manning, A. / Salomons, A. (2009): Job Polarization in Europe. American Economic Review. Vol. 99. No. 2. pp. 58–63.

Gungör, N. D. (2010): Education, Human Capital Inequality and Economic Growth: Evidence from Turkey. Regional and Sectoral Economic Studies. Vol. 10. No. 2. pp. 53-71.

Grigoli, F. / Robles, A. (2017): Inequality overhang. IMF. Working Paper 17/76.

Halter, D. / Oechslin, M. / Zweimüller, J. (2014): Inequality and Growth: The Neglected Time Dimension. Journal of Economic Growth, Vol. 19. No. 1. pp. 81–104.

Hartmann, D. / Guevara, M.R. / Jara-Figueroa, C. / Aristaran, M. / Hidalgo, C.A. (2017): Linking Economic Complexity, Institutions, and Income Inequality. World Development. Vol. 93, pp. 75–93.

Hausmann, R. / Hidalgo, C.A. (2009): The building blocks of economic complexity. PNAS. Vol. 106. No. 26. pp. 10570-10575.

Herzer, D. / Vollmer, S. (2012): Inequality and growth: evidence from panel cointegration. The Journal of Economic Inequality. Vol. 10. No. pp. 489–503.

Kierzenkowski, R. /I. Koske (2012): Less Income Inequality and More Growth – Are they Compatible? Part 8. The Drivers of Labour Income Inequality – A Literature Review. OECD Economics Department Working Papers. No. 931.

Kolev, G. / Niehues, J. (2016): The Inequality-Growth Relationship: An Empirical Reassessment. IW-Report 7/2016.

Kuznets, S. (1955): Economic Growth and Income Inequality. The American Economic Review. Vol. 45 . No. 1. pp. 1-28.

Li, T. / Lai, J. T. / Wang, Y. / Zhao, D. (2016): Long-run relationship between inequality and growth in post-reform China: New evidence from dynamic panel model. International Review of Economics and Finance. Vol. 41. pp. 238–252.

Malinen, T. (2016): Does income inequality contribute to credit cycles? The Journal of Economic Inequality. Vol. 14. No.3. pp. 309-325.

Morrisson, C. / Murtin, F. (2013): The Kuznets curve of human capital inequality: 1870–2010. The Journal of Economic Inequality. Vol. 11. No. 3. pp. 283-301.

Neves, P. C. / Afonso, O. / Silva, S. T. (2016): A Meta-Analytic Reassessment of the Effects of Inequality on Growth. World Development Vol. 78. pp. 386–400.

Østby G. (2008): Polarization, Horizontal Inequalities and Violent Civil Conflict. Journal of Peace Research. Vol. 45. No. 2. pp. 143-162.

Okun, A. M. (1975): Equality and efficiency, the big tradeoff. Washington, DC. Brookings Institution.

Ostry, J.D. / Berg, A. / Tsangarides, C.G. (2014): Redistribution, Inequality, and Growth. IMF Staff Discussion Note /14/02.

Perotti, R. (1996): Growth, income distribution, and democracy: What the data say. Journal of Economic Growth. Vol. 1. No. 2. pp. 149–187.

Roodman, D. (2009): How to Do xtabond2: An Introduction to “Difference” and “System” GMM in Stata. The Stata Journal. Vol. 9. No. 1. pp. 86-136.

Shin, I. (2012): Income inequality and economic growth. Economic Modelling. Vol. 29. No. 5. pp. 2049-2057.

Udin, M. A. / Ali, M. H. / Masih, M. (2017): Political stability and growth: An application of dynamic GMM and quantile regression. Economic Modelling 64. No. C. pp. 610-625.

Woo, J. (2011): Growth, income distribution, and fiscal policy volatility. Journal of Development Economics. Vol. 96 No. 2. pp. 289–313.

3.10. Appendix

Chart 59: Frequency of Gini net observations by country. Source: SWIID

Table 35: Total year observations by country for countries with values for all variables in a given year

Country	ID	Year observations in total	Country	ID	Year observations in total
Albania	1	12	Kazakhstan	54	21
Algeria	2	17	Kenya	55	28
Angola	3	11	Latvia	56	22
Argentina	4	40	Lebanon	57	1
Australia	5	49	Lithuania	58	22
Austria	6	38	Madagascar	59	16
Azerbaijan	7	15	Malaysia	60	43
Bangladesh	8	37	Mauritania	61	21
Belarus	9	19	Mexico	62	44
Belgium	10	45	Moldova	63	20
Bolivia	11	28	Mongolia	64	19
Bosnia and Herzegovina	12	14	Morocco	65	29
Botswana	13	6	Mozambique	66	12

Brazil	14	31	Namibia	67	11
Bulgaria	15	43	Netherlands	68	44
Cambodia	16	15	New Zealand	69	50
Cameroon	17	11	Nicaragua	70	16
Canada	18	48	Nigeria	71	9
Chile	19	37	Norway	72	41
China	20	36	Pakistan	73	48
Colombia	21	48	Panama	74	49
Costa Rica	22	27	Paraguay	75	21
Cote d'Ivoire	23	23	Peru	76	29
Croatia	24	21	Philippines	77	43
Czech Republic	25	21	Poland	78	43
Denmark	26	50	Portugal	79	37
Dominican Republic	27	25	Romania	80	23
Ecuador	28	31	Senegal	81	20
Egypt	29	43	Singapore	82	50
El Salvador	30	23	Slovenia	83	22
Estonia	31	21	South Africa	84	38
Ethiopia	32	22	Spain	85	41
Finland	33	50	Sri Lanka	86	34
France	34	49	Sudan	87	1
Georgia	35	20	Sweden	88	49
Germany	36	45	Switzerland	89	22
Ghana	37	19	Syria	90	9
Greece	38	49	Tanzania	91	20
Guatemala	39	24	Thailand	92	31
Guinea	40	1	Togo	93	4
Guinea-Bissau	41	14	Trinidad and Tobago	94	17
Honduras	42	25	Tunisia	95	25
Hong Kong	43	31	Turkey	96	36
Hungary	44	43	Uganda	97	3
India	45	47	Ukraine	98	22
Indonesia	46	44	United	99	48

			Kingdom		
Iran	47	42	United States	100	51
Ireland	48	49	Uruguay	101	21
Israel	49	32	Uzbekistan	102	14
Italy	50	46	Venezuela	103	45
Jamaica	51	23	Zambia	104	19
Japan	52	47	Zimbabwe	105	20
Jordan	53	36			

Table 36: Pair-wise correlation matrix

	Gini net (after redistribution, in points)	Gini market (before redistribution, in points)	Redistribution (Gini market - Gini net)	20% poorest households	20% second poorest households	20% middle rich households	20% second richest households	20% richest households	GDP per capita (international 2010 USD, PPP)	Mean years of schooling, population aged 15+	Economic Complexity Index	Interactive term between schooling and Economic Complexity Index
Gini net (after redistribution, in points)	1											
Gini market (before redistribution, in points)	0.6008	1										
Redistribution (Gini market - Gini net)	-0.6333	0.2382	1									
20% poorest households	-0.9058	-0.6149	0.5052	1								
20% second poorest households	-0.9447	-0.5911	0.5755	0.9654	1							
20% middle rich households	-0.9319	-0.5926	0.5585	0.9056	0.9776	1						
20% second richest households	-0.7773	-0.5154	0.4455	0.6659	0.7949	0.8962	1					
20% richest households	0.9421	0.6104	-0.5536	-0.9468	-0.9915	-0.9928	-0.8604	1				
GDP per capita (international 2010 USD, PPP)	-0.6025	0.0685	0.7984	0.4556	0.5313	0.536	0.4733	-0.5249	1			

Mean years of schooling, population aged 15+	-0.56	-0.1131	0.5709	0.4516	0.5391	0.5573	0.5258	-0.5418	0.6577	1		
Economic Complexity Index	-0.5987	0.015	0.742	0.4462	0.5104	0.5076	0.426	-0.4994	0.8002	0.6761	1	
Interactive term between schooling and Economic Complexity Index	-0.6522	-0.0092	0.7836	0.5093	0.5708	0.5636	0.4616	-0.5575	0.8321	0.6285	0.9605	1

CV
Martin Maniera

Education

10/2013 – 09/2020	PhD program in Economics and Finance, University of St. Gallen, Switzerland
09/2010 – 04/2013	Master studies in Economic, University of St. Gallen
08/2009 – 12/2009	Exchange term Copenhagen Business School, Denmark
10/2007 – 07/2010	Bachelor studies in Business Administration, Catholic University of Eichstaett– Ingolstadt, Germany

Professional experience

04/2019 - ongoing	Head of Department Economic Policy, Schweizerischer Baumeisterverband, Zurich
04/2016 - 03/2019	Author and editor of teaching books, Compendio Bildungsmedien, Zurich
06/2014 – 03/2016	Financial editor, VZ Vermögenszentrum, Zug, Switzerland
10/2013 – 05/2014	Economics and financial editor (traineeship and freelance), Neue Zürcher Zeitung, Zurich, Switzerland
10/2012 – 03/2013	Trainee, Economiesuisse, Communications department, Zurich
01/2012 – 06/2012	Intern, Credit Suisse, department „Market analytics und thematic research“, Zurich
07/2011 – 09/2011	Intern, German Cental Bank, department „International and European Economy“, Frankfurt, Germany
10/2010 – 12/2011	Student research assistant, Prof. Dr. M. Bütler, Swiss Institute for Empirical Economic Research, University of St. Gallen

Social service

10/2006 - 06/2007	ENT-department, Nordstadt hospital, Hanover

Language skills

German	Mother tongue
English	Fluent
French	Advanced
Polish	Advanced