Radboud University Nijmegen, Faculty of Arts
Department of Economic, Social and Demographic History

From Class to Citizenship.

Global Inequality and Social Mobility in the Twentieth Century

Master thesis

Date: 15 August 2016

Author: Elien van Dongen

Supervisors: Prof. dr. J. Kok, Ingrid van Dijk Msc

Acknowledgements

I could not resist the temptation to end the process of writing this thesis with some words of thanks, as I wouldn't have been able to complete it without the moral support of family and friends. My first and dearest word of thanks goes to Dennis Lundström. Although I have not (yet!) been able to do the same for you, without your help and support, endless listening to my theories, practical help behind the computer, and most importantly hugs, I would not have been able to complete this thesis.

Secondly, I could not have done without the optimism and words of support from Annabelle Jansink, one of the best historians and writers that I know. If I would have had enough time to let you read my thesis, I am sure it would have been a delight to read. Similarly, the optimism and company of Evelyn Leiva Deantonia gave me the energy to write the final parts of this thesis. I also want to thank Chiara Fasotti, as there is nobody who understands the challenges that come with writing a thesis better.

My sisters, Veerle and Janne, deserve my gratefulness for all their practical support: without the dinners they cooked for me and the shopping they did for me this summer I would have surely aged five years by now. My parents deserve praise for this too, and I will certainly miss their everyday support.

I have to thank all the anonymous nerds on Stackoverflow. For every question I ever had about programming, an answer was only two clicks away. And finally, I want to thank my supervisors, Jan Kok and Ingrid van Dijk, who not only helped me with my thesis but also more generally supported me in my academic pursuit. For this, I also want to thank Angelique Janssens, my supervisor in socioeconomic history more generally, who encourages me with her endless optimism.

Contents

Introduction

In recent years, the nineteenth century ‘Great Divergence’ has become a popular topic in economic history.[1] This economic divergence between ‘the West’ and ‘the Rest’ continued during the twentieth century, and especially in this period increases in global inequality were largely due to an increasing economic inequality between countries.[2] Contrary to this global trend, national wealth and income inequalities in the western world, as well as inequalities between countries in the Atlantic economy, decreased during the first half of the twentieth century.[3] At the same time the economy of countries in ‘the Rest’ of the world lagged behind in relative, and sometimes even in absolute, terms.[4] Although historians often discuss the latter half of the twentieth century in terms of globalization, the huge economic gap between the West and the Rest persisted.[5] We saw a high inequality in which class mattered less and less, while the state issuing your passport mattered more and more. As a consequence we can now explain 66-73% of a world citizen’s income by the county he or she lives in.[6] This means that a large part of your future income is determined by the country you are born in. Interestingly, this inequality determined by citizenship seems to be stronger for those at the bottom of national income distributions.[7]

To understand the impact of changes in the global economy on the individual level of world citizens, we must combine between- and within-country inequality. Global studies of between-country inequality often compare countries at the national level and are written in a context of globalization, factor price equalization and convergence. More detailed studies into the socioeconomic status of individuals, on the other hand, are often confined to one, or a small number of, (western) countries. While ‘social mobility’ and ‘inequality of opportunity’ play an important role in the latter, they are rarely referred to in the context of international inequality studies.

Not only the size of inequality, but especially the ‘stability’ of inequality matters for the opportunity of people to change their socioeconomic position. In national studies social mobility is often measured in the

1 K. Pomeranz, The Great Divergence. China, Europe, and the Making of the Modern World Economy (Princeton 2000).
2 F. Bourguignon and Ch. Morrisson, ”Inequality among world citizens: 1820-1992” The American Economic Review 92:4 (September 2002) 727-744.
3 Th. Piketty, Capital in the 21th century (Paris 2013), see figure I.1 for income inequality, USA, figure I.2 for wealth inequality, western Europe.
4 B. Milanovic, Worlds Apart. Measuring International and Global Inequality (Princeton 2005); idem, ”A short history of global inequality. The past two centuries” Explorations in Economic History 48:4 (December 2011) 494-506.
5 Whether overall global inequality increased or decreased during the last quarter of the twentieth century is debated, see for a comprehensive overview of the literature on the topic: S. Anad and P. Segal, ”What Do we Know about Global Income Inequality?” Journal of Economic Literature 46:1 (March 2008) 57-94.
6 B. Milanovic, ”Global inequality of opportunity. How much of our income is determined by where we live?” Review of Economics and Statistics 97:2 (May 2015) 452-460. Based on a linear regression explaining annual average household per capita income in $PPP by country dummies, income data from 2009, R 2 = 0.733 unweighted, R 2 = 0.657 population-weighted.
7 L. Pritchett, ”The Cliff at the Border” in: R. Kanbur and M. Spence, Equity and Growth in a Globalizing World (Washington DC 2006) 263-286; B. Milanovic, ”Global Inequality. From Class to Location, from Proletarians to Migrants” Global Policy 3:2 (May 2012) 125-134.

form of intergenerational changes in the socioeconomic level of occupations. However, upward and downward mobility in this study are conceptualized in a broader sense, with upward mobility meaning a positive change in position within the global income distribution. Given the high influence citizenship has on ones position within the global income distribution, a good example of upward mobility then becomes migration from a poorer to a richer country.[8] The opportunity for upward mobility is then increased by liberalization of international labour migration. In this light, the pre-1930 migration wave from Europe to the 'New World' is often viewed as one of the key reasons why the Atlantic economy converged, as for example by the American economists Timothy Hatton and Jeffrey Williamson.[9] However, while the latter part of the twentieth century in many respects was an era of globalization, it was not with regard to opportunities for international labour mobility. Generally, the opportunities for labour mobility have been restricted instead of liberalized.[10]

To understand the impact of restricted immigration on the global economy at large, and on the opportunities for upward mobility of world citizens, I will try to answer the question *if and how migration flows have affected global inequality over the last fifty years*. Specifically, I will investigate the effect of immigration and emigration on the wages or incomes of specific occupational groups within countries between 1960 and 2015. A better understanding of the relationship between migration and economic growth and inequality would be of interest to the academic as well as political debate. But an answer to this question is not only of economic interest; changing trends in global inequality in the past have been linked to political turmoil, as have changes in migration flows.[11]

An analysis of the relationship between migration flows and global inequality will be done by means of OLS regression. For three occupational groups and seven world-regions separate models will be created to analyze the relationship between migration to and from different world-regions and wage growth in these occupational groups. The results of this analysis are presented in chapter three. In order to do such an analysis, however, a dataset of unskilled, skilled and high-skilled global wages was created. The construction of this database is discussed in chapter two, as is the construction of seven world-regions based on income and migration characteristics. The first chapter starts with a discussion of the literature on global inequality and migration.

8 6] C.V. Zuccotti, H.G. Ganzeboom and A. Guveli, "Has Migration Been Beneficial for Migrants and Their Children? Comparing Social Mobility of Turks in Western Europe, Turks in Turkey, and Western European" International Migration Review (November 2015). Web; J.E. Roemer, Equality of Opportunity (Harvard University Press 2000).

9 T.J. Hatton and J.G. Williamson, Global Migration and the World Economy. Two centuries of Policy and Performance (Cambridge US, London 2005).

10 M. Ruhs, The Price of Rights. Regulating International Labor Migration (New Jersey 2013); A.S. Timmer and J.G. Williamson, "Immigration Policy Prior to the 1930s. Labor Markets, Policy Interactions, and Globalization Backlash" Population and Development Review 24:4 (December 1998) 739-771.

11 B. Milanovic, Global Inequality: A New Approach for the Age of Globalization (Harvard University Press 2016).

Chapter 1: Background

A hundred years ago the world was caught up in what we now call the Great War. Hopes and dreams of eternal progress and a world too sophisticated for war were shattered. Traditional political institutions collapsed, as did the wealth of the old establishment. The Great War signalled the beginning of a gradual change in wealth distribution in the western world which was further enhanced by the Great Depression and the Second World War. High national wealth and income inequalities which had seemed impregnable dissolved.[12] A combination of misfortunes for the wealthy and a changing political environment led to the rise of the 'middle class' and a considerable improvement in welfare for the working classes. Not only their economic situation became better, disadvantaged groups in the western world successfully fought for equity in other domains such as gender and race.

From a global perspective, however, this decreasing inequality in the western world does not tell the whole story of the global economy in the twentieth century, in which an ever increasing *international* economic inequality changed the meaning of poverty and wealth. The economic inequality between countries, or 'Great Divergence', borne by the Industrial Revolution and the great leap forward of the western world during the nineteenth century, only accelerated its course during the first half of the twentieth century – despite war and depression. While 'absolute poverty' - living at subsistence levels - had been a global phenomenon during the late nineteenth and early twentieth century, which engaged workers from England to China alike, the meaning of 'being poor' now diverged along national borders. By the end of the twentieth century, to belong to the poorest 5% of Denmark's population was equivalent to being richer than 97-99% of the Chinese and Indian population - even at purchasing power parity.[13]

Global historians often discuss the twentieth century in terms of globalisation: national political institutions became connected through international and global organisations and trade agreements, private enterprises expanded their scale to become multinationals, and from the perspective of a world citizen the world became more accessible through lowered transport costs and new communication technologies. People as well as private and public institutions gathered the knowledge and economic means to move across national boundaries. Economic convergence between a group of countries after the Second World War, mostly in what we now refer to as the 'western' or 'developed' world, led to an optimistic belief among scholars and politicians in the west that this growth trajectory would be followed by the rest of the world as well.[14] Hopeful signs came from Latin America, Northern Africa and the Soviet Union, economies that seemed well on track to catch up with Europe and the Anglosphere.

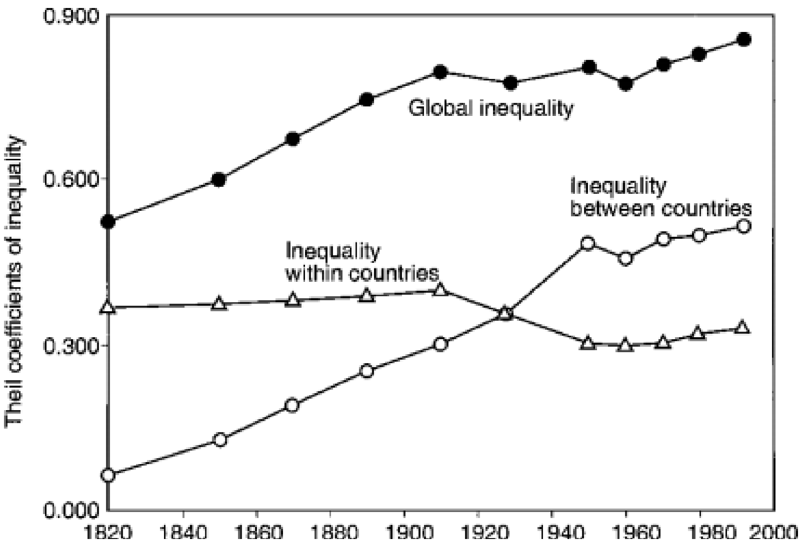
12 This process is summarized empirically in Thomas Piketty, *Capital in the 21th century* (Paris 2013), figure I.1 (income - USA) and figure I.2 (wealth – western Europe).

13 Branko Milanovic, “Global Inequality. From Class to Location, from Proletarians to Migrants” *Global Policy* 3:2 (May 2012) Web. 127-128.

14 As in various interpretations of Simon Kutznets, “Economic Growth and Income Inequality” *American Economic Review* 45 (March 1955) 1-28; Robert Solow, “A contribution to the theory of economic growth” *Quarterly Journal of Economics* 70:1 (February 1956) and; Paul Samuelson, “International Trade and the Equalisation of Factor Prices” *Economic Journal* (June 1948) 163-184.

However, during the subsequent decades most of these areas fell behind. New hopeful signs from East Asia kept the dream alive, but on their own do not provide much reason to believe in a natural development towards global economic convergence.[15] Until 1970 inequality between countries mostly increased, and even during the last decades of the twentieth century the huge gap between 'the west' and the rest persisted.[16] While only about one-fifth of global inequality could be subscribed to between-country inequality in 1820, as much as four-fifth of global inequality was due to between-country inequality in 2002 (see figure 1).[17]

18



Global economic inequality, and especially a growing inequality between countries, is thus one of the major phenomena that characterize the twentieth century's global economy. We saw a high and increasing inequality in which class mattered less and less, while the state issuing your passport mattered more and more. As a consequence we can now 'explain' (in a regression sense) 66-73% of a person's income by

15 The converging regions are based on the pre-1950 and post-1950 'convergence club' in: S. Dowrich and J. Bradford DeLong, "Globalization and Convergence" in: Michael D. Bordo, Alan M. Taylor and Jeffrey G. Williamson (eds.), *Globalization in Historical Perspective* (Chicago 2003) 199-203.

16 Whether overall global inequality increased or decreased during the last quarter of the twentieth century is debated, see for an overview of the literature on the topic: S. Anand and P. Segal, "What Do We Know about Global Income Inequality?" *Journal of Economic Literature*, 46: 1 (2008) 57-94; M. Roser, https://ourworldindata.org.

17 Branko Milanovic, "A short history of global inequality: the past two centuries' *Explorations in Economic History* 48:4 (December 2011) 494-506, 500. The results are slightly different for different inequality measures: the Gini coefficient ranges from 28% in 1820 to 85% in 2002 (using 2005 PPPs), while the Theil index ranges from 8% in 1820 to 66% in 2002 (figure 1), reaching its maximum in 1980 at 74% and decreasing afterwards due to increasing within-country inequality. These between-country inequality measures are population-weighted. The between-country data is based on Angus Maddison's 'Historical Statistics for the World Economy' (version 2004), newest version here. Global inequality data is based on François Bourguignon and Christian Morrisson, 'Inequality among world citizens: 1820-1992' *The American Economic Review* 92:4 (September 2002) 727-744.

18 P.H. Lindert and J.G. Williamson, "Does Globalization Make the World More Unequal" in: M.D. Bordo, A.M. Taylor and J.G. Williamson (eds.), *Globalization in Historical Perspective* (Chicago 2003) 227-275, 230.

knowing the country he or she lives in.[19] This means that a large part of your future income, save for overall growth of your country's economy or international migration, is determined by the country you are born in.

Importantly, former World Bank economist Branko Milanovic shows that this inequality determined by citizenship is not the same for all citizens of a country, but holds more strongly for those at the bottom of national income distributions. An unskilled worker from Beijing, Delhi or Nairobi can improve his income by a factor 11 by relocating to New York or London, while an engineer would 'only' make 3 times as much after relocating.[20] This international inequality – and especially its concentration at the level of unskilled workers – represents a severe form of inequality of opportunity: a disparity that cannot be resolved through personal effort. As it is larger among groups that are lower educated, explanations based on human capital and productivity arguments become problematic. This is shown by Harvard economist Lant Pritchett, too, in his measurements of the "cliff at the [US] border" – his name for the international gaps in the wages of equivalent labour. Pritchett does not only measure this citizenship cliff for otherwise – including ethnically – equivalent workers, but also compares the gains from international mobility to a lifetime of benefits through an international development program. He finds that, for example, the gains of earning a US wage belonging to his level of productivity for a Bangladeshi would in four weeks compensate for a lifetime of microcredit.[21]

Much of the literature on economic inequality is focussed on either one of two forms of global inequality: between-country or within-country inequality. Global studies of between-country inequality usually compare countries at the national level and are written in a context of globalisation, factor price equalization and convergence. More detailed studies into the socioeconomic status of individuals, on the other hand, are often confined to one, or a small number of, (western) countries.[22] While 'social mobility' and 'inequality of opportunity' play an important role in the latter, and are from various perspectives seen as undesirable within national boundaries, they are rarely referred to in the context of international inequality studies. To understand the impact of changes in the global economy on the individual level of world citizens, we must combine these two forms of inequality and study inequality between world citizens, as was most vividly propagated recently by Lant Pritchett and Branko Milanovic.[23]

19 B. Milanovic, "Global inequality of opportunity. How much of our income is determined by where we live?" *Review of Economics and Statistics* 97:2 (May 2015) 452-460. Based on a linear regression explaining annual average household per capita income in $PPP by country dummies, income data from 2009, R^2=0.733 unweighted, R^2=0.657 population-weighted.

20 This is due to the *negative* correlation between national income and economic inequality: higher income, in today's world, is associated with lower economic inequality. There is no reason to expect causation here, but it warrants scepticism about the assertion that inequality stimulates growth. B. Milanovic, "Global inequality. From Class to Location, from Proletarians to Migrants" *Global Policy* 3:2 (May 2012) 125-134.

21 Pritchett, "The Cliff at the Border".

22 Renowned, recent, examples include: Joseph Stiglitz, *The Price of Inequality. How Today's Divided Society Endangers Our Future* (New York 2012); Thomas Piketty, *Capital in the 21th century* (Paris 2013).

23 See for example: B. Milanovic, *Worlds Apart: Measuring International and Global Inequality* (Princeton 2005); L. Pritchett, "The Cliff at the Border" in: R. Kanbur and M. Spence, *Equity and Growth in a Globalizing World* (2006) 263-286.

Inequality of opportunity in the age of globalization

In particular, this thesis will focus on inequality of opportunity and 'social mobility' among world citizens. Not only the size of inequality, but especially the 'stability' of inequality matters for the opportunity of people to change their socioeconomic position. In recent studies 'social mobility' is often measured in the form of "career and generational changes in the socioeconomic levels of occupations" within one country, or a small group of countries.[24] Along with this habit, the meaning of the concept has in practice become confined to national borders. The opportunities for social mobility are then said to increase through for example equal access to education and an increased share of *earned* income (e.g. wages) as opposed to *owned* income (e.g. inheritance).

However, upward and downward mobility in this study are conceptualised in a broader sense, with upward mobility meaning a positive change in position within the global income distribution, and downward mobility meaning a negative change of this form. Given the high influence citizenship has on ones position within the global income distribution, a good example of upward mobility then becomes migration from a poorer to a richer country. The opportunity for upwards mobility is then increased by liberalisation of international labour migration (e.g. the European Union's Schengen zone) and decreased by restrictions on labour migration (e.g. late twentieth century approach of the 'West' towards prospective economic migrants from the 'Rest').

As in the *Encyclopaedia Britannica* entry on social mobility: "Throughout history international migration has been an important factor in upward mobility. One instance may be seen in the 19th-century migration of members of the working and peasant classes from Europe to the United States."[25] The pre-1930 migration wave from Europe to the 'New World' is indeed often viewed as one of the key reasons why the Atlantic economy converged, as for example by the American economists Timothy Hatton and Jeffrey Williamson.[26] According to their model, migration accounted for at least 70 percent of Atlantic convergence.[27] In this sense, globalisation during the latter half of the twentieth century bears little resemblance to Atlantic integration; while the period in many respects was an era of globalisation, it was not with regard to opportunities for international labour mobility. Generally, the opportunities for labour mobility have been restricted instead of liberalised: both by restrictive policies in rich countries, and by the increasing number of independent countries in the world.[28]

24 "Social Mobility" *Encyclopaedia Britannica*, Web.
25 "Social Mobility" *Encyclopaedia Britannica*, Web.
26 T.J. Hatton and J.G. Williamson, *Global Migration and the World Economy. Two Centuries of Policy and Performance* (Cambridge US, London 2005).
27 Williamson and Taylor tried to quantify the effect of migration on Atlantic convergence by developing a counterfactual model assuming no migration. When fully integrated capital markets are assumed – that offset the effect of migration as the capital flows to the New World would have been smaller without migration – migration still accounted for 70 percent of Atlantic convergence. Under the assumption of completely disintegrated capital markets, migration accounted for 125 percent of convergence. As in reality capital markets were not fully integrated, the contribution of migration was probably larger than 70 percent. A.M. Taylor and J.G. Williamson, "Convergence in the age of mass migration", *European Review of Economic History* 1:1 (April 1997) 27-63.
28 Pritchett, "The Cliff at the Border", figure 11.1; Branko Milanovic, *Global Inequality. A New Approach for the Age of Globalization* (Harvard Press 2016), 143-147 (figure 3.5).

'Upward' and 'downward' mobility are ambiguous concepts; we can define them in an absolute sense, based on income per capita with or without purchasing power corrections, or in a relative sense. A relative approach, with the nation-state as reference group, has been common in the literature. Thus, common approaches to social mobility suffer from twofold epistemological nationalism; both in choice of subject, and in choice of reference group. Especially in migration studies this approach is problematic; it is rather unlikely that in the event of international migration the reference group of the migrants – those who they compares their income to – suddenly shifts from the country of origin to the destination country. If and when, and under what circumstances, this happens, is a whole field of study in itself. In this study relative mobility of a national occupational group with respect to other groups in the same nation, as well as with respect to other nations, will be analysed.

Research question: migration and global inequality

To understand the impact of restricted immigration on the global economy at large, and on the opportunities for upward mobility of world citizens, I will try to answer the question *if and how migration flows have affected global inequality over the last fifty years.* Specifically, I will investigate the effect of immigration and emigration on the wages or incomes of specific *occupational groups* within countries between 1960 and 2015. Thus, instead of representing the entire world population, this study is concerned only with the global labour force – and the effect of international labour mobility on their wage incomes. Wage incomes are problematic in studies concerning earlier periods – as much of the remuneration for labour was not monetary, and in studies focussed on accurately capturing overall inequality in one society (for example with the Gini coefficient), as capital income is distributed more unevenly than labour income. However, in this study the focus is on the effect of international labour mobility on income growth. Therefore ideally only labourers would be studied. As we cannot directly find the total incomes of those global citizens who are active on the labour market, wages seem the most accurate approximation. From a classical economic perspective, we may expect labour mobility – leading to labour scarcity or abundancy – to have a direct effect on labour costs and thus wages. Studying the effect of migration on capital incomes would however be a useful addition for future studies.

A better understanding of the relationship between migration and economic growth of specific socioeconomic groups in different countries (inequality) would be of interest to the academic as well as public debate. But an answer to this question is not only of economic interest; changing trends in global inequality in the past have been linked to political turmoil, as have changes in migration flows. By knowing how these processes affect different groups economically, policy makers will be in a better position to avoid adverse effects and exploit the growth potential of migration.

To study the relationship between migration and global inequality, three broad occupational groups are created for each available country: unskilled workers, skilled workers and highly skilled workers. Social mobility, in this thesis, is thus not measured directly with micro-level data, but approximated by using wage data for larger national occupational groups. Therefore, variations within such groups cannot be taken into account. Since our main objective is to analyse the effects of migration on national occupational groups at large, this is not necessarily a disadvantage – although micro-level studies on these relationships would form a valuable complement to the study conducted here. This is a data-driven but also a necessary restriction on the debt of the study; to create a comprehensible overview of global income mobility, we need to organize world citizens into a limited number of groups, around which the historical narrative can revolve.

The effect of migration on the income of these socioeconomic groups will be analysed by creating a regression model. Other relevant factors, such as conflict, remittances, education and urbanisation, will be taken into account and – importantly - migration rates for different world-regions will enter the model separately. These world-regions are created on the basis of international economic inequality and immigration patterns: each world-region contains countries with similar incomes (measured by GDP per capita at PPP) and has similar bilateral immigration flows.[29] By distinguishing between migrant flows from different world-regions, we take into account political, cultural and economic differences between these world-regions.

Separate explanatory models are created for the various world-regions, in which migration from all these world regions can have a different effect on wage development. Earlier studies into the relationship between migration and income or inequality usually create one model including all observations. They subsequently attempt to control for political and cultural differences by adding indicators to the analysis that should capture these differences. As it is hard to determine which indicators actually capture such differences, a more straightforward way of taking them into account is creating separate units of analysis, which is done here through the clustered world-regions. An obvious drawback of this method is that it gives less clarity about the source of differences between these clusters. A major advantage, however, is that separate models allow variation to vary between clusters: if we expect the trends and effects of the explanatory factors to be different in a different cultural, political and economic context, we should expect the remaining variation to be different as well. This assumption therefore leads to a more accurate representation of trends within clusters.

Finally, by comparing the different occupational groups and different income groups from the first analysis conclusions can be drawn on the relationship between global inequality and ‘social mobility’ in the form of migration. Hereby we can see what, during the last 50 years, the effect of migration on economic inequality as well as growth has been. Although the results will depend on the choice of data, the main advantage of this analysis will be the spatial and temporal comparability: even if the available data is not ideal, similar data will be used for all countries in the analysis. Furthermore, instead of country aggregates and estimates, data from the *International Labour Organisation* (ILO) on national wages at the occupational level will be used.[30] This makes it possible to look at the effect of migration on specific socioeconomic groups within countries, as well as the aggregate effects on a global or regional level.

29 Clustering is done with the DBSCAN method in scikit learn (Python).

The focus of previous studies into the economic effects of migration has been on immigration into the western world, both because of data availability and because studies were often conducted by western policy institutes. In such studies often no significant effect of migration was found – although there are both examples of positive and of negative effects.[31] However, to get a better idea of the effect of migration on the global economy and find characteristics that are less context specific, we need to widen our geographical scope. Moreover, it is likely that the economic effects of migration on poor countries are (relatively) larger than the effects on rich countries – as the entire economy of poorer countries is smaller. This is especially true for remittances, which are usually high (as a share of GDP) in countries with opportunities to migrate to richer neighbouring countries (e.g. in 2013: Lesotho 20%, Tajikistan 49%, Nepal 29%, Haiti 21%, Lebanon 18%).[32]

Factor price equalization and capital labour substitution: an economic approach to migration and economic convergence

Next to the societal and historical relevance of the relationship between migration and global inequality, studying this relationship empirically is also of major relevance to economic theory, especially in light of existing assumptions on the behaviour of different factors of production. Based on factor price theory, it is traditionally assumed that migration – just like liberalisation of other factors, by for example free trade agreements – leads to economic convergence.[33] More precisely, we assume 'old-style' inequality to rise in the receiving country, and this same form of inequality to decrease in the sending country. That is, inequality between those earning a capital income (e.g. landowners, shareholders) and those earning a wage through labour will increase in the area where labour becomes more abundant, while capital becomes more scarce; the receiving country. Conversely, in the sending country we assume that the value of capital (which becomes less scarce) will decrease, while the value of labour (which becomes scarcer due to emigration) increases. Under the assumption of rational agents living in a free world, we assume that wage earners will move from high inequality (and low wage) countries to low inequality (and high wage) countries. Thereby labour and capital will become distributed more evenly over the world or region in which migration takes place, and thus capital and labour will be valued more similarly, leading to convergence.[34] This theory is not

30 In a standardized form, the 'Occupational Wages around the World' (OWW) Database: R.H. Oostendorp, "The Occupational Wages around the World (OWW) Database: Update for 1983-2008" *NBER* (May 2012).

31 See for an example of negative effects: G.J. Borjas, "The Economics of Immigration" *Journal of Economic Literature* 32 (December 1994) 1667-1717; idem, "The Economic Benefits of Immigration" *The Journal of Economic Perspectives* 9:2 (Spring 1995) 3-22; idem, "Immigration and Globalization: A Review Essay" *Journal of Economic Literature* 53:4 (December 2015) 961-974.

32 D. Ratha et al., "Migration and Development Brief" *World Bank* 24 (April 2015).

33 That is, in turn, under the assumption of limited and constant capital labour substitution, cf. Heckscher-Ohlin model.

34 It should be noted that the model is generally applied to trade in goods, not migration. However, both are seen as factors of production, with the main difference that in the traditional model equalization is achieved by trade instead of migration. There are many problems with the assumption underlying the basic model, an important one not mentioned here is treated by Robert C. Allen in "Technology and the Great Divergence: Global Economic Development since 1820" *Explorations in Economic History* 49:1 (January 2012) 1-16. Allen shows how technological development was local instead of neutral, in the sense that subsequent technologies were only economically beneficial for countries with a high enough capital to labour ratio (the frontrunners that developed them) and could thus not successfully be

only subscribed to widely among neoclassical economist, but also by critics of traditional economic theory such as Thomas Piketty.[35]

As not all individuals act rationally, or primarily out of economic interest, all of the time, and the world is not free but restricted and controlled by various legislative and governing bodies, it is not self-evident that this theory would hold true empirically. Additionally, as the old distinction between capital-owners and wage-earners gradually disappeared in the post-1945 period, the result of factor price equalization becomes less obvious.[36] However, because capital is still largely in the hand of the income-rich – and more unevenly distributed over the population than wages – the effects may still exist to some extent.[37] In the case of the Atlantic economy, we saw that large migration flows have in the past led to economic convergence.

However, even the results Williamson and Hatton have obtained on the Atlantic economy in the period 1870-1913 are somewhat mixed. While the ‘New World’ on the receiving side of migration, and the ‘European periphery’ on the sending side, behaved as the above theory would predict, the results for the ‘European core’ are mixed.[38] In the European periphery inequality decreased both at the bottom and at the top of the income distribution; the gap between wages of urban workers and GDP per capita decreased (at the bottom), while at the same time the wage-rent ratio increased (at the top, or ‘old-style’ inequality). The reverse happened in the New World.[39] In the European core, however, inequality at the bottom increased, and the increase in wage-rent ratio may be due to other influences than migration, most importantly a changing composition of wealth (away from land rents) and increased foreign investment (in, amongst others, land in

implemented elsewhere. A perhaps more fundamental critique on the model is given by (amongst others) Thomas Piketty, who questions whether the so-called ‘elasticity’ between capital and labour is constant at about one. That is, whether one factor can be replaced by the other, but only to such an extent that the contribution of capital to entire income remains constant (as in the Cobb-Douglas production function). See: Piketty, *Capital,* 263-264; Loukas Karabarbounis and Brent Neiman, “The Global Decline of the Labor share” *Quarterly Journal of Economics* 129:1 (2013) 61-103; Michael Elsby, Bart Hobijn and Ayşegül Şahin, “The Decline oof US Labor Share” *Brookings Papers on Economic Activity* (October 2013) Web.

35 Piketty, *Capital*, 643-644.

36 Thomas Piketty, *Capital,* chapter 8, figure 8.3 and 8.4.

37 Christoph Lakner and Anthony Atkinson, “Wages, Capital and Top Incomes: The Factor Income Composition of Top Incomes in the USA, 1960-2005” Web (2013) Web.

38 ‘European periphery’ = Denmark, Finland, Norway, Sweden, Italy, Portugal, Spain, Austria, Ireland. Note that Eastern Europe is largely absent. ‘European core’ = Belgium, France, Germany, Great Britain, The Netherlands, Switzerland. ‘New World’ = Argentina, Australia, Canada, United States. Wage to income inequality was rising more rapidly in the US than elsewhere in the New World, while rent to wage inequality rose slower there than in other New World countries.

39 In the European periphery wages grew on average 1.73% per year, while GDP per worker hour grew 1.6% per year over the period 1870-1913. In the New World wages grew 1.14% per year, while GDP per worker hour grew 1.77% on average per year. In 1870-1913 wage-rent ratios increased by on average 2.32% per year for the European periphery, while decreasing 3.03% per year in the New World. See Hatton and Williamson, *Global Migration and the World Economy,* 101-125, 107 for data.

the New World).[40] These mixed results on the European core do not necessarily disqualify the theory, since migration flows from this area were rather low as compared to those from the European periphery.

In light of the perception of poverty and wealth as relative, instead of absolute, phenomena, there are some further questions that we should answer, even if factor price theories hold true, to be able to say what labour mobility means for global inequality between world citizens. Firstly, is this convergence largely the result of economic decline in the receiving areas, or of economic growth in the sending areas; what happens to the global economy at large? As migration from poor to rich was the most common form of migration in the world during the latter half of the twentieth century, in this study this translates to the question whether the economic benefits of migration over the last fifty years have been larger for the 'developing' (sending) countries, than the economic decline in the 'developed' (receiving) countries. An important additional question, taking into account the heterogeneity within nations, is whether migration has reduced inequality more in the 'developing' (sending) countries, than that it increased inequality in the 'developed' (receiving) countries. All of these questions, of course, are based on the premise that there is a measurable effect of migration on economic growth, which is what we should find out first.

The wealth and poverty of nations: why some are so rich and some so poor

Over the past decades, an extensive body of literature has treated the question why certain parts of the world became – and remained – so much richer than others. To study the relationship between migration and global inequality, we need to take into account other factors that might have affected global inequality simultaneously. Three lines of reasoning in this respect are discussed here, but there exist many other explanations 'exogenous' to economics, most notably from an environmental (as for instance in the works of Jared Diamond) or cultural (post-colonial or Weberian) perspective.

Inspired by institutional economics the concept of 'conditional convergence' has been gaining influence during the last decades. The concept is used to show that countries in the post-1945 period did converge economically if other conditions are held constant ('ceteris paribus'). These supposedly exogenous

40 Between 1870 and 1913 wages of urban workers in the European core increased by on average 0.90% yearly, while yearly growth in GDP per worker hour was much higher at 1.46%, on the other side the wage to rent ratio increased here by on average 2.32% per year. The reason why comparison of this last measure between the New World and European core is problematic is derived from the work of Thomas Piketty. In *Capital in the 21th century* he shows that the value of land relative to national income decreased in both the European core and the New World (USA), but much less so in the US (from 25% to 20% of total wealth), followed by Germany (from 40 to 20% of total wealth), then France (from 40 to 15% of wealth) and lastly Britain (from 30 to 5% of total wealth). This, together with spiking foreign investment from the European core in the New World, make the wage to (land) rent measure problematic. The total value of land in the US was worth clos to 100% of national income during the entire period 1870-1910, with only slight variation over the years. Total wealth increased from 400 to 500% of national (yearly) income. In contrast, value of land became one-sixth of its 1870 value in Britain by 1910 as it decreased from double to less than one-third of national income, while total wealth was stable at around 700% of national (yearly) income. In France, too, the decrease in land value as proportional to national income was sharp; from 300% to 100% in the pre-WWI period. Here, too, total wealth only changed slightly from just over 700% to just under 700% of national income. The German developments were similar to Britain and France, with land value decreasing with almost 150% down from 275% of national income and total wealth decreasing slightly from 700 to about 650% of national income. Thomas Piketty, *Capital in the 21th century* (Paris 2013) 140,169, 180.

conditions include all kinds of political and cultural factors, so that such theories of ‘conditional convergence’ are quite meaningless for real-world prospects on economic development (i.e. controlled for suppressive regime, North and South Korea may be found to be converging during the last decades of the twentieth century). Moreover, there is a causality problem in this type of explanation. In for example the conceptualization of ‘conditional convergence’ by the American economist Robert Barro variables such as life expectancy, male secondary-school attendance and fertility rates – variables that are generally highly correlated with economic development – are controlled for.[41] Thereby, the concept of economic progress is deprived of all its welfare components and becomes largely empty. Looking at economic development at a given level of health and educational development makes international inequality into a chicken-and-egg-problem.

Another line of theory on global inequality is related to politics and specifically to openness to trade and democracy. In the work of the economists Steve Dowrick and Bradford DeLong ‘openness’ is defined by five conditions that, if a country fulfils at least one of them, make a country closed. These conditions refer mostly to trade barriers and other protectionist and monopolistic measures, but also to the existence of a large black market and a ‘socialist economic system’. This openness dummy explained some of the economic divergence in the period 1960-1980, but its validity seems ambiguous (why does socialism necessarily make a country closed?) and mostly applicable to the exact time period for which it was developed (Cold War) while not being generalizable. This is confirmed by Dowrick and DeLong themselves, as their ‘openness’ measure did a poor job in explaining economic divergence after the 1980s.[42]

Finally, global inequality is often related to technological development and differences in productivity. People in ‘the West’ are then assumed to earn more than people in ‘the Rest’ because they are more productive; because they have better technologies, because they outsource all non-productive jobs, because they have more ‘human capital’ or even because they are more hard-working. And indeed, workers in ‘the West’ are often employed in more productive (i.e. generating more output per hour of labour) jobs than workers in ‘the Rest’; the capital to labour ratio has been much higher in the West throughout the twentieth century.[43] But since international inequality is most severe at the level of unskilled workers, where it cannot be explained by skill premiums or higher productivity, technological development is not a sufficient explanation for global inequality. In the work of Lant Pritchett, it is shown that controlled for ethnicity, education, age, gender, urban residence and other productivity differentials, wages disparities between the US and 42 developing countries are still substantial. The average person in any of these 42 countries can *multiply* his or her income by 3.41 by moving to the US; they could earn PPP $10,000 more yearly.[44]

41 Steve Dowrich and J. Bradford DeLong, “Globalization and Convergence” in: Michael D. Bordo, Alan M. Taylor, and Jeffrey G. Willamson (eds.), *Globalization in Historical Perspective* (Chicago 2003) 191-226, 203-206.
42 Ibidem, 209-216.
43 Robert C. Allen, “Technology and the Great Divergence: Global Economic Development since 1820” *Explorations in Economic History* 49:1 (January 2012) 1-16
44 Pritchett, “The cliff at the border”, 271-274. The other non-quantifiable productivity differentials are approximated by dividing each country’s wage difference with the US by the wage difference between the US and Puerto Rico (1.5), as labour migration has since long been liberalized between the US and Puerto Rico. Without this correction the 42 developing countries could on average multiply their wage by 5.11 by doing similar work in the US instead of at home.

Moreover, these results would likely be more dramatic if similar comparisons would be made between developing and northern and western European countries, as European countries have higher wages at the bottom of the income distribution than the US, which is the group that migrants from developing countries would usually be compared to due to their years of schooling.

Overall, the economic literature on global inequality is substantial, but popular theories often do not align well with the existing data. The debate would therefore benefit from more empirical studies (such as the one conducted here), new perspectives, and more data-driven theorizing.

In the following chapters I will first discuss the data and methodology used in my analysis, followed by results and a conclusion. In the results chapter, the relationships between wage growth at different skill levels, migration and other covariates will be discussed. The conclusion consists of a summary of these results in the form of global trends and answers to the specific questions raised in this chapter.

For some countries (Yemen, Nigeria) the wages would be more than ten times as high in the US as at home. Note that Pritchett's analysis is only related to productivity and does not try to explain wage difference between countries in general. Pritchett therefore attempts to correct for all variables related to productivity, but *not* for other macroeconomic factors (such as scarcity of labour vs. scarcity of capital) in the countries that could also explain wage differences. Such variables, however, are irrelevant from the individual perspective and are assumed to be solvable by liberalisation of labour mobility.

Chapter 2: Data and Methodology

In this chapter the main data sources used and their role in this thesis will be outlined. First the the central elements of the analysis, migration flows and occupational wages, are described. An account and explanation of the ways in which it was adapted before use in the analysis is given. Subsequently the same is done for the various auxiliary indicators used in the regression model. After setting out its elements, the analytical tool in this analysis – a variation on the ordinary least squares (OLS) regression model – and its use here are described.

Global bilateral migration flows

New data on global bilateral migration flows, created by statistician Guy Abel, will be related to wage, income and wage and inequality data for over 180 countries during the period 1960-2015.[45] A major advantage of migration flows, as compared to migrant stocks, is that they suffer less from country specific measuring methods and conceptualization. Especially for an analysis spanning a large geographical area this is important, as there are large international differences in the time span for which previous migrants are considered migrants, leading to international comparability issues in migrant stock data. Migrant stock data is available from, among others, the United Nations Population Division and the World Bank. Guy Abel has combined these sources to estimate migration flows for 10 – and from 1990 onwards 5 – year periods during the past fifty years, adjusting migrant stock variations for demographical factors such as fertility and mortality.

However, this is not the only advantage of this new dataset over traditional migrant stock data. Migrant flows are presented bilaterally instead of per country. This makes it possible to distinguish migration patterns by country of origin and destination. That is, whereas most sources of migrant stock data only provides information on the overall number of people entering or leaving a specific country, bilateral flows show which countries are similar with respect to the nationalities they receive and the countries to which their inhabitants move.

Clustering of world-regions

To be able to use these advantages, while not creating explanatory models with unreasonably many covariates (e.g. each bilateral flow), clusters of countries will be created based on national income and migration patterns. The exact construction of the clusters based on migration can be found in attachment 2.4. In short, eight migration clusters are constructed in an automated way by placing emphasize on the direction rather than the sum of all emigration or immigration flows in a country. Therefore, countries with emigration

45 G.J. Abel, "Estimating global migration flow tables using place of birth data" *Demographic Research* 28:28 (March 2013) 505-546. G.J. Abel and N. Sander, "Quantifying Global International Migration Flows" *Science* 343:6178 (March 2014) 1520-1522.

to (and immigration from) similar other countries will likely be clustered together. The size of each bilateral flow is still taken into account, but instead of on a continuous scale presented as a categorical variable. To further distinguish clusters based on national income countries are labelled ‘low’, ‘mid’ or ‘high’ income, and for each income category the available migration clusters become separate clusters. This could lead to $8 \cdot 3 = 24$ clusters, but in reality gives a total of 15 clusters: there are, for example, not eight but only three migration clusters including high-income countries. A more detailed description of the construction of the clusters combining migration patterns and national income can be found in attachment 2.7. The countries included in each cluster can also be found in attachment 2.7, while some general features of the clusters will be discussed in the results chapter.

Population-weighted average migration rates for each *income* cluster to and from a specific country are used as explanatory variables for the wage development in that specific country. That is, immigration and emigration per income cluster, giving a total of six migration covariates. Migration rates are not further differentiates as an initial analysis showed that differentiating more clusters as explanatory variables did not add explanatory power to the model. However, more clusters are differentiated to create separate explanatory models for: models are created separately, but with largely the same covariates, for seven clusters. These clusters are combinations of the earlier mentioned fifteen clusters that were found to have a similar relationship between wage growth and migration flows (overall emigration and immigration, only two covariates). Using more than seven different clusters would also make the sample size for certain clusters too small to be able to add extra covariates to the model and differentiate between types of migration.

In the regression analysis the log of migration rates per cluster plus one is used. This is done because the distribution of migration rates is heavily skewed towards zero - there are many country combinations for which bilateral flows are zero. Therefore, the unexplained variation in the regression also is larger around zero (heteroscedasticity) leading to less reliable results. By taking the log of migration rates extreme values become less extreme and therefore heteroscedasticity is reduced. The constant one is added to the rates for several reasons: in order to have the model not automatically predict zero wage growth if any migration variable is zero, to avoid extreme variation for migration rates between zero and one, and to not have minus infinity values in the analysis. This also makes interpretation of the coefficients easier: now zero migration rates are still zero in the log migration rates.

Moreover, it may be expected that migration does not have an immediate, but rather a lagged effect on wage growth: not all migrants immediately enter the labour market, and wages do not immediately adapt to a changing labour supply. Thus, in the regression model migration rates are used from a five-year period six years before the respective five-year wage period. This has the added benefit of diminishing causality issues.

Wages and incomes

Instead of country aggregates and estimates, where possible wage data from the *International Labour Organisation* (ILO) as available on LABORSTA on national wages at the occupational level will be used. The ILO October Inquiry contains yearly data on occupational wages for 159 occupations in 171 countries from 1983 onwards and is currently the most thorough international source on occupational wages. Using this data makes it possible to look at the effect of migration on specific occupational groups, as well as the aggregate effects on a national, regional or global level. The main advantage of these data is that similar data can be used for all countries in the analysis, especially at each year. Temporal differences in data sources will also be geographically consistent.[46] As presented on LABORSTA and ILOSTAT, the October Inquiries have various comparability issues. Therefore, a standardized version created by Remco Oostendorp and Richard Freeman is used for the years up to 2008; the Occupational Wages around the World (OWW) Database.[47]

From this data source, wages in contemporary local currencies (LCU) were used and where necessary converted to the local currency in 2008. The combined wage dataset is later adjusted using inflation at consumer prices to obtain wages in constant 2005 LCU, and subsequently divided by the 2005 PPP conversion factor to obtain wages in constant 2005 international dollars.[48] However, this conversion factor combined with consumer price indices (inflation) does often not give an internationally comparable indication of real wages at constant prices. The absolute values obtained can therefore not be considered as accurate. In this context, however, only real wage change over time is relevant, and for such change wage time-series at constant prices are far better comparable internationally than series at current prices: a growth rate of 6% with 3% inflation should not be treated in the same way as a growth rate of 6% with 9% inflation. The PPP conversion is only done to facilitate finding outliers in the data series, but is irrelevant for the analysis itself since a constant factor does not affect growth rates.

Wage and income indicators for three different occupational groups are created for as many countries and years as possible. However, the October Inquires contain few data for some world-regions, such as former Soviet states, and are only available from 1983 onwards. Data from other sources is therefore also included, and has first been standardized and was then converged onto the October Inquiries.[49] Combining

46 See for consistency of the October Inquiry data and the definition of the three occupational groups based thereon attachment 2.5.

47 Database by R.b. Freeman and R.H. Oostendorp at www.nber.org/oww/. Documentation: R.C. Oostendorp, “The Occupational Wages around the World (OWW) Database: Update for 1983-2008” Web. (May 2012).

48 Consumer Price Index and PPP conversion factor for private consumption are mostly from the World Bank’s World Development Indicators, but where World Bank data is not available inflation at consumer prices is derived from the CIA World Factbook (historical series at indexmundi.com), and if none of the other sources is available the implicit price deflator from the UN is used. Full documentation of the various sources used per country-year is available upon request.

49 Standardizing the ILO dataset consisted of combining data from different sources (where certain sources such as labour force surveys were consistently preferred over other sources such as commercial surveys), with different types (e.g. including or excluding overtime pay), different time units (ranging from hourly to annual), different local currencies over time, and different versions of the classifications of occupational categories/ economic activities. Full documentation on how the dataset is standardized is available upon request. For definitions of the original ILO dataset see here.

the various datasets had to be done manually for most of the data, but was done using a linear regression where enough overlapping years were available. If few overlapping years were available one source was added to the other by using the ratio between the two sources for the closest overlapping year available. If no overlapping years were available between the data sources they were combined using a ‘neutral’ assumption on wage growth for the time gap between the different datasets: each occupational group’s wage was assumed to grow at the same rate as national GDP in current local currency. If all of these methods gave ‘unrealistic’ results combining the sources was done manually on a case-by-case basis.[50] All wages from the ILOSTAT dataset were appended to the previous series using GDP growth between 2008 and 2009.[51]

The ILO wage dataset for the years 1969-2008 contains wages per ‘economic activity’. These economic activities are well differentiated for manufacturing, so that each economic activity includes wages for a specific task. Non-manufacturing activities, however, are rather broad. To select specific economic activities from the ILO wage database which represent wages for unskilled and skilled workers, the mean and variance of the ratio between each economic activity and the general categories ‘manufacturing’ and ‘construction’ were assessed both over years and over countries. Economic activities were not included in an occupational category if they varied much between countries or years. Moreover, they were not included if they were available for too few country-years to create an (un)skilled worker that is comparable between countries and years.[52]

The first occupational group, *unskilled worker*, has the widest geographical and temporal coverage and is based on wages in specific manufacturing sectors (textiles and wood industries) from the ILO as available for the years 1969-2008, and occupational wages from the ILO October Inquiries for each year 1983-2015 - until 2008 from LABORSTA, from 2009 onwards from ILOSTAT. In the years 1983-2008 either the ILO or LABORSTA data were used depending on which wages were available for most years, to make the concept and change over time as consistent as possible. The ‘secondary’ data series was then added to the ‘primary’ series using one of the methodologies described above (regression, GDP growth or ratio).

The wage of an unskilled worker as based on the October Inquiries consists of 27 different occupational categories, mostly in basic services (e.g. waiter, grocery store re-stocker or postman) and manufacturing (unskilled labourers, packers and basic machine operators). For the years after 2008 the ILO has only made wages of general occupational categories available for public use, the wage of the unskilled worker is therefore equivalent here to the wages in the general category 'elementary occupations'. In

50 To perform any of the three operations described before for each series the original two series and the resulting series, if the operation would be applied, were printed. If these results contained outliers or irregular developments (e.g. a ten year period with every other year from a different data source giving alternating high wage growth and decline which was clearly due to data source differences instead of economic context) the adaptation was not accepted an an automatic note was created to adapt the series manually later. A full documentation of combining the datasets is available upon request.

51 GDP (growth) in current LCU from the World Bank World Development Indicators.

52 Full documentation on the selection process of categories included in the ‘unskilled worker’ and ‘skilled worker’ category based on ILO wage statistics is available upon request. The categories eventually selected can be found in attachment 2.1 and 2.2.

attachment 2.1 the exact definition of the 'unskilled worker' based on the wages in manufacturing and the October Inquiries can be found.

The second occupational group, *skilled worker, i*s approximated by wages in the 'Transport, Storage and Communication' and 'Mining and Quarrying' sector from the ‘wages by economic activity’ database of the ILO, available in LABORSTA.[53] Conceptually, this is not an ideal definition. However, the position of these sectors within national income distributions is relatively stable over the period 1969-1990 and between countries, as compared to that of other sectors. This means that the sectors are likely to contain similar (or at least similarly paid) occupations for different countries, and thus represent the same economic category for different countries. The sectors both have somewhat higher wages than the overall manufacturing sector in almost all countries covered by ILO data, with wages in the sector 'Transport, Storage and Communication' on average 1.2 times those in the manufacturing sector and wages in the sector 'Mining and Quarrying' 1.35 times those in manufacturing over all countries and the period 1970-1990. They will include unskilled as well as skilled and high-skilled occupations. Furthermore, wages for these sectors are available for relatively many countries and years, which is of major importance for comparability. For later years these sectors are no longer appropriate to use for approximating the skilled worker's wage; the size as well as the wages of the mining and quarrying sector then decreases in *most* countries, while under the influence of new communication technologies wages in that sector divert from those in the transport and storage sector.

A better definition of the wage of a skilled worker is derived from the October Inquiries for each year between 1983 and 2008. It consists of 25 occupational categories in various sectors and includes skilled construction workers (e.g. carpenter, plumber, electrician), professional nurses, elementary school teachers, office clerks and stenographer-typists. When both sources are available, from 1983 onwards this data is always used and only supplemented by the economic activities data using regression, GDP growth or the closest ratio as before. From 2009 onwards wages of four major occupational categories from ILOSTAT are combined to form the skilled worker's wage.[54] In attachment 2.2 the exact definition of the 'skilled worker' based on the wages for general economic activities (only used for the years 1970-1980), the October Inquiries and ILOSTAT can be found.

The third occupational group, *highly skilled workers*, is only included for 1983 and subsequent years and is entirely based on the October Inquiries of ILO. For each year between 1983 and 2008 a combination of 26 occupational categories is used to create this group, including higher professionals (e.g. third level teacher, dentist, engineer, computer programmer) and executive officials in both the private and public sector. From 2009 it consists of technicians and other higher professionals, legislators, senior officials and managers. The exact definition of the 'highly skilled worker' for various years can be found in attachment 2.3.

53 Sectors based on the International Standard Industrial Classification of all Economic Activities (ISIC), revision 2 and 3 by the United Nations Statistics Division.
54 Classification of the occupational groups for 2010 is based on the ISCO-88 skill levels. See here.

For the regression analysis log wage ratios over mostly (depending on data availability) five year periods are used. Log wage ratios are the logarithm of the percentage change between start and end year plus one, or equivalently the difference between log wage at the end year and log wage at the start year of a period:

$$\text{y (dependent variable)} = \log(\text{wage end year}) - \log(\text{wage start year}) = \log(\text{wage end / wage start})$$

If for a given start year the wage for the same skill level five years later is not available, instead the closest available year is used, with the restriction that the end year must be at least two years after the start year. By using changes over five year periods, instead of changes between years, short-term fluctuations have less effect on the eventual results. This choice is preferred over shorter periods as it is expected that migration will not be able to explain short-term fluctuations in a labour market, but rather have an effect on its gradual and long-term development. Note that ‘short term’ here would refer to one year, and therefore would not capture effects of seasonal migration. Moreover, we should not expect to capture effects of short term migration in the form of seasonal workers, as seasonal workers’ remuneration is probably not well represented in the used labour statistics (ILO wage database) and under-represented in the migration database as well. By creating a growth rates over longer periods with variable start and end, as opposed to for example using benchmark years (which may be represented by other years), as much of the available data as possible can be used, and periods are always exactly as long for all indicators: all covariates are adapted so that they represent change over the exact same period as wages.

For example, the USSR may have unskilled wages for 1969 until 1976, then have a gap until 1988. In this case 1969-1974, 1970-1975 etc. are included as five year periods, but only one variable, the log wage growth between 1976 and 1988, is created for the period without other data. All other indicators will also be taken for the period 1976-1988 and therefore be proportional to this longer period of wage growth.[55] An immigration indicator will include all immigration from a certain cluster to the USSR between 1976 and 1988, while the international conflict indicator will be the sum of all conflicts between 1976 and 1988, etc.

Economic inequality: ginis

One of the covariates which will be used to control the relationship between wage growth in a specific occupational category and migration is the Gini (national) inequality measure from the Standardized World Income Inequality Database (SWIID).[56] It has the widest geographical and temporal coverage of various inequality databases currently available, and is created with the purpose to improve international comparability of the Gini measure. During the period 1969-1985, it covers between 45 and 60 countries. For the subsequent period its coverage for the countries with occupational wages is nearly complete. Therefore,

55 It is assumed here that wage growth is exponential, as national income growth is usually assumed to be, and thus log wage growth should be linear with time. More on this in the description of the regression model.

56 Frederick Solt, “The Standardized World Income Inequality Database” *Working paper SWIID Version 5.0* (October 2014). Web.

an analysis controlling for inequality can only be done for a subsample which mostly includes years after 1985, especially for the non-western world-regions.

In the eventual analysis the log ratio between a country's gini at the end of a wage growth period against that country's gini at the beginning of a wage growth period is used. The ratio is taken so that the effect of changes in economic inequality on changes in wages is measured. The logarithm is applied so that economic inequality and wages are on the same scale, as it is not expected that there is an exponential relationship between economic inequality (measured as gini) and wages.

Remittances

An important control factor for the effect of migrant flows on the income of the international labour force are remittances. As the dispersion of remittances over national income distributions is unknown, remittances could not be directly included in the income of the different occupational groups. Therefore, national data on remittances received and paid as a percentage of GDP from the World Bank (based on IMF balance of payments data) is included in the analysis as a covariant where possible. However, data on remittances is first available for 1970, and does not become widely available until 1980.[57] Separate analyses including and excluding remittances are conducted for a wider set of country-years for which remittances are not available, and for a limited set of country-years for which it is possible to include remittances.

In the regression remittances are summed over the wage period, as this is equivalent to taking the mean of remittances over a period and multiplying that with the number of years between start and end. No logarithm is taken of remittances, since whether or not to take a logarithm is an irrelevant decision: the results would be nearly the same. As the remittances data all fall in the interval $[0, 1]$, taking the logarithm of remittances $+1$ (one would need to be added for the same reasons as explained in the case of migration) would give a relationship between y = wages and x = remittances of: $\mathrm{y} = (\mathrm{x}{+}1)^{\beta} \cdot \alpha$, with β the coefficient for remittances and α a constant derived from the model intercept. $x + 1$ would here be linearly distributed between $[1, 2]$. If no logarithm is taken, on the other hand, the relationship is of the form: $\mathrm{y} = (\exp(\mathrm{x}))^{\beta} \cdot \alpha$, where $\exp(x)$ is nearly linearly distributed on the interval $[1, 2.5]$ for x in $[0, 1]$.

Education

To make optimal use of the data on migration flows, we have to consider in which ways migration flows from different countries have different effects on wages in the various occupational groups. One indicator of this could be schooling; if a country of origin has, on average, fewer years of schooling than a destination country, then a migrant is (again on average) more likely to end up working in lower occupations in the destination country. Thus, his/her arrival is more likely to affect wages in those occupational categories

57 Metadata remittances paid and received as % of GDP from the World Development Indicators. Paid remittances converted from current US$ to % of GDP by dividing by GDP in current US$.

negatively, while it possibly has a positive effect on wages in other occupational categories. Conversely, emigration from a country with high levels of educational participation and duration may have a more negative effect on high-skilled wages than emigration from a country with lower educational levels. To control for this effect a 'schooling' indicator is created based on the Barro-Lee dataset on educational attainment.[58] This indicator is the product of the percentage of the population that is not known to have completed any form of education with the average years of schooling at any educational level. It should capture an elementary form of educational inequality as well as average educational level:

$$\text{school} = (1 - \%\text{prim} - \%\text{sec} - \%\text{tert}) \cdot \text{mean school}$$

In the Barro-Lee dataset these indicators are presented at five year intervals, therefore a linear interpolation is done for the intermediate years. In the eventual analysis for each wage growth period the log ratio of the schooling indicator at the end of that period and the schooling indicator at the beginning of that period is used, equivalent to what is done for ginis; wage *change* is set out against schooling *change*, and not an absolute value for schooling. As these periods are usually five years, data is not used at a higher level of precision than five year intervals and the linear interpolation is of little consequence for the eventual analysis.

Urbanization

Wage data in this thesis are largely composed of occupational wages in the industrial and services sectors. Therefore, the effect of migration on such wages might be different for highly urbanized countries than for countries with a larger share of the population living in rural areas. The urban population as a share of the total population, the urbanization ratio, will therefore be included as a covariant in the analysis. Data on urbanization is based on the United Nations World Urbanization Prospects and retrieved from the World Bank World Development Indicators dataset. This data covers all countries in the analysis for the entire period 1960-2010.

A re-expression of the urbanization ratio in terms of change over each five-year wage period is done in the same way as for ginis and schooling, using the log-ratio of urbanization at end and start year.

Internal and international armed conflict

Voluntary labour migration can be expected to have different effects on wages than forced displacement as a consequence of war or other disaster. As violent conflict is one of the main causes of forced displacement, these different forms of migration will be distinguished by adding a covariant to the analysis indicating internal and international conflict. This indicator comes in the form of a yearly binary conflict indicator for both internal and international conflict, based on Peter Brecke's Conflict Catalog as

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

presented on CLIO-INFRA. Conflicts are defined here as armed conflicts, revolts, revolutions, uprisings, rebellions and civil wars with at least 32 fatalities.[59]

In the analysis the indicator will be added in summed form; denoting the number of years with internal or international conflicts over a specific time interval. It will be added both as an interaction term with immigration and emigration, and independently. The independent indicator will denote conflict in or including the same country as the occupational groups, while the interaction terms contain conflict information for the sending and receiving regions respectively. For the period 1960-2000 the Conflict Catalog has good coverage, but it can therefore only be applied on a subset of the wage data. As the availability of this indicator is so different from that of other covariates, which usually have high availability after 1990, no models can be made that include both conflict and for example remittances.

The categorical conflict indicator is treated in a similar way as other (binary) dummies in the model.

Finally, a 'health' covariate consisting of average life expectancy was initially also included, but had no added value in explaining wage growth and is therefore not discussed in detail here.

59 A description of the Conflict Catalog ("Violent Conflicts 1400 A.D. to the Present in Different Regions of the World") is presented in: Peter Brecke, "The Long-Term Patterns of Violent Conflict in Different Regions of the World" *Uppsala Conflict Data Conference* (June 2001) Web.

Regression analysis

The various data sources described above are combined in a regression model for each of the seven income-migration clusters and for each occupational group separately. In these models wage growth for a certain occupational group in a certain cluster is explained by immigration, emigration, internal and international conflict, remittances, schooling, urbanization and economic inequality. By comparing the results from similar models but for different occupational groups an assessment is made of how migration affects economic (wage) inequality within each cluster.

The regression model has the following general form:[60]

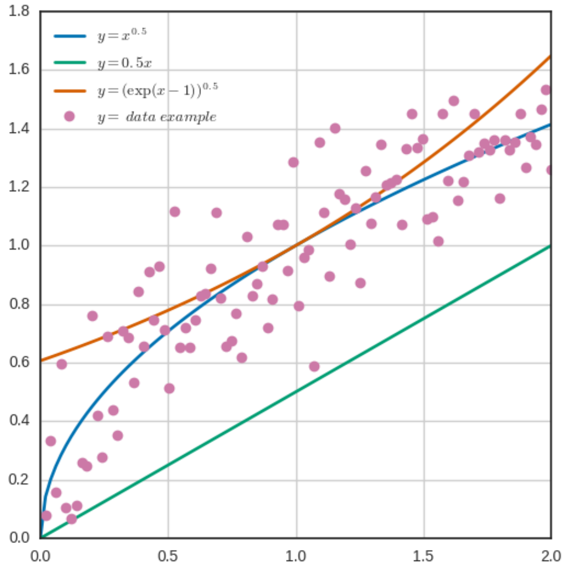
$$
\begin{aligned}
log\left(\frac{y^k_{n_2}}{y^k_{n_1}}\right) &= \beta_0 + \sum_{i=1}^{6} \beta_i \cdot \left(\sum_{j=n_1}^{n_2} log(migr_{j,i})\right) + \beta_7 \cdot log\left(\frac{gini_{n_2}}{gini_{n_1}}\right) \\
&+ \sum_{i=8}^{11} \beta_i \cdot \left(\sum_{j=n_1}^{n_2} conflict_{j,i}\right) + \beta_{12} \cdot \left(\sum_{j=n_1}^{n_2} remit\_in_{j,i}\right) + \beta_{13} \cdot \left(\sum_{j=n_1}^{n_2} remit\_out_{j,i}\right) \\
&+ \beta_{14} \cdot log\left(\frac{school_{n_2-6}}{school_{n_1-6}}\right) + \beta_{15} \cdot log\left(\frac{urban_{n_2}}{urban_{n_1}}\right) + \sum_{i=16}^{18} \beta_i \cdot period_i \ + \sigma^2_k
\end{aligned}
$$

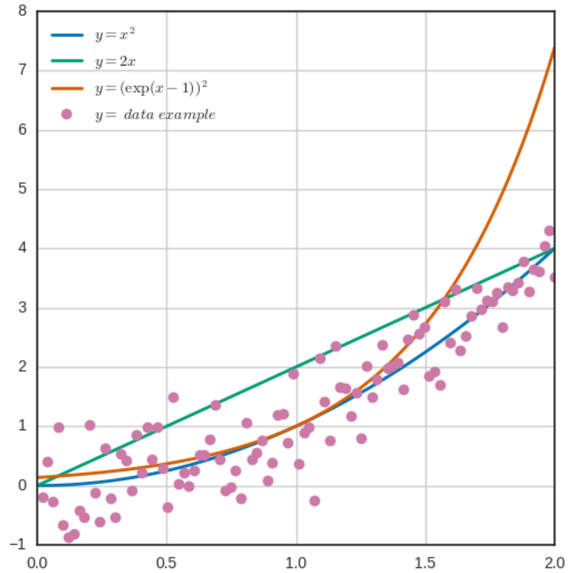
The dependent variable $y^k_{n_2}$ is the 'wage at the end year of period (n_2) from type *k*', with the type being one of the three skill levels, and for this skill level one of the seven clusters. $migr_i$ refers to one of the six migration types: immigration from low-income countries, immigration mid-income, immigration high-income, emigration to low-income countries, emigration mid-income and emigration high-income. $conflict_i$ refers to one of the four conflict types: current internal conflict, current international conflict, internal conflict (lagged), international conflict (lagged), where lagged means at the time of migration and 6 years before wage growth. $school$ refers to lagged schooling as well, so changes in schooling from 1968 to 1974 are used independently or interacted with migration (at the same time) to explain wage growth from 1974 to 1980, while changes in gini are used from the period 1974-1980. $period_i$ is one of the three decades: 1970s, 1990s or 2000s; the 1980s are used as baseline. The wage growth ratios are considered to belong to a certain decade if the first year (n_1) belongs to that decade.

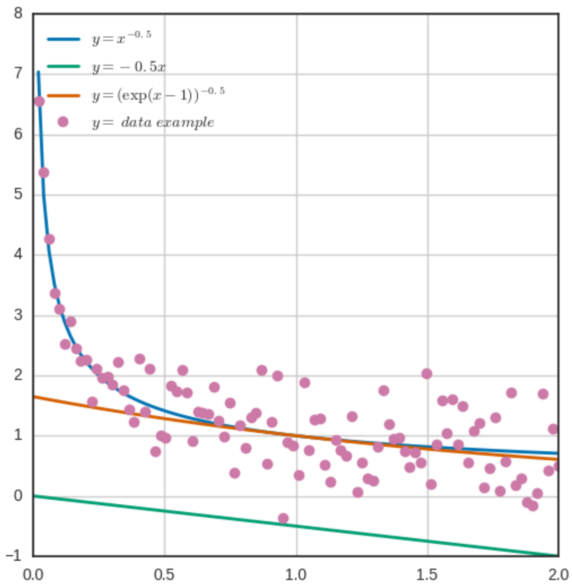
Depending on cluster, a basic model and two extended models are analysed in detail: all indicators described above are considered, but only the 'best' models are analysed. This selection is made based on data availability, Akaike information criterion (AIC, for non-nested models) and F-tests for nested models. The basic model includes only migration and period dummies, while the extended models include either one or two 'sets' of covariates from the following four: remittances, conflict, gini + urbanization, and schooling.

60 In subsequent formulas σ^2_k is omitted, so that y always refers to predicted, as opposed to actual, wage growth.

On top of the above covariates, several interactions between migration and conflict, remittances and schooling are added to the models. For this reason, all covariates are standardized to have $\mu = 0$ and $\sigma = 1$. With this set-up the coefficient for the intercept, β_0, can be interpreted as the expected geometric mean of wage growth within a cluster and occupational category when all covariates are at their (arithmetic) mean. If covariates would not be standardized, the coefficient for the intercept would instead refer to expected wage growth if all covariates were zero, while most covariates are never zero. Thus, the value of β_0 would be meaningless.







Wage growth is re-expressed using a logarithmic transformation because it is expected to have an exponential instead of linear development over time. For example, a twice as long period is not expected to have a twice as large growth. Assume that growth is 2% per year, then growth over a period of 10 years is 22% and not 21%, which is twice the growth over a 5 year period:

$(1.02)^{10} = 1.22 \neq 1.208 = 1 + 2 \cdot ((1.02)^5 - 1)$.

However, if after this transformation all covariates would be expressed in their original form, this would imply assuming each unit change in these covariates would mean a fixed growth of wage growth. For example, it would be assumed that if urbanization went from 2 to 3% (everything else constant) and its coefficient in the regression was 0.1, then wage growth would increase by $e^{0.1} - 1 = 0.105$ and go from its base value of for example 1.05 (5% growth), to $1.05 \cdot 1.105 = 1.16$ (16% growth), while the exact same would be true for urbanization going from 95 to 96%.

Instead, all continuously distributed variables are added to the model in logarithmic form as well – except for remittances, because for remittances taking the logarithm is equivalent to the original. However, period dummies and conflict are added without a logarithmic re-expression: as explained before wage growth is expected to be exponential over time, while for conflict, too, it is expected that the effect of 6 instead of 5 years conflict has approximately an equally large effect on wage growth as 2 instead of 1 year of conflict.

When both the dependent variable, y, and the independent variables x_i are re-expressed using a logarithm, the relationship between y and x in their original form (before re-expression) is of the form:

$$y = {x_1}^{\beta_1} \cdot {x_2}^{\beta_2} \cdot {x_3}^{\beta_3} + \text{rest}$$

Where one of the covariates x_i may, for example, refer to gini change over a certain period, so that if gini change over period one is twice as large as gini change over period two (everything else constant), wage growth is expected to be 2^{β_i} times as large in period two as in period one. This is similar to a simple linear model, as a fixed increase of x gives a fixed increase of y in both cases. In a linear model we even have the exact same increase for both y and x: if between datapoint 0 and 1 $x_1(0) = a \text{ and } x_1(1) = 2 \cdot a$, with all other covariates equal, then $y_0(0) = b \text{ and } y_0(1) = 2 \cdot b$.

In the model described above, just like in linear regression, coefficients <0 mean a negative correlation between a covariate and the dependent variable: if $\beta < 0$ wages are expected to decline when the covariate increases, while if $\beta > 0$ wages are expected to increase when the covariate increases. Further, if $|\beta| < 1$ we expect that a unit increase (or decline) in the covariate gives an increase (or decline) of < 1 in wages, while if $|\beta| > 1$ we expect that a unit increase in the covariate gives an increase of > 1 in wages.

The figures at the left of this page give a visual impression of the effect of the re-expressions used for a fictive dependent variable y and an independent variable x with neutral intercept and coefficient $\beta = 0.5, \ \beta = 2 \text{ and } \beta = -0.5$ respectively. The blue line shows how this relationship translates to a regression line when both dependent and independent variable have been re-expressed through a logarithm, the green line shows a traditional regression with no re-expression of either x or y, while the orange line shows what the relationship looks like if y is re-expressed through a logarithm but x is not re-expressed. Most logarithmic re-expressions in this thesis add one to the original data series before taking the logarithm, therefore, in the figure one is distracted for the x-series that is not re-expressed (red line), to make the series more comparable to the real data. In this example the fictive data would be best explained when both x and y are re-expressed using a logarithm (blue line). In the bottom figure you also see one reason why x should be larger than zero if log re-expression is used, since we see our predicted y (blue) going to infinity when x goes to zero.

Moreover, a main difference between linear regression and the model used here, with both dependent and independent variables re-expressed using logarithms, is that if any covariate x can be equal to zero, then y is predicted to be zero whenever x is zero. Thus, one was added to all migration covariates (so that their minimum value is one), and all other covariates are inherently larger than zero.

Note that the relationships in the figures refer to the relationship between original variables, and not the relationships between variables in their re-expressed form: the latter relationship is always like the green line, since an OLS regression model is learned. An interval of $x \in [0, 2]$ is chosen because most covariates take on values between zero and some small positive number.

Finally, since all covariates are standardized (after all other manipulations such as re-expression using logarithm), and wages are centred, our coefficients, for one covariate x, refer to the relationship:

$(\text{y-}\mu_y) = ((x - \mu_x)/\sigma_x)^{\beta} \cdot\alpha$

where α is a constant referring to all other covariates (kept constant), so that the predicted value of y before standardization is:

$\hat{\text{y}} = \alpha \cdot (\frac{x-\mu_x}{\sigma_x})^{\beta} + \mu_y$

where, in those instances where x is not re-expressed logarithmically, $x, \mu_x, \sigma_x \text{ refer to } \exp(x), \exp(\mu_x), \exp(\sigma_x)$.

Standardization has the advantage of making all coefficients comparable: if $\beta_1 > \beta_2$, then the expected effect of x_1 is larger than the expected effect of x_2. Moreover, it is essential in the interaction terms. With standardized values both terms in the interaction have equal force, while if one interaction term would have a larger range than the other, then its value would weight more heavily in the interaction.

Now that we have seen the general features of the model used, we can discuss the results that it yielded in the analysis of the relationship between migration and wage growth and infer which consequences this has for the relationship between migration and global inequality.

Chapter 3: Results

This thesis assesses if and how migration flows have affected global inequality during the latter half of the twentieth century. To answer this question, the relationship between wage growth of different occupational groups and multiple types of migration flows is analyzed. In this chapter the results of this analysis will be discussed. First, the selection of models per cluster will be explained. Secondly, some general results are presented, such as how an analysis separated by clusters performs as compared to an analysis where all global data is combined. The relative quality of the results for different clusters will also be addressed here.

Once the main setup and general results are presented, we can consider basic results for emigration and immigration separately. Not all covariates will come to the fore in this discussion, but their basic features are presented in a comprehensive table. In this section the assumption derived from factor price theory, that emigration generally increases wages and immigration decreases wages, is tested. Do we see this trend in real world data, or is it obscured by other influences? An assessment will be made on whether, for each cluster, there are more and stronger negative or positive effects of migration.

After this discussion, the results will be aggregated in a general conclusion. Here we will first assess whether the effects of migration appear to differ depending on the income level of the sending or receiving country. If any general trends can be distinguished, these will be used to answer the question how migration flows have affected global inequality.

Model selection

For each cluster and occupational category two models beside a base model including only migration were selected for further analysis. In this section the considerations in this selection process are outlined. Although a high number of observations were gathered overall, the number of observations per cluster varies. For clusters with relatively few observations, it was not possible to include all covariates in one model without greatly increasing the risk of overfitting. To avoid overfitting, generally no models were used where the number of covariates was greater than one-tenth of the number of observations.

Although no conventional form of cross-validation was done, for each cluster different models – including different explanatory variables – were tested. The observations available for remittances data often had little overlap with the observations available for conflict data, as remittances were mainly available for the later period (for all but the 'Western World' clusters), while the conflict indicator stops in the year 2000. Migration coefficients and their p-values were compared between models including conflict and models including remittances (and usually one other model). Only correlations between a specific type of migration and wage growth that were consistent over all models are considered 'valid'. Correlations are considered 'consistent over all models' if they are in the same direction in all models and significant in at least two out of three models, or if they were significant in the 'best' model (with significantly better AIC than the other models for that cluster-wage group) and there was a valid explanation for them not being significant in other models. Such a 'valid explanation' is usually that a migration covariate is significant in one direction independently, but significant in another direction in interaction with some other explanatory variable.

For certain clusters, mainly the 'Western World' and 'European Frontier', it would have been possible to include more covariates at once than was done now. To be able to compare the results between different clusters, however, models should be as similar as possible. Therefore similar models were used for all clusters were possible.

For each extension of the 'base' model, which includes only migration and period effects, an F-test for nested models was done to test whether the extension has significantly more explanatory power than the base model. This F-test is similar to the F-statistic generally used for OLS regression to assess whether the entire model is significantly different from the null hypothesis. The basic F-test for OLS has as its null hypothesis that the dependent variable is best explained by a constant line (the intercept), and a model is 'significant' if this null hypothesis can be rejected with 95% (or some other threshold) certainty. In my test for nested models the null hypothesis was instead that the dependent variable is best explained by the base model with only migration and period effects. This hypothesis is again rejected if we know with 95% certainty that there is some added explanatory power in the extra covariates of the extended model. The test statistic used is:

$$F(q, n-p) = \frac{SS_{res_0} - SS_{res}}{SS_{res}} \cdot \frac{n-p}{q}$$

Where SS_{res_0} is the residual sum of squares of the base model, SS_{res} the residual sum of squares of the extended model, $n-p$ the number of observations minus the number of covariates in the base model and q the number of covariates in the extended model minus the number of covariates in the base model. $F(q, n-p)$ has a standard F-distribution so that it's p-value can be calculated as for a basic OLS F-test.

The main results of these tests were that almost any extension of the base model was statistically significant – and unsurprisingly so as zero migration never explains any fluctuation in wage growth, so that all base models had large variance of the residual around zero. Notably exceptions were an extension by adding internal and international conflict and its interactions with migration, which had no added value in explaining wage development in the Western World for any occupational group. Schooling had no added value in explaining wage development of *high-skilled workers* in the Western World, possibly because the schooling indicator does not capture differences between 'good' and 'better' levels of schooling, but rather whether relatively basic schooling is available. Schooling had no added value in the unskilled wage model for South and Southeast Asia either.

While remittances improved the explanatory models for all clusters, urbanization and economic inequality (gini) had relatively little added value for several clusters: South and Southeast Asia (high-skilled) and the 'Geographical Mix' cluster (both unskilled and skilled). Even where urbanization, gini and schooling did improve the base model significantly, in most clusters they had less added value as conflict and remittances. Non-nested models, i.e. the models with different covariates added, are compared using the 'Akaike information criterion' (AIC). The AIC is defined as $2k - 2log(L)$, where k is the number of parameters of a function and L the likelihood function. If the AIC is considerably lower for model A than for model B, than model A generally does a 'better' job at explaining the dependent variable than model B. Its main advantage over the F-test is that it 'penalizes' a large number of covariates, as those make the AIC larger and thus worse.

The AIC values for all models explored in this thesis can be found in attachment 3.1. Here the results of an analysis done which included remittances and urbanization and gini are not shown, but the AICs for this combination were generally lower than those for a combination of remittances and schooling. As the models including conflict or remittances (a combination was not possible due to varying availability) had most low AIC values, these were generally chosen for further analysis. Whenever this was possible based on number of observations and 'at all feasible', a model including both remittances and schooling was used instead of a model including only remittances. At all feasible here means that the model with only schooling added does a better job in explaining wage growth than the base model.

Finally, for all models residual plots were analyzed to control for any non-randomness in the errors of the predicted wage growth. No clear patterns could be distinguished in the residuals beside the earlier mentioned problem of heteroskedasticity (the errors not being distributed equally widely around zero for all values of the explanatory variable) in the case of migration. Heteroskedasticity has been reduced by logarithmic re-expression, but cannot entirely be avoided.

General results

In this section we will look at a rather overwhelming table that includes the most basic features of the selected models for each cluster. This table will be explained with a focus on comparing global wage growth models to the models for separate clusters.

The below table contains three columns for each cluster: one for every occupational group. The 'European frontier' cluster is presented twice. For this cluster results in the conflict model and the remittances model were rather different. As all observations for the conflict model are from the period before 2000, while all observations for the successor states of Yugoslavia and the USSR in the remittances model are from the period after 1991, these deviations between the two models were not considered as model inconsistency but rather a direct consequence of the very different subsamples contained in the models. Therefore both models and presented separately.

In each column of the table the coefficients that were 'consistent and significant' (consistent in the sense of the previous section, surviving 'cross-validation'), are marked with an orange or blue color

TABLE 3.1: SUMMARY OF REGRESSION MODELS

	Cluster	0 – 'Africa'			1 – 'South&Southeast Asia'			2 – 'Geographical Mix'			3 - 'European frontier 1'			3 - 'European frontier 2'		
	Wage growth of .. labourers	unskill.	skill.	high-sk.	unskill.	skill.	high-sk.	unskill.	skill.	high-sk.	unskill.	skill.	high-sk.	unskill.	skill.	high-sk.
Covariates	Gini			too												
	Urban			few												
	School (-6)			data												
Conflict	Internal															
	International	conflict														
	Internal (-6)	no measurable effect														
	International (-6)															
Remittances	Paid															
	Received															
Immigration	Low-income															
(from)	Mid-income															
	High-income															
Emigration	Low-income															
(to)	Mid-income															
	High-income															
Migration *	School6*Immig Low															
Education	School6*Emig Mid															
	School6*Emig High															
Migration *	Internal6*Emig Low															
Conflict	Internal6*Emig Mid													later data (mostly post-		
	Internal6*Emig High													1991) remittances and		
	International6*Emig Mid													schooling available		
	International6*Emig High															
Migration *	Paid*Immig Low															
Remittances	Paid*Immig Mid										earlier data (at most until					
	Paid*Immig High										2000) conflict available					
	Received*Emig Low															
	Received*Emig Mid															
	Received*Emig High															
	No. of datapoints (migration and wages)	**234**	**366**	**135**	**333**	**419**	**173**	**337**	**440**	**173**	**610**	**704**	**201**	**610**	**704**	**201**
	Mean 5y-wage growth (%)	14.556	10.472	17.073	8.492	15.031	14.596	11.363	11.168	25.953	9.432	19.264	25.605	22.968	26.571	17.145
Main statistics	R^2	0.537	0.160	0.522	0.390	0.286	0.602	0.347	0.249	0.477	0.247	0.218	0.553	0.194	0.203	0.574
Remittances	R^2 adjusted	0.457	0.063	0.432	0.304	0.200	0.505	0.264	0.174	0.347	0.215	0.190	0.480	0.138	0.148	0.512
and schooling	p-value F-statistic	<<0.05	0.047	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05
	AIC (=2k – 2ln(L))	119.6	249.4	154.9	103.2	204.5	154.3	118.7	253.1	87.9	467.9	506.6	0.5	37.3	23.4	-89.0
	k=#covar.; L=likelihood	If no overfitting, then the lower the AIC the better the model (comparatively)									*statistics for conflict					
	Color coding:	negative correlation			positive correlation			no significant and/or consistent correlation						Not in model(s)		

SUMMARY OF REGRESSION MODELS – CONTINUED

	Cluster	4 - 'Lat. America&Caribbean'			5 - 'Arabic World'			6 - 'Western World'			X- Global		
	Wage growth of .. labourers	unskill.	skill.	high-sk.	unskill.	skill.	high-sk.	unskill.	skill.	high-sk.	unskill.	skill.	high-sk.
Covariates	Gini				all models are			?	?	?			
	Urban				unreliable and								
	School (-6)				inconsistent						life +	life +	
Conflict	Internal												
	International	conflict						conflict					
	Internal (-6)	no measurable effect						no measurable effect					
	International (-6)												
Remittances	Paid												
	Received												
Immigration	Low-income												
(from)	Mid-income												
	High-income												
Emigration	Low-income												
(to)	Mid-income												
	High-income												
Migration *	School6*Immig Low												
Education	School6*Emig Mid				schooling			schooling					
	School6*Emig High				no measurable effect			no measurable effect					
Migration *	Internal6*Emig Low												
Conflict	Internal6*Emig Mid												
	Internal6*Emig High												
	International6*Emig Mid												
	International6*Emig High												
Migration *	Paid*Immig Low												
Remittances	Paid*Immig Mid						too						
	Paid*Immig High						few						
	Received*Emig Low						data						
	Received*Emig Mid												
	Received*Emig High												
	No. of datapoints (migration and wages)	**437**	**419**	**191**	**117 (86)**	**153(100)**	**71 (49)**	**1053**	**1024**	**463**	**3121**	**3525**	**1407**
	Mean 5y-wage growth (%)	7.853	9.776	17.017	3.318	6.820	-13.345	6.844	5.608	4.673	9.565	10.305	11.860
Main statistics	R^2	0.410	0.501	0.652	0.698	0.398	0.700	0.142	0.159	0.264	0.066	0.060	0.119
Remittances	R^2 adjusted	0.337	0.435	0.572	0.623	0.274	0.563	0.125	0.142	0.235	0.056	0.051	0.103
and schooling	p-value F-statistic	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05	<<0.05
	AIC (=2k – 2ln(L))	47.3	99.2	47.4	6.9	-23.7	-0.4	-850.9	-1123.0	-428.7	493.4	971.2	857.6
	k=#covar.; L=likelihood				*statistics for conflict			*statistics for remittances					
	Color coding:	negative correlation			positive correlation			no significant and/or consistent correlation					

depending on whether they were negative of positive. If certain covariates did had no significant or no consistent correlation with wage growth in the respective cluster and occupational group then their cell is left blank. If coefficients were not included because entire models including those coefficients were either not significant, not reliable (e.g. too few observations), had no significant coefficients and little explanatory power, or only less good than other possible models, their cells are colored gray.

Below these color markings the number of observations for each cluster-occupation is given *for the base model.* Thus, this number can – and usually is – lower in an extended model if a covariate is not available for all observations in the base model. For the 'Arabic World' cluster the number of observations in the conflict model is given in parenthesis to indicate why all models for this cluster are deemed 'unreliable'. Number of observations for the remittances model is not given, but varies around fifty. All color codings for the Arabic World are shown in a lighter color because of their unreliability as well. In this cluster coefficients often changed sign or experienced other dramatical changes when minor adaptations were made to the explanatory model. It's R^2 and R^2 adjusted were unrealistically high in most models. Combined, these results show that models for the Arabic World were almost certainly severely overfitting any existing trends. Therefore results for this cluster will not be discussed in the subsequent part of this chapter.

The final part of each column contains some main statistics for the remittances and schooling model for the respective cluster and occupational group. First R^2 and its adjusted version, indicating approximately how much of the variation in wage growth is explained by the covariates in the model, are presented. We will rely on the adjusted R^2 to get an indication of the explained variance of each model, as this measure penalizes the number of covariates used and thus is less likely to reward overfitting. However, as $R^2 - R^2 adj.$ increases when the number of observations get smaller or the number of covariates gets larger, the differences between the two measures gives us an indication of the risk of overfitting (as too many covariates for a certain number of observations necessarily leads to overfitting).

The p-value of the F-statistic for each model is given, but is not very informative because all models are significant, $<< 0.05$ is written instead of the actual value when this p-value is smaller than 0.00001. Finally the AIC, used to select models, is also included.

The last three columns of the table include coefficients and statistics on a global model for each occupational group, in which all observations are combined. As this model has a sufficient amount of observations all possible combinations of covariates were assessed. Global base models (including only migration and period effects) had an adjusted R^2 ranging from 0.5% (for unskilled wage growth) to 1.6% (for skilled wage growth), extremely low likelihood and high AIC. When gini, urbanization and life expectancy were included all models still had a very low explained variance (<3%) and high AIC (far over 1000). Conflict models had slightly higher R^2 values, with a highest value of 6.4% for high-skilled wages, but higher AICs than gini models. The statistics presented in table 3.1 are for global models including remittances and schooling, which seem a considerable improvement on the other global models in terms of AIC as well as R^2.

The clustered models all seem to have much stronger explanatory power than these global models, although it should be noted that in these models the ratio between number of covariates and number of observations is always considerably higher. Thus, both AIC and R^2 will be higher regardless of the 'real' quality of the models. To take this into account the R^2 values for the global model were multiplied with $\sqrt{Nr.\ global/Nr.\ cluster}$ - the square root of the number of observations globally, divided by the number of observations per cluster. This seems a rather extreme adjustment, assuming a strong inherent decline in R^2 when more observations are added. The value of the global R^2, adjusted to the number of observations for each cluster, was then compared to the R^2 for that cluster. Even after this adjustment, only the skilled wage model for the African cluster explained less of the variation in wage growth than the global model. Thus, it seems likely that splitting up the analysis in clusters adds explanatory power to the models and that relationships between migration and wage growth are local rather than global.

The scope of this thesis does not permit a thorough discussion and research into possible justifications for each correlation presented in table 3.1. Instead, with a focus on first emigration and then immigration results will be discussed in a more general way. The focus will be on whether or not there exist differences in the effects of migration between occupational groups, and whether differences exist between migration from different income clusters. After this discussion, an attempt is made to answer the general research question. Full summary statistics for the various models on which this analysis is based can be found in the attachments 3.2 to 3.9.

Emigration

One of the assumptions made going into the analysis, was that emigration might have a positive effect on wage growth as it decreases the labour force and thus makes workers more valuable. If we look at the global model, most emigration covariates indeed have a positive effect on wages. Independently, as well as in relation with conflict, emigration only has positive coefficients. It seems to matter little whether people leave a conflicted or peaceful area, wages become higher with more people emigrating. However, in interaction with schooling emigration to mid- and high-income countries has negative effects on skilled and high-skilled wages. In the global model emigration to high-income countries also negatively affects unskilled and skilled wage growth when interacted with received remittances. In most clusters, however, there is no significant effect of received remittances interacted with emigration to high-income countries. This global effect seems entirely due to a strong such effect in Latin America and the Caribbean.

The negative global effect of more emigration to high-income countries interacted with a higher level of schooling on high-skilled wages is only found in the South and Southeast Asia cluster, while the same interaction actually appears to have a positive effect in Latin America. These differences could for example be due to different types of migration; temporary emigration for educational purposes versus a form of permanent ‘brain drain’, although the current analysis cannot determine this. Similarly, the negative global effect of more emigration to mid-income countries interacted with increases in schooling levels on skilled wages is only found for the ‘European frontier’.

Three out of four independent positive emigration effects found in the global model are negative in at least one of the clusters. In both the Latin American and European Frontier cluster, more of the emigration effects found are negative than positive. Independent emigration effects are all negative in the ‘Geographical Mix’ cluster too, which includes mostly Latin American countries, but there all interactions between emigration and conflict, schooling and received remittances are positive. In the South and Southeast Asian cluster most positive effects of emigration are found. However, here emigration to high-income countries has a negative effect on unskilled wage growth. As its effects on both skilled and high-skilled wage growth are positive, it might be a driver of inequality in this cluster.

In the Western World no emigration effects on unskilled wages are found, and in other clusters too emigration seems to affect unskilled wages less often than skilled and high-skilled wages. Both in the Western World and European Frontier emigration to mid-income countries appears to have a negative effect on high-skilled wages, but a positive effect on high-skilled wages arises for emigration to high-income countries – both independently and in interaction with received remittances. These results seem rather contradictory, and would possibly be easier to explain if the relationship would be in the other direction: people emigrating more often to mid-income countries when wages decline, and more often to high-income countries when wages improve. However, since the emigration rates are from six years before wage development such a reversed explanation seems unlikely.

Surprisingly, the correlation between emigration interacted with internal or international conflict and wage growth seems to be rather more than less positive than the effect of emigration independently - which in the model including the interaction term consists of emigration from non-conflicted areas. Although in the global model clearly positive effects of emigration dominate, when looking at separate clusters emigration seems to influence wages in various ways, and no universal positive effect can be distinguished.

Overall, there is no clear difference between the occupational clusters in the effect that emigration has on their wage development. Similarly, emigration to regions with different income levels does not seem to affect wage development differently. The clusters differ from each other in which groups are affected negatively and positively by emigration. Slightly more often significant results are measured for the high-skilled category, with the majority of these results being positive. Skilled wages experience almost equally many positive as negative effects from emigration, while unskilled wages again more often are affected positively by emigration. Emigration to mid-income countries negatively affects wage growth more often than that it has a positive effect, while both emigration to low-income and emigration to high-income countries are more often positively correlated to wage growth.

Immigration

For immigration interactions with conflict were not analyzed as the correct data was not created; interactions between conflict in the destination countries and immigration was considered relatively meaningless, but immigration clusters contain many different origins, of which some would be in conflict at any point in time, while others from the same cluster would not experience conflict. The most reasonable way to take the interaction between conflict in countries of origin and immigration into account would be to separate migration further; not only based on income, but also based on conflict.

In the global model no effects of immigration from low-income countries are found, while immigration from mid-income countries has a positive effect on skilled wages and immigration from high-income countries has a negative effect on both unskilled and high-skilled wages. No forms of immigration seem to be related to high-skilled wage growth. No interaction effects of immigration with schooling or paid remittances were found in the global model.

The negative effects found globally of immigration from high-income countries seem to exist largely for immigration between high-income countries, as these same effects were found for the Western World. In the Western World immigration from high-income countries also had a *positive* effect on high-skilled wage growth, so that migration between countries in the western world might cause rising inequality within the western world. This idea is further supported by the negative effect of immigration from low-income countries interacted with paid remittances on unskilled wages: the same type of immigration interacted with remittances is positively correlated with high-skilled wage growth. Thus, in the western world all immigration effects are positive for high-skilled wage development, while most of them are negative for wage growth in the other two groups.

The African cluster is the only cluster where such a contradictory effect of the same type of immigration on different wage groups also occurred: here immigration from mid-income countries is negatively correlated with unskilled wage growth, while it is positively correlated with skilled wage growth. However, here the positive effect on skilled wage growth is more than nullified by the negative interaction term for this type of immigration with paid remittances. The independent term is only 0.074, while the interaction term is -0.173.

Contrary to the global and western effects of immigration from high-income countries, such immigration has positive effects on both unskilled and skilled wages in the Geographical Mix cluster – the cluster including amongst others China and low-income Latin American countries. These positive effects of immigration from high-income countries in the Geographical Mix however had a smaller effect on wage growth than the negative effects of immigration from high-income countries interacted with paid remittances, so that the overall effect of immigration from high-income countries might be similar to that in the western world and Latin America. Such counteracting effects were not found in any other clusters. Usually immigration effects interacted with paid remittances were only significant where immigration effects themselves were not.

In the European Frontier immigration from high-income countries had a positive effect on unskilled wages in the earlier decades (conflict model), but a negative effect in the 1990s and 2000s (remittances model). In these later years it also had a negative effect on high-skilled wages, while immigration from low-income countries affected all wages negatively in this cluster. The European Frontier around the turn of the twenty-first century is unique in this respect: no other clusters show a negative correlation between immigration from low-income countries and wage growth at any level. In South and Southeast Asia immigration from low-income countries is even positively correlated to skilled wage growth.

While no interaction effects were found for immigration globally, higher immigration from low-income countries together with higher levels of schooling had a positive effect on unskilled and skilled wages in Latin America. Such an effect could be explained if higher levels of schooling mean that immigrants from low-income countries form a complement instead of competition to the existing labour force in a country. A similar positive correlation was found but then for high-skilled labourers in South and Southeast Asia.

Overall, the effects of immigration seem more variated than those of emigration and it can certainly not be said that they usually affect wages in a negative way. Most negative immigration effects were due to immigration from high-income countries, while immigration from low-income countries only appears to affect wages in the European Frontier cluster negatively. Immigration from mid-income countries usually has no significant effect on wages, although interacted with paid remittances it appears to often affect wages in a positive way. However, negative immigration effects were most often found on wages of unskilled workers, and immigration from high-income countries was most likely to affect unskilled wages negatively. Especially in higher income countries immigration from other high- or mid-income countries thus generally seems to affect the lower income groups negatively, possibly increasing inequality in those higher income countries.

Other covariates

The relationship between migration and wage flows showed quite a lot of variation between clusters. The analysis by clusters showed that only looking at a global model could lead to misguided conclusions since seemingly global trends are often the result of a relationship that exists in few world-regions. For most of the other covariates, however, such a clear distinction between effects for different clusters did not exist. This is not entirely surprising, as the clusters are created in such a way that they differ between each other and are similar among themselves in terms of migration. Thus, they are designed in such a way that if any particular migration features are related to wage growth in a specific way, this should come to the fore in one of these clusters.

One of the few other covariates that did not have the same effect in all clusters is received remittances. These are positively related to wage growth in most clusters, but negatively in the western world. Intuitively, a negative correlation was expected here: if more income is generated by other means, as for example remittances, the need for higher wages would be reduced. Apparently, this is not true for South and Southeast Asia (one of the regions in the world which receives the highest remittances as a share of GDP), or for the European frontier (with similarly high rates of received remittances).

Paid remittances are rarely directly related to wage growth, but they are in two of the clusters with a relatively high income – Latin America and the Western World – and there they affect only unskilled wages negatively. Somehow they globally were found to affect skilled wages negatively.

Violent conflict at the time of migration, and especially international conflict, appears to often be related positively to wage growth. This is again an unexpected result, but it is largely explained by the effect of current international conflict. Current conflict usually is negatively related to wage growth. One can imagine a scenario here where wages increase when a country recently (but no longer) had been involved in a violent international conflict, but decrease again if and when conflict reappears. Another interesting result considering conflict is that it appears to affect unskilled wages more often than skilled wages, and skilled wages again more often than high-skilled wages.

Finally, where these indicators are significantly related to wage growth urbanization and schooling always affect it positively. Economic inequality appears to nowhere affect the different occupational groups differently, but instead is negatively correlated with wage growth almost everywhere. The exception here again being the western world, which shows peculiar results: increases in inequality appear to affect unskilled and skilled wages positively, while they are negatively correlated to high-skilled wage growth.

Conclusion and Discussion

In this thesis the effect of migration on global inequality was studied. Based on existing literature, a general assumption could be made on this relationship. Factor price theory, and basic notions on supply and demand, would suggest that migration should be a driver of global convergence when decisions to migrate are taken on an economic basis. However, not all forms of migration are driven economically, and economical preferences to migrate can not always be realized in a world where labour mobility is often restricted to national borders through legislation. It was therefore difficult to make assumptions on the global relationship between migration and global inequality based on literature alone, and unclear whether any such universal relationship exists. In the analysis of this thesis, the world was thus not taken as one unit of analysis, but divided into world-regions.

To measure the effect of migration on global inequality empirically, an analysis was done of the relationship between migration and wage growth. Both wage growth and migration were disaggregated into three different types. To measure the relationship between migration and between-country inequality, different types of global bilateral migration flows were distinguished based on the the average income of countries of origin – in case of immigration, and countries of destination – in case of emigration. To measure the effect of migration on within-country inequality, the wage incomes were broken down into three occupational groups within each country.

As the position in the global income distribution of an unskilled worker in for example Bangladesh is rather different from the position of an unskilled worker in Austria, wage data was further disaggregated into world-regions, or clusters. These clusters were created to be similar in terms of national income and immigration patterns. Each of the occupational groups in each of these world-regions was analyzed separately, so that the units of analysis differ both in terms of within as between-country inequality.

The regression models suggest that migration flows have a considerable effect on wage growth, as generally speaking a considerable proportion of the variation in wage growth could be explained by migration and correlated factors, most importantly remittances. This simple finding constitutes one of the main conclusions of this thesis, as the existence and size of such a relation was not clear before.

In a comparison of the relationship between immigration and wage growth in each specific occupational group, we saw that wages of unskilled workers were more often negatively related to immigration than positively, and that these negative effects usually were due to immigration from high-income countries. Relationships between immigration and skilled or high-skilled wages were less often found, although in the western world specifically immigration from high-income countries appears to increase within-country inequality, as it has a negative correlation with both skilled and unskilled wages, but a positive correlation with high-skilled wages.

Immigration from countries with a lower income did not show any global trend in its effect on wage growth, although specifically in the 'European frontier' area immigration from low-income countries may effect wages negatively. Further, most negative effects of immigration were found in the Western World and European Frontier, while lower income clusters often were more positively affected by immigration. It would be interesting to study these geographical differences more thoroughly, and consider the various possible cultural, political, demographical and market explanations.

Although a global model showed that emigration overall was positively correlated with wage growth at any occupational level, both independently and in interaction with conflict, these results became less obvious when distinguished along geographical lines. While in some clusters the effects appear to be positive, as in South and South-East Asia and the Western World, there were suggestions of negative effects as well, as in Latin America. In contrast to immigration, emigration more often had a significant effect on high-skilled wages than on the other occupational groups, and this effect was usually positive.

To contextualize and verify the found correlations, they would have to be related to other factors which can affect wage growth that lay beyond the scope of this thesis. For example, national unemployment rates could be used as an indication of labour demand in different clusters. Differences in labour demand might explain several of the differences in results between world-regions.

Overall, the results of the models created in this thesis suggest that migration flows might drive within-country inequality: immigration appears to affect unskilled wages negatively more often than is does in other occupational groups, especially in higher-income world-regions. Emigration, on the other hand, affects high-skilled wages positively more often than skilled and unskilled wages. In terms of between-country inequality, the results might indicate that migration diminishes between-country inequality, as immigration has more negative effects on wage growth in the European Frontier region and the Western World than in other clusters.

However, to draw any definitive conclusions, further research should be done on the correlations found in this thesis. Here, a great number of explanatory analyses was done, but for definitive conclusions each of the found results should be validated by contextual analysis, studying the cultural, political and economic context of each of the clusters in more depth. This thesis thus presents an overview of general trends that come to the fore in a quantitative analysis of global wage and migration data, but cannot analyze the stories around these relationships in depth, and therefore not verify whether the correlations found are actual causal relationships. There might be, and probably are, numerous other factors that influence these relationships that could not be taken into account here. Furthermore, in its current form the results show too much general variation to unambiguously answer the question how migration affects global inequality.

The most important conclusion of this thesis is therefore that the relationship between migration and wage growth varies greatly between world regions, and is probably influenced by more factors than those controlled for here: economic inequality, urbanization, schooling, life expectancy, remittances and internal and international conflict. Future studies into the effects of migration on the economy, especially studies considering human development, should thus differentiate between world regions, as it seems unlikely that all influential factors can be controlled for in quantitative ways.

In the current political debate, another interesting result is that immigration appears to increase economic inequality in the Western World, but in particular when immigrants come from high-income countries. Moreover, the departure of these same high-income migrants has a positive effect on skilled and high-skilled wages in their country of origin. Thus, migration appears to increase inequality, in particular in the form of skill premiums, in the Western World.

This thesis has been rather ‘experimental’. An attempt was made to combine large amounts of wage data, and then analyze them through first a set of geographically and occupationally segregated models, culminating in a global explanatory model on the relationship between migration and global inequality or economic growth. Analyzing such large amounts of models and summarizing and presenting them in a comprehensive way proved a great challenge. Further, the large amount of results and great variation in them made it impossible to do any in-depth analysis of the correlations found. The study should therefore be considered as mainly exploratory, and further analysis is needed to distinguish actual causal relationships – if this is at all possible.

Outside of the performed analysis, the most valuable result of this thesis, I believe, is in its margins: a database was created including wages for three aggregated categories of unskilled, skilled and high-skilled workers in 195 countries over the period 1969-2015. To this extend, wage data from three databases of the International Labour Organization were standardized and combined. This database could be used to study how wage earners globally, as well as locally, are affected by their cultural, political and economical context. Wage data might be preferable over production (GDP) data in development studies, as it disregards capital income and other forms of income that end up largely in the hands of the few - or foreign.

Bibliography

Abel, G.J., “Estimating global migration flow tables using place of birth data” *Demographic Research* 28:18 (March 2013) 505-546.

Abel, G.J. and N. Sander, “Quantifying Global International Migration Flows” *Science* 343:6178 (March 2014) 1520-1522.

Allen, R.C., ‘Technology and the great divergence: Global economic development since 1820’ *Explorations in Economic History* 49 (2012) 1-16.

Anad, S. and P. Segal, “What do we know about global income inequality?” *Journal of Economic Literature* 46:1 (March 2008) 57-94.

Beauchemin, C., “A Manifesto for Quantitative Multi-sited Approaches to International Migration” *International Migration Review* 48:4 (November 2014) 921-938.

Bertoli, S. and J. Fernándes-Huertas Moraga, "The Size of the Cliff at the Border" *Regional Science and Urban Economics* 51 (March 2015) 1-6.

Bourguignon, F. and Ch. Morrisson, "Inequality among world citizens: 1820-1992" *American Economic Review* (2002) 727-744.

Borjas, G.J., “The Economics of Immigration” *Journal of Economic Literature* 32 (December 1994) 1667-1717.

Borjas, G.J., “The Economic Benefits of Immigration”, *The Journal of Economic Perspectives* 9:2 (Spring 1995) 3-22.

Borjas, G.J., “Immigration and Globalization: A Review Essay” *Journal of Economic Literature* 53:4 (December 2015) 961-974.

Dowrich, S. and J. Bradford DeLong, "Globalization and Convergence" in: Michael D. Bordo, Alan M. Taylor and Jeffrey G. Williamson (eds.), *Globalization in Historical Perspective* (Chicago 2003) 191-226.

Fanjul, G. and L. Pritchett, "Goldilocks Globalization: Searching for "Just Right" Regulation of Cross-Border Labor Flows" *Web* (May 2010) Web.

Hatton, T.J., “Emigration from the UK, 1870-1913 and 1950-1998” *European Review of Economic History* 8 (2004) 149-171.

Hatton, T.J. and J.G. Williamson, *Global Migration and the World Economy. Two Centuries of Policy and Performance* (Cambridge US, London 2005).

Hatton, T.J. and J.G. Williamson, "A Dual policy paradox: Why have trade and immigration policies always differed in labor-scarce economies" *National Bureau of Economic Research No. w11866* (2005).

Held, D. et al., *Global Transformations. Politics, Economics and Culture* (Cambridge, Oxford 1999).

Lindert, P.H. and J.G. Williamson, "Does globalization make the world more unequal?" in: Michael D. Bordo, Alan M. Taylor and Jeffrey G. Williamson (eds.), *Globalization in Historical Perspective* (Chicago 2003) 227-270.

Maddison, A., ‘Historical Statistics for the World Economy’ (version 2004) Web.

McKeown, A., "Global migration 1846-1940" Journal of World History 15 (2004) 155-189.

Milanovic, B., *Worlds Apart: Measuring International and Global Inequality* (Princeton 2005).

Milanovic, B., “Global Inequality And the Global Inequality Extraction Ratio: The Story Of the Past Two Centuries” *World Bank Policy Research Working Paper* (September 2009) Web.

Milanovic, B., “Global inequality recalculated and updated: the effect of new PPP estimates on global inequality and 2005 estimates” *Journal of Economic Inequality* (October 2010) Web.

Milanovic, B., “A short history of global inequality: The past two centuries” *Explorations in Economic History* 48:4 (December 2011) 494-506. Web.

Milanovic, B., “Global Inequality. From Class to Location, from Proletarians to Migrants” *Global Policy* 3:2 (May 2012) Web.

Milanovic, B. and Ch. Lakner, “Global Income Distribution. From the Fall of the Berlin Wall to the Great Recession” *Policy Research Working Paper. World Bank Development Research Group. Poverty and Inequality Team* 6719 (2013). Web.

Milanovic, B., “Global inequality of opportunity. How much of our income is determined by where we live?” *Review of Economics and Statistics* 97:2 (May 2015) 452-460. Web.

Milanovic, B., *Global Inequality: A New Approach for the Age of Globalization* (Harvard University Press 2016).

Oostendorp, R.C., “The Occupational Wages around the World (OWW) Database: Update for 1983-2008” Web. (May 2012).

Piketty, Th., *Capital in the 21th century* (Paris 2013).

Pomeranz, *The Great Divergence. China, Europe and the Making of the Modern World Economy* (Princeton 2000).

Pritchett, L., “Divergence, Big Time” *Journal of Economic Perspectives* 11:3 (Summer 1997) 3-17.

Pritchett, L., "The Cliff at the Border" in: Ravi Kanbur and Michael Spence, *Equity and Growth in a Globalizing World* (2006) 263-286. Web.

Ratha, D. et al., “Migration and Development Brief” *World Bank* 24 (April 2015).

Ruhs, M., *The Price of Rights. Regulating International Labor Migration* (New Jersey 2013).

Taylor, A.M. and J.G. Williamson, ‘Convergence in the age of mass migration’ *European Review of Economic History* 1:1 (April 1997) 27-63.

Timmer, A.S. and J.G. Williamson, "Immigration policy prior to the 1930s: Labor markets, policy interactions, and globalization backlash" *Population and Development Review* (1998) 739-771.

Vries, P., “The California School and beyond: how to study the Great Divergence?” *History Compass* 8:7 (2010) 730-751.

Zanden, J.L. van et al., “The Changing Shape of Global Inequality 1820-2000; Exploring a New Dataset” *Review of Income and Wealth* 60:2 (June 2014) 279-297.

Zuccotti, C.V., H.G. Ganzeboom and A. Guveli, “Has Migration Been Beneficial for Migrants and Their Children? Comparing Social Mobility of Turks in Western Europe, Turks in Turkey, and Western European” *International Migration Review* (November 2015). Web.

Attachments

Attachment 2.1

Unskilled worker: wages in manufacturing

The ILO wages in manufacturing are classified according to the International Standard Industrial Classification of all Economic Activities (ISIC), revision 2, 3 and 4 by the United Nations Statistics Division. Revision 2 is most commonly used for the period 1969-1995, while revision 3 becomes dominant after 1995. As wages in manufacturing are mostly used to construct the unskilled worker's wage for the period 1969-1990 and only play a supplementary role afterwards, the construction is based on ISIC-Rev. 2. For later years only those categories that are largely contained in the revision 2 definition are included in the 'unskilled worker' category.

The unskilled worker's wage is composed of wages in the production of 'textile, wearing apparel and leather' and 'wood and wood products, including furniture'. However, because of the above mentioned approach furniture is only included in the Rev. 2 version. Furthermore, in the eventual category wages in textiles industries are weighted twice as much as wages in wood industries where both are available.

ISIC-Rev. 2 classification	**ISIC-Rev. 2 code**	**Included?**
32 Textile, Wearing Apparel and Leather Industries	02_	YES
321 Manufacture of textiles	02C	YES
322 Manufacture of wearing apparel, except footwear	02H	YES
323 Manufacture of leather and products of leather, leather substitutes and fur, except footwear and	02L	YES
324 Manufacture of footwear, except vulcanized or moulded rubber or plastic footwear	02Q	YES
33 Manufacture of Wood and Wood Products, Including Furniture	03_	YES
331 Manufacture of wood and wood and cork products, except furniture	03G	YES
332 Manufacture of furniture and fixtures, except primarily of metal	03H	YES
321�322,324	02E	YES
323�324	02M	YES
321�322	02D	YES
322324	02K	YES
322�324	02J	YES
323355	02N	YES
33�34	03A	YES
ISIC-Rev. 3 classification	**ISIC-Rev. 3 code**	
17 Manufacture of Textiles	03_	YES
18 Manufacture of Wearing Apparel; Dressing and Dyeing of Fur	04_	YES
19 Tanning and Dressing of Leather; Manufacture of Luggage, Handbags, Saddlery, Harness and Footwear	05_	YES
20 Manufacture of Wood and of Products of Wood and Cork, except Furniture; Manufacture of articles o	06_	YES
36 Manufacture of Furniture; Manufacturing NEC	22_	NO
17,21,22,24	03D	YES
18�19	04A	YES
17�18	03C	YES
36�37	22A	NO
20�21	06C	YES
20,36	06A	YES
17�19	03B	YES
ISIC-Rev. 4 classification	**ISIC-Rev. 4 code**	
13 Manufacture of textiles	04_	YES
14 Manufacture of wearing apparel	05_	YES
15 Manufacture of leather and related products	06_	YES
16 Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles o	07_	YES
31 Manufacture of furniture	22_	NO
13�15	04B	YES
31�32	22A	NO

Unskilled worker: October Inquiry occupational wages (LABORSTA)

Occupations contained in the unskilled worker category for the period 1983-2008.

Industry	**Industry code**	**Occupation**	**Occupation code**
Communication	NH	127 Postman	127
Construction	LA	90 Labourer	90
Electric light and power	KA	80 Labourer	80
Iron and steel basic industries	IA	65 Labourer	65
Manufacture of footwear	DD	35 Shoe sewer (machine)	35
Manufacture of footwear	DD	34 Laster	34
Manufacture of footwear	DD	33 Clicker cutter (machine)	33
Manufacture of industrial chemicals	GA	56 Labourer	56
Manufacture of leather and leather products (except footwear)	DC	32 Leather goods maker	32
Manufacture of other chemical products	GB	58 Packer	58
Manufacture of other chemical products	GB	59 Labourer	59
Manufacture of wearing apparel (except footwear)	DB	30 Sewing-machine operator	30
Manufacture of wearing apparel (except footwear)	DB	29 Garment cutter	29
Manufacture of wooden furniture and fixtures	EB	41 Wooden furniture finisher	41
Medical and dental services	PD	155 Auxiliary nurse	157
Plantations	AB	4 Plantation worker	4
Printing, publishing and allied industries	FB	51 Labourer	51
Public administration	PA	142 Office clerk	144
Restaurants and hotels	MC	100 Room attendant or chambermaid	100
Restaurants and hotels	MC	99 Waiter	99
Retail trade (grocery)	MB	96 Salesperson	96
Retail trade (grocery)	MB	95 Cash desk cashier	95
Sanitary services	PB	144 Refuse collector	146
Slaughtering, preparing and preserving meat	CA	21 Packer	21
Spinning, weaving and finishing textiles	DA	28 Labourer	28
Spinning, weaving and finishing textiles	DA	27 Cloth weaver (machine)	27
Spinning, weaving and finishing textiles	DA	25 Thread and yarn spinner	25

2010: October Inquiry (ILOSTAT)

The unskilled worker is defined as a worker in 'elementary occupations' according to the ISCO-88 classification of occupational categories. The major group 'service workers and shop and market sales workers' has not been included because it combines occupations from class one (unskilled) and class two (skilled).

Occupational category	**ISCO version**	**skill class**
Elementary occupations	ISCO-88	1

Attachment 2.2

Skilled worker: wages by economic activity (LABORSTA)

ISIC-Rev. 2 classification	ISIC-Rev. 2 code	Contained in	Included?
2 Mining and Quarrying	02_		YES
7 Transport, Storage and Communication	07_		YES
21	02B	2 Mining and Quarrying	YES
22	02C	2 Mining and Quarrying	YES
29	02D	2 Mining and Quarrying	YES
72	07A	7 Transport, Storage and Communication	YES

ISIC-Rev. 3 classification	ISIC-Rev. 3 code		
C Mining and Quarrying	03_		Regression only
I Transport, Storage and Communications	09_		Regression only
11	03Q	C Mining and Quarrying	Regression only
14	03S	C Mining and Quarrying	Regression only
64	09E	I Transport, Storage and Communications	Regression only
C�E	03B	C Mining and Quarrying	Regression only
I�J	09A	I Transport, Storage and Communications	Regression only
C�K	03N	C Mining and Quarrying	Regression only
C,E	03E	C Mining and Quarrying	Regression only
C�O	03G	C Mining and Quarrying	Regression only
C�X	03P	C Mining and Quarrying	Regression only
C�I	03K	C Mining and Quarrying	Regression only
C�D	03D	C Mining and Quarrying	Regression only
C�F	03F	C Mining and Quarrying	Regression only

ISIC-Rev. 4 classification	ISIC-Rev. 4 code		
B Mining and quarrying	02_		Regression only
H Transportation and storage	08_		Regression only
J Information and communication	10_		NO
B�C	02C	B Mining and quarrying	Regression only
B�F	02D	B Mining and quarrying	Regression only
B�U	02A	B Mining and quarrying	Regression only

Economic activities contained in the skilled worker category. Mining and Quarrying makes up a small part of the economy only in most countries, therefore, wages in the sector ‘Transport, Storage and Communication’ are weighted three times as much in the eventual data as mining wages.

Skilled worker: October Inquiry occupational wages (LABORSTA)

Occupations contained in the skilled worker category for the period 1983-2008. Concentrated on skilled labourers in construction and 'white collar' workers.

Industry	Industry code	Occupation	Occupation code
Banks	OA	130 Stenographer-typist	130
Banks	OA	131 Bank teller	131
Banks	OA	132 Book-keeping machine operator	132
Construction	LA	88 Construction carpenter	88
Construction	LA	84 Building painter	84
Construction	LA	85 Bricklayer (construction)	85
Construction	LA	82 Plumber	82
Construction	LA	81 Building electrician	81
Education services	PC	150 First-level education teacher	152
Electric light and power	KA	77 Office clerk	77
Electric light and power	KA	78 Electric power lineman	78
Insurance	OB	134 Stenographer-typist	134
Insurance	OB	135 Card- and tape-punching-machine operator	135
Manufacture of industrial chemicals	GA	55 Mixing- and blending-machine operator	55
Manufacture of metal products (except machinery and equipment)	JA	67 Welder	67
Medical and dental services	PD	158 Ambulance driver	160
Medical and dental services	PD	157 Medical X-ray technician	159
Medical and dental services	PD	154 Professional nurse (general)	156
Medical and dental services	PD	156 Physiotherapist	158
Passenger transport by road	NB	110 Automobile mechanic	110
Passenger transport by road	NB	111 Motor bus driver	111
Printing, publishing and allied industries	FB	47 Hand compositor	47
Printing, publishing and allied industries	FB	45 Stenographer-typist	45
Printing, publishing and allied industries	FB	49 Printing pressman	49
Public administration	PA	143 Fire-fighter	145

2010: Skilled workers October Inquiry occupational wages (ILOSTAT)

Due to the inclusion of 'plant and machine operators' this conceptualization probably yields lower wages than previous years, but data is aligned with the previous series using GDP growth.

Occupational category	ISCO version	skill class
Craft and related trades workers	ISCO-88	2
Skilled agricultural and fishery workers	ISCO-88	2
Skilled agricultural, forestry and fishery workers	ISCO-08	2
Clerks	ISCO-88	2
Clerical support workers	ISCO-08	2
Plant and machine operators and assemblers	ISCO-88	2
Plant and machine operators, and assemblers	ISCO-08	2

Attachment 2.3

High-skilled worker: October Inquiry occupational wages (LABORSTA)

Occupations contained in the high-skilled worker category for the period 1983-2008. Professions which generally require tertiary education.

Industry	**Industry code**	**Occupation**	**Occupation code**
Air transport	NF	118 Air transport pilot	118
Banks	OA	129 Accountant	129
Coalmining	BA	11 Coalmining engineer	11
Crude petroleum and natural gas production	BB	14 Petroleum and natural gas engineer	14
Education services	PC	145 Mathematics teacher (third level)	147
Education services	PC	146 Teacher in languages and literature (third level)	148
Education services	PC	147 Teacher in languages and literature (second level)	149
Education services	PC	148 Mathematics teacher (second level)	150
Education services	PC	149 Technical education teacher (second level)	151
Electric light and power	KA	76 Power distribution and transmission engineer	76
Electric light and power	KA	79 Power-generating machinery operator	79
Insurance	OB	133 Computer programmer	133
Insurance	OB	136 Insurance agent	136
Manufacture of industrial chemicals	GA	52 Chemical engineer	52
Manufacture of industrial chemicals	GA	54 Supervisor or general foreman	54
Manufacture of industrial chemicals	GA	53 Chemistry technician	53
Maritime transport	ND	114 Ship's chief engineer	114
Medical and dental services	PD	152 General physician	154
Medical and dental services	PD	153 Dentist (general)	155
Printing, publishing and allied industries	FB	44 Journalist	44
Public administration	PA	139.a Government executive official: a) central government	139
Public administration	PA	139.b Government executive official: b) regional or provincial government	140
Public administration	PA	138 Computer programmer	138
Public administration	PA	139.c Government executive official: c) local authority	141
Railway transport	NA	102 Railway services supervisor	102
Supporting services to air transport	NG	124 Air traffic controller	124

2010: High-skilled worker October Inquiry occupational wages (ILOSTAT)

This classification is largely consistent with the above classification for the years 1983-2008.

Occupational category	**ISCO version**	**skill class**
Managers	ISCO-88	N.A.
Legislators, senior officials and managers	ISCO-08	N.A.
Professionals	ISCO-88	4
Technicians and associate professionals	ISCO-88	3

Attachment 2.4

The global bilateral migration flows dataset of Guy Abel provides absolute numbers of migrants between two countries over a 10 (or 5) year period. These absolute numbers are converted to emigration and immigration rates per 1,000 by: dividing by the total population of the destination country for immigration rates, dividing by the total population of the origin country for emigration rates – and multiplying with 1000.

In the eventual analysis mostly migration flows over a five year period are used, although depending on availability of wage data this period can also be longer or shorter (but is the same as the period over which wage growth is calculated). This wage growth period can start in any year between 1969 and 2013. Therefore, approximations of migration flows per year were needed. These are obtained by linear interpolation between the existing datapoints, where each ten-year flow is taken as the one year flow for the middle year per 10,000 inhabitants. For the years 1990-2000 data from both Abel's old dataset as his new dataset is available. To align these two data sources to each other, a weighted average of both sources is taken for the years 1990-1995 (with a 5/6th weight for the old data source in 1990, decreasing linearly to a 1/6th weight for 1990 and a 5/6th weight for the new data source in 1995).

To create migration clusters the original data without linear interpolation are used, these are first converted into categorical data. Each of the migration ratios has been classified to belong to one in four categories: low, moderate, high and major bilateral flow. The categories are based on percentiles to make the number of non-zero datapoints comparable over the different years and types of migration.

Categorization migration flows

Low bilateral flow: about 0 to 1 per 100.000 of the population

Immigration range cut-off 1th category: 6.556-12.37 (per million) for the 1960s and 1990s respectively.

Emigration range cut-off 1th category: 5.429-11.03 (per million) for the 1960s and 1990s respectively.

Moderate bilateral flow: about 1 to 8 per 100.000

Immigration range cut-off 2nd category: 50.36-114.6 (per million) for 1960s and 2000s respectively.

Emigration range cut-off 2nd category: 43.98-85.94 (per million) for 1960s and 2000s respectively.

High bilateral flow: about 8 to 75 per 100.000

Immigration range cut-off 3th category: 618.9-998.8 (per million) for 1970s and 2000s respectively

Emigration range cut-off 3th category: 639.2-886.7 (per million) for 1960s and 2000s respectively.

Major bilateral flow: over 75 per 100.000 (relatively high variation between decades and types)

The low bilateral flow is assigned value 0. The moderate, high and major bilateral flow are assigned the squareroot of the average value of its components

Example: for immigration in decade X the average value of all bilateral flows deemed 'moderate' is calculated and squared and assigned to all these bilateral flows. In this way different magnitudes of migration flows for the different years/types are maintained. Square root is used to mute the effect of major flows (in a similar way as a logarithmic scale for GDP is often used).

Clustering migration flows

For the categorized datasets thus obtained cosine similarities are calculated. If the adapted dataset, and not cosine similarities, would be used to create clusters, then countries with little migration would be more likely to be clustered than countries with higher migration. On top of this, similarities would be less sensitive to the set of countries with which bilateral flows exists. Cosine similarities place more emphasize on the ‘direction’ of the migration data, i.e. with which countries the flows are high or low, while using the original data would place emphasize mainly on the total size of the migration flows.[61]

To make these cosine similarities into a (semi-)metric, i.e. a set of comparable values, they are distracted from 1. Subsequently the dataset now obtained is used to make density based clusters by the DBSCAN method.[62] In this clustering method nearest neighbours are calculated (countries similar in terms of migration) and clustered based on some parameters given to the method. First, the method finds some ‘core countries’, which are surrounded by relatively many other countries with similar migration rates. Therefore one of the parameters that has to be given to the method is the minimum number of countries in its neighbourhood for a country to be seen as a ‘core country’ (able to form a cluster), which I have set to four for all decades and both emigration and immigration to make the clustering comparable.

Another parameter is epsilon (eps), the maximum distance for countries to be considered as laying in the same neighbourhood as a core country. I have adapted this parameter slightly for each period and type of migration in order to get a ‘good’ clustering (trade-off between many unclassified countries and all countries being placed in one cluster, see below). This parameter determines whether you get clusters, how many clusters you get, how many countries remain unclassified etc.

In the below table these parameters and the eventual number of clusters and number of non-classified countries for each year and type of migration are given.

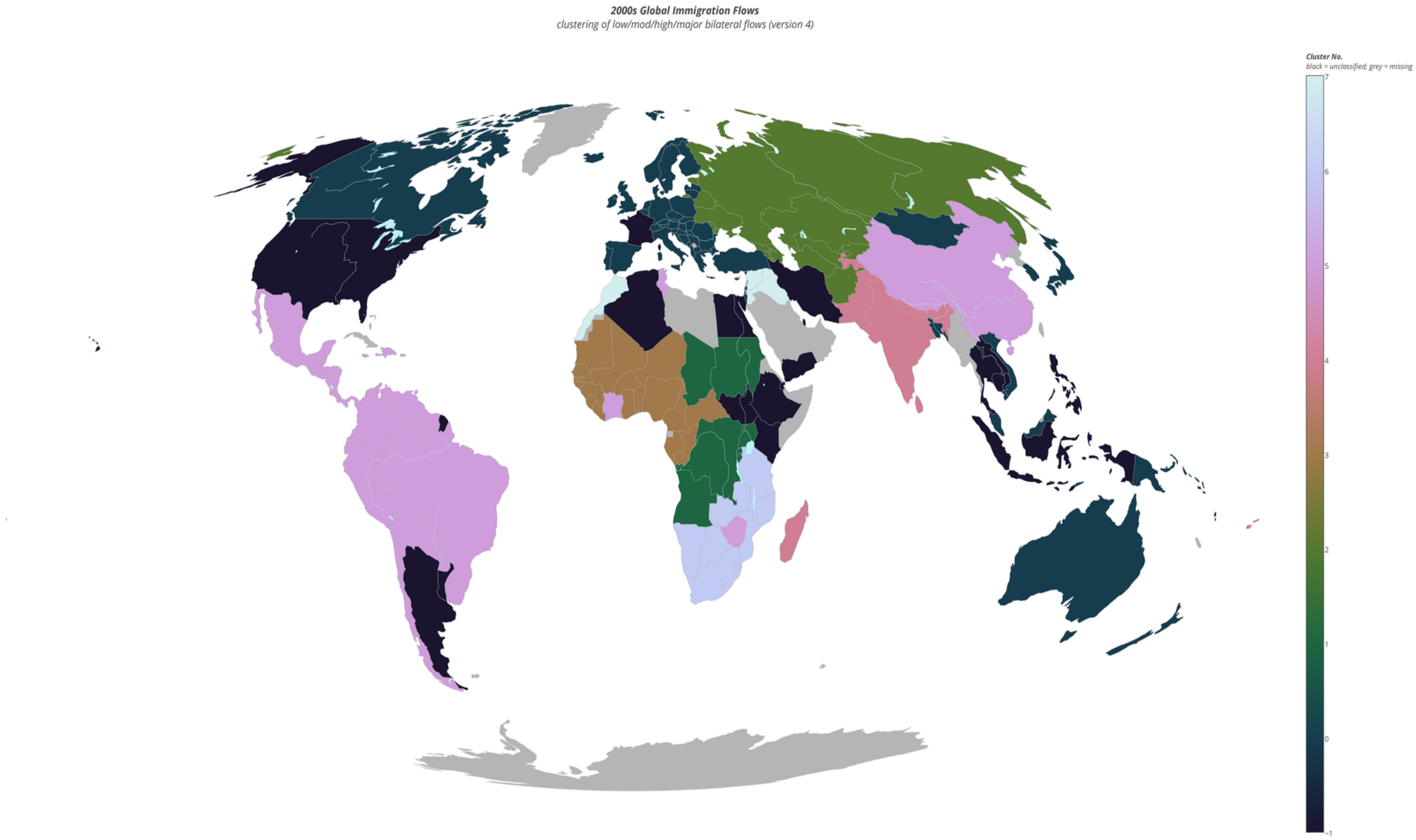
<u>Selected clusters immigration and emigration 1960-2010</u>

year	*Migration type*	*Eps*	*# Non-classified*	*No. of clusters*
1960s	emigration	1.48	31	9
1970s	emigration	1.415	27	8
1980s	emigration	1.44	29	7
1990s	emigration	1.48	19	6
2000s	emigration	1.445	24	5
1960s	immigration	1.445	18	9
1970s	immigration	1.425	22	9
1980s	immigration	1.40	16	11
1990s	immigration	1.38	32	8
2000s	immigration	1.49	19	8

In the regression models, one clustering is used for all years. For this the immigration cluster for the 2000s is selected. On the one hand because Abel his database for this later database is of higher quality than the old database, on the other hand because GDP at PPP is available for more countries for this later decade. The emigration clustering for the 2000s gives one large cluster containing almost all countries, as both people from the ‘western world’ and from many countries outside of it emigrate more to the western world than to anywhere else. Thus, the differences in origin are more variable than the differences in destination and therefore more relevant to control for in the analysis.

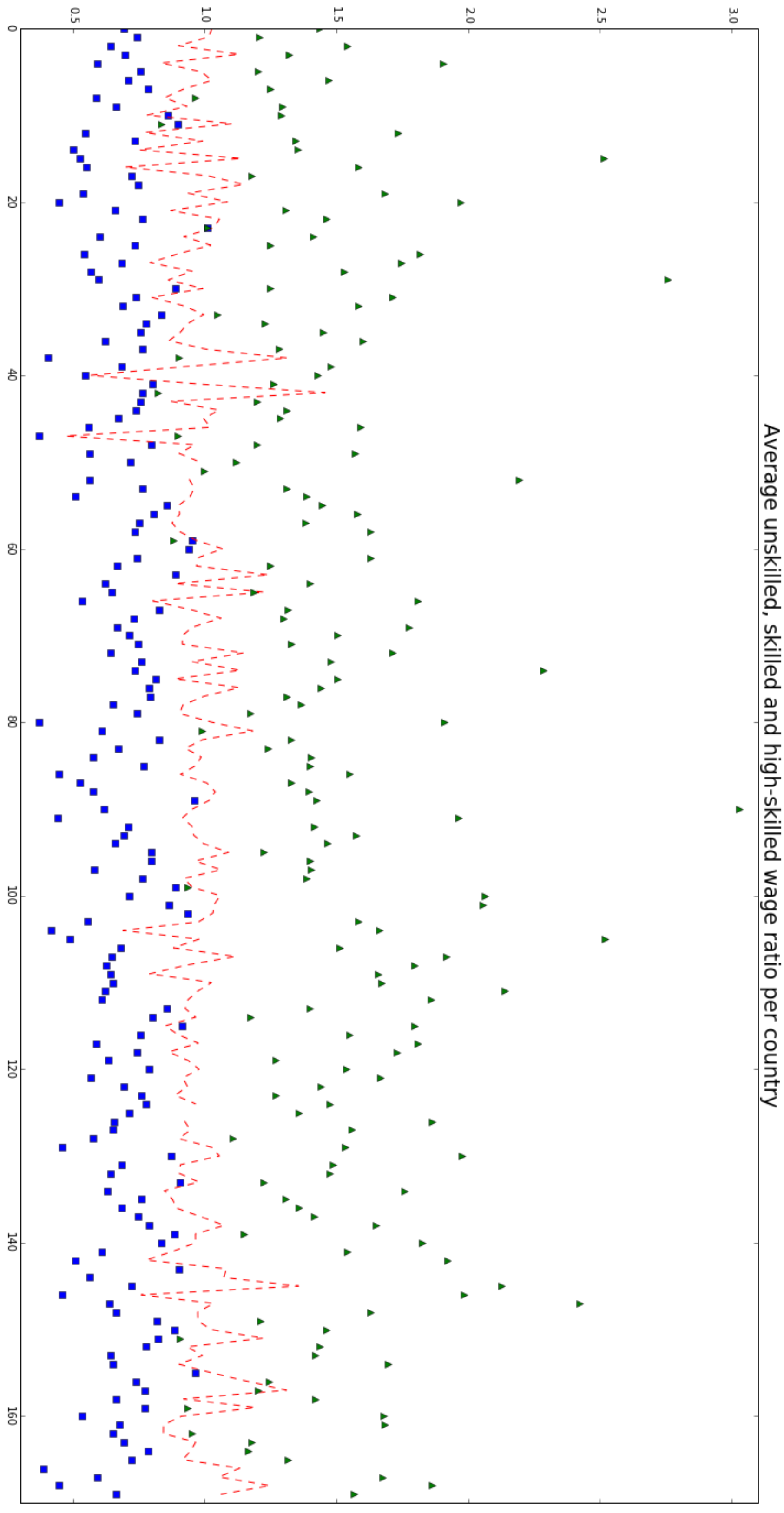
In the below world map the 2000s immigration cluster used for the regression model is shown in its original form, countries that were not classified by the automated clustering were assigned manually to clusters based on geographical proximity, migration rates (if available) and cultural ties. These clusters were subsequently further separated on the basis of national income, as is described in attachment 2.6.

61 Documentation of cosine similarities: http://scikit-learn.org/stable/modules/metrics.html.
62 Documentation of the DBSCAN method: http://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html.


2000s Global Immigration Flows
clustering of low/mod/high/major bilateral flows (version 4)
Cluster No.
black = unclassified; grey = missing
7
6
5
4
3
2
1
0
-1

Attachment 2.5

In this figure average unskilled, skilled and high-skilled wages from the October Inquiries are represented for each country, with countries on the x-axis and average wage ratios on the y-axis. Wage ratios are calculated as the ratio between each wage category and the average of all wages in the sectors construction and manufacturing. Unskilled wages are presented in blue, skilled wages in red, high-skilled wages in green.



Outlier countries:

In Liberia (91) (at 3x), Cameroon (30) (at 2.8x) and Burundi (16) (at 2.5x) high-skilled wages lay very far above the country average. The country average in these countries is very low so this is probably not a data problem.

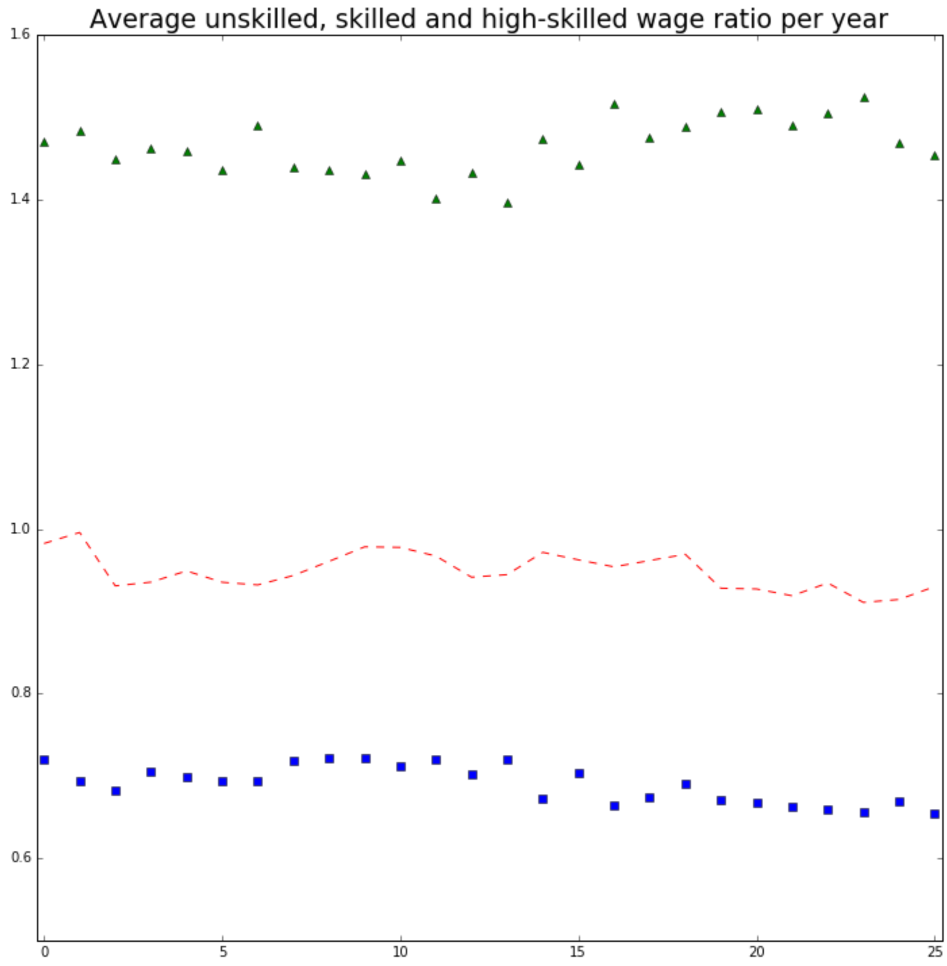
In Ethiopia (48) and Equatorial Guinea (60) all categories are below the manufacturing and construction country wage. Also in Equatorial Guinea, high-skilled wage (at 0.88) is below unskilled and skilled (at 0.95).

In Dominica (39) (at 0.8x), the Dominican Republic (43) (at 0.8x), Kyrgyzstan (82) (at 1x), Tonga (152) (at 0.9x) and Uruguay (160) (at 0.9x) the high-skilled wage is below or at the country average, while the skilled wage is quite far above this, and ranges from 1.2 to 1.5 times the country average.

In several other countries high-skilled wages are at the same level, or slightly below, skilled wages. This is probably due to the large share of public sector occupations in this category and not a data problem.

Djibouti (41) has the exact same wage for unskilled and skilled. Belarus (24) has almost the exact same wages for all categories (data pre-1991).

For other countries the definition seems quite consistent; variation in the range reflects differences in national wage inequality, most important is that skilled wage is between 0.8 and 1.1 for most countries.



In this figure averages over all countries for each year of unskilled, skilled and high-skilled wages as a ratio of manufacturing and construction wages are presented. Years are on the x-axis (with 0=1983, 25=2008), ratios on the y-axis. Blue are unskilled wages, red skilled wages, green high-skilled wages.

Overall the relative position, and thus meaning, of the occupational categories is consistent over time.

Over the years 1983-2008 wages in unskilled and skilled occupations decrease slightly (as a percentage of the average main sectors country wage), while wages in high-skilled occupations increase (except for 2007 and 2008, but this can be due to the limited data availability in those years).

Attachment 2.6

National income classification

Three national income classes where created, over which countries were distributed in such a way that each national income class contained equally many wage data points: this is beneficial for the eventual analysis and any choice of cut-off income would be arbitrary. As national income the average of GDP per capita at PPP (World Bank) was used from the years 2001 and 2011, since the immigration clusters are also based on immigration in this decade. Following this method, all countries with an annual GDP per capita up to $8,410 (international dollar) were classified as 'low income', or equivalently all countries with its average inhabitant living off at most $23.03 international dollar a day were classified as low income, where median income should be lower. The country with the highest income which was classified as low income was Jordan, with many other countries in the Middle East around the cut-off between low and mid-income. The country with the lowest income in the dataset was Burundi, with an annual GDP per capita of $643.77, or $1.76 international per day.

Mid-income countries have a GDP per capita at PPP ranging between $8,958 and $23,221 annually, equivalent to $24.53-$63.58 international dollar a day. The country with the lowest income classified as mid-income was the Dominican Republic, while the country with the highest income was Malta. Finally, high-income countries have a GDP per capita at PPP ranging from $23,655 to $111,617 annually, equivalent to 64.76-305.59 international dollar a day.

Since most countries in the Middle East and northern Africa together form a immigration cluster, and their incomes were around the threshold between low-income and mid-income, these countries were combined in one income-migration cluster as 'low to mid income' - although in the regression analysis they are included in the low-income migration rates as the majority of the countries in this cluster were classified as low income.[63]

Some other countries were re-assigned as well: Thailand was the only country from East- and Southeast Asia which was classified mid-income, while all other countries in the same migration cluster were located around the borders of Europe (from Portugal to Turkey to Poland). Therefore, Thailand was re-classified as 'low-income' in order to assign it the cluster which geographically was more fitting. Gabon, similarly, was the only African country, and the only country from any of the three African migration clusters, with a GDP per capita classified as 'mid-income'. In this case the choice to reclassify as low-income is justified by the composition of Gabon's GDP, which largely consists of oil production and hardly reaches the population.[64]

Finally, countries that could not be assigned to one of the migration clusters on the basis of their migration rates were usually assigned to a cluster with neighbouring countries, or a cluster with countries that they were connected to in other ways (e.g. island dependencies of the UK and France were generally assigned to the European cluster). A full documentation of this reassignment is available upon request.

63 Countries with a GDP per capita higher than the low-income threshold now classified as low income are Algeria and Iran.

64 Gabon's human development index is far lower than average at their GDP per capita. UN Development Programme, Human Development Indicators 2013, as cited in: Tom Burghis, *The Looting Machine* (London 2015), p. 212.

Migration-income clustering

The resulting income-migration clusters are:

Nr of datapoints	Cluster	Income and migration cluster	Description
488	**0**	low income – cluster 0	'From the Bay of Bengal to the Solomon Sea': Southeastern Asia and part of Melanesia, also including Albania.
187	**1**	low income – cluster 1	'From the Horn of Africa to the South Atlantic': north-east African countries (Chad, Sudan, Eritrea) through Congo (DM) to Angola, including Rwanda, Burundi and Uganda
223	**2**	low income – cluster 2	'Around the Caspian and Black Sea': former Soviet from Moldova and Ukraine to Kyrgyzstan, excluding Azerbaijan (middle income) and Tajikistan (Indian cluster).
273	**3**	low income – cluster 3	'Around the Gulf of Guinea': western and middle Africa from Mauritania to Niger to the Republic of Congo, including Cabo Verde, very low incomes
454	**4**	low income – cluster 4	'Indian Ocean and South Pacific': non-Arabic parts of Southern Asia, but including Pakistan, including Madagascar and (independent) parts of Melanesia.
950	**5**	low income – cluster 5	'Geographical mix': mostly poorer Latin American countries, but also China and various African (from Tunesia to Zimbabwe) – all have low immigration rates and relatively low incomes.
185	**6**	low income – cluster 6	'Foot of Africa': from Namibia and Lesotho to Tanzania
250	**7**	low to mid income – cluster 7	'Arabic world': northern African and Middle East (from Algeria to Iran)
1159	**8**	mid income – cluster 0	'European frontiers': Former Yugoslavia and European part of former Soviet, former western colonies (island states), Portugal and Malta, Turkey
334	**9**	mid income – cluster 2	'Inner former Soviet': Russia, Belarus, Kazakhstan, including Azerbaijan ad Estonia
1078	**10**	mid income – cluster 5	'Latin America and the Caribbean': characterized mainly by low immigration rates, also includes Hungary.
163	**11**	mid income – cluster 6	'Around South Africa': South Africa, Botswana and Seychelles
2645	**12**	high income – cluster 0	'Western world': western Europe, Anglosphere, Asian tigers
136	**13**	high income – cluster 5	'Caribbean': US and British Virgin Islands, Puerto Rico, Netherlands Antilles
91	**14**	high income – cluster 7	'Around the Persian Gulf': Qatar, Kuwait, Bahrain

Only cluster 5 does not cover a clear geographical region, but instead covers several countries with very low immigration rates – from China to Latin America. Since immigration rates are so low, emigration clusters were taken into account here as well. But in all emigration clusters the countries from cluster 5 fall either in the large global cluster with high emigration rates to the 'western world', or the Western part of Latin America (from Mexico to Argentina) forms an independent emigration cluster with China, India and other East Asian countries: although total emigration rates from China are low while those from Latin America are high, these countries form both an emigration and immigration cluster together because the main direction of emigration is the same. Therefore, this 'geographical mix' cluster is maintained, as it is similar in terms of emigration, immigration and national income.

Some other clusters do contain ‘odd’ countries, as for example Hungary in the ‘Latin American’ cluster, but clustering was not done on the basis of geographical location and is also not intended as such. It is rather surprising that clusters to such a large extend do align with geographical world-regions.

As several clusters include too few datapoints to be split up in three groups (unskilled, skilled and high-skilled wage growth) again before analysis, and especially contain too few high-skilled wage datapoints, these clusters are combined again in the following way:

0 - ‘Africa’: clusters 1, 3, 6 and 11 (mostly low, some mid-income).

1 - ‘South and Southeast Asia’: clusters 0 and 4 (low income).

2- ‘Geographical mix’: cluster 5 (low income).

3 - ‘European frontier’: clusters 2, 8 and 9 (mostly mid, some low-income).

4 - ‘Latin America and the Caribbean’: cluster 10 (mid income).

5 - ‘Arabic World’: clusters 7 and 14 (low, mid and high income).

6 - ‘Western World’: clusters 12 and 13 (high income).

Thus, while the covariates in the regression analysis are emigration and immigration rates separated by the *income* cluster of the countries of origin and destination respectively, the final units of analysis often contain countries from multiple income clusters.

The final form of these combinations was based on a regression analysis that was done for each of the 15 clusters separately, with only (total) immigration, (total) emigration and periods as covariates. In these models the clusters that are now combined were most similar in terms of the relationship between migration and wage growth.

Countries per cluster and occupational category:

[font makes comparison easier since all letters have equal length]

```
Cluster  0  - countries with available unskilled wages:
['"Mali"', '"Gambia"', '"Rwanda"', '"Chad"', '"Gabon"', '"Kenya"', '"Ghana"',
'"Burundi"', '"Lesotho"', '"Liberia"', '"Zambia"', '"Swaziland"', '"Burkina
Faso"', '"Central African Republic"', '"Cameroon"', '"Benin"', '"Togo"',
'"Eritrea"', '"Sudan"', '"Ethiopia"', '"Niger"', '"Botswana"', '"Nigeria"',
'"Malawi"', '"Angola"', '"Cabo Verde"', '"South Africa"', '"Sierra Leone"',
'"Mozambique"', '"Senegal"']
Total nr. of countries available:  30

Cluster  0  - countries with available skilled wages:
['"Mali"', '"Gambia"', '"Rwanda"', '"Chad"', '"Gabon"', '"Kenya"', '"Ghana"',
'"Burundi"', '"Lesotho"', '"Zambia"', '"Swaziland"', '"Burkina Faso"', '"Central
African Republic"', '"Cameroon"', '"Benin"', '"Togo"', '"Eritrea"', '"Sudan"',
'"Ethiopia"', '"Niger"', '"Botswana"', '"Nigeria"', '"Guinea"', '"Malawi"',
'"Angola"', '"Cabo Verde"', '"Tanzania, United Republic of"', '"South Africa"',
'"Sierra Leone"', '"Mozambique"', '"Senegal"']
Total nr. of countries available:  31

Cluster  0  - countries with available high-skilled wages:
['"Mali"', '"Rwanda"', '"Chad"', '"Gabon"', '"Kenya"', '"Ghana"', '"Burundi"',
'"Lesotho"', '"Liberia"', '"Zambia"', '"Swaziland"', '"Burkina Faso"', '"Central
African Republic"', '"Cameroon"', '"Benin"', '"Togo"', '"Eritrea"', '"Sudan"',
'"Ethiopia"', '"Niger"', '"Botswana"', '"Nigeria"', '"Malawi"', '"Angola"',
'"Cabo Verde"', '"South Africa"', '"Sierra Leone"', '"Mozambique"', '"Senegal"']
Total nr. of countries available:  29
```

```
Cluster  1  - countries with available unskilled wages:
['"Bangladesh"', '"Madagascar"', '"Viet Nam"', '"India"', '"Pakistan"',
'"Thailand"', '"Fiji"', '"Cambodia"', '"Myanmar"', '"Mongolia"', '"San Marino"',
'"Sri Lanka"', '"Tonga"', '"Nepal"', '"Maldives"', '"Papua New Guinea"',
'"Indonesia"', '"Philippines"', '"R\xef\xbf\xbdunion"']
Total nr. of countries available:  19

Cluster  1  - countries with available skilled wages:
['"Bangladesh"', '"Fiji"', '"Madagascar"', '"Viet Nam"', '"Philippines"',
'"Myanmar"', '"Cambodia"', '"Thailand"', '"Tajikistan"', '"Samoa"',
'"Indonesia"', '"India"', '"Albania"', '"Solomon Islands"', '"Sri Lanka"',
'"Nepal"', '"Mongolia"', '"Pakistan"', '"Tonga"', '"San Marino"', '"Maldives"',
'"Papua New Guinea"']
Total nr. of countries available:  22

Cluster  1  - countries with available high-skilled wages:
['"Madagascar"', '"Viet Nam"', '"India"', '"Pakistan"', '"Fiji"', '"Cambodia"',
'"Albania"', '"Tonga"', '"Mongolia"', '"San Marino"', '"Myanmar"', '"Sri
Lanka"', '"Thailand"', '"Tajikistan"', '"Maldives"', '"Papua New Guinea"',
'"Indonesia"', '"Philippines"', '"Bangladesh"']
Total nr. of countries available:  19

Cluster  2  - countries with available unskilled wages:
['"Nicaragua"', '"Bolivia"', '"Haiti"', '"Comoros"', '"Peru"', '"Paraguay"',
'"Jamaica"', '"China"', '"C\xef\xbf\xbdte d\'Ivoire"', '"Honduras"', '"French
Guiana"', '"Guadeloupe"', '"El Salvador"', '"Guatemala"', '"Zimbabwe"',
'"Belize"', '"Tunisia"', '"Ecuador"', '"Guyana"', '"Saint Vincent and the
Grenadines"']
Total nr. of countries available:  20

Cluster  2  - countries with available skilled wages:
['"Nicaragua"', '"Bolivia"', '"Haiti"', '"Comoros"', '"Peru"', '"Paraguay"',
'"Jamaica"', '"China"', '"C\xef\xbf\xbdte d\'Ivoire"', '"Honduras"', '"French
Guiana"', '"Guadeloupe"', '"Tunisia"', '"Guatemala"', '"Zimbabwe"', '"Belize"',
'"El Salvador"', '"Ecuador"', '"Guyana"', '"Saint Vincent and the Grenadines"']
Total nr. of countries available:  20

Cluster  2  - countries with available high-skilled wages:
['"Nicaragua"', '"Bolivia"', '"Haiti"', '"Peru"', '"Paraguay"', '"Guadeloupe"',
'"China"', '"C\xef\xbf\xbdte d\'Ivoire"', '"Honduras"', '"Comoros"',
'"Tunisia"', '"Guatemala"', '"Zimbabwe"', '"Belize"', '"El Salvador"',
'"Ecuador"', '"Guyana"', '"Saint Vincent and the Grenadines"']
Total nr. of countries available:  18

Cluster  3  - countries with available unskilled wages:
['"Latvia"', '"Croatia"', '"Armenia"', '"Portugal"', '"Georgia"', '"Bulgaria"',
'"Estonia"', '"Macedonia, The former Yugoslav Rep. of"', '"Lithuania"',
'"Kazakhstan"', '"French Polynesia"', '"Romania"', '"Slovakia"', '"Mauritius"',
'"Ukraine"', '"Malta"', '"Czech Republic"', '"Poland"', '"Azerbaijan"',
'"Turkey"', '"Moldova, Republic of"', '"Malaysia"', '"Belarus"', '"Russian
Federation"', '"Kyrgyzstan"', '"Serbia"']
Total nr. of countries available:  26

Cluster  3  - countries with available skilled wages:
['"Latvia"', '"Croatia"', '"Armenia"', '"Portugal"', '"Georgia"', '"Bulgaria"',
'"Guam"', '"Estonia"', '"Macedonia, The former Yugoslav Rep. of"',
'"Lithuania"', '"Kazakhstan"', '"Uzbekistan"', '"French Polynesia"',
'"Romania"', '"Slovakia"', '"Mauritius"', '"Montenegro"', '"Ukraine"',
'"Malta"', '"Czech Republic"', '"Poland"', '"Azerbaijan"', '"Turkey"',
'"Moldova, Republic of"', '"Malaysia"', '"Belarus"', '"Russian Federation"',
'"Kyrgyzstan"', '"Serbia"']
```

```
Total nr. of countries available:  29

Cluster  3  - countries with available high-skilled wages:
['"Poland"', '"Azerbaijan"', '"Turkey"', '"Moldova, Republic of"', '"Malaysia"',
'"Portugal"', '"Bulgaria"', '"Lithuania"', '"Estonia"', '"Latvia"',
'"Mauritius"', '"Kazakhstan"', '"Ukraine"', '"Belarus"', '"Russian Federation"',
'"Macedonia, The former Yugoslav Rep. of"', '"Kyrgyzstan"', '"French
Polynesia"', '"Slovakia"', '"Romania"', '"Czech Republic"']
Total nr. of countries available:  21

Cluster  4  - countries with available unskilled wages:
['"Costa Rica"', '"Chile"', '"Hungary"', '"Dominican Republic"', '"Brazil"',
'"Grenada"', '"Bahamas"', '"Argentina"', '"Cuba"', '"Suriname"', '"Uruguay"',
'"Trinidad and Tobago"', '"Martinique"', '"Netherlands Antilles"', '"Venezuela,
Bolivarian Rep. of"', '"Colombia"', '"Barbados"', '"Mexico"', '"Panama"',
'"Saint Lucia"']
Total nr. of countries available:  20

Cluster  4  - countries with available skilled wages:
['"Costa Rica"', '"Chile"', '"Mexico"', '"Barbados"', '"Brazil"', '"Grenada"',
'"Panama"', '"Argentina"', '"Cuba"', '"Suriname"', '"Uruguay"', '"Trinidad and
Tobago"', '"Martinique"', '"Netherlands Antilles"', '"Venezuela, Bolivarian Rep.
of"', '"Colombia"', '"Dominican Republic"', '"Hungary"', '"Bahamas"', '"Saint
Lucia"']
Total nr. of countries available:  20

Cluster  4  - countries with available high-skilled wages:
['"Costa Rica"', '"Chile"', '"Grenada"', '"Barbados"', '"Brazil"', '"Hungary"',
'"Panama"', '"Argentina"', '"Cuba"', '"Suriname"', '"Uruguay"', '"Trinidad and
Tobago"', '"Netherlands Antilles"', '"Venezuela, Bolivarian Rep. of"',
'"Colombia"', '"Dominican Republic"', '"Mexico"', '"Bahamas"', '"Saint Lucia"']
Total nr. of countries available:  19

Cluster  5  - countries with available unskilled wages:
['"Bahrain"', '"Kuwait"', '"Algeria"', '"Egypt"', '"West Bank and Gaza Strip"',
'"Jordan"', '"Syrian Arab Republic"', '"Yemen"', '"Iran, Islamic Rep. of"',
'"Qatar"']
Total nr. of countries available:  10

Cluster  5  - countries with available skilled wages:
['"Bahrain"', '"Kuwait"', '"Algeria"', '"Egypt"', '"West Bank and Gaza Strip"',
'"Jordan"', '"Syrian Arab Republic"', '"Yemen"', '"Qatar"']
Total nr. of countries available:  9

Cluster  5  - countries with available high-skilled wages:
['"Bahrain"', '"Algeria"', '"Egypt"', '"West Bank and Gaza Strip"', '"Jordan"',
'"Yemen"', '"Qatar"']
Total nr. of countries available:  7

Cluster  6  - countries with available unskilled wages:
['"Hong Kong, China"', '"United States"', '"Finland"', '"United Kingdom"',
'"Macau, China"', '"Puerto Rico"', '"Greece"', '"Austria"', '"Brunei
Darussalam"', '"Slovenia"', '"Canada"', '"Spain"', '"Australia"', '"New
Zealand"', '"Denmark"', '"Netherlands"', '"Korea, Republic of"', '"New
Caledonia"', '"Cyprus"', '"Singapore"', '"Germany"', '"Ireland"', '"Sweden"',
'"Norway"', '"Switzerland"', '"France"', '"Belgium"', '"Italy"', '"Japan"',
'"Luxembourg"', '"Israel"', '"Iceland"']
Total nr. of countries available:  32

Cluster  6  - countries with available skilled wages:
['"Hong Kong, China"', '"United States"', '"Australia"', '"Macau, China"',
'"Taiwan, China"', '"Puerto Rico"', '"Greece"', '"Austria"', '"Brunei
```

```
Darussalam"', '"Slovenia"', '"Canada"', '"Spain"', '"Finland"', '"Norway"',
'"New Zealand"', '"New Caledonia"', '"Netherlands"', '"Korea, Republic of"',
'"Cyprus"', '"Virgin Islands (US)"', '"Singapore"', '"Germany"', '"Ireland"',
'"Sweden"', '"Denmark"', '"United Kingdom"', '"Switzerland"', '"France"',
'"Belgium"', '"Italy"', '"Japan"', '"Luxembourg"', '"Israel"', '"Iceland"']
Total nr. of countries available:  34

Cluster  6  - countries with available high-skilled wages:
['"Hong Kong, China"', '"United States"', '"Australia"', '"Macau, China"',
'"Puerto Rico"', '"Austria"', '"Slovenia"', '"Canada"', '"United Kingdom"',
'"Finland"', '"New Zealand"', '"Denmark"', '"Spain"', '"Netherlands"',
'"Iceland"', '"Cyprus"', '"Singapore"', '"Germany"', '"Ireland"', '"Sweden"',
'"Norway"', '"Switzerland"', '"Belgium"', '"Italy"', '"Japan"', '"Luxembourg"',
'"Israel"', '"Korea, Republic of"']
Total nr. of countries available:  28
```

Attachment 3.1

high-skilled

AIC for:	**base**	**+conflict**	**+remit**	**+urban&gini**	**+school**	**+school&remit**
0 - Africa	**263.5**	231.1	**154.9** (too few data)	53.60 (too few data)	189.6	144.0 (too few data)
1 – Asia	**345.4**	**229.5**	217.8	256.8	270.6	**154.3**
2 - Mix	**163.3**	**88.71**	100.8	107.7	100.9	**87.87**
3 - Frontier	**28.50**	**0.5014**	12.96	27.33	14.41	**-89.03**
4 – Latin	**368.7**	277.2	254.6	**159.8**	139.5	**47.37**
5 - Arabic	**31.47**	**-0.4187**	**-4.231** (too few data)	14.15	31.66	-6.030
6 - Western	**-384.6**	-319.1	**-428.7**	-418.3	-302.5	-448.3

skilled

AIC	**base**	**+conflict**	**+remit**	**+urban&gini**	**+school**	**+school&remit**
0 - Africa	**421.5**	396.9	266.2	**134.1**	363.2	**249.4**
1 - Asia	**619.8**	**450.3**	211.0	390.3	519.8	**204.5**
2 - Mix	**418.6**	**305.4**	257.5	307.6	283.5	**253.1**
3 - Frontier	**692.5**	**506.6**	60.72	596.4	519.2	**23.36**
4 - Latin	**625.4**	500.9	351.2	**168.8**	162.7	**99.21**
5 - Arabic	**-18.28**	**-23.74**	**-40.84** (too few data)	6.248	-7.596	-33.54
6 - Western	**971.9**	901.2	**-1123.0**	601.0	889.6	-1101.0

unskilled

AIC	**base**	**+conflict**	**+remit**	**+urban&gini**	**+school**	**+school&remit**
0 - Africa	**262.0**	243.7	139.5	**109.7**	204.0	**119.6**
1 - Asia	**528.5**	**357.7**	143.7	280.1	441.4	**103.2**
2 - Mix	**270.4**	**180.5**	130.9	110.4	179.2	**118.7**
3 - Frontier	**684.1**	**467.9**	109.3	611.5	469.0	**37.33**
4 - Latin	**652.1**	567.4	344.2	**219.0**	180.9	**47.27**
5 - Arabic	**84.82**	**6.948**	**18.64** (too few data)	39.42	60.98	19.92
6 - Western	**948.8**	903.4	**-850.9**	609.0	883.1	-887.2

Red values are not considered improvements of the base model in column 1-4, as the number of parameters is probably under-penalized by AIC (cf. AICc) and these models barely have a lower AIC than the base model. Red values are not considered an improvement of the remittances only model in column 5. Bold values are chosen for further analysis.
The combination urban-gini-remittances was also explored, instead of school-remittances, but school-remittances has a better AIC for almost all clusters and skill levels. Where this was not the case, a model including urbanization and gini only was usually analysed beside conflict or remittances.
The values above are not entirely comparable because the models with added covariates are applied to subsets of the data where this covariate is available; the F-tests for nested models are more accurate, but resulted in close to all extensions being a significant improvement on the base model, and do not provide additional comparative information. The F-test also does not take into account number of covariates.

Attachment 3.2

Summary statistics base model: unskilled wage growth

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

```
Percentage of the data from decennium:
y1970s    0.07265
y1980s    0.50000
y1990s    0.34188
y2000s    0.08547
Mean wage growth over period (log):  0.0634724867712
Cluster:  0                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.148
Model:                            OLS   Adj. R-squared:                  0.114
Method:                 Least Squares   F-statistic:                     4.331
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           3.25e-05
Time:                        23:36:50   Log-Likelihood:                -121.01
No. Observations:                 234   AIC:                             262.0
Df Residuals:                     224   BIC:                             296.6
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.1306      0.159     -0.823      0.411      -0.444       0.182
cl0im          0.0100      0.014      0.691      0.490      -0.019       0.039
cl1im         -0.0476      0.021     -2.243      0.026      -0.089      -0.006
cl2im          0.0402      0.120      0.335      0.738      -0.196       0.276
cl0em          0.0067      0.008      0.857      0.392      -0.009       0.022
cl1em          0.0875      0.017      5.156      0.000       0.054       0.121
cl2em         -0.0636      0.059     -1.080      0.281      -0.180       0.053
y1970s        -0.1178      0.114     -1.037      0.301      -0.342       0.106
y1990s         0.0053      0.069      0.077      0.938      -0.130       0.140
y2000s         0.1008      0.104      0.967      0.335      -0.105       0.306
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.147147
y1980s    0.291291
y1990s    0.312312
y2000s    0.249249
Mean wage growth over period (log):  0.0513777380693
Cluster:  1                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.057
Model:                            OLS   Adj. R-squared:                  0.030
Method:                 Least Squares   F-statistic:                     2.153
Date:                Sun, 14 Aug 2016   Prob (F-statistic):             0.0250
Time:                        23:36:50   Log-Likelihood:                -254.24
No. Observations:                 333   AIC:                             528.5
Df Residuals:                     323   BIC:                             566.6
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.4190      0.265     -1.578      0.116      -0.941       0.103
cl0im          0.0042      0.028      0.152      0.879      -0.050       0.059
cl1im         -0.1656      0.146     -1.135      0.257      -0.453       0.121
cl2im          0.1369      0.099      1.379      0.169      -0.058       0.332
cl0em          0.0212      0.016      1.294      0.197      -0.011       0.053
cl1em         -0.0765      0.038     -1.993      0.047      -0.152      -0.001
cl2em         -0.0500      0.034     -1.451      0.148      -0.118       0.018
y1970s        -0.1121      0.097     -1.157      0.248      -0.303       0.079
y1990s        -0.0002      0.086     -0.002      0.998      -0.170       0.170
y2000s         0.1551      0.085      1.826      0.069      -0.012       0.322
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.151335
y1980s    0.281899
y1990s    0.359050
y2000s    0.207715
Mean wage growth over period (log):  0.0838548151868
Cluster:  2                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.255
Model:                            OLS   Adj. R-squared:                  0.235
Method:                 Least Squares   F-statistic:                     12.45
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           5.35e-17
Time:                        23:36:50   Log-Likelihood:                -125.21
No. Observations:                 337   AIC:                             270.4
Df Residuals:                     327   BIC:                             308.6
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.2319      0.037     -6.206      0.000      -0.305      -0.158
cl0im         -0.0252      0.017     -1.483      0.139      -0.059       0.008
cl1im          0.0554      0.025      2.180      0.030       0.005       0.105
cl2im         -0.1385      0.037     -3.713      0.000      -0.212      -0.065
cl0em          0.0138      0.011      1.279      0.202      -0.007       0.035
cl1em         -0.0447      0.015     -2.997      0.003      -0.074      -0.015
cl2em          0.0424      0.011      3.747      0.000       0.020       0.065
y1970s         0.1751      0.064      2.749      0.006       0.050       0.300
y1990s         0.2195      0.053      4.172      0.000       0.116       0.323
y2000s         0.1754      0.059      2.990      0.003       0.060       0.291
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.124590
y1980s    0.186885
y1990s    0.337705
y2000s    0.350820
Mean wage growth over period (log):  0.137575168591
Cluster:  3                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.073
Model:                            OLS   Adj. R-squared:                  0.059
Method:                 Least Squares   F-statistic:                     5.262
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           6.43e-07
Time:                        23:36:50   Log-Likelihood:                -332.04
No. Observations:                 610   AIC:                             684.1
Df Residuals:                     600   BIC:                             728.2
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0120      0.042     -0.288      0.774      -0.094       0.070
cl0im         -0.0108      0.008     -1.281      0.201      -0.027       0.006
cl1im          0.0217      0.014      1.508      0.132      -0.007       0.050
cl2im          0.0418      0.029      1.434      0.152      -0.015       0.099
cl0em          0.0242      0.013      1.798      0.073      -0.002       0.051
cl1em         -0.0020      0.013     -0.152      0.879      -0.028       0.024
cl2em          0.0193      0.022      0.872      0.383      -0.024       0.063
y1970s         0.1784      0.063      2.822      0.005       0.054       0.302
y1990s        -0.1381      0.064     -2.145      0.032      -0.265      -0.012
y2000s         0.2660      0.052      5.101      0.000       0.164       0.368
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.208238
y1980s    0.306636
y1990s    0.283753
y2000s    0.201373
Mean wage growth over period (log):  0.0755948697656
Cluster:  4                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.148
Model:                            OLS   Adj. R-squared:                  0.130
Method:                 Least Squares   F-statistic:                     8.239
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           2.47e-11
Time:                        23:36:50   Log-Likelihood:                -316.05
No. Observations:                 437   AIC:                             652.1
Df Residuals:                     427   BIC:                             692.9
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0316      0.066      0.479      0.632      -0.098       0.161
cl0im          0.0483      0.031      1.582      0.114      -0.012       0.108
cl1im          0.0893      0.028      3.178      0.002       0.034       0.145
cl2im         -0.1567      0.033     -4.789      0.000      -0.221      -0.092
cl0em          0.0705      0.027      2.639      0.009       0.018       0.123
cl1em         -0.0418      0.025     -1.681      0.094      -0.091       0.007
cl2em          0.0499      0.022      2.316      0.021       0.008       0.092
y1970s         0.1011      0.071      1.434      0.152      -0.037       0.240
y1990s         0.1413      0.068      2.080      0.038       0.008       0.275
y2000s         0.2154      0.078      2.777      0.006       0.063       0.368
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.085470
y1980s    0.188034
y1990s    0.418803
y2000s    0.307692
Mean wage growth over period (log):  0.0958748872738
Cluster:  5                            OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.358
Model:                           OLS   Adj. R-squared:                  0.304
Method:                Least Squares   F-statistic:                     6.617
Date:               Sun, 14 Aug 2016   Prob (F-statistic):           1.79e-07
Time:                       23:36:50   Log-Likelihood:                -32.411
No. Observations:                117   AIC:                             84.82
Df Residuals:                    107   BIC:                             112.4
Df Model:                          9
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0924      0.084      1.097      0.275      -0.075       0.259
cl0im          0.0234      0.023      1.021      0.309      -0.022       0.069
cl1im         -0.0818      0.049     -1.678      0.096      -0.178       0.015
cl2im          0.0507      0.022      2.304      0.023       0.007       0.094
cl0em         -0.0403      0.015     -2.768      0.007      -0.069      -0.011
cl1em          0.0651      0.026      2.470      0.015       0.013       0.117
cl2em          0.0002      0.024      0.010      0.992      -0.047       0.047
y1970s        -0.8421      0.133     -6.323      0.000      -1.106      -0.578
y1990s        -0.2121      0.100     -2.117      0.037      -0.411      -0.014
y2000s        -0.0190      0.103     -0.186      0.853      -0.222       0.184
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.200380
y1980s    0.266857
y1990s    0.272555
y2000s    0.260209
Mean wage growth over period (log):  0.0918471227779
Cluster:  6                            OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.046
Model:                           OLS   Adj. R-squared:                  0.037
Method:                Least Squares   F-statistic:                     5.546
Date:               Sun, 14 Aug 2016   Prob (F-statistic):           1.72e-07
Time:                       23:36:50   Log-Likelihood:                -464.39
No. Observations:               1053   AIC:                             948.8
Df Residuals:                   1043   BIC:                             998.4
Df Model:                          9
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.1244      0.026      4.717      0.000       0.073       0.176
cl0im          0.0073      0.006      1.208      0.228      -0.005       0.019
cl1im         -0.0124      0.009     -1.450      0.147      -0.029       0.004
cl2im         -0.0176      0.009     -1.927      0.054      -0.036       0.000
cl0em         -0.0060      0.008     -0.727      0.467      -0.022       0.010
cl1em          0.0361      0.012      3.084      0.002       0.013       0.059
cl2em         -0.0284      0.019     -1.492      0.136      -0.066       0.009
y1970s        -0.0712      0.035     -2.031      0.042      -0.140      -0.002
y1990s        -0.1012      0.038     -2.635      0.009      -0.177      -0.026
y2000s        -0.2034      0.036     -5.697      0.000      -0.273      -0.133
==============================================================================
```

Attachment 3.3

Summary statistics base model: skilled wage growth

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

```
Percentage of the data from decennium:
 y1970s    0.240437
y1980s    0.412568
y1990s    0.275956
y2000s    0.071038

Mean wage growth over period (log):  0.00927323756686
Cluster:  0                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.126
Model:                            OLS   Adj. R-squared:                  0.104
Method:                 Least Squares   F-statistic:                     5.687
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           2.26e-07
Time:                        23:45:14   Log-Likelihood:                -200.77
No. Observations:                 366   AIC:                             421.5
Df Residuals:                     356   BIC:                             460.6
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0168      0.121     -0.139      0.890      -0.255       0.221
cl0im          0.0020      0.010      0.208      0.835      -0.017       0.021
cl1im          0.0080      0.016      0.496      0.620      -0.024       0.040
cl2im          0.1587      0.094      1.690      0.092      -0.026       0.343
cl0em          0.0206      0.006      3.272      0.001       0.008       0.033
cl1em         -0.0115      0.014     -0.821      0.412      -0.039       0.016
cl2em         -0.0479      0.053     -0.912      0.362      -0.151       0.055
y1970s        -0.2433      0.059     -4.127      0.000      -0.359      -0.127
y1990s         0.0478      0.061      0.781      0.435      -0.073       0.168
y2000s         0.1081      0.094      1.144      0.253      -0.078       0.294
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.167064
y1980s    0.260143
y1990s    0.288783
y2000s    0.284010
Mean wage growth over period (log):  0.0793877242628
Cluster:  1                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.044
Model:                            OLS   Adj. R-squared:                  0.023
Method:                 Least Squares   F-statistic:                     2.095
Date:                Sun, 14 Aug 2016   Prob (F-statistic):             0.0289
Time:                        23:45:14   Log-Likelihood:                -299.88
No. Observations:                 419   AIC:                             619.8
Df Residuals:                     409   BIC:                             660.1
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0329      0.131     -0.251      0.802      -0.290       0.224
cl0im          0.0597      0.025      2.371      0.018       0.010       0.109
cl1im         -0.0667      0.067     -0.993      0.321      -0.199       0.065
cl2im         -0.0310      0.068     -0.457      0.648      -0.164       0.102
cl0em         -0.0064      0.014     -0.457      0.648      -0.034       0.021
cl1em         -0.0079      0.026     -0.300      0.765      -0.060       0.044
cl2em          0.0511      0.029      1.743      0.082      -0.007       0.109
y1970s        -0.1592      0.080     -1.985      0.048      -0.317      -0.002
y1990s        -0.0914      0.074     -1.230      0.219      -0.237       0.055
y2000s         0.0980      0.071      1.371      0.171      -0.043       0.239
==============================================================================


Percentage of the data from decennium:
 y1970s    0.184091
y1980s    0.270455
y1990s    0.329545
y2000s    0.215909
Mean wage growth over period (log):  0.0475636102121
Cluster:  2                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.139
Model:                            OLS   Adj. R-squared:                  0.121
Method:                 Least Squares   F-statistic:                     7.730
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           1.44e-10
Time:                        23:45:15   Log-Likelihood:                -199.31
No. Observations:                 440   AIC:                             418.6
Df Residuals:                     430   BIC:                             459.5
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.2679      0.036     -7.396      0.000      -0.339      -0.197
cl0im         -0.0226      0.015     -1.511      0.132      -0.052       0.007
cl1im          0.0713      0.025      2.850      0.005       0.022       0.120
cl2im         -0.1128      0.037     -3.035      0.003      -0.186      -0.040
cl0em          0.0078      0.011      0.709      0.478      -0.014       0.030
cl1em         -0.0385      0.015     -2.501      0.013      -0.069      -0.008
cl2em          0.0326      0.011      2.901      0.004       0.011       0.055
y1970s         0.1738      0.056      3.088      0.002       0.063       0.284
y1990s         0.2572      0.053      4.863      0.000       0.153       0.361
y2000s         0.2676      0.055      4.906      0.000       0.160       0.375
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.117898
y1980s    0.237216
y1990s    0.325284
y2000s    0.319602
Mean wage growth over period (log):  0.188963185405
Cluster:  3                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.053
Model:                            OLS   Adj. R-squared:                  0.041
Method:                 Least Squares   F-statistic:                     4.313
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           1.77e-05
Time:                        23:45:15   Log-Likelihood:                -336.27
No. Observations:                 704   AIC:                             692.5
Df Residuals:                     694   BIC:                             738.1
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.1117      0.034      3.296      0.001       0.045       0.178
cl0im         -0.0094      0.008     -1.128      0.260      -0.026       0.007
cl1im          0.0199      0.011      1.795      0.073      -0.002       0.042
cl2im          0.0117      0.022      0.539      0.590      -0.031       0.054
cl0em          0.0206      0.012      1.722      0.085      -0.003       0.044
cl1em          0.0060      0.011      0.557      0.578      -0.015       0.027
cl2em          0.0180      0.018      0.979      0.328      -0.018       0.054
y1970s        -0.1370      0.054     -2.551      0.011      -0.242      -0.032
y1990s        -0.1432      0.055     -2.622      0.009      -0.250      -0.036
y2000s         0.0888      0.043      2.046      0.041       0.004       0.174
==============================================================================


Percentage of the data from decennium:
 y1970s    0.150358
y1980s    0.274463
y1990s    0.322196
y2000s    0.252983
Mean wage growth over period (log):  0.120617500436
Cluster:  4                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.170
Model:                            OLS   Adj. R-squared:                  0.152
Method:                 Least Squares   F-statistic:                     9.322
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           6.49e-13
Time:                        23:45:15   Log-Likelihood:                -302.70
No. Observations:                 419   AIC:                             625.4
Df Residuals:                     409   BIC:                             665.8
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0457      0.063      0.722      0.471      -0.079       0.170
cl0im         -0.0014      0.030     -0.046      0.963      -0.061       0.058
cl1im          0.1491      0.035      4.322      0.000       0.081       0.217
cl2im         -0.1028      0.023     -4.465      0.000      -0.148      -0.058
cl0em          0.0946      0.025      3.726      0.000       0.045       0.144
cl1em         -0.0652      0.022     -3.031      0.003      -0.108      -0.023
cl2em         -0.0195      0.018     -1.079      0.281      -0.055       0.016
y1970s         0.1381      0.082      1.694      0.091      -0.022       0.298
y1990s         0.2543      0.069      3.699      0.000       0.119       0.389
y2000s         0.3014      0.074      4.081      0.000       0.156       0.447
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.124183
y1980s    0.176471
y1990s    0.352941
y2000s    0.346405
Mean wage growth over period (log):  0.0661607153458
Cluster:  5                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.119
Model:                            OLS   Adj. R-squared:                  0.064
Method:                 Least Squares   F-statistic:                     2.152
Date:                Sun, 14 Aug 2016   Prob (F-statistic):             0.0288
Time:                        23:45:15   Log-Likelihood:                 19.142
No. Observations:                 153   AIC:                            -18.28
Df Residuals:                     143   BIC:                             12.02
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.1648      0.054     -3.079      0.002      -0.271      -0.059
cl0im          0.0304      0.010      3.163      0.002       0.011       0.049
cl1im         -0.0617      0.026     -2.407      0.017      -0.112      -0.011
cl2im          0.0034      0.010      0.337      0.736      -0.017       0.024
cl0em         -0.0135      0.006     -2.193      0.030      -0.026      -0.001
cl1em          0.0179      0.010      1.870      0.064      -0.001       0.037
cl2em          0.0234      0.011      2.100      0.038       0.001       0.045
y1970s         0.0880      0.069      1.274      0.205      -0.049       0.225
y1990s         0.0813      0.060      1.348      0.180      -0.038       0.201
y2000s         0.1253      0.066      1.910      0.058      -0.004       0.255
==============================================================================


Percentage of the data from decennium:
 y1970s    0.174805
y1980s    0.264648
y1990s    0.286133
y2000s    0.274414
Mean wage growth over period (log):  0.0922367678813
Cluster:  6                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.048
Model:                            OLS   Adj. R-squared:                  0.040
Method:                 Least Squares   F-statistic:                     5.728
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           8.82e-08
Time:                        23:45:15   Log-Likelihood:                -475.97
No. Observations:                1024   AIC:                             971.9
Df Residuals:                    1014   BIC:                             1021.
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.1338      0.027      4.997      0.000       0.081       0.186
cl0im         -0.0006      0.006     -0.095      0.925      -0.013       0.012
cl1im          0.0011      0.009      0.123      0.902      -0.016       0.018
cl2im         -0.0129      0.009     -1.381      0.167      -0.031       0.005
cl0em          0.0027      0.007      0.391      0.696      -0.011       0.016
cl1em          0.0173      0.012      1.462      0.144      -0.006       0.041
cl2em         -0.0185      0.018     -1.006      0.315      -0.054       0.018
y1970s        -0.0965      0.038     -2.540      0.011      -0.171      -0.022
y1990s        -0.1642      0.038     -4.285      0.000      -0.239      -0.089
y2000s        -0.2042      0.037     -5.529      0.000      -0.277      -0.132
==============================================================================
```

Attachment 3.4

Summary statistics base model: high-skilled wage growth

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.518519
y1990s    0.392593
y2000s    0.088889
Mean wage growth over period (log):  0.118785965552
Cluster:  0                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.070
Model:                            OLS   Adj. R-squared:                  0.011
Method:                 Least Squares   F-statistic:                     1.180
Date:                Sun, 14 Aug 2016   Prob (F-statistic):              0.316
Time:                        23:47:43   Log-Likelihood:                -122.73
No. Observations:                 135   AIC:                             263.5
Df Residuals:                     126   BIC:                             289.6
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.1813      0.361     -0.502      0.617      -0.897       0.534
cl0im          0.0365      0.035      1.044      0.299      -0.033       0.106
cl1im         -0.1235      0.053     -2.343      0.021      -0.228      -0.019
cl2im         -0.0866      0.277     -0.312      0.755      -0.635       0.462
cl0em         -0.0044      0.017     -0.259      0.796      -0.038       0.029
cl1em          0.0838      0.032      2.647      0.009       0.021       0.147
cl2em          0.0222      0.103      0.215      0.830      -0.182       0.227
y1990s        -0.1446      0.123     -1.171      0.244      -0.389       0.100
y2000s        -0.0801      0.205     -0.391      0.697      -0.486       0.326
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.242775
y1990s    0.393064
y2000s    0.364162
Mean wage growth over period (log):  0.137507015086
Cluster:  1                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.156
Model:                            OLS   Adj. R-squared:                  0.115
Method:                 Least Squares   F-statistic:                     3.784
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           0.000424
Time:                        23:47:43   Log-Likelihood:                -163.72
No. Observations:                 173   AIC:                             345.4
Df Residuals:                     164   BIC:                             373.8
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.6832      0.455     -1.500      0.135      -1.582       0.216
cl0im          0.1095      0.055      2.009      0.046       0.002       0.217
cl1im         -0.7553      0.242     -3.116      0.002      -1.234      -0.277
cl2im          0.4402      0.188      2.342      0.020       0.069       0.811
cl0em          0.0436      0.028      1.537      0.126      -0.012       0.100
cl1em          0.1195      0.053      2.249      0.026       0.015       0.224
cl2em         -0.0192      0.055     -0.347      0.729      -0.128       0.090
y1990s        -0.3743      0.139     -2.694      0.008      -0.649      -0.100
y2000s         0.0036      0.135      0.027      0.979      -0.264       0.271
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.046243
y1980s    0.242775
y1990s    0.445087
y2000s    0.265896
Mean wage growth over period (log):  0.155984928161
Cluster:  2                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.375
Model:                            OLS   Adj. R-squared:                  0.345
Method:                 Least Squares   F-statistic:                     12.33
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           9.52e-14
Time:                        23:47:43   Log-Likelihood:                -72.649
No. Observations:                 173   AIC:                             163.3
Df Residuals:                     164   BIC:                             191.7
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.2170      0.065     -3.331      0.001      -0.346      -0.088
cl0im         -0.0524      0.021     -2.501      0.013      -0.094      -0.011
cl1im         -0.0940      0.053     -1.786      0.076      -0.198       0.010
cl2im          0.0636      0.050      1.262      0.209      -0.036       0.163
cl0em         -0.0290      0.012     -2.372      0.019      -0.053      -0.005
cl1em         -0.0478      0.023     -2.050      0.042      -0.094      -0.002
cl2em          0.0493      0.014      3.518      0.001       0.022       0.077
y1990s         0.2240      0.075      2.988      0.003       0.076       0.372
y2000s        -0.1304      0.085     -1.533      0.127      -0.298       0.038
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.084577
y1990s    0.373134
y2000s    0.542289
Mean wage growth over period (log):  0.206131473303
Cluster:  3                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.297
Model:                            OLS   Adj. R-squared:                  0.267
Method:                 Least Squares   F-statistic:                     10.12
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           9.36e-12
Time:                        23:47:43   Log-Likelihood:                -5.2508
No. Observations:                 201   AIC:                             28.50
Df Residuals:                     192   BIC:                             58.23
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0150      0.064     -0.234      0.815      -0.142       0.112
cl0im         -0.0464      0.011     -4.292      0.000      -0.068      -0.025
cl1im          0.0522      0.014      3.649      0.000       0.024       0.080
cl2im         -0.0510      0.034     -1.516      0.131      -0.117       0.015
cl0em          0.0744      0.017      4.376      0.000       0.041       0.108
cl1em         -0.0007      0.015     -0.047      0.963      -0.031       0.029
cl2em         -0.0129      0.030     -0.427      0.670      -0.072       0.047
y1990s         0.0122      0.090      0.136      0.892      -0.165       0.189
y2000s         0.2008      0.075      2.664      0.008       0.052       0.349
==============================================================================


Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.340314
y1990s    0.293194
y2000s    0.366492
Mean wage growth over period (log):  0.157150718482
Cluster:  4                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.296
Model:                            OLS   Adj. R-squared:                  0.265
Method:                 Least Squares   F-statistic:                     9.560
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           5.13e-11
Time:                        23:47:43   Log-Likelihood:                -175.35
No. Observations:                 191   AIC:                             368.7
Df Residuals:                     182   BIC:                             398.0
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0899      0.106     -0.847      0.398      -0.299       0.120
cl0im         -0.0088      0.048     -0.183      0.855      -0.103       0.086
cl1im          0.2084      0.054      3.874      0.000       0.102       0.315
cl2im         -0.2030      0.048     -4.264      0.000      -0.297      -0.109
cl0em          0.1723      0.046      3.752      0.000       0.082       0.263
cl1em         -0.1254      0.050     -2.533      0.012      -0.223      -0.028
cl2em          0.0537      0.035      1.555      0.122      -0.014       0.122
y1990s         0.4933      0.119      4.158      0.000       0.259       0.727
y2000s         0.4367      0.148      2.960      0.003       0.146       0.728
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.253521
y1990s    0.408451
y2000s    0.338028
Mean wage growth over period (log):  -0.0478953672622
Cluster:  5                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.400
Model:                            OLS   Adj. R-squared:                  0.323
Method:                 Least Squares   F-statistic:                     5.177
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           5.73e-05
Time:                        23:47:43   Log-Likelihood:                -6.7360
No. Observations:                  71   AIC:                             31.47
Df Residuals:                      62   BIC:                             51.84
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.4691      0.106     -4.423      0.000      -0.681      -0.257
cl0im         -0.0430      0.024     -1.810      0.075      -0.090       0.004
cl1im          0.0598      0.063      0.956      0.343      -0.065       0.185
cl2im          0.0798      0.034      2.346      0.022       0.012       0.148
cl0em         -0.0270      0.019     -1.438      0.155      -0.065       0.011
cl1em         -0.0387      0.026     -1.483      0.143      -0.091       0.013
cl2em          0.0013      0.032      0.041      0.967      -0.062       0.065
y1990s         0.1416      0.103      1.381      0.172      -0.063       0.347
y2000s         0.5316      0.130      4.099      0.000       0.272       0.791
==============================================================================


Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.280778
y1990s    0.347732
y2000s    0.371490
Mean wage growth over period (log):  0.0493197241304
Cluster:  6                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.120
Model:                            OLS   Adj. R-squared:                  0.105
Method:                 Least Squares   F-statistic:                     7.759
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           8.93e-10
Time:                        23:47:43   Log-Likelihood:                 201.28
No. Observations:                 463   AIC:                            -384.6
Df Residuals:                     454   BIC:                            -347.3
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0309      0.019     -1.658      0.098      -0.068       0.006
cl0im         -0.0004      0.004     -0.098      0.922      -0.008       0.007
cl1im          0.0069      0.007      0.989      0.323      -0.007       0.021
cl2im          0.0055      0.005      1.065      0.288      -0.005       0.016
cl0em          0.0184      0.007      2.475      0.014       0.004       0.033
cl1em         -0.0491      0.010     -4.848      0.000      -0.069      -0.029
cl2em          0.0418      0.013      3.142      0.002       0.016       0.068
y1990s        -0.0302      0.021     -1.462      0.144      -0.071       0.010
y2000s        -0.0436      0.022     -2.021      0.044      -0.086      -0.001
==============================================================================
```

Attachment 3.5

Summary statistics conflict model: unskilled wage growth, clusters 1, 2, 3, 5 only.

InterConflict = internal conflict, OuterConflict = international conflict, added 6 = six years before wage growth – at the same time as migration.

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

```
Percentage of the data from decennium:
 y1970s    0.190661
y1980s    0.373541
y1990s    0.404669
y2000s    0.031128
Mean wage growth over period (log):  0.00583900083266
Cluster:  1                             OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.243
Model:                           OLS   Adj. R-squared:                  0.189
Method:                Least Squares   F-statistic:                     4.509
Date:               Mon, 15 Aug 2016   Prob (F-statistic):           3.90e-08
Time:                       18:33:08   Log-Likelihood:                -160.83
No. Observations:                257   AIC:                             357.7
Df Residuals:                    239   BIC:                             421.5
Df Model:                         17
Covariance Type:           nonrobust
========================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept             -0.3879      0.263     -1.473      0.142      -0.907       0.131
InterConflict          0.0329      0.042      0.793      0.429      -0.049       0.115
OuterConflict         -0.0333      0.028     -1.191      0.235      -0.088       0.022
cl0im                 -0.0532      0.035     -1.539      0.125      -0.121       0.015
cl1im                  0.0570      0.149      0.381      0.703      -0.237       0.352
cl2im                  0.0580      0.099      0.584      0.559      -0.137       0.253
InterConflict6        -0.1491      0.079     -1.888      0.060      -0.305       0.006
cl0em                  0.0228      0.026      0.862      0.389      -0.029       0.075
InterConflict6:cl0em   0.0100      0.014      0.707      0.480      -0.018       0.038
OuterConflict6         0.2300      0.046      4.956      0.000       0.139       0.321
cl1em                 -0.2559      0.045     -5.718      0.000      -0.344      -0.168
cl2em                  0.0063      0.048      0.131      0.896      -0.088       0.100
InterConflict6:cl1em  -0.1469      0.070     -2.108      0.036      -0.284      -0.010
InterConflict6:cl2em   0.1788      0.048      3.690      0.000       0.083       0.274
OuterConflict6:cl1em   0.1386      0.029      4.754      0.000       0.081       0.196
OuterConflict6:cl2em  -0.0503      0.036     -1.402      0.162      -0.121       0.020
y1970s                -0.0076      0.088     -0.086      0.931      -0.182       0.167
y1990s                 0.0465      0.085      0.548      0.584      -0.121       0.213
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.182879
y1980s    0.330739
y1990s    0.451362
y2000s    0.035019
Mean wage growth over period (log):  0.106345372182
Cluster:  2                             OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.259
Model:                            OLS   Adj. R-squared:                  0.206
Method:                 Least Squares   F-statistic:                     4.919
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           4.58e-09
Time:                        18:33:08   Log-Likelihood:                -72.265
No. Observations:                 257   AIC:                             180.5
Df Residuals:                     239   BIC:                             244.4
Df Model:                          17
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               -0.1222      0.077     -1.593      0.112      -0.273       0.029
InterConflict           -0.0820      0.033     -2.466      0.014      -0.147      -0.017
OuterConflict           -0.1138      0.053     -2.161      0.032      -0.218      -0.010
cl0im                    0.0243      0.019      1.299      0.195      -0.013       0.061
cl1im                    0.0330      0.046      0.711      0.478      -0.059       0.125
cl2im                   -0.1370      0.054     -2.538      0.012      -0.243      -0.031
InterConflict6           0.1640      0.047      3.480      0.001       0.071       0.257
cl0em                    0.0052      0.011      0.485      0.628      -0.016       0.026
InterConflict6:cl0em     0.0533      0.022      2.422      0.016       0.010       0.097
OuterConflict6           0.0119      0.062      0.191      0.849      -0.111       0.134
cl1em                   -0.1049      0.026     -4.091      0.000      -0.155      -0.054
cl2em                    0.0795      0.030      2.615      0.009       0.020       0.139
InterConflict6:cl1em     0.0150      0.018      0.842      0.401      -0.020       0.050
InterConflict6:cl2em    -0.1074      0.028     -3.797      0.000      -0.163      -0.052
OuterConflict6:cl1em    -0.1473      0.079     -1.875      0.062      -0.302       0.007
OuterConflict6:cl2em     0.2149      0.097      2.222      0.027       0.024       0.405
y1970s                   0.0285      0.069      0.414      0.679      -0.107       0.164
y1990s                   0.1807      0.054      3.328      0.001       0.074       0.288
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.185366
y1980s    0.270732
y1990s    0.487805
y2000s    0.056098
Mean wage growth over period (log):  0.0901338525988
Cluster:  3                             OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.247
Model:                            OLS   Adj. R-squared:                  0.215
Method:                 Least Squares   F-statistic:                     7.580
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           2.66e-16
Time:                        18:33:08   Log-Likelihood:                -215.96
No. Observations:                 410   AIC:                             467.9
Df Residuals:                     392   BIC:                             540.2
Df Model:                          17
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept                0.0700      0.074      0.952      0.341      -0.075       0.215
InterConflict            0.3749      0.087      4.312      0.000       0.204       0.546
OuterConflict           -0.1889      0.025     -7.534      0.000      -0.238      -0.140
cl0im                   -0.0045      0.010     -0.437      0.662      -0.025       0.016
cl1im                   -0.0167      0.018     -0.906      0.365      -0.053       0.020
cl2im                    0.1212      0.035      3.480      0.001       0.053       0.190
InterConflict6          -0.2353      0.197     -1.196      0.232      -0.622       0.151
cl0em                    0.0710      0.043      1.653      0.099      -0.013       0.155
InterConflict6:cl0em     0.1361      0.124      1.098      0.273      -0.108       0.380
OuterConflict6           0.0425      0.029      1.447      0.149      -0.015       0.100
cl1em                   -0.0504      0.039     -1.287      0.199      -0.128       0.027
cl2em                    0.1637      0.069      2.365      0.019       0.028       0.300
InterConflict6:cl1em    -0.2342      0.114     -2.059      0.040      -0.458      -0.011
InterConflict6:cl2em     0.5060      0.188      2.688      0.007       0.136       0.876
OuterConflict6:cl1em    -0.0331      0.014     -2.312      0.021      -0.061      -0.005
OuterConflict6:cl2em     0.0910      0.040      2.289      0.023       0.013       0.169
y1970s                   0.1000      0.066      1.514      0.131      -0.030       0.230
y1990s                  -0.2574      0.066     -3.888      0.000      -0.388      -0.127
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.116279
y1980s    0.255814
y1990s    0.569767
y2000s    0.058140
Mean wage growth over period (log):  0.0326378252111
Cluster:  5                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.698
Model:                            OLS   Adj. R-squared:                  0.623
Method:                 Least Squares   F-statistic:                     9.266
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           7.71e-12
Time:                        18:33:08   Log-Likelihood:                 14.526
No. Observations:                  86   AIC:                             6.948
Df Residuals:                      68   BIC:                             51.13
Df Model:                          17
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               -0.0399      0.099     -0.402      0.689      -0.238       0.158
InterConflict           -0.0526      0.041     -1.288      0.202      -0.134       0.029
OuterConflict            0.2088      0.044      4.762      0.000       0.121       0.296
cl0im                    0.0179      0.058      0.310      0.757      -0.097       0.133
cl1im                   -0.1116      0.066     -1.695      0.095      -0.243       0.020
cl2im                    0.0633      0.033      1.907      0.061      -0.003       0.130
InterConflict6          -0.0424      0.146     -0.291      0.772      -0.333       0.248
cl0em                   -0.0144      0.045     -0.316      0.753      -0.105       0.076
InterConflict6:cl0em     0.0629      0.129      0.489      0.627      -0.194       0.320
OuterConflict6           0.1281      0.084      1.533      0.130      -0.039       0.295
cl1em                    0.1389      0.134      1.038      0.303      -0.128       0.406
cl2em                   -0.0744      0.165     -0.451      0.653      -0.403       0.255
InterConflict6:cl1em     0.1201      0.395      0.304      0.762      -0.668       0.908
InterConflict6:cl2em    -0.1521      0.459     -0.331      0.741      -1.068       0.764
OuterConflict6:cl1em     0.0939      0.090      1.047      0.299      -0.085       0.273
OuterConflict6:cl2em    -0.1510      0.137     -1.098      0.276      -0.425       0.123
y1970s                  -0.7330      0.198     -3.695      0.000      -1.129      -0.337
y1990s                  -0.0170      0.077     -0.222      0.825      -0.170       0.136
==============================================================================
```

Attachment 3.6

Summary statistics conflict model: skilled wage growth, clusters 1, 2, 3, 5 only.

InterConflict = internal conflict, OuterConflict = international conflict, added 6 = six years before wage growth – at the same time as migration.

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

```
Percentage of the data from decennium:
 y1970s    0.224359
y1980s    0.349359
y1990s    0.387821
y2000s    0.038462

Mean wage growth over period (log):  0.0425047001858
Cluster:  1                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.262
Model:                            OLS   Adj. R-squared:                  0.219
Method:                 Least Squares   F-statistic:                     6.136
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           3.28e-12
Time:                        18:43:59   Log-Likelihood:                -207.13
No. Observations:                 312   AIC:                             450.3
Df Residuals:                     294   BIC:                             517.6
Df Model:                          17
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept                0.0054      0.172      0.031      0.975      -0.333       0.343
InterConflict            0.0280      0.045      0.624      0.533      -0.060       0.116
OuterConflict            0.1338      0.027      4.923      0.000       0.080       0.187
cl0im                    0.0194      0.035      0.553      0.581      -0.050       0.088
cl1im                    0.0722      0.104      0.692      0.490      -0.133       0.278
cl2im                   -0.0392      0.083     -0.470      0.638      -0.203       0.125
InterConflict6          -0.2598      0.085     -3.047      0.003      -0.428      -0.092
cl0em                   -0.0103      0.020     -0.521      0.603      -0.049       0.029
InterConflict6:cl0em     0.0493      0.014      3.420      0.001       0.021       0.078
OuterConflict6           0.0606      0.042      1.456      0.147      -0.021       0.142
cl1em                   -0.1159      0.042     -2.740      0.007      -0.199      -0.033
cl2em                    0.1306      0.040      3.262      0.001       0.052       0.209
InterConflict6:cl1em    -0.2363      0.076     -3.116      0.002      -0.386      -0.087
InterConflict6:cl2em     0.1659      0.048      3.457      0.001       0.071       0.260
OuterConflict6:cl1em     0.1396      0.027      5.179      0.000       0.087       0.193
OuterConflict6:cl2em    -0.0626      0.034     -1.821      0.070      -0.130       0.005
y1970s                  -0.0924      0.079     -1.164      0.245      -0.249       0.064
y1990s                  -0.0568      0.078     -0.730      0.466      -0.210       0.096
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.227139
y1980s    0.321534
y1990s    0.412979
y2000s    0.038348

Mean wage growth over period (log):  0.0271764089381
Cluster:  2                            OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.201
Model:                           OLS   Adj. R-squared:                  0.159
Method:                Least Squares   F-statistic:                     4.761
Date:               Mon, 15 Aug 2016   Prob (F-statistic):           4.70e-09
Time:                       18:43:59   Log-Likelihood:               -134.68
No. Observations:                339   AIC:                             305.4
Df Residuals:                    321   BIC:                             374.2
Df Model:                         17
Covariance Type:           nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               -0.2737      0.072     -3.824      0.000      -0.415      -0.133
InterConflict            0.0390      0.033      1.199      0.231      -0.025       0.103
OuterConflict            0.0186      0.038      0.488      0.626      -0.056       0.094
cl0im                   -0.0060      0.016     -0.374      0.708      -0.038       0.026
cl1im                    0.1039      0.038      2.710      0.007       0.028       0.179
cl2im                   -0.1790      0.056     -3.223      0.001      -0.288      -0.070
InterConflict6        9.437e-05      0.037      0.003      0.998      -0.072       0.072
cl0em                    0.0076      0.011      0.673      0.501      -0.015       0.030
InterConflict6:cl0em     0.0115      0.024      0.487      0.626      -0.035       0.058
OuterConflict6          -0.1167      0.067     -1.746      0.082      -0.248       0.015
cl1em                   -0.0588      0.027     -2.164      0.031      -0.112      -0.005
cl2em                    0.0318      0.031      1.013      0.312      -0.030       0.093
InterConflict6:cl1em     0.0388      0.018      2.135      0.034       0.003       0.074
InterConflict6:cl2em    -0.0233      0.024     -0.960      0.338      -0.071       0.024
OuterConflict6:cl1em    -0.0559      0.088     -0.634      0.526      -0.229       0.118
OuterConflict6:cl2em    -0.0423      0.103     -0.411      0.681      -0.244       0.160
y1970s                   0.1191      0.061      1.940      0.053      -0.002       0.240
y1990s                   0.2583      0.055      4.705      0.000       0.150       0.366
==============================================================================


Percentage of the data from decennium:
 y1970s    0.168016
y1980s    0.331984
y1990s    0.451417
y2000s    0.048583

Mean wage growth over period (log):  0.176173000221
Cluster:  3                            OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.218
Model:                           OLS   Adj. R-squared:                  0.190
Method:                Least Squares   F-statistic:                     7.805
Date:               Mon, 15 Aug 2016   Prob (F-statistic):           2.57e-17
Time:                       18:43:59   Log-Likelihood:               -235.29
No. Observations:                494   AIC:                             506.6
Df Residuals:                    476   BIC:                             582.2
Df Model:                         17
Covariance Type:           nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept                0.1789      0.075      2.392      0.017       0.032       0.326
InterConflict            0.1337      0.086      1.556      0.120      -0.035       0.303
OuterConflict           -0.1793      0.022     -8.328      0.000      -0.222      -0.137
cl0im                   -0.0061      0.010     -0.607      0.544      -0.026       0.014
cl1im                    0.0080      0.013      0.601      0.548      -0.018       0.034
cl2im                    0.0259      0.027      0.976      0.329      -0.026       0.078
InterConflict6          -0.0093      0.230     -0.040      0.968      -0.462       0.444
cl0em                    0.0454      0.046      0.995      0.320      -0.044       0.135
InterConflict6:cl0em     0.0639      0.130      0.492      0.623      -0.191       0.319
OuterConflict6           0.0994      0.042      2.366      0.018       0.017       0.182
cl1em                   -0.0343      0.043     -0.801      0.424      -0.118       0.050
cl2em                    0.0775      0.079      0.984      0.326      -0.077       0.232
InterConflict6:cl1em    -0.1208      0.123     -0.981      0.327      -0.363       0.121
InterConflict6:cl2em     0.2418      0.222      1.091      0.276      -0.194       0.677
OuterConflict6:cl1em    -0.0055      0.007     -0.808      0.419      -0.019       0.008
OuterConflict6:cl2em     0.0188      0.022      0.860      0.390      -0.024       0.062
y1970s                  -0.1926      0.055     -3.510      0.000      -0.300      -0.085
y1990s                  -0.1736      0.059     -2.926      0.004      -0.290      -0.057
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.19
y1980s    0.27
y1990s    0.50
y2000s    0.04

Mean wage growth over period (log):  0.0659774620226
Cluster:  5                             OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.398
Model:                           OLS   Adj. R-squared:                  0.274
Method:                Least Squares   F-statistic:                     3.194
Date:               Mon, 15 Aug 2016   Prob (F-statistic):           0.000223
Time:                       18:43:59   Log-Likelihood:                 29.872
No. Observations:                100   AIC:                            -23.74
Df Residuals:                     82   BIC:                             23.15
Df Model:                         17
Covariance Type:           nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               -0.0571      0.073     -0.786      0.434      -0.202       0.087
InterConflict           -0.0347      0.038     -0.926      0.357      -0.109       0.040
OuterConflict           -0.0327      0.037     -0.877      0.383      -0.107       0.041
cl0im                    0.0009      0.046      0.019      0.985      -0.091       0.093
cl1im                   -0.0113      0.053     -0.214      0.831      -0.117       0.094
cl2im                    0.0165      0.027      0.620      0.537      -0.036       0.069
InterConflict6           0.1947      0.085      2.281      0.025       0.025       0.365
cl0em                    0.0302      0.029      1.027      0.307      -0.028       0.089
InterConflict6:cl0em     0.1789      0.084      2.132      0.036       0.012       0.346
OuterConflict6           0.0346      0.047      0.731      0.467      -0.060       0.129
cl1em                    0.0365      0.025      1.473      0.144      -0.013       0.086
cl2em                    0.0195      0.026      0.750      0.455      -0.032       0.071
InterConflict6:cl1em    -0.1239      0.055     -2.262      0.026      -0.233      -0.015
InterConflict6:cl2em     0.1295      0.046      2.824      0.006       0.038       0.221
OuterConflict6:cl1em     0.0355      0.017      2.146      0.035       0.003       0.068
OuterConflict6:cl2em    -0.0662      0.022     -3.061      0.003      -0.109      -0.023
y1970s                   0.2965      0.103      2.892      0.005       0.093       0.500
y1990s                  -0.0189      0.068     -0.276      0.783      -0.155       0.117
==============================================================================
```

Attachment 3.7

Summary statistics conflict model: high-skilled wage growth, clusters 1, 2, 3, 5 only.

InterConflict = internal conflict, OuterConflict = international conflict, added 6 = six years before wage growth – at the same time as migration.

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.362069
y1990s    0.586207
y2000s    0.051724
Mean wage growth over period (log):  0.113198126289
Cluster:  1                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.442
Model:                            OLS   Adj. R-squared:                  0.358
Method:                 Least Squares   F-statistic:                     5.283
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           1.18e-07
Time:                        18:40:58   Log-Likelihood:                -98.765
No. Observations:                 116   AIC:                             229.5
Df Residuals:                     100   BIC:                             273.6
Df Model:                          15
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               -1.0835      0.717     -1.511      0.134      -2.506       0.339
InterConflict            0.0813      0.074      1.101      0.274      -0.065       0.228
OuterConflict            0.0433      0.048      0.895      0.373      -0.053       0.139
cl0im                    0.1271      0.085      1.499      0.137      -0.041       0.295
cl1im                   -0.6445      0.396     -1.628      0.107      -1.430       0.141
cl2im                    0.1191      0.272      0.439      0.662      -0.420       0.658
InterConflict6          -0.2425      0.128     -1.900      0.060      -0.496       0.011
cl0em                    0.0565      0.061      0.924      0.358      -0.065       0.178
InterConflict6:cl0em     0.0134      0.030      0.447      0.656      -0.046       0.073
cl1em                   -0.0368      0.093     -0.393      0.695      -0.222       0.149
cl2em                   -0.0610      0.131     -0.465      0.643      -0.321       0.199
InterConflict6:cl1em    -0.1236      0.114     -1.085      0.281      -0.350       0.102
InterConflict6:cl2em     0.2203      0.088      2.497      0.014       0.045       0.395
OuterConflict6:cl1em     0.1142      0.044      2.599      0.011       0.027       0.201
OuterConflict6:cl2em    -0.0852      0.077     -1.112      0.269      -0.237       0.067
y1990s                  -0.0583      0.147     -0.397      0.692      -0.349       0.233
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.308333
y1990s    0.641667
y2000s    0.050000
Mean wage growth over period (log):  0.202365469733
Cluster:  2                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.681
Model:                            OLS   Adj. R-squared:                  0.635
Method:                 Least Squares   F-statistic:                     14.81
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           1.60e-19
Time:                        18:40:58   Log-Likelihood:                -28.355
No. Observations:                 120   AIC:                             88.71
Df Residuals:                     104   BIC:                             133.3
Df Model:                          15
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
Intercept                -0.0405      0.135     -0.299      0.765      -0.309       0.228
InterConflict             0.1606      0.084      1.902      0.060      -0.007       0.328
OuterConflict             0.9227      0.251      3.669      0.000       0.424       1.421
cl0im                    -0.0302      0.022     -1.390      0.167      -0.073       0.013
cl1im                    -0.1066      0.068     -1.571      0.119      -0.241       0.028
cl2im                     0.0206      0.070      0.295      0.769      -0.118       0.159
InterConflict6           -0.1736      0.113     -1.533      0.128      -0.398       0.051
cl0em                    -0.0261      0.012     -2.228      0.028      -0.049      -0.003
InterConflict6:cl0em      0.0480      0.029      1.629      0.106      -0.010       0.106
cl1em                    -0.0049      0.047     -0.105      0.917      -0.098       0.088
cl2em                     0.0338      0.056      0.606      0.546      -0.077       0.144
InterConflict6:cl1em      0.0562      0.036      1.558      0.122      -0.015       0.128
InterConflict6:cl2em      0.0522      0.047      1.101      0.273      -0.042       0.146
OuterConflict6:cl1em      0.2000      0.171      1.169      0.245      -0.139       0.539
OuterConflict6:cl2em     -0.2314      0.219     -1.055      0.294      -0.667       0.204
y1990s                    0.2235      0.078      2.878      0.005       0.069       0.378
==============================================================================


Percentage of the data from decennium:
 y1970s    0.00
y1980s    0.14
y1990s    0.75
y2000s    0.11

Mean wage growth over period (log):  0.227969498353
Cluster:  3                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.553
Model:                            OLS   Adj. R-squared:                  0.480
Method:                 Least Squares   F-statistic:                     7.515
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           5.72e-10
Time:                        18:40:58   Log-Likelihood:                 14.749
No. Observations:                 100   AIC:                            0.5014
Df Residuals:                      85   BIC:                             39.58
Df Model:                          14
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
Intercept                 0.0227      0.091      0.249      0.804      -0.159       0.204
InterConflict             0.2013      0.109      1.855      0.067      -0.015       0.417
OuterConflict            -0.0055      0.022     -0.249      0.804      -0.050       0.039
cl0im                    -0.0566      0.013     -4.230      0.000      -0.083      -0.030
cl1im                     0.0274      0.020      1.391      0.168      -0.012       0.067
cl2im                     0.0756      0.045      1.670      0.099      -0.014       0.166
InterConflict6           -0.4834      0.318     -1.519      0.132      -1.116       0.149
cl0em                    -0.0690      0.116     -0.595      0.554      -0.300       0.162
InterConflict6:cl0em     -0.4976      0.333     -1.495      0.139      -1.159       0.164
cl1em                     0.2734      0.119      2.296      0.024       0.037       0.510
cl2em                    -0.3813      0.152     -2.510      0.014      -0.683      -0.079
InterConflict6:cl1em      0.5600      0.342      1.638      0.105      -0.120       1.240
InterConflict6:cl2em     -0.2479      0.398     -0.623      0.535      -1.040       0.544
OuterConflict6:cl1em      0.2085      0.067      3.098      0.003       0.075       0.342
OuterConflict6:cl2em     -0.5471      0.184     -2.971      0.004      -0.913      -0.181
y1990s                   -0.1029      0.076     -1.350      0.181      -0.254       0.049
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.367347
y1990s    0.591837
y2000s    0.040816

Mean wage growth over period (log):  -0.14323340851
Cluster:  5                            OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.700
Model:                           OLS   Adj. R-squared:                  0.563
Method:                Least Squares   F-statistic:                     5.124
Date:               Mon, 15 Aug 2016   Prob (F-statistic):           4.50e-05
Time:                       18:40:59   Log-Likelihood:                 16.209
No. Observations:                 49   AIC:                           -0.4187
Df Residuals:                     33   BIC:                             29.85
Df Model:                         15
Covariance Type:           nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
Intercept               -0.1059      0.204     -0.519      0.607      -0.521       0.309
InterConflict           -0.1479      0.060     -2.485      0.018      -0.269      -0.027
OuterConflict            0.0316      0.066      0.480      0.634      -0.102       0.165
cl0im                   -0.0184      0.103     -0.179      0.859      -0.228       0.191
cl1im                    0.0385      0.167      0.231      0.819      -0.301       0.378
cl2im                    0.0621      0.103      0.601      0.552      -0.148       0.272
InterConflict6           0.7908      0.247      3.204      0.003       0.289       1.293
cl0em                   -0.0624      0.058     -1.076      0.290      -0.180       0.056
InterConflict6:cl0em    -0.1112      0.141     -0.790      0.435      -0.397       0.175
cl1em                    0.7707      0.298      2.584      0.014       0.164       1.378
cl2em                   -0.9000      0.361     -2.491      0.018      -1.635      -0.165
InterConflict6:cl1em     1.4450      0.756      1.912      0.065      -0.092       2.982
InterConflict6:cl2em    -1.6176      0.810     -1.998      0.054      -3.265       0.029
OuterConflict6:cl1em     1.1020      0.438      2.516      0.017       0.211       1.993
OuterConflict6:cl2em    -1.2383      0.481     -2.576      0.015      -2.216      -0.260
y1990s                   0.0201      0.096      0.209      0.836      -0.176       0.216
==============================================================================
```

Attachment 3.8

Summary statistics remittances model, unskilled wage growth [clusters 5 and 6 only]

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

RemitOut = remittances paid, RemitIn = remittances received, added 6 = six years before wage growth, at the same time as migration.

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.169811
y1990s    0.396226
y2000s    0.433962

Mean wage growth over period (log):  0.196758035703
Cluster:  5                               OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.442
Model:                           OLS   Adj. R-squared:                  0.237
Method:                Least Squares   F-statistic:                     2.152
Date:               Sun, 14 Aug 2016   Prob (F-statistic):             0.0307
Time:                       23:08:35   Log-Likelihood:                 5.6796
No. Observations:                 53   AIC:                             18.64
Df Residuals:                     38   BIC:                             48.20
Df Model:                         14
Covariance Type:           nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         -0.0917      0.423     -0.217      0.830      -0.949       0.765
RemitOut          -0.7581      0.486     -1.561      0.127      -1.741       0.225
cl0im              0.0320      0.059      0.541      0.592      -0.088       0.152
RemitOut:cl0im    -0.0359      0.127     -0.284      0.778      -0.292       0.220
cl1im              0.1326      0.223      0.595      0.555      -0.318       0.584
RemitOut:cl1im     0.2810      0.408      0.688      0.495      -0.545       1.107
cl2im             -0.0651      0.121     -0.540      0.593      -0.309       0.179
cl0em              0.0953      0.172      0.553      0.584      -0.254       0.444
RemitIn            0.6337      0.188      3.369      0.002       0.253       1.015
cl1em             -0.0962      0.057     -1.693      0.099      -0.211       0.019
RemitIn:cl1em     -0.0525      0.040     -1.311      0.198      -0.134       0.029
cl2em             -0.0833      0.090     -0.926      0.360      -0.265       0.099
RemitIn:cl2em     -0.0294      0.031     -0.956      0.345      -0.092       0.033
y1970s         -4.333e-16   2.48e-16     -1.746      0.089   -9.36e-16    6.91e-17
y1990s             0.3384      0.174      1.944      0.059      -0.014       0.691
y2000s             0.3061      0.157      1.953      0.058      -0.011       0.623
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.128477
y1980s    0.242384
y1990s    0.296689
y2000s    0.332450

Mean wage growth over period (log):  0.0661961835619
Cluster:  6                            OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.142
Model:                           OLS   Adj. R-squared:                  0.125
Method:                Least Squares   F-statistic:                     8.152
Date:               Sun, 14 Aug 2016   Prob (F-statistic):           2.58e-17
Time:                       23:08:35   Log-Likelihood:                 441.47
No. Observations:                755   AIC:                            -850.9
Df Residuals:                    739   BIC:                            -776.9
Df Model:                         15
Covariance Type:           nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         -0.0009      0.014     -0.065      0.948      -0.029       0.027
RemitOut          -0.0199      0.010     -2.001      0.046      -0.039      -0.000
cl0im              0.0056      0.003      1.794      0.073      -0.001       0.012
RemitOut:cl0im    -0.0056      0.003     -2.216      0.027      -0.011      -0.001
cl1im             -0.0028      0.004     -0.665      0.506      -0.011       0.005
RemitOut:cl1im     0.0087      0.004      2.298      0.022       0.001       0.016
cl2im             -0.0130      0.004     -2.995      0.003      -0.022      -0.004
cl0em              0.0025      0.004      0.554      0.580      -0.006       0.011
RemitIn           -0.0559      0.027     -2.082      0.038      -0.109      -0.003
cl1em             -0.0089      0.007     -1.269      0.205      -0.023       0.005
RemitIn:cl1em      0.0288      0.019      1.494      0.136      -0.009       0.067
cl2em              0.0240      0.013      1.858      0.064      -0.001       0.049
RemitIn:cl2em     -0.0542      0.032     -1.671      0.095      -0.118       0.009
y1970s             0.0295      0.017      1.696      0.090      -0.005       0.064
y1990s            -0.0260      0.017     -1.572      0.116      -0.058       0.006
y2000s            -0.0768      0.016     -4.777      0.000      -0.108      -0.045
==============================================================================
```

Attachment 3.9

Summary statistics remittances model, skilled wage growth [clusters 5 and 6 only]

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

RemitOut = remittances paid, RemitIn = remittances received, added 6 = six years before wage growth, at the same time as migration.

```
Percentage of the data from decennium:
 y1970s    0.0500
y1980s    0.1625
y1990s    0.3500
y2000s    0.4375

Mean wage growth over period (log):  0.103234185523
Cluster:  5                            OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.461
Model:                           OLS   Adj. R-squared:                  0.334
Method:                Least Squares   F-statistic:                     3.646
Date:               Sun, 14 Aug 2016   Prob (F-statistic):           0.000143
Time:                       22:58:53   Log-Likelihood:                 36.418
No. Observations:                 80   AIC:                            -40.84
Df Residuals:                     64   BIC:                            -2.724
Df Model:                         15
Covariance Type:           nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         -0.4116      0.186     -2.207      0.031      -0.784      -0.039
RemitOut          -0.1800      0.192     -0.939      0.351      -0.563       0.203
cl0im             -0.0066      0.024     -0.272      0.786      -0.055       0.042
RemitOut:cl0im    -0.0327      0.037     -0.882      0.381      -0.107       0.041
cl1im             -0.0104      0.046     -0.226      0.822      -0.102       0.081
RemitOut:cl1im     0.2176      0.111      1.955      0.055      -0.005       0.440
cl2im              0.0431      0.030      1.450      0.152      -0.016       0.103
cl0em             -0.0543      0.067     -0.813      0.419      -0.188       0.079
RemitIn           -0.0425      0.068     -0.627      0.533      -0.178       0.093
cl1em             -0.0205      0.018     -1.129      0.263      -0.057       0.016
RemitIn:cl1em      0.0027      0.008      0.351      0.727      -0.013       0.018
cl2em              0.1048      0.023      4.515      0.000       0.058       0.151
RemitIn:cl2em     -0.0209      0.009     -2.229      0.029      -0.040      -0.002
y1970s             0.3421      0.200      1.711      0.092      -0.057       0.741
y1990s             0.1127      0.100      1.129      0.263      -0.087       0.312
y2000s             0.2891      0.097      2.981      0.004       0.095       0.483
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.120924
y1980s    0.233696
y1990s    0.304348
y2000s    0.341033

Mean wage growth over period (log):  0.0545635467983
Cluster:  6                           OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.159
Model:                            OLS   Adj. R-squared:                  0.142
Method:                 Least Squares   F-statistic:                     9.083
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           1.35e-19
Time:                        22:58:53   Log-Likelihood:                 577.63
No. Observations:                 736   AIC:                            -1123.
Df Residuals:                     720   BIC:                            -1050.
Df Model:                          15
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept          0.0138      0.012      1.124      0.262      -0.010       0.038
RemitOut           0.0100      0.008      1.210      0.227      -0.006       0.026
cl0im              0.0043      0.003      1.636      0.102      -0.001       0.010
RemitOut:cl0im     0.0035      0.002      1.713      0.087      -0.001       0.008
cl1im             -0.0002      0.003     -0.067      0.947      -0.007       0.007
RemitOut:cl1im    -0.0020      0.003     -0.647      0.518      -0.008       0.004
cl2im             -0.0134      0.004     -3.793      0.000      -0.020      -0.006
cl0em              0.0081      0.003      2.327      0.020       0.001       0.015
RemitIn           -0.0262      0.023     -1.130      0.259      -0.072       0.019
cl1em             -0.0241      0.006     -4.204      0.000      -0.035      -0.013
RemitIn:cl1em     -0.0183      0.016     -1.173      0.241      -0.049       0.012
cl2em              0.0240      0.011      2.226      0.026       0.003       0.045
RemitIn:cl2em     -0.0099      0.026     -0.375      0.708      -0.062       0.042
y1970s             0.0054      0.015      0.367      0.714      -0.024       0.034
y1990s            -0.0567      0.014     -4.133      0.000      -0.084      -0.030
y2000s            -0.0903      0.014     -6.650      0.000      -0.117      -0.064
==============================================================================
```

Attachment 3.10

Summary statistics remittances model, high-skilled wage growth [clusters 0, 5 and 6 only]

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

RemitOut = remittances paid, RemitIn = remittances received, added 6 = six years before wage growth, at the same time as migration.

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.584270
y1990s    0.348315
y2000s    0.067416

Mean wage growth over period (log):  0.157626883799
Cluster:  0                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.522
Model:                            OLS   Adj. R-squared:                  0.432
Method:                 Least Squares   F-statistic:                     5.775
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           1.96e-07
Time:                        21:47:49   Log-Likelihood:                -62.458
No. Observations:                  89   AIC:                             154.9
Df Residuals:                      74   BIC:                             192.2
Df Model:                          14
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept          0.9937      0.707      1.406      0.164      -0.415       2.402
RemitOut           0.0502      0.376      0.133      0.894      -0.700       0.800
cl0im             -0.0134      0.057     -0.235      0.815      -0.127       0.100
RemitOut:cl0im    -0.0547      0.056     -0.983      0.329      -0.166       0.056
cl1im              0.1709      0.162      1.057      0.294      -0.151       0.493
RemitOut:cl1im    -0.0606      0.096     -0.631      0.530      -0.252       0.131
cl2im             -0.1320      0.403     -0.328      0.744      -0.934       0.670
cl0em             -0.0181      0.020     -0.900      0.371      -0.058       0.022
RemitIn            0.6486      0.546      1.187      0.239      -0.440       1.737
cl1em             -0.0279      0.088     -0.316      0.753      -0.204       0.148
RemitIn:cl1em      0.1147      0.036      3.195      0.002       0.043       0.186
cl2em              0.4536      0.177      2.566      0.012       0.101       0.806
RemitIn:cl2em      0.2652      0.362      0.732      0.467      -0.457       0.987
y1990s            -0.0116      0.171     -0.068      0.946      -0.353       0.330
y2000s             0.1451      0.290      0.501      0.618      -0.432       0.723
==============================================================================
Omnibus:                        5.336   Durbin-Watson:                   2.030
Prob(Omnibus):                  0.069   Jarque-Bera (JB):                6.481
Skew:                           0.236   Prob(JB):                       0.0391
Kurtosis:                       4.235   Cond. No.                         119.
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.216216
y1990s    0.405405
y2000s    0.378378
Mean wage growth over period (log):  -0.0630470160046
Cluster:  5                              OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.711
Model:                            OLS   Adj. R-squared:                  0.527
Method:                 Least Squares   F-statistic:                     3.863
Date:                Sun, 14 Aug 2016   Prob (F-statistic):            0.00234
Time:                        21:47:49   Log-Likelihood:                 17.116
No. Observations:                  37   AIC:                            -4.231
Df Residuals:                      22   BIC:                             19.93
Df Model:                          14
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         -0.3225      0.525     -0.614      0.546      -1.412       0.767
RemitOut           1.2563      0.443      2.834      0.010       0.337       2.175
cl0im             -0.1765      0.060     -2.935      0.008      -0.301      -0.052
RemitOut:cl0im     0.0600      0.083      0.725      0.476      -0.112       0.232
cl1im              0.0766      0.210      0.366      0.718      -0.358       0.511
RemitOut:cl1im    -0.3955      0.287     -1.379      0.182      -0.990       0.199
cl2im              0.0344      0.102      0.338      0.738      -0.177       0.246
cl0em              0.0683      0.217      0.314      0.756      -0.383       0.519
RemitIn           -0.6509      0.230     -2.831      0.010      -1.128      -0.174
cl1em             -0.0421      0.056     -0.748      0.462      -0.159       0.075
RemitIn:cl1em      0.0587      0.041      1.431      0.167      -0.026       0.144
cl2em              0.0025      0.107      0.024      0.981      -0.219       0.224
RemitIn:cl2em      0.0481      0.023      2.072      0.050   -3.43e-05       0.096
y1990s             0.5845      0.178      3.275      0.003       0.214       0.955
y2000s             0.6807      0.162      4.191      0.000       0.344       1.018
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.248663
y1990s    0.344920
y2000s    0.406417
Mean wage growth over period (log):  0.0456676011459
Cluster:  6                              OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.264
Model:                            OLS   Adj. R-squared:                  0.235
Method:                 Least Squares   F-statistic:                     9.185
Date:                Sun, 14 Aug 2016   Prob (F-statistic):           2.62e-17
Time:                        21:47:49   Log-Likelihood:                 229.34
No. Observations:                 374   AIC:                            -428.7
Df Residuals:                     359   BIC:                            -369.8
Df Model:                          14
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         -0.0251      0.025     -1.017      0.310      -0.074       0.023
RemitOut           0.0369      0.020      1.813      0.071      -0.003       0.077
cl0im              0.0008      0.005      0.166      0.868      -0.009       0.011
RemitOut:cl0im     0.0132      0.006      2.369      0.018       0.002       0.024
cl1im              0.0005      0.007      0.083      0.934      -0.012       0.014
RemitOut:cl1im    -0.0069      0.007     -0.943      0.346      -0.021       0.007
cl2im              0.0099      0.005      1.968      0.050    5.54e-06       0.020
cl0em              0.0282      0.013      2.151      0.032       0.002       0.054
RemitIn           -0.0912      0.034     -2.651      0.008      -0.159      -0.024
cl1em             -0.0855      0.012     -6.913      0.000      -0.110      -0.061
RemitIn:cl1em     -0.1014      0.027     -3.746      0.000      -0.155      -0.048
cl2em              0.1208      0.020      5.967      0.000       0.081       0.161
RemitIn:cl2em      0.1207      0.047      2.570      0.011       0.028       0.213
y1990s            -0.0446      0.021     -2.117      0.035      -0.086      -0.003
y2000s            -0.0248      0.023     -1.066      0.287      -0.070       0.021
==============================================================================
```

Attachment 3.11

Summary statistics urbanization and gini model, unskilled wage growth [clusters 0, 4 and 6 only]

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

Urban = urbanization, Gini = gini-coefficient for economic inequality

```
Percentage of the data from decennium:
 y1970s    0.024194
y1980s    0.395161
y1990s    0.467742
y2000s    0.112903
Mean wage growth over period (log):  0.0679317467785
Cluster:  0                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.223
Model:                            OLS   Adj. R-squared:                  0.154
Method:                 Least Squares   F-statistic:                     3.235
Date:                Mon, 15 Aug 2016   Prob (F-statistic):            0.00106
Time:                        01:10:16   Log-Likelihood:                -43.830
No. Observations:                 124   AIC:                             109.7
Df Residuals:                     113   BIC:                             140.7
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.2998      0.177     -1.692      0.093      -0.651       0.051
Urban          0.0466      0.023      1.985      0.050    8.06e-05       0.093
Gini          -0.0865      0.040     -2.144      0.034      -0.166      -0.007
cl0im          0.0318      0.019      1.668      0.098      -0.006       0.070
cl1im         -0.0552      0.024     -2.336      0.021      -0.102      -0.008
cl2im         -0.0086      0.123     -0.070      0.944      -0.252       0.235
cl0em          0.0140      0.013      1.068      0.288      -0.012       0.040
cl1em          0.0014      0.022      0.065      0.948      -0.042       0.045
y1970s        -0.0749      0.222     -0.337      0.737      -0.516       0.366
y1990s         0.0780      0.095      0.822      0.413      -0.110       0.266
y2000s         0.1981      0.122      1.629      0.106      -0.043       0.439
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.215385
y1980s    0.267692
y1990s    0.313846
y2000s    0.203077
Mean wage growth over period (log):  0.0255111851006
Cluster:  4                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.108
Model:                            OLS   Adj. R-squared:                  0.080
Method:                 Least Squares   F-statistic:                     3.808
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           7.21e-05
Time:                        01:10:16   Log-Likelihood:                -98.512
No. Observations:                 325   AIC:                             219.0
Df Residuals:                     314   BIC:                             260.6
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0461      0.070     -0.655      0.513      -0.185       0.092
Urban         -0.0805      0.046     -1.753      0.080      -0.171       0.010
Gini          -0.0931      0.030     -3.067      0.002      -0.153      -0.033
cl0im          0.1023      0.026      3.904      0.000       0.051       0.154
cl1im         -0.0482      0.023     -2.076      0.039      -0.094      -0.003
cl2im         -0.0288      0.025     -1.166      0.245      -0.077       0.020
cl0em         -0.0560      0.032     -1.737      0.083      -0.119       0.007
cl1em          0.0394      0.021      1.914      0.057      -0.001       0.080
y1970s         0.0648      0.056      1.157      0.248      -0.045       0.175
y1990s         0.0528      0.058      0.909      0.364      -0.062       0.167
y2000s         0.0683      0.064      1.066      0.287      -0.058       0.194
==============================================================================

Percentage of the data from decennium:
 y1970s    0.170601
y1980s    0.262876
y1990s    0.298283
y2000s    0.268240
Mean wage growth over period (log):  0.0841226483721
Cluster:  6                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.094
Model:                            OLS   Adj. R-squared:                  0.084
Method:                 Least Squares   F-statistic:                     9.555
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           2.96e-15
Time:                        01:10:16   Log-Likelihood:                -293.50
No. Observations:                 932   AIC:                             609.0
Df Residuals:                     921   BIC:                             662.2
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.1323      0.027      4.931      0.000       0.080       0.185
Urban          0.1332      0.029      4.550      0.000       0.076       0.191
Gini           0.0909      0.017      5.409      0.000       0.058       0.124
cl0im          0.0036      0.006      0.635      0.525      -0.008       0.015
cl1im         -0.0003      0.008     -0.034      0.973      -0.017       0.016
cl2im         -0.0179      0.008     -2.134      0.033      -0.034      -0.001
cl0em         -0.0027      0.009     -0.320      0.749      -0.020       0.014
cl1em          0.0101      0.010      1.023      0.306      -0.009       0.030
y1970s        -0.0305      0.036     -0.840      0.401      -0.102       0.041
y1990s        -0.1084      0.035     -3.103      0.002      -0.177      -0.040
y2000s        -0.1542      0.032     -4.758      0.000      -0.218      -0.091
==============================================================================
```

Attachment 3.12

Summary statistics urbanization and gini model, skilled wage growth [clusters 0, 4 and 6 only]

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

Urban = urbanization, Gini = gini-coefficient for economic inequality

```
Percentage of the data from decennium:
 y1970s    0.110429
y1980s    0.331288
y1990s    0.466258
y2000s    0.092025
Mean wage growth over period (log):  0.0885515866355
Cluster:  0                           OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.221
Model:                            OLS   Adj. R-squared:                  0.170
Method:                 Least Squares   F-statistic:                     4.321
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           2.53e-05
Time:                        01:06:27   Log-Likelihood:                -56.027
No. Observations:                 163   AIC:                             134.1
Df Residuals:                     152   BIC:                             168.1
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0272      0.138     -0.197      0.844      -0.300       0.245
Urban          0.0410      0.015      2.727      0.007       0.011       0.071
Gini          -0.1427      0.034     -4.145      0.000      -0.211      -0.075
cl0im          0.0010      0.015      0.067      0.946      -0.029       0.031
cl1im          0.0303      0.023      1.306      0.194      -0.016       0.076
cl0em          0.0286      0.011      2.536      0.012       0.006       0.051
cl1em         -0.0366      0.019     -1.967      0.051      -0.073       0.000
cl2em          0.0723      0.071      1.019      0.310      -0.068       0.212
y1970s        -0.2379      0.109     -2.192      0.030      -0.452      -0.023
y1990s        -0.0596      0.084     -0.708      0.480      -0.226       0.107
y2000s         0.2110      0.111      1.898      0.060      -0.009       0.431
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.110320
y1980s    0.234875
y1990s    0.355872
y2000s    0.298932
Mean wage growth over period (log):  0.0848559091156
Cluster:  4                             OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.179
Model:                            OLS   Adj. R-squared:                  0.148
Method:                 Least Squares   F-statistic:                     5.877
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           4.97e-08
Time:                        01:06:28   Log-Likelihood:                -73.413
No. Observations:                 281   AIC:                             168.8
Df Residuals:                     270   BIC:                             208.8
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.1902      0.069      2.759      0.006       0.054       0.326
Urban          0.0286      0.040      0.720      0.472      -0.050       0.107
Gini          -0.1147      0.037     -3.120      0.002      -0.187      -0.042
cl0im          0.0259      0.029      0.883      0.378      -0.032       0.084
cl1im         -0.0054      0.032     -0.168      0.867      -0.069       0.058
cl0em          0.1541      0.039      3.962      0.000       0.078       0.231
cl1em          0.0010      0.023      0.041      0.967      -0.045       0.047
cl2em         -0.0755      0.017     -4.459      0.000      -0.109      -0.042
y1970s        -0.0458      0.078     -0.592      0.555      -0.198       0.107
y1990s         0.0002      0.058      0.003      0.998      -0.115       0.115
y2000s         0.1299      0.062      2.097      0.037       0.008       0.252
==============================================================================

Percentage of the data from decennium:
 y1970s    0.163494
y1980s    0.251960
y1990s    0.304591
y2000s    0.279955
Mean wage growth over period (log):  0.0801187144312
Cluster:  6                             OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.088
Model:                            OLS   Adj. R-squared:                  0.077
Method:                 Least Squares   F-statistic:                     8.480
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           2.75e-13
Time:                        01:06:28   Log-Likelihood:                -289.52
No. Observations:                 893   AIC:                             601.0
Df Residuals:                     882   BIC:                             653.8
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.1153      0.028      4.120      0.000       0.060       0.170
Urban          0.0905      0.029      3.129      0.002       0.034       0.147
Gini           0.1045      0.017      6.080      0.000       0.071       0.138
cl0im          0.0004      0.006      0.061      0.952      -0.011       0.012
cl1im         -0.0040      0.008     -0.476      0.634      -0.020       0.012
cl0em          0.0097      0.009      1.069      0.285      -0.008       0.027
cl1em          0.0104      0.011      0.935      0.350      -0.011       0.032
cl2em         -0.0201      0.020     -0.993      0.321      -0.060       0.020
y1970s        -0.0117      0.038     -0.311      0.756      -0.086       0.062
y1990s        -0.1350      0.036     -3.734      0.000      -0.206      -0.064
y2000s        -0.1453      0.034     -4.261      0.000      -0.212      -0.078
==============================================================================
```

Attachment 3.13

Summary statistics urbanization and gini model, high-skilled wage growth [clusters 4 and 6 only]

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

Urban = urbanization, Gini = gini-coefficient for economic inequality

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.244094
y1990s    0.330709
y2000s    0.425197

Mean wage growth over period (log):  0.0696293776064
Cluster:  4                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.137
Model:                            OLS   Adj. R-squared:                  0.062
Method:                 Least Squares   F-statistic:                     1.836
Date:                Mon, 15 Aug 2016   Prob (F-statistic):             0.0616
Time:                        01:01:35   Log-Likelihood:                -68.914
No. Observations:                 127   AIC:                             159.8
Df Residuals:                     116   BIC:                             191.1
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.2038      0.155     -1.315      0.191      -0.511       0.103
Urban          0.1589      0.082      1.932      0.056      -0.004       0.322
Gini          -0.1680      0.085     -1.977      0.050      -0.336       0.000
cl0im          0.0097      0.056      0.174      0.862      -0.101       0.120
cl1im          0.0012      0.057      0.020      0.984      -0.113       0.115
cl2im         -0.1180      0.050     -2.381      0.019      -0.216      -0.020
cl0em          0.0776      0.082      0.943      0.348      -0.085       0.241
cl1em          0.0012      0.051      0.023      0.982      -0.100       0.102
cl2em         -0.0350      0.047     -0.737      0.463      -0.129       0.059
y1990s         0.3510      0.130      2.693      0.008       0.093       0.609
y2000s         0.1329      0.151      0.878      0.382      -0.167       0.433
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.276256
y1990s    0.363014
y2000s    0.360731

Mean wage growth over period (log):  0.0425407655081
Cluster:  6                            OLS Regression Results
==============================================================================
Dep. Variable:         Growthratelog   R-squared:                       0.138
Model:                           OLS   Adj. R-squared:                  0.117
Method:                Least Squares   F-statistic:                     6.815
Date:               Mon, 15 Aug 2016   Prob (F-statistic):           6.84e-10
Time:                       01:01:35   Log-Likelihood:                 220.15
No. Observations:                438   AIC:                            -418.3
Df Residuals:                    427   BIC:                            -373.4
Df Model:                         10
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0498      0.021     -2.366      0.018      -0.091      -0.008
Urban          0.0693      0.025      2.804      0.005       0.021       0.118
Gini          -0.0350      0.011     -3.056      0.002      -0.057      -0.012
cl0im          0.0059      0.004      1.383      0.167      -0.002       0.014
cl1im          0.0058      0.007      0.822      0.411      -0.008       0.020
cl2im          0.0020      0.005      0.418      0.676      -0.008       0.012
cl0em         -0.0174      0.011     -1.614      0.107      -0.038       0.004
cl1em         -0.0394      0.010     -3.957      0.000      -0.059      -0.020
cl2em          0.0521      0.014      3.619      0.000       0.024       0.080
y1990s        -0.0199      0.020     -0.997      0.319      -0.059       0.019
y2000s        -0.0454      0.021     -2.168      0.031      -0.087      -0.004
==============================================================================
```

Attachment 3.14

Summary statistics schooling and remittances model, unskilled wage growth [clusters 0-4 only]

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

RemitOut = remittances paid, RemitIn = remittances received, School = schooling indicator, added 6 = six years before wage growth, at the same time as migration.

```
Percentage of the data from decennium:
 y1970s    0.029412
y1980s    0.529412
y1990s    0.338235
y2000s    0.102941

Mean wage growth over period (log):  0.135896566561
Cluster:  0                           OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.537
Model:                            OLS   Adj. R-squared:                  0.457
Method:                 Least Squares   F-statistic:                     6.677
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           9.72e-12
Time:                        19:21:05   Log-Likelihood:                -38.779
No. Observations:                 136   AIC:                             119.6
Df Residuals:                     115   BIC:                             180.7
Df Model:                          20
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept          0.2988      0.333      0.896      0.372      -0.361       0.959
School6:cl0im     -0.0245      0.018     -1.378      0.171      -0.060       0.011
cl0im             -0.0136      0.032     -0.426      0.671      -0.077       0.050
cl1im             -0.0397      0.050     -0.789      0.432      -0.139       0.060
cl2im              0.1417      0.182      0.778      0.438      -0.219       0.502
RemitOut:cl0im     0.0458      0.034      1.367      0.174      -0.021       0.112
RemitOut:cl1im    -0.0393      0.080     -0.490      0.625      -0.198       0.120
RemitOut:cl2im    -0.1120      0.186     -0.601      0.549      -0.481       0.257
RemitIn           -0.0682      0.499     -0.137      0.892      -1.057       0.920
cl0em             -0.0008      0.025     -0.031      0.975      -0.050       0.049
cl1em              0.0958      0.036      2.686      0.008       0.025       0.166
cl2em             -0.0275      0.132     -0.208      0.835      -0.289       0.234
RemitIn:cl0em     -0.0662      0.061     -1.081      0.282      -0.187       0.055
RemitIn:cl1em      0.0766      0.033      2.310      0.023       0.011       0.142
RemitIn:cl2em     -0.4027      0.237     -1.699      0.092      -0.872       0.067
School6            0.0200      0.134      0.149      0.882      -0.245       0.285
School6:cl1em     -0.0078      0.024     -0.325      0.746      -0.055       0.040
School6:cl2em      0.0224      0.074      0.304      0.762      -0.123       0.168
y1970s            -0.0890      0.204     -0.437      0.663      -0.492       0.314
y1990s             0.0925      0.100      0.921      0.359      -0.106       0.291
y2000s             0.0867      0.135      0.644      0.521      -0.180       0.354
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.080247    y1980s    0.320988
y1990s    0.246914     y2000s    0.351852
Mean wage growth over period (log):  0.0815099501094
Cluster:  1                              OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.390
Model:                            OLS   Adj. R-squared:                  0.304
Method:                 Least Squares   F-statistic:                     4.509
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           4.12e-08
Time:                        19:21:05   Log-Likelihood:                -30.616
No. Observations:                 162   AIC:                             103.2
Df Residuals:                     141   BIC:                             168.1
Df Model:                          20
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept          1.5025      0.734      2.046      0.043       0.051       2.954
School6:cl0im     -0.0420      0.038     -1.113      0.268      -0.117       0.033
cl0im              0.0692      0.042      1.660      0.099      -0.013       0.152
cl1im              0.5133      0.400      1.282      0.202      -0.278       1.305
cl2im              0.0151      0.267      0.057      0.955      -0.513       0.544
RemitOut:cl0im     0.2110      0.091      2.325      0.021       0.032       0.390
RemitOut:cl1im    -0.7710      0.368     -2.094      0.038      -1.499      -0.043
RemitOut:cl2im     1.1321      0.603      1.879      0.062      -0.059       2.323
RemitIn            0.4425      0.163      2.717      0.007       0.121       0.764
cl0em             -0.0480      0.033     -1.461      0.146      -0.113       0.017
cl1em              0.1619      0.067      2.408      0.017       0.029       0.295
cl2em             -0.1941      0.069     -2.820      0.005      -0.330      -0.058
RemitIn:cl0em     -0.2047      0.042     -4.910      0.000      -0.287      -0.122
RemitIn:cl1em      0.2966      0.095      3.123      0.002       0.109       0.484
RemitIn:cl2em      0.0532      0.077      0.690      0.491      -0.099       0.206
School6           -0.0659      0.096     -0.689      0.492      -0.255       0.123
School6:cl1em     -0.0827      0.042     -1.973      0.050      -0.166       0.000
School6:cl2em      0.1401      0.045      3.083      0.002       0.050       0.230
y1970s            -0.0303      0.104     -0.290      0.772      -0.237       0.176
y1990s            -0.0050      0.089     -0.056      0.955      -0.182       0.172
y2000s            -0.1209      0.089     -1.352      0.179      -0.298       0.056
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.050562    y1980s    0.292135
y1990s    0.426966     y2000s    0.230337
Mean wage growth over period (log):  0.107625123911
Cluster:  2                              OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.347
Model:                            OLS   Adj. R-squared:                  0.264
Method:                 Least Squares   F-statistic:                     4.174
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           1.45e-07
Time:                        19:21:05   Log-Likelihood:                -38.344
No. Observations:                 178   AIC:                             118.7
Df Residuals:                     157   BIC:                             185.5
Df Model:                          20
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         -0.3272      0.139     -2.361      0.019      -0.601      -0.053
School6:cl0im     -0.0191      0.020     -0.938      0.350      -0.059       0.021
cl0im             -0.1177      0.078     -1.508      0.133      -0.272       0.036
cl1im             -0.0585      0.080     -0.727      0.468      -0.217       0.100
cl2im              0.2544      0.115      2.215      0.028       0.028       0.481
RemitOut:cl0im     0.0608      0.036      1.708      0.090      -0.010       0.131
RemitOut:cl1im     0.1215      0.101      1.205      0.230      -0.078       0.321
RemitOut:cl2im    -0.3585      0.194     -1.846      0.067      -0.742       0.025
RemitIn           -0.0380      0.046     -0.831      0.407      -0.128       0.052
cl0em              0.0834      0.072      1.154      0.250      -0.059       0.226
cl1em              0.0346      0.040      0.875      0.383      -0.044       0.113
cl2em             -0.0423      0.030     -1.398      0.164      -0.102       0.017
RemitIn:cl0em      0.0261      0.034      0.756      0.451      -0.042       0.094
RemitIn:cl1em     -0.0541      0.035     -1.538      0.126      -0.124       0.015
RemitIn:cl2em      0.0513      0.017      3.086      0.002       0.018       0.084
School6           -0.0656      0.078     -0.844      0.400      -0.219       0.088
School6:cl1em     -0.0380      0.037     -1.039      0.301      -0.110       0.034
School6:cl2em     -0.0510      0.039     -1.305      0.194      -0.128       0.026
y1970s             0.0968      0.130      0.747      0.456      -0.159       0.353
y1990s             0.4164      0.068      6.140      0.000       0.282       0.550
y2000s             0.3550      0.098      3.613      0.000       0.161       0.549
==============================================================================
```

Percentage of the data from decennium:

y1970s	0.035370	y1980s	0.070740
y1990s	0.327974	y2000s	0.565916

Mean wage growth over period (log): 0.206750546089

Cluster: 3 OLS Regression Results

Dep. Variable:	Growthratelog	R-squared:	0.194
Model:	OLS	Adj. R-squared:	0.138
Method:	Least Squares	F-statistic:	3.481
Date:	Mon, 15 Aug 2016	Prob (F-statistic):	1.51e-06
Time:	19:21:05	Log-Likelihood:	2.3350
No. Observations:	311	AIC:	37.33
Df Residuals:	290	BIC:	115.9
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0825	0.068	-1.217	0.225	-0.216	0.051
School6:cl0im	-0.0370	0.011	-3.229	0.001	-0.060	-0.014
cl0im	-0.0365	0.010	-3.582	0.000	-0.057	-0.016
cl1im	0.0380	0.020	1.908	0.057	-0.001	0.077
cl2im	-0.0760	0.032	-2.381	0.018	-0.139	-0.013
RemitOut:cl0im	-0.0505	0.032	-1.565	0.119	-0.114	0.013
RemitOut:cl1im	0.0417	0.048	0.864	0.388	-0.053	0.137
RemitOut:cl2im	-0.0025	0.059	-0.042	0.966	-0.119	0.114
RemitIn	-0.0030	0.050	-0.060	0.952	-0.101	0.095
cl0em	-0.0016	0.015	-0.106	0.915	-0.031	0.028
cl1em	-0.0373	0.017	-2.236	0.026	-0.070	-0.004
cl2em	0.0847	0.038	2.243	0.026	0.010	0.159
RemitIn:cl0em	0.0380	0.023	1.655	0.099	-0.007	0.083
RemitIn:cl1em	-0.0133	0.016	-0.822	0.412	-0.045	0.019
RemitIn:cl2em	0.0168	0.051	0.330	0.741	-0.083	0.117
School6	0.2178	0.046	4.744	0.000	0.127	0.308
School6:cl1em	-0.0448	0.023	-1.940	0.053	-0.090	0.001
School6:cl2em	0.0940	0.050	1.895	0.059	-0.004	0.192
y1970s	-0.1278	0.098	-1.299	0.195	-0.321	0.066
y1990s	0.2947	0.092	3.218	0.001	0.114	0.475
y2000s	0.3348	0.082	4.101	0.000	0.174	0.495

Percentage of the data from decennium:

y1970s	0.208238	y1980s	0.306636
y1990s	0.283753	y2000s	0.201373

Mean wage growth over period (log): 0.0755948697656

Cluster: 4 OLS Regression Results

Dep. Variable:	Growthratelog	R-squared:	0.410
Model:	OLS	Adj. R-squared:	0.337
Method:	Least Squares	F-statistic:	5.625
Date:	Thu, 18 Aug 2016	Prob (F-statistic):	2.13e-10
Time:	01:19:45	Log-Likelihood:	-3.6373
No. Observations:	174	AIC:	47.27
Df Residuals:	154	BIC:	110.5
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0608	0.164	-0.371	0.711	-0.385	0.263
School6:cl0im	0.1203	0.062	1.938	0.054	-0.002	0.243
RemitOut	-0.9645	0.393	-2.453	0.015	-1.741	-0.188
cl0im	0.0213	0.059	0.360	0.719	-0.096	0.138
cl1im	-0.0639	0.068	-0.945	0.346	-0.198	0.070
RemitOut:cl0im	-0.3357	0.215	-1.564	0.120	-0.760	0.088
RemitOut:cl1im	-0.1819	0.217	-0.837	0.404	-0.611	0.247
cl2im	0.0034	0.068	0.050	0.960	-0.132	0.139
cl0em	-0.1257	0.071	-1.776	0.078	-0.266	0.014
School6	0.4953	0.156	3.173	0.002	0.187	0.804
RemitIn	0.3649	0.342	1.067	0.288	-0.311	1.041
cl1em	0.2432	0.089	2.733	0.007	0.067	0.419
cl2em	0.0030	0.062	0.049	0.961	-0.119	0.125
School6:cl1em	0.0383	0.062	0.619	0.537	-0.084	0.160
School6:cl2em	-0.1131	0.069	-1.645	0.102	-0.249	0.023
RemitIn:cl1em	0.2872	0.179	1.609	0.110	-0.066	0.640
RemitIn:cl2em	-0.1276	0.065	-1.969	0.051	-0.256	0.000
y1970s	0.2201	0.112	1.964	0.051	-0.001	0.442
y1990s	0.0866	0.063	1.375	0.171	-0.038	0.211
y2000s	0.4514	0.092	4.907	0.000	0.270	0.633

Attachment 3.15

Summary statistics schooling and remittances model, skilled wage growth [clusters 0-4 only]

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

RemitOut = remittances paid, RemitIn = remittances received, School = schooling indicator, added 6 = six years before wage growth, at the same time as migration.

Percentage of the data from decennium:

y1970s	0.093264
y1980s	0.497409
y1990s	0.305699
y2000s	0.103627

Mean wage growth over period (log): 0.0995881830104

Cluster: 0

OLS Regression Results

Dep. Variable:	Growthratelog	R-squared:	0.160
Model:	OLS	Adj. R-squared:	0.063
Method:	Least Squares	F-statistic:	1.643
Date:	Mon, 15 Aug 2016	Prob (F-statistic):	0.0477
Time:	19:07:56	Log-Likelihood:	-103.70
No. Observations:	193	AIC:	249.4
Df Residuals:	172	BIC:	317.9
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.4179	0.333	1.255	0.211	-0.239	1.075
School6	-0.0694	0.134	-0.519	0.605	-0.334	0.195
School6:cl0im	0.0041	0.017	0.239	0.811	-0.030	0.038
cl0im	-0.0217	0.026	-0.841	0.401	-0.073	0.029
cl1im	0.0740	0.037	1.976	0.050	8.54e-05	0.148
cl2im	0.2042	0.182	1.124	0.262	-0.154	0.563
RemitOut:cl0im	0.0435	0.037	1.191	0.235	-0.029	0.116
RemitOut:cl1im	-0.1728	0.077	-2.245	0.026	-0.325	-0.021
RemitOut:cl2im	0.2898	0.187	1.553	0.122	-0.079	0.658
RemitIn	0.0925	0.546	0.169	0.866	-0.986	1.171
cl0em	-0.0152	0.029	-0.526	0.600	-0.072	0.042
cl1em	-0.0866	0.036	-2.377	0.019	-0.158	-0.015
cl2em	-0.0543	0.136	-0.399	0.691	-0.323	0.215
RemitIn:cl0em	-0.1286	0.065	-1.966	0.051	-0.258	0.001
RemitIn:cl1em	0.0152	0.033	0.462	0.645	-0.050	0.080
RemitIn:cl2em	-0.4284	0.267	-1.606	0.110	-0.955	0.098
School6:cl1em	-0.0130	0.018	-0.738	0.462	-0.048	0.022
School6:cl2em	-0.0083	0.073	-0.113	0.910	-0.153	0.137
y1970s	-0.2078	0.139	-1.498	0.136	-0.482	0.066
y1990s	-0.0102	0.107	-0.096	0.924	-0.221	0.201
y2000s	0.1363	0.140	0.971	0.333	-0.141	0.413

Percentage of the data from decennium:
y1970s 0.058824 y1980s 0.256684
y1990s 0.229947 y2000s 0.454545
Mean wage growth over period (log): 0.140028180754
Cluster: 1 OLS Regression Results

Dep. Variable:	Growthratelog	R-squared:	0.286
Model:	OLS	Adj. R-squared:	0.200
Method:	Least Squares	F-statistic:	3.327
Date:	Mon, 15 Aug 2016	Prob (F-statistic):	1.07e-05
Time:	19:07:56	Log-Likelihood:	-81.238
No. Observations:	187	AIC:	204.5
Df Residuals:	166	BIC:	272.3
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.1611	0.434	0.371	0.711	-0.695	1.018
School6	0.0153	0.107	0.143	0.887	-0.196	0.227
School6:cl0im	0.0108	0.047	0.229	0.819	-0.082	0.104
cl0im	0.2020	0.048	4.170	0.000	0.106	0.298
cl1im	0.0483	0.186	0.260	0.795	-0.319	0.416
cl2im	-0.3457	0.200	-1.726	0.086	-0.741	0.050
RemitOut:cl0im	-0.0705	0.085	-0.826	0.410	-0.239	0.098
RemitOut:cl1im	0.2269	0.208	1.088	0.278	-0.185	0.638
RemitOut:cl2im	-0.3427	0.343	-1.000	0.319	-1.019	0.334
RemitIn	0.2368	0.066	3.607	0.000	0.107	0.366
cl0em	-0.0567	0.036	-1.588	0.114	-0.127	0.014
cl1em	-0.0087	0.043	-0.202	0.840	-0.094	0.076
cl2em	0.0693	0.071	0.974	0.332	-0.071	0.210
RemitIn:cl0em	-0.0012	0.031	-0.040	0.968	-0.062	0.060
RemitIn:cl1em	0.0459	0.041	1.110	0.269	-0.036	0.128
RemitIn:cl2em	-0.0449	0.080	-0.561	0.575	-0.203	0.113
School6:cl1em	-0.0228	0.033	-0.698	0.486	-0.087	0.042
School6:cl2em	-0.0702	0.048	-1.472	0.143	-0.164	0.024
y1970s	-0.0782	0.150	-0.522	0.602	-0.374	0.218
y1990s	-0.0454	0.106	-0.430	0.667	-0.254	0.163
y2000s	-0.0357	0.106	-0.337	0.736	-0.245	0.174

Percentage of the data from decennium:
y1970s 0.031674 y1980s 0.280543
y1990s 0.407240 y2000s 0.280543
Mean wage growth over period (log): 0.105876008047
Cluster: 2 OLS Regression Results

Dep. Variable:	Growthratelog	R-squared:	0.249
Model:	OLS	Adj. R-squared:	0.174
Method:	Least Squares	F-statistic:	3.320
Date:	Mon, 15 Aug 2016	Prob (F-statistic):	7.55e-06
Time:	19:07:56	Log-Likelihood:	-105.55
No. Observations:	221	AIC:	253.1
Df Residuals:	200	BIC:	324.5
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.3214	0.165	-1.950	0.053	-0.646	0.004
School6	0.0611	0.087	0.698	0.486	-0.111	0.234
School6:cl0im	0.0097	0.024	0.395	0.693	-0.039	0.058
cl0im	-0.1123	0.071	-1.573	0.117	-0.253	0.028
cl1im	-0.0468	0.091	-0.517	0.606	-0.225	0.132
cl2im	0.2347	0.130	1.806	0.072	-0.022	0.491
RemitOut:cl0im	0.0813	0.036	2.254	0.025	0.010	0.152
RemitOut:cl1im	0.3305	0.118	2.809	0.005	0.098	0.562
RemitOut:cl2im	-0.4448	0.231	-1.926	0.056	-0.900	0.011
RemitIn	-0.0443	0.071	-0.622	0.535	-0.185	0.096
cl0em	0.0817	0.061	1.350	0.179	-0.038	0.201
cl1em	-0.0825	0.044	-1.866	0.064	-0.170	0.005
cl2em	0.0518	0.037	1.396	0.164	-0.021	0.125
RemitIn:cl0em	0.0045	0.053	0.084	0.933	-0.101	0.110
RemitIn:cl1em	-0.0395	0.063	-0.627	0.531	-0.164	0.085
RemitIn:cl2em	-0.0010	0.025	-0.042	0.966	-0.050	0.048
School6:cl1em	0.1233	0.047	2.611	0.010	0.030	0.216
School6:cl2em	-0.0759	0.053	-1.419	0.157	-0.181	0.030
y1970s	0.0693	0.187	0.371	0.711	-0.299	0.437
y1990s	0.2069	0.092	2.258	0.025	0.026	0.388
y2000s	0.2724	0.116	2.339	0.020	0.043	0.502

```
Percentage of the data from decennium:
 y1970s    0.019169     y1980s    0.057508
y1990s    0.335463      y2000s    0.587859
Mean wage growth over period (log):  0.235636254737
Cluster:  3                             OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.203
Model:                            OLS   Adj. R-squared:                  0.148
Method:                 Least Squares   F-statistic:                     3.715
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           3.59e-07
Time:                        19:07:56   Log-Likelihood:                 9.3196
No. Observations:                 313   AIC:                             23.36
Df Residuals:                     292   BIC:                             102.0
Df Model:                          20
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept          0.0807      0.072      1.125      0.261      -0.060       0.222
School6            0.1933      0.048      4.038      0.000       0.099       0.288
School6:cl0im     -0.0121      0.013     -0.906      0.366      -0.039       0.014
cl0im             -0.0352      0.011     -3.341      0.001      -0.056      -0.014
cl1im              0.0014      0.019      0.076      0.939      -0.035       0.038
cl2im             -0.0162      0.028     -0.585      0.559      -0.071       0.038
RemitOut:cl0im    -0.0884      0.033     -2.697      0.007      -0.153      -0.024
RemitOut:cl1im     0.1113      0.048      2.308      0.022       0.016       0.206
RemitOut:cl2im    -0.0722      0.056     -1.291      0.198      -0.182       0.038
RemitIn           -0.0223      0.045     -0.501      0.617      -0.110       0.065
cl0em              0.0286      0.014      2.109      0.036       0.002       0.055
cl1em             -0.0168      0.017     -0.969      0.333      -0.051       0.017
cl2em              0.0462      0.037      1.261      0.208      -0.026       0.118
RemitIn:cl0em      0.0082      0.021      0.387      0.699      -0.034       0.050
RemitIn:cl1em      0.0017      0.012      0.139      0.889      -0.022       0.026
RemitIn:cl2em      0.0019      0.048      0.040      0.968      -0.092       0.096
School6:cl1em     -0.0553      0.023     -2.399      0.017      -0.101      -0.010
School6:cl2em      0.0758      0.054      1.416      0.158      -0.030       0.181
y1970s            -0.3613      0.123     -2.948      0.003      -0.603      -0.120
y1990s             0.1775      0.092      1.939      0.053      -0.003       0.358
y2000s             0.1703      0.087      1.961      0.051      -0.001       0.341
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.028902    y1980s    0.219653
y1990s    0.312139     y2000s    0.439306
Mean wage growth over period (log):  0.0932704225055
Cluster:  4                             OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.501
Model:                            OLS   Adj. R-squared:                  0.435
Method:                 Least Squares   F-statistic:                     7.617
Date:                Mon, 15 Aug 2016   Prob (F-statistic):           1.11e-14
Time:                        19:07:56   Log-Likelihood:                -28.605
No. Observations:                 173   AIC:                             99.21
Df Residuals:                     152   BIC:                             165.4
Df Model:                          20
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         -0.0011      0.151     -0.008      0.994      -0.299       0.297
School6            0.2650      0.106      2.494      0.014       0.055       0.475
School6:cl0im      0.1232      0.049      2.536      0.012       0.027       0.219
cl0im              0.0713      0.055      1.294      0.198      -0.038       0.180
cl1im              0.0337      0.068      0.492      0.624      -0.102       0.169
cl2im             -0.1357      0.099     -1.368      0.173      -0.332       0.060
RemitOut:cl0im    -0.0511      0.262     -0.195      0.846      -0.569       0.466
RemitOut:cl1im     0.4892      0.323      1.512      0.133      -0.150       1.128
RemitOut:cl2im    -0.8728      0.348     -2.506      0.013      -1.561      -0.185
RemitIn           -0.1752      0.325     -0.539      0.591      -0.817       0.467
cl0em              0.4104      0.109      3.773      0.000       0.195       0.625
cl1em             -0.3163      0.105     -3.013      0.003      -0.524      -0.109
cl2em             -0.1389      0.052     -2.695      0.008      -0.241      -0.037
RemitIn:cl0em      0.7195      0.266      2.701      0.008       0.193       1.246
RemitIn:cl1em     -0.9613      0.252     -3.808      0.000      -1.460      -0.463
RemitIn:cl2em     -0.1418      0.061     -2.314      0.022      -0.263      -0.021
School6:cl1em      0.0790      0.075      1.052      0.295      -0.069       0.227
School6:cl2em     -0.2534      0.091     -2.772      0.006      -0.434      -0.073
y1970s             0.4881      0.157      3.108      0.002       0.178       0.798
y1990s             0.1760      0.090      1.965      0.051      -0.001       0.353
y2000s             0.5329      0.114      4.686      0.000       0.308       0.758
==============================================================================
```

Attachment 3.16

Summary statistics schooling and remittances model, high-skilled wage growth [clusters 1-4 only]

cl0im= immigration low income, cl1im = immigration mid income, cl2im = immigration high income

cl0em = emigration low income, cl1em = emigration mid income, cl2em = emigration high income

RemitOut = remittances paid, RemitIn = remittances received, School = schooling indicator, added 6 = six years before wage growth, at the same time as migration.

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.193878
y1990s    0.306122
y2000s    0.500000

Mean wage growth over period (log):  0.136240085277
Cluster:  1                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.602
Model:                            OLS   Adj. R-squared:                  0.505
Method:                 Least Squares   F-statistic:                     6.202
Date:                Tue, 16 Aug 2016   Prob (F-statistic):           2.90e-09
Time:                        14:47:43   Log-Likelihood:                -57.133
No. Observations:                  98   AIC:                             154.3
Df Residuals:                      78   BIC:                             206.0
Df Model:                          19
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept          0.8472      1.053      0.804      0.424      -1.250       2.944
School6:cl0im      0.2650      0.067      3.942      0.000       0.131       0.399
cl0im              0.0201      0.083      0.242      0.809      -0.146       0.186
cl1im             -0.2578      0.632     -0.408      0.684      -1.516       1.000
cl2im              0.9756      0.792      1.232      0.222      -0.601       2.552
RemitOut:cl0im    -0.0168      0.194     -0.087      0.931      -0.402       0.369
RemitOut:cl1im    -1.7224      1.078     -1.598      0.114      -3.869       0.424
RemitOut:cl2im     2.8644      1.746      1.641      0.105      -0.611       6.340
RemitIn           -0.4107      0.339     -1.210      0.230      -1.087       0.265
cl0em              0.1023      0.058      1.778      0.079      -0.012       0.217
cl1em              0.0163      0.125      0.131      0.896      -0.232       0.265
cl2em              0.0515      0.136      0.380      0.705      -0.218       0.321
RemitIn:cl0em     -0.3609      0.099     -3.638      0.000      -0.558      -0.163
RemitIn:cl1em      0.2561      0.226      1.131      0.262      -0.195       0.707
RemitIn:cl2em      0.4973      0.129      3.843      0.000       0.240       0.755
School6            0.5852      0.173      3.381      0.001       0.241       0.930
School6:cl1em      0.1472      0.065      2.279      0.025       0.019       0.276
School6:cl2em     -0.3045      0.111     -2.746      0.007      -0.525      -0.084
y1990s            -0.2502      0.197     -1.270      0.208      -0.642       0.142
y2000s            -0.0681      0.230     -0.296      0.768      -0.525       0.389
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.156250
y1990s    0.552083
y2000s    0.291667
Mean wage growth over period (log):  0.230735616797
Cluster:  2                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.477
Model:                            OLS   Adj. R-squared:                  0.347
Method:                 Least Squares   F-statistic:                     3.653
Date:                Tue, 16 Aug 2016   Prob (F-statistic):           2.85e-05
Time:                        14:47:43   Log-Likelihood:                -23.935
No. Observations:                  96   AIC:                             87.87
Df Residuals:                      76   BIC:                             139.2
Df Model:                          19
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept          0.2768      0.291      0.952      0.344      -0.303       0.856
School6:cl0im      0.0623      0.032      1.973      0.052      -0.001       0.125
cl0im             -0.0398      0.118     -0.337      0.737      -0.274       0.195
cl1im              0.1328      0.145      0.918      0.362      -0.155       0.421
cl2im             -0.2462      0.273     -0.900      0.371      -0.791       0.298
RemitOut:cl0im     0.0771      0.064      1.206      0.232      -0.050       0.204
RemitOut:cl1im     1.5110      0.501      3.015      0.003       0.513       2.509
RemitOut:cl2im    -2.6585      0.868     -3.062      0.003      -4.388      -0.929
RemitIn            0.1974      0.133      1.480      0.143      -0.068       0.463
cl0em              0.1194      0.107      1.112      0.270      -0.094       0.333
cl1em              0.0572      0.065      0.881      0.381      -0.072       0.187
cl2em             -0.1930      0.088     -2.201      0.031      -0.368      -0.018
RemitIn:cl0em      0.2066      0.071      2.919      0.005       0.066       0.348
RemitIn:cl1em     -0.1110      0.067     -1.663      0.100      -0.244       0.022
RemitIn:cl2em     -0.0857      0.063     -1.359      0.178      -0.211       0.040
School6            0.1850      0.120      1.542      0.127      -0.054       0.424
School6:cl1em     -0.0722      0.065     -1.108      0.271      -0.202       0.058
School6:cl2em      0.1473      0.081      1.807      0.075      -0.015       0.310
y1990s             0.0877      0.138      0.636      0.526      -0.187       0.362
y2000s            -0.2343      0.195     -1.201      0.234      -0.623       0.154
==============================================================================
```

```
Percentage of the data from decennium:
 y1970s    0.000000
y1980s    0.340314
y1990s    0.293194
y2000s    0.366492
Mean wage growth over period (log):  0.157150718482
Cluster:  4                            OLS Regression Results
==============================================================================
Dep. Variable:          Growthratelog   R-squared:                       0.652
Model:                            OLS   Adj. R-squared:                  0.572
Method:                 Least Squares   F-statistic:                     8.117
Date:                Thu, 18 Aug 2016   Prob (F-statistic):           1.53e-11
Time:                        01:20:45   Log-Likelihood:                -4.6870
No. Observations:                  97   AIC:                             47.37
Df Residuals:                      78   BIC:                             96.29
Df Model:                          18
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         -0.7362      0.380     -1.939      0.056      -1.492       0.020
School6:cl0im     -0.2300      0.134     -1.720      0.089      -0.496       0.036
RemitOut           0.9695      0.834      1.162      0.249      -0.692       2.630
cl0im             -0.1658      0.110     -1.505      0.136      -0.385       0.054
cl1im              0.1100      0.095      1.163      0.248      -0.078       0.298
RemitOut:cl0im    -0.8262      0.359     -2.300      0.024      -1.541      -0.111
RemitOut:cl1im     1.5007      0.410      3.664      0.000       0.685       2.316
cl2im              0.1627      0.115      1.411      0.162      -0.067       0.392
cl0em             -0.4339      0.111     -3.924      0.000      -0.654      -0.214
School6           -0.4351      0.344     -1.264      0.210      -1.120       0.250
RemitIn           -0.3734      0.385     -0.970      0.335      -1.140       0.393
cl1em              0.2939      0.104      2.830      0.006       0.087       0.501
cl2em              0.4726      0.113      4.183      0.000       0.248       0.697
School6:cl1em     -0.1639      0.084     -1.954      0.054      -0.331       0.003
School6:cl2em      0.2958      0.117      2.528      0.013       0.063       0.529
RemitIn:cl1em      0.0002      0.212      0.001      0.999      -0.423       0.423
RemitIn:cl2em     -0.2038      0.117     -1.745      0.085      -0.436       0.029
y1990s             0.3051      0.131      2.329      0.022       0.044       0.566
y2000s             0.9439      0.173      5.458      0.000       0.600       1.288
==============================================================================
```