Note: work is in progress, please do not quote!

The Social Inclusion and Inequality Nexus: EU versus non-EU migrants

Amjad Naveed (PhD)
(amjadn@btech.au.dk)
Department of Business Development and Technology
Aarhus University, Denmark

Cong Wang (PhD)
Bond Business School
Australia

Abstract:

This study examines how social inclusion of migrants affects income inequality in European Countries. Both positive and negative impacts of social inclusion and integrations on income inequality exists in the literature. However, there is a lack of evidences whether the impacts of social inclusion of EU and non-EU nationals on income inequality are different. Therefore, this study specifically considers two types of migration flow: 1) migration flow from EU, 2) migration flow from non-EU. The main hypothesis is to test whether the social inclusion from different types of migration flow reduces income inequality. Using data from 33 mainly European countries over the period 2003-2015 and controlling for savings rate, arable land rate and age-dependency ratio, our results indicate that there is a significant negative relationship between social inclusion and income inequality. In particular, we find that social inclusion from non-EU migrants significantly reduces income inequality compare with EU migrants.

JEL Codes: C31, D24, O49

Keywords: migration flow; social inclusion; income inequality; European Union

Note: work is in progress, please do not quote!

1. Introduction

This paper examines the relationship between social inclusion and income inequality across European (EU) countries. Specifically, we examine how social inclusion of immigrants from EU and non-EU affects income inequality in European Countries. Existing literature concerning integration and inequality have focused on political, economic, and social dimension of social inclusion from immigrants. For instance, Kosonen (1995), Boje et. al. (1999), Beckfield (2006) and Busemeyer & Tober (2015) observe the same relationship and argue that social inclusion from immigrants increases income inequality. On the other hand, Moses (1995) claim that social integrations promotes the process of integration, which ultimately reduces income inequality. The sociological approach also promotes the view that regional integration is a process of social inclusion that diminish the income gap between rich and poor (Therborn 1999). However, none of the previous studies has differentiate the impact from EU and non-EU immigrants on income inequality. Therefore, this study specifically considers the impact of social inclusion from EU and non-EU immigrants flow to the EU countries.

According to the standard model of migration, income differential among countries stimulates migration from underdeveloped to developed countries (Grubel and Scott, 1966; Bhagwati and Hamada, 1974; Haque and Kim, 1995). Skilled migration is another source raising income and employment to the less developed areas whereas unskilled migration may have opposite effect where brain drain occurs (Naveed et al., 2017). Moreover, Bhagwati and Hamada (1974) show that skilled migrants can easily be integrated in the host country if they have ample access to the opportunities of social interaction. Consequently, social inclusion helps and supports immigrants with both social and economic integration. For instance, being social and making connections with the locals not only provides opportunity to discuss about employment but also boost pleasant interactions between migrants and locals, which ultimately reduces income inequality between and across the immigrant's groups (EU and non-EU). Besides non-EU immigrants flow to EU, there is rising flow of immigration from new EU member states, which is a serious challenged for income inequality in the Europe

(Kahanec and Pytliková, 2017). It means there are two different types of migration flow exists: 1) EU migration (mainly from underdeveloped areas of EU to developed part of EU), 2) non-EU migration (mainly from non-EU developing countries to the EU). The process of integration and social inclusion may differ for these two types of immigration flow. Consequently, they may have differential impact on oncome inequality. A key European goal is to provide equal employment opportunities and reducing income gap between rich and poor (Beckfield, 2006; Busemeyer & Tober, 2015). Therefore, analysing this relationship between social integration (social inclusion) and income inequality is in the interest of both academics and policy makers. The analysis in this paper is interesting and unique in the sense that there is no such study exists which have discussed whether EU or non-EU migration helps to maintain local growth and reduce income inequality. This study specifically address this issue which is largely been ignored in the literature. The main hypothesis is to test whether the social inclusion from migration flow (with different social and ethnic background) helps to reduce income inequality.

For empirical analysis, we use the data for 33 mainly European countries (with the exception of Turkey) from Eurostat over the period 2003-2015. The dependent variable is the Gini coefficient (measure of income inequality) and main explanatory variable is social inclusion (SI) with three dimensions: SI-EU, SI-nonEU and SI-total. The measure of social inclusion is constructed by using four sub-indicators from Eurostat definition (for detail, see Section 3). We estimate a panel data model using System GMM, and control for savings rate, arable land rate and age-dependency ratio. Our results indicate that there is a significant negative relationship between social inclusion and income inequality. In particular, we find that social inclusion from non-EU migrants has strong effect on reducing income inequality compared to EU migrants. However, the effects from sub-indicators of social inclusion on inequality are mixed. From policy point of view, the results of this study are important which highlights that social inclusion promotes integration that ultimately reduces income inequality, which is in line with the agenda of EU policy.

The remainder of the paper is organized as follows. Section 2 provides the theoretical foundation and discusses the literature. The data and variables, identification issues and empirical findings are presented in Section 3. The robustness check for the main results is offered in Section 4. Section 5 concludes the study.

2. Theoretical Foundation and Literature

The theoretical foundation of this study is based on the standard models of migration where income differential across countries is the main source of migration. Typically migrants moves from rich to poor countries. The movement is not only taking place across countries but also economically less developed region to developed regions within the country. Such process of migration increases the growth at the destination location while it reduces may hampers growth where brain drain occurs (Grubel and Scott, 1966; Bhagwati and Hamada, 1974; Haque and Kim, 1995). However, regarding its positive or negative effects, different argument has been discussed in the literature. For instance, the sociological approach considers that migration and integration reduce income inequality (Therborn 1999). While opposite effect of migration on integration and income inequality are exposed by different scholars. For instance, Beckfiled (2006) provide a detail analysis about the link between regional integration and income inequality with a specific focus on political and economic integration in EU.

Similarly, Boje et al. (1999) claimed that income inequality will rise in the host country when high rate of migration flow occurs Reduction in different forms of welfare attributes are also results of migration which further affect purchasing power of natives. Kosonen (1995) and Moses (1995) give a different type of explanation that income gap between local and migrants is due to globalization process, which only occurs at the initial phase of the process but it may reduce at the later stage. Along the same line of reasoning, Walby (1999) claims that globalization is the main reason for increasing migration flow that reduces the state capacity in terms of providing social welfare, equal opportunities and economic rights. Therefore, social inclusion become challenge for the state at the presence of different types of practical constraints. Besides, Bhagwati and Hamada (1974) argue that migration flow of skilled person can easily be integrated given that they have enough access to the opportunities of being socially integrated. Consequently, social inclusion enhances integration that may reduce income inequality in the receiving country.

Social inclusion is not a sole term to explain rather it is connected to employment opportunities, ethnic background of immigrants, social status, level of education and government policies about welfare that may affect the income of migrants and income inequality. Recently, Frederiksen and Poulsen (2016) empirically show that skill-upgrading and high level of education are the dominant factors that cause rising income inequality in Denmark over the period 1992-2007. With regards to immigrants, Nielsen et al., (2003) and Husted et al., (2001) find that 2nd generation immigrants from developing countries have lower employment rate and hence lower income compare to native because of parental capital and neighbourhood effects. Furthermore, they also explain that 1st generation immigrants integrate partially to natives, but the integration process differs between refugees and non-refugees. Naveed and Wang (2018) briefly explain a relationship between income inequality, ethnic background and religious affiliation for global sample. Though, a micro level study by Esmaeilzadeh et al., (2018) discover that earning difference between native and immigrants is largely determined by immigrant minority status, age at entry, education, experience, and marital status.

According to normative view, social inclusion is to make sure that all actors of society including migrant and children can participate as valued, respected and contributing members of society (Mitchell and Shillington, 2002). In a theoretical study by Oxoby (2009) discuss the link between social inclusion, social cohesion and social capital with reference to the issues of poverty, unemployment, and social exclusion for the society. Their study explain that there is a trade-off between social inclusion and economic efficiency. Achieving social inclusion means reduction in poverty but not efficiency. In order to achieve balanced growth, policy makers should focus on both social and economic inclusion which are beneficial for the society.

The literature cited above shows the importance of social inclusion (or integration) of migrants on income inequality. However, there is not enough empirical evidence that how different types of social inclusion of migrants (from EU and non-EU) affect income inequality. Furthermore, whether both types of migrants have different impact on income inequality. These are very important questions which are considered in this study and tried to fill the gap in the literature. For that reason, this study considers two types of immigrant's flow (from EU and from non-EU origin) and analyses their impact on income inequality in EU countries.

3. Empirics

3.1 Data

We collect unbalanced panel data on inequality, social inclusion, savings rate, arable land rate and age-dependency ratio for 30 European countries from 2003 to 2015. Detailed data sources can be found in the appendix (A3). The detailed discussions on the variables are the following:

3.1.1 Inequality

Inequality is measured in three ways: (1) the World Bank's market Gini, which is based on gross income (income before taxes and transfers); (2) the top 10% income share; and (3) the top 20% income share. The downside of using the top 10% and 20% income share is that it does not account for income inequality at the lower end of the distribution; however, it has the advantage over the Gini in that it measures the income share of the rich elite and, therefore, the resources available to the elite to potentially hamper equality. The baseline regressions are based on World Banks's Gini data, while the results based on top 10% and top 20% income share are presented in the robustness section.

3.1.2 **Social Inclusion**

Our main variable of interest is social inclusion (SI: proxy for social integration) which make sure that everybody in the society including migrant and children can participate as valued, respected and contributing members of society and have long term positive impact on employment and poverty (Mitchell and Shillington, 2002; Oxoby, 2009). For our empirical analysis, we used three different dimensions of SI: 1) social inclusion from EU migrants (SI-EU), 2) social inclusion from non-EU migrants (SI-nonEU) and 3) social inclusion from total (EU and non-EU combined) immigrants (SI-total). Social inclusion is one of the important indicators of integration which is constructed by combining four sub-indicators according to definitions from the Eurostat (Eurostat 2018). The four sub-indicators are severe material deprivation rate, housing cost overburden rate, tenure status (tenant versus owner), and overcrowding rate. Each indicator is measured by the percentage of the total population by broad group of citizenship aged 18 and over except for the reporting country. Table A1 in the appendix explains the short description of each indicators of social inclusion variable.

Figures 1 to 3 show a two-way simple OLS regression plot between inequality and the three social inclusion measures. The data used in the graphs are unweighted averages across the time dimension (2003-2015). From these figures, it’s clear that there’s a significant negative association between inequality and all three aspects of social inclusion from different immigrant groups (EU, non-EU and overall). The higher the level of social inclusion, the lower the after-tax Gini index (hence, inequality). Whether this relationship is causal and prominent across both the time and country dimensions requires careful analysis of the panel data with further treatment for endogeneity, which we discuss later in Sections 3.1.4.

Note: work is in progress, please do not quote!

Figure 1: SI_{EU} and Inequality

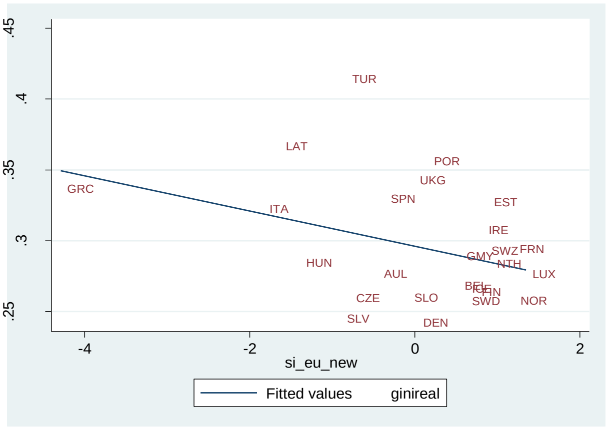


Figure 2: SI_{nonEU} and Inequality

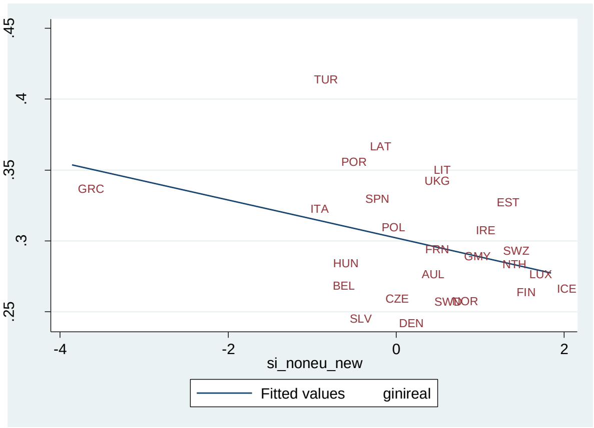
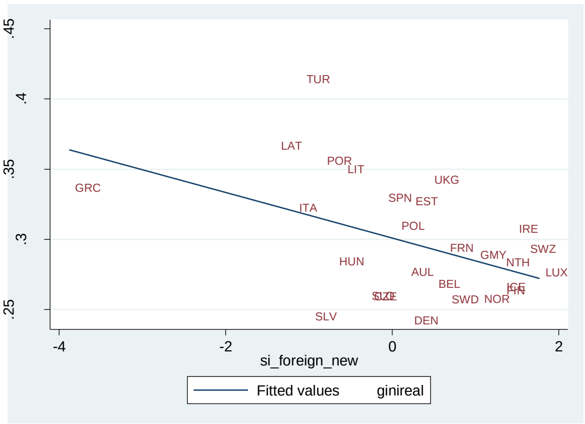


Figure 3: $SI_{foreign}$ and Inequality



3.1.3 Control Variables (Savings Rate, Age-dependency Ratio and Arable Land Rate)

Savings rate is the national savings as a percentage of Gross National Income (GNI), measured using the natural resource depletion method. This indicator is obtained from World Bank. Duesenberry (1949), in his seminal work, introduces the relative income hypothesis to rationalize the well-established differences between cross-sectional and time series properties of savings rate. He proposes that "for any given relative income distribution, the percentage of income saved by a family will tend to be a unique, invariant, and increasing function of its percentile position in the income distribution". Friedman (1957) on the other hand, proposes the permanent income hypothesis. According to this view the cross-sectional correlation between savings rate and income is driven by transitory deviations from permanent income, while in the aggregate, most transitory components cancel out, leading to the close relation between consumption and income observed in time series data. Numerous empirical studies have also established that there's a potential negative relationship between income inequality and savings rate (see e.g., Dynan et al. (2004), Mayer (1966, 1972), Alvarez-Cuadrado and Vilalta (2012)). We therefore decided to control savings rate as a potential determinant of inequality in our core regressions concerning the relationship between women empowerment and inequality.

Age dependency ratio, taken from the World Bank, measures the ratio of non-working-age population (dependents including retired population and minors) to working-age population. In the current literature, demographic change such as aging is found to have significant impact on income inequality. For e.g., Guerin (2013) in a policy briefing for the European Commission note that "income inequality in Europe is sensitive to population aging, since the elderly face high poverty risks and represent a growing share of the population. As pension system come under increased strain, meeting the Europe 2020 poverty targets will partly hinge on sustainable pensions." Other studies including Deaton and Paxson (1997), Schultz (1997) and Lam and Levison (1992) contend that population ageing leads to an increase in consumption or income inequality, as aging has led to a decline in the share of resources going to the elderly (Gruber and Wise (2001)), and that a rise in the overall dependency ratio is leading to a decline in social transfers (Razin et al. (2002)). We therefore decide to include aging (measured as the age dependency ratio) as a control variable for investigating the impact of women empowerment on inequality. According to the above studies, we expect age dependency ratio to positively impact inequality.

Finally, we also control the ratio of arable land area to total population (arable land area per head) in our core regression. Natural resource scarcity has long been argued as a potential determinant of income inequality in the current literature. For example, Gylfason and Zoega (2002) argue that "increased dependence on natural resources and natural resource scarcity tend to go along with less rapid economic growth and greater inequality in the distribution of income across countries". Arable land resource has long been a point of contention in the economic development in many developing countries, especially those in Africa where the economies are still largely agrarian. Abundant arable land resource can reduce the chance of unequal distribution of land, which further reduces the chance of income inequality.

3.1.4 Identification Issue

It's highly plausible that social inclusion is endogenous in its relationship with inequality, since it's highly likely that there exists a third variable that potentially influences both social inclusion and inequality such as religion. For e.g., On the one hand, Bouma (2016) in an editorial for Social Inclusion contends that there exists both theoretical arguments and empirical evidence that religion shapes national values even in secular societies where certain religious belief is dominant, and such shaped national values in turn influences social norms and consequently determines whether a society is tolerant and inclusive towards "outsiders" such as foreign immigrants who may have different sets of values to the natives. On the other hand, Naveed and Wang (2018) provides empirical evidence that there exists a causal relationship between religion and income inequality. In particular, they find that Islam and Judaism reduce income inequality while in general Christianity and Buddhism increases inequality.

To address the potential endogeneity issue on social inclusion, we estimate our core model using the system GMM estimator. The system GMM estimator (see Blundell-Bond (1998)), uses internal time lags as instruments for each endogenous regressor and treat all explanatory variables (including control variables) as being endogenous.

Table 1 provides summary statistics and definitions of the key variables in the core regressions. The notable differences between means and standard deviations for our inequality variables ($gini, top10$ and $top20$), social inclusion variables ($SI_{EU}, SI_{nonEU}, and\ SI_{foreign}$) and the age dependency ratio suggest high variations in at least the cross-country dimension. On the contrary, the other two control variables (sav and $arable$) have low variations, suggesting a potential stationary time-series for both types of variables. After checking for unit root and co-integration, we dismiss the need for a co-integration-based error correction model for our core model. A correlation table for all core variables and list of countries involved in core regressions can be found in the appendix (A1 and A2 respectively).

Table 1: Summary Statistics and Variable Definition

Variables	Definition	Obs	Mean	Std. Dev.	Min	Max
$gini$	GINI index	292	31.52	3.69	23.72	42.18
$ginireal$	After tax GINI index	281	29.91	4.34	23.01	43.60
$top10$	Income share held by highest 10%	279	24.76	2.34	20.14	31.73
$top20$	Income share held by highest 20%	292	39.65	0.77	34.04	48.37
$herf$	Herfindahl index	416	0.07	0.03	0.03	0.3
age	Age dependency ratio (% of working-age population)	408	48.58	3.89	38.09	59.17
sav	Adjusted savings: natural resources depletion (% of GNI)	374	0.87	2.20	0.00	17.06
$arable$	Arable land (per capita arable land area)	371	23.99	12.93	1.20	58.89
SI_{EU}	Social inclusion – immigrants from EU	289	-0.025	1.42	-5.54	2.60
SI_{nonEU}	Social inclusion – immigrants from non-EU	314	0.07	1.36	-5.53	2.97
$SI_{foreign}$	Social inclusion – all foreign immigrants	333	0.02	1.38	-5.40	2.12

Notes: The summary statistics given above for each variable covers the largest sample possible (unbalanced panel). The largest time span covered is from 2003 to 2015 for all variables. $overcrowd$ $housing$ $tenant$ and $deprived$ are subcomponents of social inclusion. Three versions of the four variables are collected for the EU, non-EU and total foreign immigrants population respectively. The descriptive stats for the 12 variables are not reported, they are available upon request. The definition of the four subcomponents are the following: $overcrowd$ is the overcrowding rate by broad group of citizenship for population aged 18 and over. $housing$ is housing cost overburden rate by broad group of citizenship for population aged 18 and over. $tenant$ is percent of population aged 18 and over who are renting a property (holding tenant status). $deprived$ is severe material deprivation rate by broad group of citizenship for population aged 18 and over.

3.2 OLS and System GMM results

For the panel OLS regressions with fixed effects, we attempt to estimate the following equation:

$$inequal_{i,t} = \alpha_0 + \alpha_1 SI_{i,t} + \alpha_2 X_{i,t} + \delta_i + b_t + \epsilon_{i,t} \quad (1)$$

Where $inequal_{i,t}$ is inequality, SI is social inclusion, specifically, we measure social inclusions among the EU foreigner (SI_EU), non-EU foreigner (SI_nonEU) and the overall foreign population ($SI\_foreign$) in the core regressions. $X_{i,t}$ is a set of control variables including savings rate, age dependency ratio and arable land rate. α_0 is constant across time and cross-section, δ_i is the country dummy, b_t is the time dummy, and ϵ is the error term.

Columns (1) to (3) in Table 3 show results from the pooled OLS model. It's quite evident that social inclusion has strong influence on inequality. All three types of social inclusion from three different immigrants groups (EU, non-EU and overall foreign) have significant and negative (reducing) impact on inequality. The size of the coefficients across all three aspects of social inclusion are also comparable to one another, suggesting that the equal importance of the impacts of all three aspects of social inclusion on inequality. In other words, having an inclusive society that has accepting attitudes towards EU immigrants, non-EU immigrants and all immigrants are equally important for reducing inequality. There's no priority in choosing one of the three. The significant F-test statistics from Column (1) to (3) suggest the joint significance of our core social inclusion variables and the control variables in the pooled OLS model.

For the system GMM regressions with fixed effects, we estimate the following equation:

$$inequal_{i,t} = \alpha_3 + \alpha_4 inequal_{i,t-1} + \alpha_5 SI_{i,t} + \alpha_6 X_{i,t} + b_t + u_{i,t} \quad (2)$$

Where $inequal_{i,t-1}$ is the lag of $inequal_{i,t}$ country dummies are not included because System GMM already includes country specific fixed effects.

Columns (3) to (6) in Table 2 show the results from the system GMM regressions. These results are consistent with the OLS results in terms of significance and the signs of the coefficients. The significant negative impact of all three aspects of social inclusion on inequality are confirmed here. However, we do notice that the size of the coefficients differs in some magnitude for the three social inclusion variables, with the negative impacts of social inclusion on inequality from both the EU migrants and the total foreign immigrants, being significantly smaller under the system GMM estimator than they are under the OLS. We believe these findings are the results of measurement errors from the social inclusion variables, which are also a source of endogeneity. This further justifies the use of system GMM as an estimation technique which treats all regressors endogenously. In all system GMM regressions, our instrument matrix collapsing technique ensures that we use less number of instruments than the number of countries. P-values from Hansen-J and Difference in Hansen tests indicate that the instruments (first lag of endogenous variables plus year dummies) are not over-identified and exogenous. As expected, p-values from AR(1) and AR(2) tests suggest that there are high first order autocorrelation and no evidence for significant second-order autocorrelation. Moreover, the coefficients on SI_{nonEU} is 3 to 4 times larger than those on SI_{EU} and$SI_{foreign}$. This suggests that having a society that's inclusive and accepting non-EU immigrants matters much more for reducing income inequality.

From all columns in Table 2, the control variables have expected signs and significant. Furthermore, the results are consistent across OLS and system GMM models. The positive impact of age-dependence ratio, as well as the negative impacts of savings rate and arable land rate on inequality confirm the findings of Guerin (2013), Dynan et al. (2004), Mayer (1966, 1972) and Gylfason and Zoega (2002).

4. Robustness Checks

So far, we have attempted to explain inequality by social inclusion using only one particular measure of inequality (Gini) and having only looked at the impact from three generalized measures of social inclusion. There are certainly opportunities to explore different types of measures of inequality to ensure that our results are not driven by the specific choice of measures of inequality included. There's also the prospect to explore the impacts of subcomponents of social inclusion on income inequality.

To address these issues, we perform two types of robustness checks in Tables 3 and 4. First, we use the income share of the top 10% and 20% as alternative measures of inequality in Columns (1) to (6) in Table 3, we also use the gross Gini (which is a less favourable measure of Gini since it does not take into account income redistribution) in Columns (7) to (9) and the Herfindahl index in Columns (10) to (12) as further alternative measures of income inequality. Secondly, we investigate the impact of social inclusion on inequality by examining the impacts of the subcomponents, i.e. overcrowding rate, housing cost, tenant status and material deprivation rate. This is to ensure that our core results are not driven by the use of principal component analysis (PCA) as an approach to combine the four sub-indicators of social inclusion.

Columns (1) to (3) and Columns (4) to (6) in Table 4 report the results estimated using the core model (system GMM) for top 10% and top 20% respectively. The significant negative (decreasing) effect of social inclusion on income inequality is confirmed, albeit when top 20% is used to measure inequality, the impacts of social inclusion (across all three aspects) on inequality are weaker. This suggests that social inclusion has higher impact on reducing income inequality between the top elite and the rest of society. The gross Gini results also confirm the core results, suggesting our core results are not driven by the choice of using the after-tax Gini measure.

Moreover, the coefficients of almost all social inclusion variables are also considerably smaller when the dependent variable is changed to Herfindahl index. While the Herfindahl index is an effective proxy for inequality (see e.g., Nembua (2007), Rhoades (1995)), the concept nevertheless is based on market share, hence, measures inequality stemming from market concentration. The smaller coefficients of social inclusion in the Herfindahl index regressions therefore seem to suggest that the impact of social inclusion on income inequality is much bigger than its impact on market concentration. Nevertheless, the Herfindahl index regressions confirm the importance of social inclusion from different immigrant groups in impacting different types of inequality.

Most importantly, in all cases where top 10%, top 20%, the Herfindahl index and gross Gini are used as measures of inequality, the non-EU immigrant inclusion seems to matter more for the reduction of

Note: work is in progress, please do not quote!

Table 2: Social Inclusion and Inequality: OLS, Panel Fixed Effect and System GMM Results

	Pooled OLS Regressions			System GMM Regressions		
	(1) *ginireal*	(2) *ginireal*	(3) *ginireal*	(4) *ginireal*	(5) *ginireal*	(6) *ginireal*
l.ginireal				0.979*** (25.93)	1.132*** (11.64)	0.948*** (13.55)
SI_EU	-0.009*** (-6.77)			-0.002** (-2.36)		
SI_nonEU		-0.008*** (-5.07)			-0.009*** (-3.44)	
SI_foreign			-0.012*** (-7.53)			-0.003* (-1.94)
age	0.002*** (2.99)	0.001 (0.81)	0.001*** (2.81)	0.000 (1.43)	-0.001 (-0.34)	0.000 (0.98)
sav	-0.003*** (-5.47)	-0.003*** (-5.61)	-0.003*** (-5.36)	0.001 (1.40)	0.000 (0.35)	-0.000 (0.31)
arable	-0.001** (-2.56)	-0.000 (-0.93)	-0.000 (-0.93)	-0.002* (-1.77)	-0.001 (-1.53)	-0.000 (-1.08)
Observations	198	206	220	182	200	187
No. of Countries	25	26	27	25	27	26
No. of Instruments				13	13	13
AR(1)				(0.017)	(0.018)	(0.038)
AR(2)				(0.736)	(0.542)	(0.247)
Hansen Test of Over Identification				3.52 (0.897)	10.65 (0.223)	10.39 (0.239)
Difference in Hansen Test				5.22 (0.266)	9.33 (0.230)	4.07 (0.396)
Country fixed effect	No	No	No	Yes	Yes	Yes
Time fixed effects	No	No	No	Yes	Yes	Yes
Log likelihood						
Wald Chi^2						
F	27.19 (0.000)	19.70 (0.000)	29.55 (0.000)			
R^2	0.21	0.12	0.21			

Notes: The regressions are estimated using pooled OLS with heteroscedasticity consistent standard errors, as well as two-step system GMM. The year coverage ranges from 2003 to 2015, z-values are in the parentheses**.** For system GMM, which uses internal lags as instruments and treat all explanatory variables as endogenous, we report the Hansen test of over identification, the null hypothesis is that the instruments are valid (i.e. not over-identified). We also report the number of instruments (lags) used in the system GMM process and ensure that the number of instrument is less than the number of countries. $l.ginireal$ is the lagged dependent variables. See notes in Tables 1 for other information including variable definitions. Significance at the 10%, 5% and 1% levels are indicated by *, ** and ***.

Note: work is in progress, please do not quote!

Table 3: Robustness to Alternative Measures of Inequality

	$top10$ Sample			$top20$ Sample			$gini$ Sample			Herfindahl Index Sample		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$l.top10$	0.809***	0.714***	0.789***	0.473***	1.195***	0.614*	0.901***	0.918***	0.700**	0.992***	0.266*	0.928***
	(22.88)	(7.07)	(8.43)	(9.05)	(2.70)	(1.67)	(5.20)	(4.74)	(2.35)	(12.31)	(1.86)	(9.43)
$l.top20$												
$l.gini$												
$l.herf$												
SI_EU	-0.570***			-0.426***			-0.502***			-0.003**		
	(-3.11)			(-3.03)			(-2.59)			(-2.06)		
SI_nonEU		-0.654***			-0.781***			-0.927**			-0.012*	
		(-5.80)			(-3.00)			(-2.03)			(-1.84)	
$SI\_foreign$			-1.178*			-0.772*			-1.861**			-0.011***
			(-1.95)			(-1.81)			(-1.96)			(-2.78)
age	0.097***	0.139***	0.088	0.446***	-0.206	0.350	0.053	0.135*	0.192	0.000	0.001***	-0.000
	(5.14)	(2.64)	(1.40)	(9.58)	(-0.45)	(0.95)	(0.49)	(1.65)	(0.83)	(0.24)	(3.97)	(-0.35)
sav	0.009	-0.114	0.015	-0.475***	-0.388	-0.680	-0.470	-0.241	-1.583	-0.001	-0.001***	-0.001
	(0.55)	(-0.98)	(0.12)	(-3.19)	(-0.87)	(-1.57)	(-1.37)	(-1.13)	(-0.78)	(-0.21)	(-2.57)	(-0.83)
$arable$	-0.006	0.010	0.028	-0.011	0.070	-0.039	0.020	-0.115	0.028	-0.000	0.000	0.000
	(-0.68)	(0.53)	(1.09)	(-0.32)	(0.41)	(-0.34)	(1.09)	(-0.96)	(1.04)	(-0.74)	(0.09)	(0.54)
Observations	154	170	180	162	181	192	162	181	192	198	218	232
No. Countries	27	29	30	27	29	30	27	29	30	27	29	30
No.Instruments	16	12	12	17	12	12	9	13	10	9	16	11
AR(1)	(0.058)	(0.035)	(0.033)	(0.026)	(0.086)	(0.065)	(0.008)	(0.014)	(0.002)	(0.012)	(0.087)	(0.051)
AR(2)	(0.141)	(0.109)	(0.114)	(0.189)	(0.318)	(0.272)	(0.167)	(0.115)	(0.397)	(0.698)	(0.271)	(0.561)
Hansen Test	13.81	2.88	1.91	15.99	1.75	2.15	4.02	4.74	3.15	2.82	10.47	2.60
	(0.244)	(0.896)	(0.965)	(0.192)	(0.973)	(0.951)	(0.403)	(0.785)	(0.676)	(0.589)	(0.489)	(0.857)
Difference in Hansen Test	3.52	2.24	1.51	7.75	1.07	1.50	2.93	1.70	2.86	2.65	3.94	2.53
	(0.833)	(0.693)	(0.679)	(0.458)	(0.785)	(0.682)	(0.570)	(0.791)	(0.581)	(0.448)	(0.684)	(0.772)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The regressions are estimated using the two-step system GMM estimator. The year coverage ranges from 2003 to 2015, z-values in the parentheses**.** For system GMM, which uses internal lags as instruments and treat all explanatory variables as endogenous, we report the Hansen test of over identification, the null hypothesis is that the instruments are valid (i.e. not over-identified). We also report the number of instruments (lags) used in the system GMM process and ensure that the number of instrument is less than the number of countries. The $l.$ variables are the lagged relevant dependent variables. See notes in Tables 1 and 2 for other information including variable definitions. Significance at the 10%, 5% and 1% levels are indicated by *, ** and ***.

inequality than social inclusion from EU-immigrant. These findings echo those found in Table 2 using the same estimation technique: system GMM, even though the OLS results in Table 2 suggest this is not the case.

Therefore, we can at least conclude here that using the same core estimation technique (System GMM), the results are consistent regardless how we measure inequality. The size and the sign of the coefficients on control variables are also consistent across the regressions using top 10%, top 20%, gross Gini and the Herfindahl index. In general, we do not find significant differences between our core results and the results derived from these four alternative measures of inequality.

Turning to the analysis of this relationship using the subcomponents of social inclusions in Table 4, the four subcomponents results for each of the three social inclusion measures: SI_{EU}, SI_{nonEU} and $SI_{foreign}$ are reported in Columns (1) to (4), (5) to (8) and (9) to (12) respectively. Generally speaking, while all four subcomponents matter for inequality reduction among EU and non-EU migrants, the magnitude of influence differs across the two groups. In particular, overcrowding rate is the most prominent determinant of inequality among intra-EU migrants, while housing cost matters the most for inequality reduction among non-EU migrants. When considering the migrant population as a whole, however, the influence of these four components on inequality seems to be quite the same. These findings seem to suggest that reducing housing cost for non-EU migrants is the key to inequality reduction, while EU migrants care more about overcrowding rate, albeit to a lesser degree compared with housing cost for non-EU migrants.

Overall, these results suggest that our core results are not driven by the use of a particular statistical method (PCA) to combine determinants of social inclusion to form a general measure. The results also give interesting insights to policy makers on the difference between EU migrants and non-EU migrants in terms of perception of quality of life, and how the policy makers may utilise these results to make

Note: work is in progress, please do not quote!

Table 4: Robustness to Subcomponents of Social Inclusion

	Subcomponents for SI_EU				Subcomponents for SI_nonEU				Subcomponents for $SI\_foreign$			
	(1) *ginireal*	(2) *ginireal*	(3) *ginireal*	(4) *ginireal*	(5) *ginireal*	(6) *ginireal*	(7) *ginireal*	(8) *ginireal*	(9) *ginireal*	(10) *ginireal*	(11) *ginireal*	(12) *ginireal*
$l.ginireal$	0.979*** (8.28)	0.952*** (7.67)	0.940*** (25.78)	0.784*** (18.65)	0.910*** (7.76)	1.001*** (21.48)	0.733*** (7.96)	0.910*** (18.65)	0.814*** (5.71)	1.045*** (6.01)	0.723*** (11.24)	0.965*** (22.13)
$overcrowd$	-0.011** (-2.02)				-0.005* (-1.83)				-0.006* (-1.86)			
$housing$		-0.003** (-2.26)				-2.26** (-2.26)				-0.009** (-2.04)		
$tenant$			-0.007*** (-3.95)				-0.003*** (-3.33)				-0.008** (-2.52)	
$deprived$				-0.003* (-1.75)				-0.004*** (-2.68)				-0.006*** (-3.42)
age	0.000 (0.03)	0.001 (0.77)	0.0004** (2.26)	0.001*** (5.30)	0.001 (0.72)	-0.000 (-0.19)	0.002*** (3.83)	0.000 (1.21)	0.001 (1.20)	0.000 (0.17)	0.002*** (4.89)	0.000 (0.47)
sav	0.001 (0.81)	-0.001 (-1.16)	-0.001*** (-3.91)	-0.002* (-1.75)	-0.000 (-0.02)	0.000 (0.11)	0.000 (1.56)	-0.001 (-1.46)	0.000 (0.45)	-0.001 (-1.06)	-0.002*** (-6.38)	0.001 (1.32)
$arable$	0.000 (0.67)	-0.001** (-2.32)	-0.000 (-0.43)	0.000 (1.26)	-0.000 (-0.08)	0.000 (1.03)	0.000 (0.47)	0.000 (1.23)	0.000 (0.65)	-0.000 (-0.45)	0.000 (1.07)	0.000 (0.43)
Observations	199	191	189	199	207	199	194	207	220	212	207	220
No. of Countries	25	25	25	25	26	26	26	26	27	27	27	27
No. of Instruments	9	13	17	13	13	17	17	16	13	13	17	16
AR(1)	(0.027)	(0.008)	(0.003)	(0.005)	(0.011)	(0.006)	(0.030)	(0.005)	(0.007)	(0.016)	(0.016)	(0.009)
AR(2)	(0.357)	(0.906)	(0.393)	(0.834)	(0.293)	(0.766)	(0.234)	(0.612)	(0.516)	(0.941)	(0.627)	(0.918)
Hansen Test of Over Identification	5.26 (0.261)	4.76 (0.783)	11.04 (0.525)	0.21 (0.834)	6.43 (0.599)	7.76 (0.803)	13.44 (0.338)	9.05 (0.617)	3.55 (0.896)	4.21 (0.838)	14.14 (0.292)	10.59 (0.478)
Difference in Hansen Test	5.26 (0.154)	3.54 (0.471)	6.57 (0.584)	4.23 (0.836)	1.57 (0.815)	5.51 (0.702)	5.88 (0.660)	4.32 (0.634)	2.01 (0.733)	0.62 (0.961)	6.95 (0.542)	0.74 (0.994)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The regressions are estimated using the two-step system GMM estimator. The year coverage ranges from 2003 to 2015, z-values are in the parentheses**.** For system GMM, which uses internal lags as instruments and treat all explanatory variables as endogenous, we report the Hansen test of over identification, the null hypothesis is that the instruments are valid (i.e. not over-identified). We also report the number of instruments (lags) used in the system GMM process and ensure that the number of instrument is less than the number of countries. $l.ginireal$ is the lagged dependent variables. See notes in Tables 1 and 2 for other information including variable definitions. Significance at the 10%, 5% and 1% levels are indicated by *, ** and ***.

more tailored policies towards the two groups of immigrants for the purpose of inequality reduction in the whole society in general.

5. Conclusions

In sum, this paper is motivated by theoretical literature that assert a significant negative relationship between social inclusion and inequality (see e.g. Mitchell and Shillington, 2002; Oxoby, 2009) and aims to provide some empirical analysis on the relationship between these two variables. By performing OLS and system GMM regressions with fixed effects, this paper has found evidence that support the following:

First, we identify that increasing social inclusion among EU and non-EU migrants reduces income inequality significantly. This relationship is significant regardless if we use estimation techniques that don't treat endogeneity (pooled OLS) or the one that treats endogeneity (System GMM), suggesting the relationship between the two variables is not only an association but also causal. This effect of social inclusion on inequality is also robust after controlling for the potential effect of savings rate, arable land resources and age dependency ratio on inequality.

Second, the significant decreasing effect of social inclusion on inequality is robust to alternative measures of inequality ($top\ 10, top\ 20, herf$ and $gini$), and subcomponents of social inclusion ($overcrowd, tenant$, $deprived$ and $housing$), suggesting our core results are not driven by a particular measure of inequality and the particular measure of social inclusion that we construct. The relative sizes of the coefficients on all three groups of immigrants' social inclusion in all four alternative measures of inequality regressions are also consistent with those found in the core System GMM regressions.

Thirdly we find that the effect of social inclusion on inequality differs across the immigrant groups. In particular, overcrowding rate seems to matter the most for EU migrants in terms of inequality reduction, while housing cost matters the most for non-EU migrants for the same purpose. We think policy makers may use these findings to make tailored policies against these two distinct immigrant groups for inequality reduction purposes. For example, the policy makers may create affordable housing schemes to encourage the increase of social inclusion among non-EU migrants, reducing income inequality for the whole society during the process.

Finally, although this paper has a number of important implications which help the understanding of the relationship between income inequality and social inclusion, however, there are various limitations to the current study and therefore, potential directions for future work. The key limitation stems from the fact that our core sample of 30 European countries with a short time series from 2003 to 2015 for the measurement of social inclusion is not large enough, which restricts the usable sample size of our control variables, which have much larger number of observations. Another limitation, which is also connected to the first one, is the measurement of social inclusion. In particular, the lack of observations on the sub-indicators of social inclusion can certainly be further improved in the future, when the availability of both the scale and the scope of public data on these social measures improve. To better understand the relationship between all types of inclusions and income inequality, future work could focus on combining the different types of inclusions, including, social, political and economic inclusions to quantify the overall impact of inclusions on income inequality.

References

Alvarez – Cuadrado, F. and Long, N.V., 2012, "Envy and Inequality", Scandinavian Journal of Economics, 114(3), 949-973.

Beckfield, J. (2006). European integration and income inequality. American Sociological Review, 71(6), 964-985.

Bhagwati, J., & Hamada, K. (1974). The brain drain, international integration of markets for professionals and unemployment: a theoretical analysis. Journal of Development Economics, 1(1), 19-42.

Blundell, R. and Bond, S.R., 1998, “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models”, Journal of Econometrics, 87, 115-143.

Boje, T., Van Steenbergen, B., & Walby, S. (Eds.). (2013). European societies: fusion or fission?. Routledge.

Bouma, G., 2016, “Religious Diversity and the Challenge of Social Inclusion”, *Social Inclusion*, 4(2), 1-2.

Busemeyer, M. R., & Tober, T. (2015). European integration and the political economy of inequality. European Union Politics, 16(4), 536-557.

Deaton, A.S. and Paxson, C.H., 1997, “The Effects of Economic and Population Growth on National Saving and Inequality”, Demography, 34(1), 97-114.

Duesenberry, J., 1949, Income, Saving and the Theory of Consumer Behavior, Harvard University Press, Cambridge, MA.

Dynan, K., Skinner, J. and Zeldes, S., 2004, “Do the Rich Save More?”, Journal of Political Economy, 112(2), 397-444.

Esmaeilzadeh, R., Ahmad, N., & Naveed, A. (2018). Do Immigrants catch-up with the Natives in terms of earnings? Evidence from Individual level data of Canada. International Migration (forthcoming).

Frederiksen, A., & Poulsen, O. (2016). Income inequality: the consequences of skill - upgrading when firms have hierarchical organizational structures. Economic Inquiry, 54(2), 1224-1239.

Friedman, M., 1957, A Theory of the Consumption Function, Princeton University Press, Princeton.

Grubel, H. B., & Scott, A. D. (1966). The international flow of human capital. The American Economic Review, 56(1/2), 268-274.

Gruber, J. and Wise, D., 2001, “An International Perspective on Policies for an Aging Society”, NBER Working Papers W8103.

Guerin, B., 2013, Demography and Inequality: How Europe's Changing Population will Impact on Income Inequality. Santa Monica, CA: RAND Corporation, 2013. Downloaded from: http://www.rand.org/pubs/research_reports/RR183.html.

Gylfason, T. and Zoega, G., 2002, “Inequality and Economic Growth: Do Natural Resources Matter?”, CESifo Working Paper No. 712(5).

Haque, N. U., & Kim, S. J. (1995). “Human capital flight”: Impact of migration on income and growth. Staff Papers, 42(3), 577-607.

Husted, L., Skyt Nielsen, H., Rosholm, M., & Smith, N. (2001). Employment and wage assimilation of male first-generation immigrants in Denmark. International Journal of Manpower, 22(1/2), 39-71.

Kahanec, M., & Pytliková, M. (2017). The economic impact of east–west migration on the European Union. Empirica, 44(3), 407-434.

Kosonen, P. (1995). European welfare state models: Converging trends. International Journal of Sociology, 25(1), 81-110.

Lam, D. and Levison, D., 1992, “Age, Experience, and Schooling: Decomposing Earnings Inequality in the United States and Brazil”, Sociological Inquiry, 62(2), 218-245.

Mayer, T., 1966, “The Prosperity to Consume Permanent Income”, American Economic Review, 56, 1158-1177.

Mayer, T., 1972, Permanent Income, Wealth, and Consumption, A Critique of the Permanent Income Theory, the Lif-Cycle Hypthesis, and Related Theories, Berkley University Press, California.

Mitchell, A., & Shillington, E. R. (2002). Poverty, inequality and social inclusion. Toronto, Ontario: Laidlaw Foundation.

Moses, J. W. (1995). The social democratic predicament in the emerging European Union: A capital dilemma1. Journal of European Public Policy, 2(3), 407-426.

Naveed, A., Javakhishvili-Larsen, N., & Schmidt, T. D. (2017). Labour mobility and local employment: building a local employment base from labour mobility?. Regional Studies, 51(11), 1622-1634.

Naveed, A., & Wang, C. (2018). Can religion explain cross-country differences in inequality? A global perspective. Social Choice and Welfare, 50(3), 481-518.

Nembua, C.C., 2007, “The Multi-Decomposition of the Hirschman-Herfindahl Index: Measuring Household Inequality in Cameroon, 1996-2001”, *Applied Economic Letters*, 14(1), 27-34.

Nielsen, H. S., Rosholm, M., Smith, N., & Husted, L. (2003). The school-to-work transition of 2nd generation immigrants in Denmark. Journal of Population Economics, 16(4), 755-786.

Oxoby, R. (2009). Understanding social inclusion, social cohesion, and social capital. International Journal of Social Economics, 36(12), 1133-1152.

Razin, A., Sadka, E. and Swagel, P., 2002, “The Aging Population and the Size of the Welfare State”, Journal of Political Economy, 110(4), 900-918.

Rhoades, S.A., 1995, “Market Share Inequality, the HHI, and other Measures of the Firm-Composition of a Market”, *Review of Industrial Organization*, 10(6), 657-674.

Schultz, T.P., 1982, “Family Composition and Income Inequality”, Pp. 137-151 in Income Therborn, G. (1999). ’Europe’as issues of sociology. European societies: fusion or fission, 19-29.

Therborn, G. (1999). ’Europe’as issues of sociology. European societies: fusion or fission, 19-29.

Eurostat (2018) www.ec.europa.eu, Accessed on April, 2018.

Walby, S. (1999). The new regulatory state: the social powers of the European Union. The British journal of sociology, 50(1), 118-140.

Appendix

A1: Correlation Table for Core Variables

	$gini$	$ginireal$	$top10$	$top20$	$herf$	age	sav	$arable$	SI_{EU}	SI_{nonEU}	$SI_{foreign}$
$gini$	1										
$ginireal$	0.9433	1									
$top10$	0.9260	0.8933	1								
$top20$	0.9867	0.9407	0.9386	1							
$herf$	-0.2361	-0.2263	-0.1904	-0.1748	1						
age	0.1446	0.0328	0.1348	0.1045	-0.6116	1					
sav	-0.2253	-0.2584	-0.2797	-0.2323	0.2253	0.1923	1				
$arable$	-0.0158	-0.0785	0.0196	-0.0308	0.0925	-0.0795	-0.2659	1			
SI_{EU}	-0.2895	-0.3476	-0.2103	-0.2383	0.2243	0.1236	0.2187	-0.1384	1		
SI_{nonEU}	-0.2675	-0.2878	-0.2267	-0.2215	0.1936	-0.0248	0.1487	-0.1183	0.8077	1	
$SI_{foreign}$	-0.3417	-0.4010	-0.2922	-0.3045	0.2068	0.0634	0.1928	-0.0881	0.9076	0.9202	1

Notes: see Table 1 for variable definitions.

A2: Countries in the Regressions (Core Sample)

Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom

A3: Data Sources

Social Inclusion Variables ($SI_{EU}, SI_{nonEU}, SI_{foreign}$) from Eurostat. **Control variables** ($age, sav, arable$) from World Bank data's World Development Indicators (WDI). **Inequality variables** ($gini, top10, top20$) from World Bank data's World Development Indicators (WDI).

Table A1: Social Inclusions sub-indicators and short description

Indicators	Short Description
Material Deprivation rate by sex.	The percentage of population with an enforced lack of at least three out of nine material deprivation items in the 'economic strain and durables' dimension.
Living conditions: Housing cost overburden rate	The percentage of the population living in a household where the total housing costs (net of housing allowances) represent more than 40% of the total disposable household income (net of housing allowances)
Living conditions: Overcrowding rate	The percentage of the population living in an overcrowded household (excluding the single-person households). A person is considered as living in an overcrowded household if the household does not have at its disposal a minimum of rooms equal to: - one room for the household; - one room by couple in the household; - one room for each single person aged 18 and more; - one room by pair of single people of the same sex between 12 and 17 years of age; - one room for each single person between 12 and 17 years of age and not included in the previous category; - one room by pair of children under 12 years of age.
Living conditions: tenure status (tenant versus owner)	Distribution of population by broad group of citizenship and tenure status (owner versus tenant)

Source: Eurostat database (www.ec.europa.eu)