Working Paper Series

Is There a Relationship Between Income Inequality and Credit Cycles?

Tuomas Malinen

ECINEQ WP 2013 – 292

ECINEQ 2013 – 292
March 2013

www.ecineq.org

Is There a Relationship Between Income Inequality and Credit Cycles?*

Tuomas Malinen†
University of Helsinki and HECER

Abstract
Recent studies by Atkinson (2011); Rajan (2010); Kumhof and Ranciére (2010); Bordo and Meissner (2013) have assessed the relationship between income inequality and financial stability. Bordo and Meissner found that changes in income inequality do not have an effect on the growth of credit. We extend their study by assessing the relationship between levels of income inequality and leverage. We find that the relationship between inequality and credit is long-run, i.e. trending, in nature and that removing this relation with first differencing will lead to biased inference. In conclusion we find that income inequality is associated with increased leverage in the economy.

Keywords: top 1% income share, bank loans, unit root, cointegration
JEL Classification Codes: C23, D31, G21

* Author wishes to thank Henri Nyberg, Matthijs Lof, Pentti Saikkonen and Jari Hännikäinen for comments. Author gratefully acknowledges financial support from the OP-Pohjola Group's Research Foundation.
† **Contact details**: tuomas.malinen@helsinki.fi

1 Introduction

The economic effects of income inequality have been under intensive study within the last two decades, the relationship between income inequality and economic growth being the most studied subject (e.g. Barro 2000; Banerjee and Duflo 2003; Castelló-Climent 2010; Forbes 2000; Galor and Moav 2004; Malinen 2012; Persson and Tabellini 1994). After the financial crisis of 2007-2009, the relation between inequality and financial stability has also become under scrutiny. Rajan (2010) argues that rising inequality in developed economies caused redistribution in the form of subsidized housing finance, which led to a housing boom and later to a crash with known consequences. Kumhof and Ranciére (2010) argue that inequality raises leverage in middle-income and poor households as a result of consumption smoothing. When the income concentrates on the high income households, middle-income and poor households sustain their level of consumption by borrowing against their future incomes. If there is no recovery in the real incomes of the poor and middle income households, leverage among them will keep on rising, which will eventually lead to loan defaults thus increasing the probability of a financial crisis.

In a recent article, Bordo and Meissner (2013) set to test the hypothesis that inequality increases leverage using data on top income shares and the ratio of bank loans. They estimate the effect of change in income inequality on the change in the ratio of bank loans and find "very little evidence linking credit booms and financial crises to rising inequality". The result of Bordo and Meissner coincides with that of Atkinson (2011) who finds that there seems to be an ambiguous causal relation of income inequality on economic crises. However, as Atkinson points out, both his and Bordo and Meissner's result applies only to the relation between financial stability and changes in income inequality. Atkinson (2011, p. 35) concludes that "we have not investigated whether inequality level was relatively higher before identified macroeconomic shocks. Therefore, the level hypothesis cannot be ruled out at this stage."

Atkinson (2011) refers to the hypothesis put forth by Stiglitz (2009) and formalized by Kumhof andRanciére (2010), which states that during stagnating real incomes, poorer households borrow to maintain their rising standard of living. This creates a *trending* relation between income inequality and credit in the economy. As real income keeps on stagnating, credit acquired by lower income households keeps on growing and this trend eventually leads to defaults and to stress among financial institutions. First differencing removes this trend and focuses the analysis on the short term effects of inequality on credit. If the relationship between inequality and credit is long-run, i.e. trending, in nature, using first differenced variables may give biased information on the effect of inequality on leverage.

The analysis of this possible *long-run* relationship is complicated by the fact that bank loans tend to grow over time, whereas the generally used measures of income inequality, like the top 1% income share, are bounded from above. This creates a problem, because it is not possible for something that is not trending to have a long-run or an *equilibrium* relation with something that is upward trending, in the first two moments at least. There are two ways around this problem: the trending series can be detrended or it can be bounded using some suitable transformation. Detrending of the series is problematic, as it would remove the very thing under interest, i.e. the trend. This makes "bounding" of the series a preferable method, and there is a natural candidate by which the series can be transformed. The top 1% income share, used by Bordo and Meissner (2013), measures the share of national income concentrated on the hands of the highest percentile of income earners. As GDP is, in practice, the national income of a country, the share can be presented as $\frac{\text{income of the top 1\%}}{\text{GDP}}$. Therefore, it would be natural to convert bank loans the same way, i.e. $\frac{\text{bank loans}}{\text{GDP}}$. This transformation would make the measures comparable, as both would be expressed as a percentage of GDP, without removing the possible long-run relation that may exist between inequality and credit. In the theoretical model by Kumhof andRanciére (2010), leverage is modeled as workers debt-to-income ratio. Thus, credit-to-GDP ratio is also a more accurate statistical approximation of the measure of leverage used by Kumhof and Ranciére than the level or the first difference of credit.

In this article, we test and estimate the relationship of income inequality and credit as ratios to GDP. We use data on the income share of top 1% income earners and bank loans on eight developed economies. Results indicate that both the top 1% income share and the share of credit to real GDP follow an unit root process. The two series are also found to be cointegrated of order one implying that there is a long-run *steady-state* relation between them. The long-run elasticity of share of bank loans with respect to income inequality is estimated with panel DSUR and it is found to be positive.

The rest of the paper is organized as follows. Section 2 presents the data and gives the results of panel unit root tests. Cointegration test and estimation results are presented in section 3, and section 4 concludes.

2 Data and unit root tests

We use the top 1% income share of the population to proxy the income inequality as Bordo and Meissner (2013). Leigh (2007) has demonstrated that the top 1% income share series have a high correlation with other measures of income inequality, like the Gini index. The data on top income share is obtained from the World Top Income Database (Atkinson *et al.* 2011). The data on bank loans, real GDP per capita, investment as a share of GDP, short-term interest rates, and broad money (M2) as a share of GDP is obtained from the dataset of Schularik and Taylor (2012). The data on real GDP is taken from the Maddison dataset of the Groningen Growth and Development Centre.

Leverage is modeled as a debt to real income ratio in the theoretical model by Kumhof and Ranciére (2010). This ratio is also behind the level hypothesis introduced by Stiglitz (2009), where stagnating real incomes cause middle-income and poor households to borrow in nominal terms. Thus, to test the hypotheses by Kumhof and Ranciere and Stiglitz we use bank loans to real GDP (RGDP) as our dependent variable.

Due to limitations of the data on top 1% income share, we are able to construct a balanced panel on eight countries. The baseline dataset spans from 1959 to 2008, whereas the dataset including short-term interest rate spans from 1972 to 2008. Figure 1 presents the time series of the mean of the share of credit to real GDP and the mean of the top 1% income share in our data. Figure shows that during the period of 1959-1980 the share of income of the top 1% decreased, but at the same time the share of bank loans increased, although only marginally. After 1980 the share of income earned by the top 1% and the share of bank loans to real GDP grew at a very similar pace. This latter period gives some evidence in favor of the level hypothesis. The best way to analyze the possible relation between the two variable is to test are the different trend processes driven by the same factors.

The data on bank loans is extremely heterogeneous, as described by Schularik and Taylor (2012). Credit, money and banking institutions differ profoundly across countries and in some cases historical data on credit covers only commer-

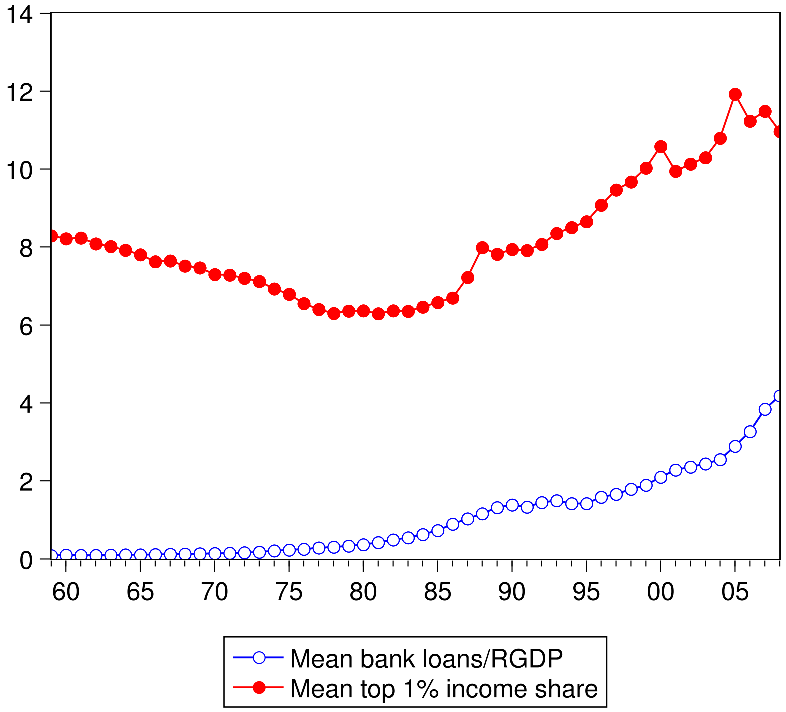

Figure 1. Means of the share of bank loans to real GDP and the top 1% income share. Sources: Atkinson *et al.* (2011); Schularik and Taylor (2012).

cial banks. As Schularik and Taylor, we tackle the issue of heterogeneity by using country-related constants. In addition, we log-linearize the share of bank loans to real GDP to diminish the cross-sectional variation in the series.[1] There are few yearly observations missing from the top 1% income share data, which we replace by averages of the values preceding and following the missing observation.

We start by testing the time series properties of the data. We use two sets of panel unit root tests to test for the possible unit roots. The first two are the so

[1] As the ratio of credit to real GDP is a percent variable, the log-linearization is done by transformation:

$$ln(x) = ln\left(\frac{1+x}{100}\right).$$

called first-generation tests, by Im et al. (2003) (IPS) and the Fisher type ADF test by Maddala and Wu (1999). These tests assume that there is no cross-sectional correlation between the units of the panel. The second generation panel unit root tests by Pesaran (2007) and Phillips and Sul (2003) allow for cross-sectional correlation within the panel. A more detailed explanation on the used test is provided in the Appendix I. Table 1 presents the results of panel unit root tests for the six included variables. According to results presented in table 1, all panel unit root

Table 1: Panel unit root tests

variable	IPS	ADF	PS	Pesaran
ln(credit/RGDP)	2.973	12.875	9.101	5.370
	(0.998)	(0.682)	(0.999)	(0.999)
top 1%	3.075	3.631	8.470	-3.891
	(0.999)	(0.997)	(0.863)	(<.001)
investments/GDP	-3.077	36.646	14.978	-1.792
	(0.001)	(0.002)	(0.380)	(0.037)
M2/GDP	3.543	8.127	25.954	-0.629
	(0.999)	(0.945)	(0.026)	(0.265)
ln(real GDP per capita)	-1.947	27.268	11.772	-0.833
	(0.0258)	(0.0386)	(0.625)	(0.203)
short term interest rate*	-2.880	34.090	40.484	-.866
	(0.002)	(0.005)	(0.002)	(0.002)

In unit root tests, the tested equation is: $\triangle y_{it} = \rho_i y_{i,t-1} + \delta_i + \eta_i t + \theta_t + \epsilon_{it}$. The p-values of the test statistics are presented in parentheses. All other test are done with the eight country panel ranging from 1959 to 2008, except tests for short term interest rates are done with a panel with yearly observations from 1972 to 2008.

tests find the share of credit to real GDP to be an unit root process, i.e. tests cannot reject the null hypothesis of unit root. Three out of four panel unit root tests find the top 1% income share and the share of broad money to GDP to be unit root processes. Two out of the four tests find the real GDP per capita to be an unit root process and one out of the four tests find the share of investments to GDP to be an unit root process. According to all tests, the short-term interest rate is a trend-stationary process.[2]

3 Cointegration test and estimations

3.1 Panel cointegration testing

According to unit root tests presented in table 1, stochastic trends would drive the time series of the top 1% income share and the share of credit to real GDP. Next we test if the stochastic trends are linear combinations of one and another, i.e. we test are the series cointegrated. To this end, we use two panel cointegration tests, where the first one is the cointegration test by Pedroni (2004) and the second is the cointegration test by Banerjee and Carrion-i-Silvestre (2011) (from now on BC). The biggest difference between these tests is that while Pedroni's test assumes uncorrelated residual structure, BC's test allows for cross-sectional correlation through common factors and it also controls for possible structural breaks in the cointegration relation. Appendix III gives more detailed description of the used tests.

The model for testing for cointegration between inequality and credit is:

$$ln(credit/RGDP)_{it} = \alpha_i + \gamma_i top1\%_{it} + \epsilon_{it}, \quad (1)$$

where the level of bank loans are explained by the level of inequality, and $(1, -\gamma_i)$ is the country-specific cointegration vector between bank loans and the top 1% income share. We include individual constants due to heterogeneity of the data on bank loans discussed in the previous section. Results of panel cointegration tests based on the model (1) are presented in table 2.[3] 15 out of the 19 test statistics

[2]According to all second generation panel unit root tests, the first-differences of log of credit to real GDP, top 1% income share, investments to GDP, money to GDP and GDP per capita are trend-stationary. Results are available upon request.

[3]The Pedroni's test was conducted with Eviews 6 and B&C's test was done with Gauss. We are grateful to Carrion-i-Silvestre for providing the program code.

Table 2: Panel cointegration test statistics for ln(credit/RGDP) and top 1% income share

Pedroni tests		
Within-dimension	Constants	Constants
panel v-statistic	-2.285	-1.835
	(0.029)	(0.074)
panel ρ-statistic	2.274	2.264
	(0.030)	(0.031)
panel PP-statistic	3.070	2.948
	(0.004)	(0.005)
panel ADF-statistic	3.205	2.786
	(0.002)	(0.008)
Between-dimension	Constants	
group ρ-statistic	2.999	
	(0.004)	
group PP-statistic	3.861	
	(<.001)	
group ADF-statistic	3.777	
	(<.001)	
BC tests		
	Constants	Trends
$Z_{\hat{t}_{NT}}(\hat{\lambda})$	0.864	-1.302
	(0.801)	(0.096)
$Z_{\hat{\rho}_{NT}}(\hat{\lambda})$	-3.354	-6.026
	(<.001)	(<.001)
	Constants, level shift	Constants, ci. vector shift
$Z_{\hat{t}_{NT}}(\hat{\lambda})$	-0.988	-2.451
	(0.162)	(0.007)
$Z_{\hat{\rho}_{NT}}(\hat{\lambda})$	-5.543	-11.55
	(<.001)	(<.001)

The null hypothesis is that the variables are not cointegrated. In the test by Pedroni, lag length were determined with Schwarz information criterion. *Constants* states that individual constants were used in the test, and *trends* that individual constants and trends were used in the test. Tests with level and cointegration vector shifts allow for structural breaks to occur in the country-specific cointegration relations.

presented in table 2 find that the series of top 1% income share and credit to real GDP are cointegrated of order one at the 5% level. Moreover, 13 out of the 15 test statistics that do not allow for level of cointegration vector shifts find the top 1% income share and share of credit to real GDP to be cointegrated at the 5% level. Therefore, we conclude that top 1% income share and credit to real GDP seem to be cointegrated, i.e. that the two series have a long-run *steady-state* relation.

3.2 Estimations

First differencing of cointegrated variables removes stochastic trends and it eliminates the long-run dependency between the variables. What remains is a short-run relation, which may or may not exist. To test for this, we first estimate variables in first differences. More precisely, we estimate a model:

$$\Delta ln(credit/RGDP)_{it} = \alpha_i + \beta_1 \Delta top_{i,t-1} + \beta_2 \Delta investments/GDP_{i,t-1} + \\ + \beta_3 \Delta M2/GDP_{i,t-1} + + \beta_4 \Delta ln(RGDP_{i,t-1} + \beta_5 \Delta stir_{i,t-1} + u_{it}, \quad (2)$$

where α_i are individual constants and u_{it} is the idiosyncratic error term. Explanatory variables are lagged with one period to control for the possible endogeneity of regressors.

Table 3 presents the results. Results of table 3 indicate that income inequality would not have a statistically significant short-run effect on nominal credit. In the last estimation, none of the parameter estimates of the explanatory variables is statistically significant at the 5% level.

The picture somewhat changes when the levels of credit to real GDP and top 1% income share are used. We use panel DSUR (dynamic seemingly unrelated regressions) estimator by Mark *et al.* (2005) to estimate the cointegration coefficient

Table 3: Regression results using first differenced variables

Dependent variable: Δln(credit/RGDP)		
	FE-OLS	FE-OLS
Δtop 1_{t-1}	-0.0002	-0.0019
	(0.0028)	(0.0029)
Δln(real GDP per capita)$_{t-1}$	-0.0022	0.1611
	(0.0693)	(0.1105)
Δinvestments/GDP$_{t-1}$	0.4223**	0.3944
	(0.1504)	(0.2479)
ΔM2/GDP $_{t-1}$	0.0652*	0.0255
	(0.0292)	(0.0253)
Δshort term interest rate$_{t-1}$	-	0.0799
		(0.1332)
countries	8	8
years	1959-2008	1972-2008
observations	384	280

Estimations are done with country fixed-effects. White heteroskedasticity-consistent standard errors are presented in parentheses.

of top 1% income share using a model:

$$ln(credit/RGDP)_{it} = \alpha_i + \gamma_1' top1_{it} + \beta_\mathbf{p} \mathbf{X_{it}} + \theta_t + u_{it}, \quad (3)$$

where α_i are individual constant, θ_t is the common time effect, $(1, -\gamma_1')$ is the cointegrating vector between bank loans and top 1% income share, $\mathbf{X_{it}}$ is the matrix of additional explanatory variables, and u_{it} is the idiosyncratic error. As the panel DSUR does not allow for cointegration between explanatory variables, all the other explanatory variables, besides top 1% income and short term interest rates, are differenced.[4] The panel DSUR estimator controls for the possible endogeneity of explanatory variables by including the leads and lags of the first differences of the explanatory variables in the estimated equation. More information about the panel DSUR can be found in the Appendix III.

[4]There is no need to take the first difference of the short term interest rate, as all the panel unit root tests presented in table 1 found the series to be trend-stationary.

Table 4 presents the results panel DSUR estimations on equation (3) using the dataset spanning from 1959 to 2008.[5] First differences of the GDP per capita and shares of M2 and investment to GDP are included as additional explanatory variables. According to the results presented in table 4, the cointegration coefficient

Table 4: DSUR estimates, 1959-2008

Dependent variable: ln(credit/RGDP)			
top 1%	0.051***	0.073***	0.017***
	(0.0037)	(0.0050)	(0.0026)
Δln(real GDP per capita)	-	-0.012*	0.071
		(0.0050)	(0.0499)
Δmoney/GDP	-	-	-0.003
			(0.0259)
Δinvestment/GDP	-	-	-0.075
			(0.1218)
countries	8	8	8
years	1959-2008	1959-2008	1960-2008
observations	400	400	392

* = p<.05, ** = p<.01, *** = p<.001. Standard errors are presented in parentheses. All DSUR estimations include individual constants and common time effects. First and second leads and lags of the first differences are used as instruments for the explanatory variables.

of top 1% income share is positive and highly statistically significant. The value of the cointegrating coefficient varies from around 0.05 to around 0.07. In the last estimation none of the parameter estimates of the first differenced explanatory variables are statistically significant at the 5% level.

Table 5 presents the results of panel DSUR estimations on equation 3 using the dataset spanning from 1972 to 2008. In addition to first differences of the GDP per capita, M2 to GDP and investment share to GDP, short-term interest rate in levels is included as an explanatory variable.[6] According to the results of

[5] DSUR estimations were done with Gauss. We are grateful to Donggyu Sul for providing the program code on his homepage.

[6] DSUR estimations were done with Gauss. We are grateful to Donggyu Sul for providing the

Table 5: DSUR estimates, 1972-2008

Dependent variable: ln(credit/RGDP)				
top 1%	0.0558***	0.0563***	0.0054***	0.019***
	(0.0039)	(0.0031)	(0.0003)	(0.00573)
Δln(real GDP per capita)	-	-0.649	-0.0381	0.0153
		(0.3511)	(0.0532)	(0.1297)
Δmoney/GDP	-	-	-0.123***	-0.1023
			(0.0345)	(0.0663)
Δinvestment/GDP	-	-	0.0245	-0.112*
			(0.0914)	(0.0396)
short term interest rate	-	-	-	-0.152***
				(0.0329)
countries	8	8	8	8
years	1972-2008	1972-2008	1972-2008	1972-2008
observations	296	296	296	296

* = p<.05, ** = p<.01, *** = p<.001. Standard errors are presented in parentheses. All DSUR estimations include individual constants and common time effects. First and second leads and lags of the first differences are used as instruments for the explanatory variables.

table 5, the cointegrating coefficient of top 1% income share is positive and highly statistically significant. The first differences of money share to GDP and the short-term interest rate have statistically significant negative parameter estimates. The negative effect of short-term interest rate to ratio of bank loans to real GDP is expected, as higher interest rates make borrowing more expensive. The negative parameter estimate of the share of M2 to GDP, on the other hand, is likely to result from reverse causality. That is, as bank loans increase, money held in deposit accounts (etc.) decreases, which will decrease the broad money in circulation.

program code on his homepage.

4 Conclusion

Income inequality is a trending variable. In the absence of wars or other major catastrophes changes in inequality are gradual, manifesting during a course of several years, even decades (Atkinson *et al.* 2011). Same applies to credit. The share of credit to GDP has been gradually growing within the last fifty years or so. Kumhof and Ranciére (2010) argue that there is a long-run relationship between income inequality and share of credit to income, where income inequality will to lead to increasing leverage in the economy.

In this study, we have tested the existence of such a long-run relationship. According to the results, there is a long-run *steady-state* relationship between income inequality and leverage in developed economies. The long-run elasticity of leverage with respect to income inequality was found to be positive. This indicates that income inequality increases leverage in the economy in accordance with the theories by Kumhof and Ranciére (2010), Rajan (2010) and Stiglitz (2009).

References

Albanesi S. (2007). Inflation and inequality. *Journal of Monetary Economics*, 54(4), 1088–1114.

Atkinson AB. (2011). Economic crises and inequality. United Nations, Human Development Report.

Atkinson AB, Piketty T, and E. Saez (2011). The World Top Incomes Database. http://g-mond.parisschoolofeconomics.eu/topincomes. Accessed 7th of June 2011.

Banerjee A, Carrion-i-Silvestre J. (2011) Cointegration in panel data with breaks and cross-section dependence. University of Birmingham Deparment of Economics discussion paper no. 11-25.

Banerjee A, Duflo E. (2003). Inequality and growth: what can the data say? *Journal of Economic Growth*, 8(3):267-299.

Barro R. (2000). Inequality and growth in a panel of countries. *Journal of Economic Growth* 5(1):5-32

Bordo MD, Meissner CM. (2013). Does inequality lead to a financial crisis? *Journal of International Money and Finance*, (forthcoming).

Castelló-Climent A. (2010). Inequality and growth in advanced economies: an empirical investigation. *Journal of Economic Inequality*, 8(3): 293–321.

Forbes K. (2000). A reassessment of the relationship between inequality and growth. *American Economic Review*, 90(4):869-887.

Galor O, Moav O (2004) From Physical to human capital accumulation: inequality and the process of development. *Review Economic Studies* 71(4):1001-1026

Im K, Pesaran H, Shin Y (2003) Testing for unit roots in heteregenous panels. *Journal of Econometrics*, 115(1):53-74

Kumhof M,Ranciére R. (2010). Inequality, leverage and crises. IMF working paper 10/268.

Leigh, A. (2007). How closely do top income shares track other measures of inequality? *Economic Journal*, 117, 619-33.

Levin A, Lin CF, Chu CS (2002) Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108(1):1-24

Maddala G, Wu S .(1999) A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Econics and Statistics*, 61(special issue):631-652

Malinen T. (2012). Estimating the long-run relationship between income inequality and economic development. *Empirical Economics*, 42(1), 209-33.

Mark N, Ogaki M, Sul D. (2005). Dynamic seemingly unrelated cointegrating regressions. *Review of Economic Studies*, 72 (3), 797-820,

Persson T, Tabellini G. (1994). Is inequality harmful for growth. *American Economic Review*, 84(3): 600–621.

Pesaran H. (2007). A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics*, 22(2), 265-315

Pedroni P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to PPP hypothesis. *Econometric Theory*, 20 (3), 597-625

Pedroni P. (1999). Critical values for cointegration tests in heterogenous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, 61(special issue): 653-670.

Phillips P.C.B, Sul D. (2003). Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence. *Econometrics Journal, 6*, 217–259.

Rajan RG. (2010). *Fault Lines: How hidden fractures still threaten the world economy*. Princeton: Princeton University Press.

Roine, J. and D. Waldenström (2011). Common trends and shocks to top incomes: a structural breaks approach. *The Review of Economics and Statistics*, 93(3), 832-46.

Schularik M, Taylor AM. (2012). Credit booms gone bust: monetary policy, leverage cycles and financial crises, 1870-2008. *American Economic Review*, 102(2): 1029–1061.

Stiglitz JE (2009). Joseph Stiglitz and why inequality is at the root of the recession. *Next Left website*, posted 9 of January 2009.

APPENDIX I: Panel unit root tests

All the used tests allow for individual unit root processes. That is, they allow the coefficient of unit root to differ across countries.

The traditional panel unit root tests, are based on the following regression:

$$\triangle y_{it} = \rho_i y_{i,t-1} + \eta_i t + \alpha_i + \theta_t + \epsilon_{it}, \tag{4}$$

where α_i are individual constants, $\eta_i t$ are individual time trends, and θ_t are the common time effects. The tests rely on the assumption that $E[\epsilon_{it}\epsilon_{js}] = 0 \; \forall \; t, s$ and $i \neq j$, which is required for calculating common time effects. Thus, if the different series are correlated, the last assumption is violated.

The second generation test are based on the regression

$$\triangle y_{it} = \rho y_{i,t-1} + \eta_i t + \alpha_i + \delta_i \theta_t + \epsilon_{it}, \tag{5}$$

where α_i are the individual constants, $\eta_i t$ are the individual time trends, and θ_t is the common time effect, whose coefficients, δ_i, are assumed to be non-stochastic, measure the impact of the common time effects of series i, and ϵ_{it} is assumed to be normally distributed with mean zero and covariance of σ^2 and independent of ϵ_{js} and θ_s for all $i \neq j$ and s, t. Cross-sectional dependence is allowed through the common time effect, which generates the correlation between cross-sectional units. The matrix δ_i gives the non-random factor loading coefficients that determine the extent of the cross-sectional correlation.

The null hypothesis in all tests is that $\rho_i = 0 \; \forall \; i$, i.e. that the process in $I(1)$ nonstationary. The alternative hypotheses are:

$$H_1 : \rho_i < 0, \quad i = 1, 2, ..., N_1, \quad \rho_i = 0, \quad i = N_1 + 1, N_1 + 2, ..., N. \tag{6}$$

For consistency of panel unit root tests it is also required that, under the alternative, the fraction of the individual processes that are stationary is non-zero, formally $lim_{N\rightarrow\infty}(N_1/N) = \gamma$, $0 < \gamma \leq 1$ (Im et al. 2003).

Appendix II: Panel cointegration tests

Panel cointegration test developed by Banerjee and Carrion-i-Silvestre (2011) is based on the normalized bias and the pseudo t-ratio test statistics by Pedroni (2004). The data generating process behind Pedroni's test statistics is given by:

$$
\begin{aligned}
y_{it} &= f_i(t) + x'_{it} + e_{it}, \\
\triangle x_{it} &= v_{it}, \qquad (7) \\
e_{it} &= \rho_i e_{i,t-1} + \epsilon_{it}\zeta_{it} = (\epsilon_{it}, v_{it})',
\end{aligned}
$$

where $f_i(t)$ includes member specific fixed effects and deterministic trends.

The data generating process is described as a partitioned vector $z'_{it} \equiv (y_{it}, x_{it})$ where the true process is generated as $z_{it} = z_{i,t-1} + \zeta_{it}$, $\zeta'_{it} = (\zeta^{y}_{it}\zeta^{X}_{it})$ (**?**). $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\zeta_{it}$ is assumed to converge to a vector Brownian motion with asymptotic covariance of Ω_i as $T \longrightarrow \infty$. The individual process is assumed to be *i.i.d.* so that $E[\zeta_{it}\zeta'_{js}] = 0$ $\forall s, t, i \neq j$.

Let $\hat{e}_{it}$ denote the estimated residuals of obtained from (7) and $\hat{\Omega}_i$ the consistent estimator of Ω_i. The two test statistics can now be defined as :

$$
\begin{aligned}
\tilde{Z}_{\hat{\rho}_{NT}-1} &\equiv \sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{e}^2_{i,t-1}\right)^{-1}\sum_{t=1}^{T}(\hat{e}_{i,t-1}\Delta\hat{e}_{it} - \hat{\lambda}_i), \\
\tilde{Z}^{*}_{\hat{t}_{NT}} &\equiv \sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{s}^{*2}_i\hat{e}^{*2}_{i,t-1}\right)^{-1/2}\sum_{t=1}^{T}(\hat{e}^{*}_{i,t-1}\Delta\hat{e}^{*}_{it}),
\end{aligned}
$$

where $\hat{\lambda}_i = 1/T\sum_{s=1}^{k_i}(1 - s/(k_i+1))\sum_{t=s+1}^{T}\hat{\mu}_{it}\hat{\mu}_{i,t-s}$, $\tilde{\sigma}^2_{NT} \equiv 1/N\sum_{i=1}^{N}\hat{L}^{-2}_{11i}\hat{\sigma}^2_i$, $\hat{s}^{*2}_i \equiv 1/t\sum_{t=1}^{T}\hat{\mu}^{*2}_{it}$, $\tilde{s}^{*2}_{NT} \equiv 1/N\sum_{i=1}^{N}\hat{s}^{*2}_i$, $\hat{L}^2_{11i} = 1/T\sum_{t=1}^{T}\hat{\vartheta}^2_{it} + 2/T\sum_{s=1}^{k_i}(1 - s/(k-i+1))\sum_{t=s+1}^{T}\hat{\vartheta}_i, \hat{\vartheta}_{i,t-s}$.

The residuals $\hat{\mu}_{it}, \hat{\mu}^*_{it}$ and $\hat{\vartheta}_{it}$ are attained from regressions: $\hat{e}_{it} = \hat{\gamma}\hat{e}_{i,t-1} + \hat{\mu}_{it}$, $\hat{e}_{it} = \hat{\gamma}_i\hat{e}_{i,t-1} + \sum_{k=1}^{K-i}\hat{\gamma}_{ik}\Delta\hat{e}_{i,t-k} + \hat{\mu}^*_{it}$, $\Delta y_{it} = \sum_{m=1}^{M}\hat{b}_{mi}\Delta x_{mi,t} = \hat{\vartheta}_{it}$. (Pedroni 1999, 2004)

The statistics pool the between dimension of the panel and they are constructed by computing the ratio of the corresponding conventional time series statistics and then by computing the standardized sum of the N time series of the panel. Pedroni (1999, 2004) shows that under the null of no cointegration the asymptotic distributions of the two statistics presented above converge to normal distributions with zero mean and variance of one as N and T sequentially converge to infinity.

Banerjee and Carrion-I-Silvestre (2006) extend the model by Pedroni (2004) to include common factors:

$$\begin{aligned} y_{i,t} &= f_i(t) + x'_{i,t} + u_{i,t}, \\ \triangle x_{i,t} &= v_{i,t}, \\ f_i(t) &= \mu_i + \beta_i t \\ u_{it} &= F'_t\pi_i + e_{it} \end{aligned} \tag{8}$$

where $e_{i,t} = \rho_i e_{i,t} + \epsilon_{i,t}$ and F'_t:s are the common factors which are used to account for the possible cross-sectional dependence.

APPENDIX III: Panel DSUR estimator

The data generation process in Mark *et al.* (2005) DSUR estimator is of the form

$$y_{it} = \alpha_i + \lambda_i t + \theta_t + \beta' x_{it} + u_{it}, \tag{9}$$

$$\triangle x_{it} = e_{it} \tag{10}$$

where there are n cointegrating regression each with T observations, $(1 - \beta')$ is the cointegration vector between y_{it} and x_{it}, and x_{it} and e_{it} are $k \times 1$ dimensional vectors. Panel DSUR eliminates the possible endogeneity between explanatory variables and the dependent variable by assuming that u_{it} is correlated at most with p_i leads and lags of $\triangle x_{it}$ (Mark *et al.* 2005). The possible endogeneity can be controlled by projecting u_{it} onto these p_i leads and lags:

$$u_{it} = \sum_{s=-p_i}^{p_i} \delta'_{i,s} \triangle x_{i,t-s} + u_{it}* = \delta'_i z_{it} + u^*_{it}. \tag{11}$$

The projection error u^*_{it} is orthogonal to all leads and lags of $\triangle x_{it}$ and the estimated equation becomes:

$$y_{it} = \alpha_i + \lambda_{it} + \theta_t + \beta' x_{it} + \delta_i z_{it} + u^*_{it}, \tag{12}$$

where $\delta'_i z_{it}$ is a vector of projection dimensions. Panel DSUR estimates a long-run covariance matrix that is used in estimation of equation (9). This makes panel DSUR more efficient when cross-sections are dependent. The efficiency of panel DSUR actually improves as the correlation between cross-sections increases. Asymptotics properties of the estimator are based on $T \longrightarrow \infty$ with N fixed.