Helsinki
Center
of
Economic
Research

Discussion Papers

From structural breaks to regime switching: the nonlinearity in the process of income inequality

Tuomas Malinen
University of Helsinki

and

Leena Kalliovirta
University of Helsinki

Discussion Paper No. 356
October 2012

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

HECER
Discussion Paper No. 356

From structural breaks to regime switching: the nonlinearity in the process of income inequality*

Abstract

We use top income data and Gaussian mixture autoregressive model to explain the dynamics of income inequality. Our results imply that the structural breaks in the top income series identified by Roine and Waldenström (2011) are actually shifts from one regime to another. We also find that these regimes have different characteristics: one has low mean and variance whereas the other has high mean and variance. These highly differing means and variances between regimes are also likely to explain the findings of a unit root in the series of bounded income inequality measures.

JEL Classification: C32, D30

Keywords: top 1% income share, regime switching, GMAR model.

Tuomas Malinen

Department of Political and Economic Studies
University of Helsinki
P.O. Box 17
FI-00014 University of Helsinki
FINLAND

e-mail: tuomas.malinen@helsinki.fi

Leena Kalliovirta

Department of Political and Economic Studies
University of Helsinki
P.O. Box 17
FI-00014 University of Helsinki
FINLAND

e-mail: leena.kalliovirta@helsinki.fi

* We thank Donald Andrews, Timothy Armstrong, Markku Lanne, Mika Meitz, Peter Phillips, Pentti Saikkonen, Rami Tabri, and seminar audience at universities of Helsinki and Yale for useful comments and suggestions. Tuomas gratefully acknowledges financial support from the OP-Pohjola Research foundation. Leena thanks the Academy of Finland and the Cowles Econometrics Program for financial support.

1 Introduction

Dynamics of income inequality have become a major interest especially in the developed world within the last decade or so. This has been due to the fact that in many developed economies income inequality has been rising again after contracting for nearly a century. In a recent paper, Roine and Waldenström (2011) analyze the process of income inequality using deterministic trends and structural breaks. Their analysis is based on the newly compiled data on top 1% income shares and it employs recently developed methods for estimating and testing for the structural breaks. They found that the trends in the top 1% income share series has several break points from which some are global or common to some country-groups and some are country-specific.

We extend the study of Roine and Waldenström (2011), hereafter RW, by analyzing the nonlinearities in the process of income inequality. The breaks in the trends identified by RW indicate, among other things, that countries have gone through different historical phases of income distribution. It is intuitive that some economic or political change may have caused these phases. RW identify the number and timing of the breaks, but they determine the number of breaks based on a "rule of thumb" instead of a statistical procedure. In contrast, we use appropriate residual-based diagnostics to determine the adequacy of the estimated models. We find that instead of deterministic trends and structural breaks the series contains regime shifts between different phases. These phases differ in levels and scales of variation.

Furthermore, the observed jumps in the series - whether breaks or regime shifts - could be the reason why many studies have been unable to reject the unit root hypothesis in the autoregressive models for different measures of income inequality (Jäntti and Jenkins 2010; Herzer and Vollmer 2012; Malinen 2012; Mocan 1999; Parker 2000).[1]

[1] This is a problematic result in empirical literature as series of commonly used measures of income inequality, like the Gini index and the top income share are bounded between 0 and 1, whereas unit root series has a time-increasing variance.

The models used by RW ignore the strong autocorrelation of the series, whereas a linear autoregressive model will be misspecified due to the observed jumps. However, our nonlinear model depicts simultaneously both the autocorrelation and the jumps in the series.

We analyze an updated version of the top 1% income share data ranging from the end of the 19th century to the beginning of the 21st century for six countries (USA, Canada, Australia, France, Finland, and Japan). To this data we employ a Gaussian mixture autoregressive (GMAR) model studied in Kalliovirta, Meitz, and Saikkonen (2012) to identify the different regimes and autoregressive dynamics in the series. We find that in all analyzed countries, the process of income inequality has consisted on at least two different regimes, where the shifts between regimes coincide with the structural breaks identified in RW. We also find that the regimes are not only characterized by different means or levels but also with different variances or scales of variation. In the two regime model, one regime represents more unevenly distributed income with large fluctuations in the income, whereas in the other regime, the income is more evenly distributed and the fluctuation is low. In our estimated GMAR models, the persistence of the series is adequately described by a simple autoregressive structure constrained to be the same in all regimes. The trend approach of RW is unable to model the series with such parsimony. For example, their model for the US series (Figure 6 in RW) uses at least 10 parameters whereas our model has only 6 parameters and we are able to model the variances too. Further, the residual diagnostics of our GMAR model imply that the other breaks found in RW for the US series are superfluous.

2 Data and method

The top 1% income share of population is used to proxy the income inequality. These shares are also the only aggregate measures of income inequality that currently contain enough observations for meaningful testing of the time series properties. Leigh (2007) has also 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).

We assume that the observed top 1% income share series y_t is generated by the following constrained version of the GMAR model

$$y_t = \sum_{m=1}^{M} s_{t,m}(\varphi_{m,0} + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \sigma_m \varepsilon_t),$$

where unobservable random variables $s_{t,m}$ indicate the regimes $m = 1,...,M$, $\varphi_{m,0}$, φ_1, φ_2, and $\sigma_m > 0$ are parameters to be estimated, and ε_t is an i.i.d. N(0,1). Random variables ε_t and $s_{t,m}$ are independent given $\{y_{t-j}, j > 0\}$, the history of the observed series y_t. For each t, exactly one of $s_{t,m}$ variables takes the value one and others are equal to zero. The conditional probabilities $P(s_{t,m} = 1|y_{t-j}, j > 0) = \alpha_{m,t}$ are time-dependent mixing weights. Also, the probability of the observation y_t being generated by the AR(2) model of the m:th regime, $\varphi_{m,0} + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \sigma_m \varepsilon_t$, is equal to $\alpha_{m,t}$. Thus, these weights have to satisfy $\sum_{m=1}^{M} \alpha_{m,t} = 1$ for all t. In particular, the mixing weights depend on the past observations, the parameters $\varphi_{m,0}$, φ_1, φ_2, and σ_m, and additional weight parameters $\alpha_m > 0$, $\sum_{m=1}^{M} \alpha_m = 1$, according to

$$\alpha_{m,t} = \frac{\alpha_m \mathsf{n}(\boldsymbol{y}_{t-1}; \mu_m, \boldsymbol{\Gamma}_m)}{\sum_{n=1}^{M} \alpha_n \mathsf{n}(\boldsymbol{y}_{t-1}; \mu_n, \boldsymbol{\Gamma}_n)},$$

where $\boldsymbol{y}_{t-1} = (y_{t-1}, y_{t-2})$, $\mu_m = \frac{\varphi_{m,0}}{1-\varphi_1-\varphi_2}$ and

$$\mathsf{n}(\boldsymbol{y}_{t-1}; \mu_m, \boldsymbol{\Gamma}_m) = \{2\pi\}^{-1/2} det(\boldsymbol{\Gamma}_m)^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y}_{t-1} - \mu_m \mathbf{1}_2)' \boldsymbol{\Gamma}_m^{-1} (\boldsymbol{y}_{t-1} - \mu_m \mathbf{1}_2)\right\}.$$

Symmetric, 2x2 Toeplitz matrix $\boldsymbol{\Gamma}_m$ has diagonal elements $\gamma_{m,0}$ and off-diagonal $\gamma_{m,1}$, and solves the Yule-Walker equations $\boldsymbol{\Gamma}_m \boldsymbol{\varphi} = \boldsymbol{\gamma}_m$, where $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$ and $\boldsymbol{\gamma}_{\boldsymbol{m}} = (\gamma_{m,1}, \gamma_{m,2})$. If $\varphi_2 = 0$, then matrix $\boldsymbol{\Gamma}_m$ simplifies to $\frac{\sigma_m^2}{1-\varphi_1^2}$. Note that the restriction $\sum_{m=1}^{M} \alpha_m = 1$ reduces the number of free weight parameters to $M - 1$. Thus, if $M = 2$ we have to estimate only one weight parameter α_1. Note also that the structure of the mixing weights yields us an alternative parameterization for the GMAR model using parameters μ_m, $\boldsymbol{\varphi}$, $\gamma_{m,0}$, and α_m. We use this parameterization in our analysis and, as suggested in Kalliovirta, Meitz, and Saikkonen (2012), estimate these parameters using maximum likelihood.

The GMAR model has several advantageous properties. First, the above GMAR model is stationary, if the usual stationarity condition of the conventional linear AR(2) model is fulfilled. Second, the stationary distribution of the GMAR model is known to be $\sum_{m=1}^{M} \alpha_m \mathsf{n}(\boldsymbol{y}_{t-1}; \mu_m, \boldsymbol{\Gamma}_m)$. Thus, we are able to make direct comparisons to the unconditional moments of the original observations (as in Table 1). This would be unavailable if any other nonlinear model had been used, because the transition from the conditional to the unconditional distribution could not be achieved. Thirdly, the GMAR model is very parsimonious, a considerable advantage when only yearly data is available. To learn more about the GMAR model and its competing nonlinear alternatives, see Kalliovirta, Meitz, and Saikkonen (2012).

3 Results

As a starting point for the analysis of each series, we estimated linear AR models. Residual diagnostics (not reported) rejected these models due to non-normality and conditional heteroskedasticity. Table 1 presents properties of the original series and the estimation results for GMAR models that pass the quantile residual diagnostics of Kalliovirta (2012). Clearly, the original series are persistent in all six countries, and the variances are also highly fluctuating from the around 25 in Japan to around 5 in Australia. In the GMAR models, there are two regimes in top 1% incomes series in all countries except Australia, where three regimes are found. The series of France and Japan require two lags in the GMAR model, whereas one lag is enough for all the other countries. When the different constants and variances of regimes are taken

Table 1: Estimation results on the top 1% income share

	USA	Canada	France	Finland	Japan	Australia
Original data						
First autocorrelation	0.956	0.957	0.966	0.954	0.975	0.911
mean	12.62	11.28	10.90	8.75	12.30	8.14
variance	14.12	9.63	15.24	10.39	25.22	5.24
GMAR model parameters						
autocorrelation (φ_1)	0.918	0.886	1.076	0.931	1.316	0.915
	(0.03)	(0.05)	(0.11)	(0.03)	(0.08)	(0.04)
autocorrelation (φ_2)			-0.128		-0.447	
			(0.11)		(0.09)	
mean 1 (μ_1)	14.8	15.0	15.9	8.3	16.6	11.4
	(1.5)	(2.1)	(2.4)	(1.4)	(1.0)	(8.4)
mean 2 (μ_2)	8.2	9.2	8.5	4.7	7.8	8.0
	(0.3)	(0.4)	(0.4)	(0.7)	(0.2)	(0.6)
mean 3 (μ_3)						4.6
						(0.2)
variance 1 (γ_1)	6.1	7.0	11	3.6	8.9	53
	(2.6)	(3.3)	(8.8)	(1.5)	(2.8)	(36)
variance 2 (γ_2)	0.2	0.6	0.5	0.6	0.4	2.2
	(0.1)	(0.2)	(0.4)	(0.4)	(0.1)	(0.7)
variance 3 (γ_3)						0.04
						(0.02)
α_1	0.66	0.21	0.55	0.81	0.81	0.08
	(0.5)	(0.3)	(0.4)	(0.3)	(0.3)	(0.1)
α_2						0.61
						(0.34)

Standard errors (in parentheses) are calculated using the Hessian. More details on the estimated models and residual diagnostics are available upon request.

into account, autocorrelation in the top 1% income series diminishes quite clearly in all other countries, except in Australia.

The regimes of GMAR models seem to be marked with quite clear and similar characteristics in all countries. In one regime, the mean and variance of the top 1% income series are clearly higher, whereas in the other regime both are considerably lower. So, it seems that, at least in these countries, income inequality has consisted on two notably different regimes. First one is a low income inequality, low income fluctuations regime and the second is a high income inequality, high income fluctuations regime. The Australian series has three regimes. The variance of the first regime that corresponds to a short period around 1951 is very high and the shortness of the period explains why the variance is so inaccurately estimated. However, the other two regimes with smaller means and variances show the same familiar characteristics found in other countries.

Thus, the dynamics of income inequality seem to follow a joined path, and we can infer that income equality creates stability in the incomes of the top 1% income earners, while income inequality creates fluctuations in the earned incomes of the same group. Further, our analysis points out that the evolution of the top 1% income series cannot be modeled adequately using a linear model. The nonlinear structure of the series with different constants and variances between regimes increases the autocorrelation of the original series. This indicates that, although the process can be *approximated* with a stochastic trend, i.e. unit root process, it is not its true form.

In Figure 1 we present the top 1% income shares and the estimated regime switching probabilities for the above mentioned six countries. In all subfigures the regime switching probability is given on the right axis while the share of total income earned by the top 1% income earners is given on the left axis. The right axis tells us the probability that the income inequality series is in the first regime. For the Australian series, both the first and the third regime probabilities are given.

In Canada, France and Japan, income inequality switches regime right after the second world war. In Canada, the probability that the income inequality series is in the first regime drops to 2% in 1946, in France the probability drops to 6% in 1945, and in Japan the probability drops to 1% in 1946. This corresponds to the global trend break point of 1946 found by RW. In Canada the probability jumps to 91% in 1996, which corresponds to the country-specific break point of 1994 found by RW. In the USA, the probability that income inequality is in the first regime drops to 12% in 1954 and to 1%

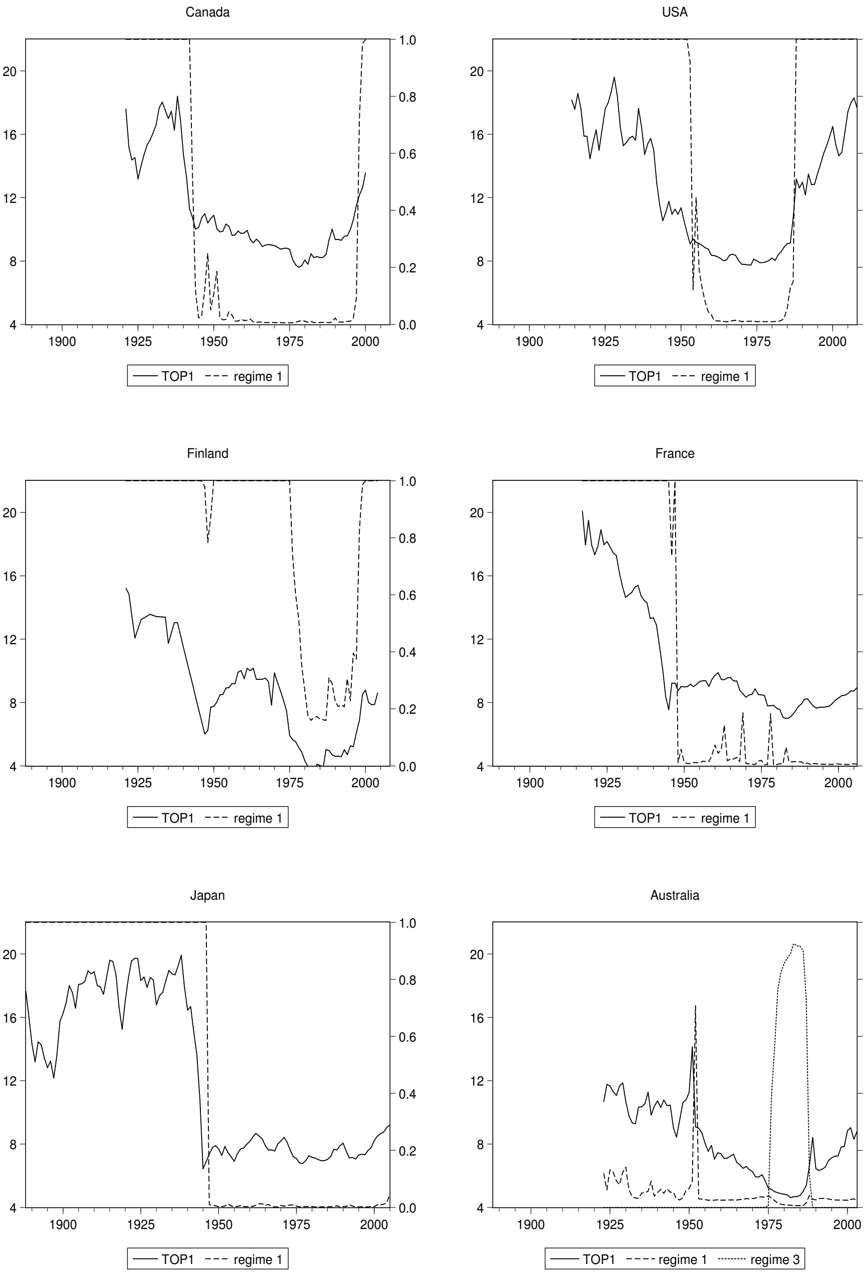

Figure 1. Top 1% income shares and the regime switching probabilities for Australia, Canada, France, Finland, Japan and the USA.

in 1961. In 1988, the probability jumps back to 99%. The change in the regime in 1954 corresponds to the common structural break in Anglo-Saxon countries in 1953, and the change in 1988 corresponds to the common post-war break in Anglo-Saxon countries in 1987, found by RW. In Finland, the probability of income inequality being in the first regime drops to 16% in 1982, which is the same year as the break in post-war data on Nordic countries found by RW. In Finland, the probability rises back to over 80% in 1998, which corresponds to the country-specific break in 1997 found by RW.

In Australia the probability that income inequality is in the second regime is around 90% in all but two occasions. In 1951, the probability drops to 30% while the probability that income inequality is in the third regime is under 0.01%. Thus, in 1951, the income inequality series of Australia is on the first regime. In 1982, the probability of the series being in the second regime drops to around 7% and the probability of the series being in the third regime rises to 92%. This regime change is likely to correspond on the structural break in the country-specific series of Australia in 1985 found by RW.[2]

The results based on GMAR models imply that some of the structural breaks found by RW are points, where the series of income inequality changes regime and the characteristics of the series change in terms of means and variances. Furthermore, according to the residual diagnostics of our GMAR models, the other structural breaks found by RW in these series are not statistically significant.

4 Conclusion

Results of this study indicate that some of the breaks estimated by Roine and Waldenström (2011) are actually shifts between higher mean, high variance regimes and lower mean, low variance regimes. These regime changes imply that on or after these points economies have gone through some major change that has fundamentally affected the distribution of income. The timing of these shifts can thus be used to assess the effects

[2] RW do not find any other breaks in the top 1% income share series of Australia, but their analysis ignores the series prior 1950.

of different policies, like the opening of international trade or major regulatory changes, on the distribution of income. In addition, some shifts are also likely be driven by economic factors, like skill-biased technological change and education (Goldin and Katz 2008; Gordon and Dew-Becker 2008).

Combining the results of this study with the recent studies on the relationship between income inequality and economic development also gives some policy implications. Studies by Malinen (2012) and Herzer and Vollmer (2012) indicate that there is an equilibrium relation between the stochastic parts of GDP per capita and income inequality. This implies that income inequality would be associated with macroeconomic instability. That is, if the stochastic elements of GDP per capita and income inequality move in tandem, fluctuation in the distribution of income will result to a fluctuation in the GDP per capita and *vice versa*.

References

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

Goldin K. and Katz LF. (2008). *The Race Between Education and Technology*. Boston: Harvard University Press.

Gordon RJ. and Dew-Becker I. (2008). Controversies about the rise of American inequality: a survey. NBER working paper no. 13982.

Herzer, D. and S. Vollmer (2012). Inequality and growth. Evidence from panel cointegration. *Journal of Economic Inequality*, (forthcoming).

Jäntti, M. and S.P. Jenkins (2010). The impact of macroeconomics conditions on income inequality. *Journal of Economic Inequality*, 8, 221-40.

Kalliovirta, L. (2012). Misspecification tests based on quantile residuals. *The Econometrics Journal*, 15, 358-93.

Kalliovirta, L., Meitz, M., and P. Saikkonen (2012). A Gaussian mixture autoregressive model for univariate time series. HECER discussion paper 352.

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

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

Mocan, H. N. (1999) Structural unemployment, cyclical unemployment, and income inequality. *The Review of Economics and Statistics*, 81 (1), 122-34.

Parker, S. (2000). Opening a can of worms: the pittfalls of time-series regression analyses of income inequality. *Applied Economics*, 32(2), 221-30.

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.