UNITED NATIONS UNIVERSITY
UNU-WIDER

This paper is a draft submission to

Inequality—Measurement, trends, impacts, and policies

5–6 September 2014 Helsinki, Finland

Reversal of the Kuznets curve: Study on the inequality–development relation using top income shares data

Elina Tuominen
elina.tuominen@uta.fi
University of Tampere, Finland

WORK IN PROGRESS - PLEASE, DO NOT CITE

Abstract

This paper uses recently published top 1% income share series in studying the inequality–development association. Top income shares data are of high quality and cover about a century for some countries and thus provide an interesting opportunity to study slow development processes. Moreover, nonlinearities have not been studied sufficiently in the empirical inequality–development literature. To address the issue of functional form, this study utilizes penalized spline methods. It is found that the association between inequality and development experiences a reversal at later stages of development and is, thus, U-shaped in many advanced countries. In addition, results support an inverse-U-shaped relation between inequality and urbanization, and positive relation between inequality and service sector. These results have an interpretation that is possible to fit into ideas presented by Kuznets who discussed shifts in the economy.

Keywords:
inequality, top incomes, economic development, nonlinearity, longitudinal data
JEL: D30, N30, O11, O15, O50

Version 1.0 *July 29, 2014*

1. Introduction

According to Kuznets, the relationship between inequality and economic development resembles an inverse-U curve as the focus of the economy shifts from agriculture to other sectors: at the earlier stages of development inequality increases but after a certain point inequality starts to decrease. However, in many empirical papers that have studied the existence of the famous Kuznets curve, the original setting of different sectors in the economy is not discussed. Previous empirical studies have presented mixed evidence on the shape of the inequality–development relation, and the debate has focused on whether the empirical results support the inverse-U association or not.

This study utilizes new inequality data on top 1% income shares (Alvaredo et al., 2013b) to study the inequality–development association. The top income share series are unprecedently long and cover the whole 20th century for some countries. During this period some countries have faced not only urbanization but also later stages of development. The focus of the paper is mainly in (currently) advanced countries but also some (currently) less-advanced countries are included.

Moreover, flexible methods are used to discuss issues related to the choice of the functional form in inequality–development studies. For this purpose, additive models with penalized regression splines are used. Additive model specifications do not imply a specific functional form and can be taken as a semi-parametric estimation approach. Formerly, nonparametric approach has been used by Frazer (2006) although he uses Gini coefficients and is only able to study a shorter time period due to data availability. However, he shows that it is reasonable to use flexible methods in studying the inequality–development relation. As an example, one can take a look at his finding in Figure 1 that illustrates the prediction with confidence intervals.

This paper finds that the so called Kuznets curve experiences a reversal during the later phases of the development process. This finding is robust to including controls for urbanization and service sector. This paper also supports the idea that top 1% income shares would be a reasonable measure for inequality. There are striking similarities in the shape of inequality–development relation when one compares the results of this paper to those in Frazer (2006).

This paper is organized as follows: Section 2 describes the famous Kuznets (1955) article and the empirical literature that has studied the existence of

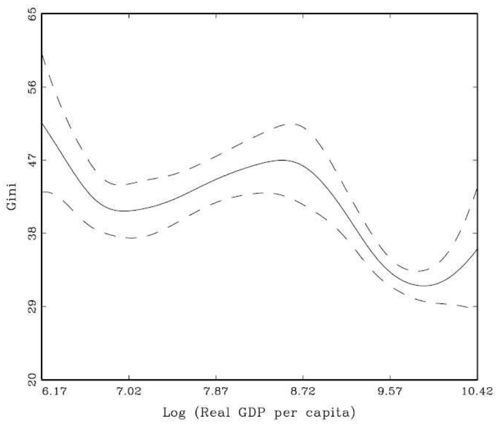


Figure 3. *Pooled nonparametric regression—Gini versus log(real GDP/captia).*

Figure 1: Illustration of the inequality–development relation in Frazer (2006, p. 1464).

the Kuznets curve. Section 3 introduces the data used in the empirical analysis, and section 4 describes the estimation framework. Section 5 provides empirical results. Section 6 discusses the main findings and concludes.

2. Related literature

Kuznets (1955) suggested that there are at least two types of forces that induce inequality in the distribution before taxes and transfers. The first is related to the cumulative effect of concentrated savings at the top of the distribution. The second is what Kuznets describes as the sectoral shift from agriculture to other sectors. Typically, average incomes per capita are lower in the rural sector than in the urban sector, and inequality is also lower in the rural sector. As the urban sector kept growing, Kuznets expected inequality to have increased. However, this was not what he found. Using income shares for different parts in the distribution, Kuznets (1955) reports a modest decline in inequality in the UK, USA and Germany during the first half of the 20th century.

On the one hand, Kuznets (1955) suggested that there are equalizing forces that hinder the concentration of saving at the top of the distribution. He mentioned political decision-making, for example inheritance taxes.[1]

[1] Role of taxation has been discussed also within top-income literature, e.g., Alvaredo, Atkinson, Piketty, and Saez (2013a) and Atkinson, Piketty, and Saez (2011).

Kuznets also suggested that there are other, less-obvious reasons that are 'characteristics of a dynamic growing economy' (Kuznets, 1955, p. 11). For example, if new generations at the top of the distribution are not able to conform into new technologies and industries their wealth is likely to deteriorate. Kuznets also suggested that the service incomes of workers at lower-income levels are likely to grow faster than the corresponding incomes of workers at upper-income levels. Also, shifts from lower-income to higher-income industries can take place.

On the other hand, Kuznets discussed the shift from agricultural to urban sectors. Kuznets (1955) provided calculations using a simple numerical example of two sectors: agricultural (A) and all others (B). In his calculations he assumed that (1) $\textit{per capita income}_A \leq \textit{per capita income}_B$, (2) *proportion of sector A declines*, and (3) $\textit{income inequality}_A \leq \textit{income inequality}_B$. His calculations revealed, for example, the following observations:

1. If the difference in per capita incomes increases between sectors A and B, or if income inequality is greater in sector B (compared to sector A) – or if both cases hold – then the relative weight of sector B causes an increase in overall income inequality.
2. If the sectoral difference in per capita incomes is constant and the two intrasectoral distributions are the same, then changes in proportions of the sectors can create differences in the overall distribution. In general, as the proportion of sector A declines from 0.8 to 0.2, inequality first increases and then decreases.
3. After the proportion of sector A falls "enough", the share of the 5th quintile declines. Kuznets reasons that, during industrialization, the non-agricultural sector raises the per capita incomes for the economy as a whole (although the per capita incomes were constant within and between both sectors). Kuznets also continues reasoning that the top income shares would fail to decline only if there were a stronger growth in per capita incomes of sector B than in those of sector A, or increasing inequality in per capita incomes of sector B.

In summary, Kuznets (1955) suggested that at earlier stages of economic development we would expect inequality to rise. Then there would be a short phase of stagnation, and thereafter inequality would decrease. Kuznets called this process a 'long swing' and proposed that it is likely to be observed for the 'old countries'. He also argued that if this process was to be seen in incomes before taxes and transfers (which he used), it would most likely be observable in net incomes. The progressive tax systems only enforce the downward trend in inequality.

Many theoretical papers have studied the Kuznets-type relation and found support for it (e.g., Robinson, 1976; Galor & Tsiddon, 1996; Aghion & Bolton, 1997; Dahan & Tsiddon, 1998). Also numerous empirical studies have investigated the inequality–developement relation. Cross-sectional study by Ahluwalia (1976) supported the inverted-U relation. However, Anand and Kanbur (1993) found that the results are not robust to different functional forms and that the inequality data are not of good quality in Ahluwalia (1976). For example, comparability with respect to income concept is not satisfied. When Anand and Kanbur (1993) constructed a somewhat consistent inequality data, their results pointed to a functional form that is actually a reversal of the originally represented Kuznets hypothesis. Other examples of contradictory results are Deininger and Squire (1998) and Barro (2000). Deininger and Squire (1998) find no support for the Kuznets hypothesis in their fixed-effect specification, but they find support for it in the pooled case. In comparison, Barro (2000) finds support for the Kuznets hypothesis using both of these specifications. The difference in these studies lies in the functional form of the GDP per capita variable.

The quality of inequality datasets has been discussed since the 1990s. Especially Deininger and Squire (1996) highlighted that the reliability of the previously used inequality data was questionable, and they constructed a new data set. However, their data could still be considered problematic. The data or its subsets have been widely used despite their problems. Atkinson and Brandolini (2001) highlighted that consistent inequality series have not been available, e.g., Gini coefficients may have been reported for different income concepts. They showed that different sources can give a very different picture of inequality.

An interesting example of fairly recent inequality–development studies is Frazer (2006). His study addresses the problem of functional form using nonparametric regression.[2] He represents his results graphically not only within each country but also across countries. The approach that Frazer uses stands between cross-section and country-specific studies. The time span covered

[2]To be more precise, Frazer (2006) uses local linear least-squares regression. He uses kernel weights from a normal density function. Some of the control variables enter his specification linearly, and thus his estimation method is semiparametric. His method follows Robinson (1988).

by Frazer is approximately 50 years, and the data are an expanded version of the Deininger–Squire data, namely the World Income Inequality Database (WIID). Frazer focuses on using Gini coefficients in his estimations.[3] He finds support for a nonlinear association that is more complex than a second degree polynomial, between Gini coefficients and economic development. However, this relation is significant and negative only at a certain phase in the development process (see illustration on this finding in Figure 1). He also studies the effect of including different control variables (one at a time) in his model. Even though the significance of his results varies depending on the added control variable the overall shape holds.

During the last two decades, at least, the assumptions behind Kuznets' observation 2 (described above) have been rightly questioned. It is not realistic to assume that inequality within two sectors and the levels of per capita incomes would stay constant as urbanization takes place.[4] Moreover, Atkinson (1995, pp. 25-26) suspects that Kuznets would not have been surprised if the inverse-U shape no longer holds. Atkinson highlights that Kuznets discussed very carefully the conflicting forces behind the inequality–development relation. Also various inequality indices have shown an upward trend in many countries during the last twenty or thirty years. The question then remains: What can we say about the advanced countries as the shift from agricultural sector to urban sector has already taken place? The data used in this study will be discussed in the following section.

3. Data

3.1. Top income shares data

The top of the distribution deserves attention because changes in the upper part affect the distribution as a whole. Moreover, many of the available Gini series have suffered from comparability problems both in time and between countries, and the series have not covered long time intervals. Using tax and population statistics, it is possible to compose long and fairly

[3] Frazer (2006) also tries to check some of his results using the income shares of the bottom 40% and top 20% but he notes that the number of observations decreases significantly at this point and he cannot get statistically significant results as he did using Gini coefficients.

[4] A recent paper by Kanbur and Zhuang (2013) investigates inequality and urbanization in Asia in the spirit of Kuznets and discusses this issue.

consistent series on top income shares. Kuznets (1953) was the first to use this kind of data to produce top income share estimates, and Piketty (2001) generalized Kuznets' approach. Following Piketty, top income share series have been constructed by different researchers.[5]

Naturally, using tax registers as a basis for computations has its limits, e.g., tax avoidance, income re-arrangement, how population statistics are used as help, and differences in tax units between countries (see, e.g., Atkinson, 2004a,b). In some countries capital gains have not been taxed or the tax has been so low that re-arrangement of income into capital gains can be seen as a way to avoid taxes. However, according to Atkinson (2004b) the observed changes in top income shares cannot be explained only by re-arrangement of income. For more information, also Piketty and Saez (2006), Leigh (2007), and Roine and Waldenström (2014) have discussed the advantages and limitations of these series. However, according to Leigh (2007) top income shares have shown similar development as various other inequality indices over time. Also Roine and Waldenström (2014) provide evidence that support this idea, and they conclude that top income shares are useful as a general measure of inequality.

Top income data can be easily accessed using the *The World Top Incomes Database* webpages by Alvaredo et al. (2013b).[6] This study uses the top 1% income shares of 26 countries from 1900 to as far as 2010. The data include, e.g., English-speaking countries, Continental and Southern European countries, Nordic countries and some 'less-advanced' countries. The top 1% income shares are presented graphically in Appendix A. For each country, there are observations from various time points but the longitudinal data are

[5]Progressive income tax systems were created in most industrial countries at the beginning of the 20th century. Tax authorities started to collect statistics based on income tax data. These statistics reported the number of taxpayers in a specific income bracket, their total income and their tax liability. Usually this information was divided into capital income, wage income, business income and so on. Before the World War II, in most countries, there was at most 10–15% of the population under income taxation. This is why it is possible to calculate the top income shares only for the top decile (or its upper part). Many of the top income share series used here have been constructed using the so-called Pareto interpolation technique, but also mean split histograms have been used (for more information see, e.g., Atkinson, 2005, 2007; Piketty, 2005).

[6]The first book on these series, edited by Atkinson and Piketty (2007), contrasts the evidence from Continental Europe and English-speaking countries. The second volume, also edited by Atkinson and Piketty, was published in 2010.

not balanced. Most drastic changes during the past century have taken place in the top 1% incomes, not, e.g., in the whole top 10% (see, e.g., Roine & Waldenström, 2014). For this reason, we focus on the top 1% income share series, denoted by *top1*. For better comparability, capital gains have been excluded from the income concept whenever possible.

In English-speaking countries, the evolution of top 1% income shares resembles U over the 20th century. There is a significant increase since the 1980s. In contrast, top 1% shares in Continental Europe and Japan have remained fairly stable during the last three decades. During the first half of the 20th century, top incomes consisted mainly of capital income. In most countries, capital incomes fell dramatically during wartime and the Great Depression. Also the distribution of earned incomes became more equal in many countries after the wars. One explanation for the extended fall in top income shares is progressive taxation.[7] In the USA and other English-speaking countries, the growth in top income shares has been explained by growth in top wages after the 1970s. As the top wages have increased, the top executives and capital owners cohabitate the top of the income distribution. In contrast, in Finland and Sweden capital incomes continue being important. One explanation for the rise in top incomes is the decrease in highest tax rates. Moreover, it has been suggested that explanations based on skill-biased technological change are not sufficient in explaining the surge in top earnings. Literature on 'college premium' does not describe well why *top1* has increased in comparison to *top10*, since basically the whole group of *top10* are college-educated. Alternative explanations, e.g., theories on executive remuneration in a hierarchical structure and superstar theory have been suggested. (Piketty & Saez, 2003, 2006; Atkinson et al., 2011; Alvaredo et al., 2013a)

On the basis of these findings an inverse-U-shaped *top1*–development relation is not expected. Atkinson et al. (2011) and Roine and Waldenström (2014) discuss the problem of fitting top income shares into the Kuznets (1955) approach where the inverse-U relation is described by shift from traditional to modern sector. One reason is that Kuznets highlighted the role of labor income in explaining the structural shift, but as it comes to top

[7] Roine et al. (2009) find that tax progressivity is in negative relation to top income shares. They do not discuss the inequality–development relation in the spirit of Kuznets (1955).

incomes we need to consider also capital income. But one cannot rule out the possible start of a new shift after the 1970s – another shift could have started because of the expansion of the information technology sector (or more broadly put: the service sector). Also Roine and Waldenström (2014) discuss the possibility of another Kuznets curve starting although they, too, acknowledge that this type of approach does not capture the importance of capital income. As Kuznets proposed, there are various contradicting forces at play.

3.2. Development and economic sectors

Level of economic development is measured in a traditional way using GDP per capita data. The GDP data are from a Maddison Project update (Bolt & van Zanden, 2013) and are available annually until 2010. Here data from 1900 onwards are used whenever available. The level of development is measured using the natural logarithm of per capita GDP in 1990 international Geary-Khamis dollars. These GDP series are plotted in Appendix B.

Models that control for the level of urbanization are studied to check the baseline results. However, the availability of data limits the time horizon that can be studied. In additional analyses, urbanization data describe the proportion of urban population (United Nations, 2012). This data are available from 1950 onwards every 5 years. The urbanization series are plotted in Appendix C.

The results are also checked by controlling for the employment in the service sector (% of total employment) (World Bank, 2014a). Unfortunately, including this variable into the models narrows the time span of the results even more, as these data are available from 1980 onwards. However, this can be seen as an extension to the study of Frazer (2006) who did not include a measure for the service sector. The series are plotted in Appendix C.

4. Estimation method

Additive models provide a flexible framework to investigate the association between inequality and development.[8] We follow here the approach

[8]Additive models are a special case of generalized additive models (GAMs). GAMs were introduced by Hastie and Tibshirani (1986, 1990). They present a GAM as a generalized linear model with a linear predictor that involves a sum of smooth functions of covariates. Some of the covariates can enter in linear form. Note here the analogy to 'generalized

presented in Wood (2006). The basic idea is that the model's predictor is a sum of linear and smooth functions of covariates:

$$\mathbf{E}(Y_i) = \boldsymbol{X}_i^* \boldsymbol{\theta} + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + ... \quad ,$$
$$\text{where} \quad Y_i \sim \text{normal distribution.}$$

In the above presentation Y_i is the response variable (here: level of top 1% income share), $\boldsymbol{X}_i^*$ is a row of the model matrix for any strictly parametric model components, $\boldsymbol{\theta}$ is the corresponding parameter vector, and the $f_\bullet$ are smooth functions of the covariates, $x_\bullet$.

The flexibility of these models comes at the cost of two problems. Firstly, one needs to be able to represent the smooth functions $f_\bullet$ in some way. One way to represent these smooths are cubic regression splines, which is the approach taken here. A cubic regression spline is a curve constructed from sections of cubic polynomials that are joined together so that the resulting curve is continuous up to second derivative. The points at which sections are joined are *knots* of the splines, and these locations must be chosen. The spline can be represented in terms of its values at the knots.[9] Secondly, one needs to choose the amount of smoothness that functions $f_\bullet$ have. One does not want to overfit, and thus the departure from smoothness is penalized. The appropriate degree of smoothness for the $f_\bullet$ can be estimated from the data. Various selection criteria are available, e.g., (generalized) cross-validation or maximum likelihood.

Illustration

Consider a model containing only one smooth function of one covariate: $y_i = f(x_i) + \epsilon_i$, where ϵ_i are i.i.d. $N(0, \sigma^2)$ random variables. To estimate function f here, f can be represented so that the model becomes a linear model. This is possible by choosing a *basis*, defining the space of functions of which f (or a close approximation to it) is an element. In practice, one chooses *basis functions* which are treated as known.

Assume that the function to be estimated is $f(x) = \sum_{j=1}^{k} \beta_j b_j(x)$, where β_j refers to coefficients that are estimated, and b_j to known basis functions.

linear models' and 'linear models'. Here we restrict ourselves to a special case: using identity link and assuming normality in errors, which brings us to additive models.

[9] Usually, there are two extra conditions specifying that the second derivative of the curve should be zero at the two end knots.

Using a chosen basis for f means that we have a linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where the model matrix $\boldsymbol{X}$ can be represented using basis functions such as those in the cubic regression spline basis.

The departure from smoothness can be penalized with $\int f''(x)^2 dx$, where $f''(x) = \sum_{j=1}^{k} \beta_j b_j''(x) = \boldsymbol{\beta}^T \boldsymbol{b}''(x)$, and $\boldsymbol{b}''(x)$ is the vector of second derivatives of the basis functions evaluated at x. The penalty $\int f''(x)^2 dx$ can be expressed as $\boldsymbol{\beta}^T \boldsymbol{S} \boldsymbol{\beta}$, where $\boldsymbol{S}$ is the *penalty matrix* that can be expressed in terms of the known basis functions.

The estimation procedure is based on the minimization of the sum of squared residuals subject to a penalizing term, i.e. min $\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda \boldsymbol{\beta}^T \boldsymbol{S} \boldsymbol{\beta}$, with respect to $\boldsymbol{\beta}$. The problem of estimating the degree of smoothness is a problem of estimating the smoothing parameter λ.[10] The penalized least squares estimator of $\boldsymbol{\beta}$, given λ, is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X} + \lambda \boldsymbol{S})^{-1} \boldsymbol{X}^T \boldsymbol{y}$. Thus, the expected value vector is estimated as $\widehat{\mathbf{E}(\boldsymbol{y})} = \hat{\boldsymbol{\mu}} = \boldsymbol{A}\boldsymbol{y}$, where $\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{X}^T \boldsymbol{X} + \lambda \boldsymbol{S})^{-1} \boldsymbol{X}^T$ is called an influence matrix.

This setting can be augmented to include various covariates and smooths. Given a basis, an additive model is simply a linear model with one or more associated penalties. To use multi-dimensional smooths, like a two-dimensional smooth $f_3(x_{3i}, x_{4i})$, one needs to know how to scale variables in this context. In this paper, tensor product smooths are used in cases of multi-dimensional smooths. Tensor product smooths are constructed using marginal smooths of lower dimension, and the obtained smooth has a penalty for each marginal basis. Tensor product bases are shortly described in Appendix D.

Practical notes

Usually in estimation, the size of the basis dimension for each smooth is not critical because it only sets an upper limit on the flexibility of a term. Smoothing parameter controls the actual effective degrees of freedom (edf). Effective degrees of freedom are defined as trace($\boldsymbol{A}$) where $\boldsymbol{A}$ is the influence matrix. It is also possible to divide the effective degrees of freedom into degrees of freedom for each smooth. The effective degrees of freedom can be used to measure the flexibility of a model. For example, a simple linear term would have one degree of freedom, and edf=2.3 can be thought of as a

[10] In the estimation, one faces a bias-variance tradeoff: on the one hand the bias should be small, but on the other hand the fit should be smooth. One needs to compromise between the two extremes. $\lambda \longrightarrow \infty$ refers to straight line estimate for f, and $\lambda = 0$ refers to an un-penalized regression spline estimate.

function that is a bit more complex than a second degree polynomial.

Confidence intervals for the model terms or parameters can be derived using Bayesian methods. Also approximate p-values for model terms can be calculated. Models can be compared using information criteria, e.g., the Akaike information criterion (AIC). When using AIC for penalized models (models including smooth terms), the degrees of freedom are the *effective* degrees of freedom, not the number of parameters. Moreover, random effects can be included in these models.

In the next section, the estimated degrees of freedom (edf) of each smooth term and model selection criteria are provided for each model. However, the focus will be on investigating graphical illustrations.[11]

5. Estimation results

The baseline model results are given first. These models refer to specifications without sectoral variables and they use as long series as possible (1900–2010) with annual data. Subsets of countries (different country groups) are also discussed. Second, the urbanization variable is included in the specifications, which means that 5-year data are used (1950–2009). Third, both urban and service sector variables are included, which leads to using 5-year data covering 1980–2009.

5.1. Baseline models

The baseline results are presented for annual data and they cover the years 1900–2010 (whenever data are available). The baseline specification is

$$\mathit{top1}_{it} = \alpha + f(\mathit{ln(GDP\ per\ capita)}_{it}) + u_i + \epsilon_{it},$$

[11]The results presented in this paper are obtained using the R software package 'mgcv' (version 1.7-21), which includes a function 'gam'. Marginal basis construction 'cr' for cubic regression splines is used here. The knots are placed evenly through the range of covariate values (default). The maximum likelihood method 'ML' is used in the selection of the smoothing parameter. The identifiability constraints (due to, e.g., model's additive constant term) are taken into account by default. The function 'gam' also allows for simple random effects: it represents the conventional random effects in a GAM as penalized regression terms. Also in this case, maximum likelihood method is used in the selection of the smoothing parameters. More details can be found in Wood (2006) and R project's web pages (`http://cran.r-project.org/`).

where α is a constant and f is a smooth function that can be described using a cubic regression spline, i refers to country and t refers to year, u_i is a country effect and $\epsilon_{it} \sim N(0, \sigma^2)$ is the error term. The country effects can be fixed (dummy for each country) or random ($u_i \sim N(0, \sigma_u^2)$).[12]

Table 1: Results on baseline models, using annual data (years 1900–2010): estimated degrees of freedom (edf) for each smooth. Constant terms are not reported. See also Figure 2.

	depending variable: $top1_t$ (N=1609)		
variable in period t	(1)	(2)	(3)
$f(log\ GDP\ pc_t)$	[edf=9.23[a]]***	[edf=9.13[a]]***	[edf=9.14[a]]***
country effects	no	fixed	random
adjusted r^2	0.395	0.702	0.702
AIC	8572.2	7457.1	7457.2

Approximate significance levels: (***) < 0.01 , (**) < 0.05 , (*) < 0.10.
These are calculated using the Bayesian estimated covariance matrix of the parameter estimators. The p-values are based on an F statistic.
[a]The basis dimension of the smooth before imposing identifiability constraints is $k = 15$.

The pooled model (i.e. without country-specific effects) is provided in the column (1) of Table 1. The models (2) and (3) in Table 1 include country effects: unsurprisingly the models fit the data better after allowing for random-effects (RE) or fixed-effects (FE). To investigate the models in Table 1, Figure 2 illustrates the predictions. The 'first' turning point or a 'plateau' is not evident after including country effects but the 'second' turning point around ln(GDP per capita)$\approx$ 9.5 is robust to adding country effects. One reason for the difference in the results at lower levels of development is that we have shorter series for some of the less-advanced countries and including simple country effects can capture the level of *top1* in these specifications. Moreover, the RE and FE specifications (models (2) and (3) in plots (b) and (c), respectively) give practically identical predictions. With country effects, one only observes the negative relation between economic development and *top1* before the turning point is reached.[13]

[12]Simple, purely parametric models with second or third degree polynomials were also checked. Simple pooled OLS results or models with country effects were not in conflict with the models in Table 1. However, additive models fitted the data better.

[13]Note: ln(GDP per capita) $\approx$ 7 corresponds to GDP per capita $\approx$ 1096 (Int. GK$) and ln(GDP per capita) $\approx$ 9.5 corresponds to GDP per capita $\approx$ 13359 (Int. GK$).

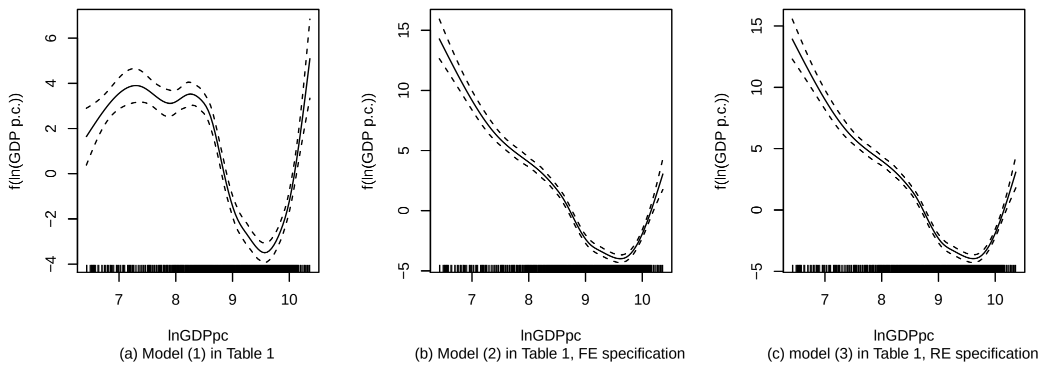


Figure 2: Illustration of the *top1*–development relation (annual data 1900–2010). See Table 1. The figures present only the smooth function $f(\ln(\text{GDP per capita}))$ and not the constant term, so the reader should focus on the shape of the prediction – not the level. The solid line represents the estimates (smooth function f). The figures also show the 95% Bayesian credible intervals (dashed) and the covariate values as a rug plot along the horizontal axes.

5.2. Baseline models for subsets of countries

To check the previous results, different subsets of the data were studied. In summary, the main conclusions hold. However, this subsection illustrates that predictions are not robust for 'less-advanced' countries (or at lower development levels).

The results concerning English-speaking, Nordic, Continental and Southern European or 'less-advanced' countries separately, are reported in Table 2. The illustrations of the smooths for each group of countries are provided in Figure 3. There seems to be a pattern that holds as countries reach a higher level of economic development. The English-speaking countries show a clearly U-shaped relation with a downward peak at ln(GDP per capita)$\approx$ 9.5. For the Nordic countries there is a negative slope that has 'recently' reached the turning point close to ln(GDP per capita)$\approx$ 9.6. The Continental and Southern European countries seem to follow this path as well, even though the positive slope is not (yet?) evident. The plot for 'less-advanced' countries shows almost a plateau following a negative slope.

Since the number of countries in the data is not large, the sensitivity to having a certain country in the sample is assessed. For this reason, each country was dropped (one at a time) from the sample and the overall results

Table 2: Results on baseline models without country effects for specific groups of countries, using annual data (years 1900–2010). Constant terms are not reported. See also Figure 3.

$top1_{it} = \alpha + f(ln(\text{GDP per capita})_{it}) + e_{it}$	N	smooth f	adjusted r^2
English-speaking[a]	459	[edf=6.96[e]]***	0.613
Nordic[b]	333	[edf=3.79[f]]***	0.641
Continental & Southern Europe[c]	360	[edf=5.35[e]]***	0.647
Less-advanced[d]	288	[edf=5.61[e]]***	0.119

Approximate significance levels: (***) < 0.01 , (**) < 0.05 , (*) < 0.10.
These are calculated using the Bayesian estimated covariance matrix of the parameter estimators. The p-values are based on an F statistic.

[a] Australia, Canada, Ireland, New Zealand, UK and USA.

[b] Denmark, Finland, Norway and Sweden.

[c] France, Germany, Italy, Netherlands, Portugal, Spain and Switzerland.

[d] Argentina, China, Colombia, India, Indonesia, Mauritius and South Africa.

[e] The basis dimension of the smooth before imposing identifiability constraints is $k = 10$.

[f] The basis dimension of the smooth before imposing identifiability constraints is $k = 5$.

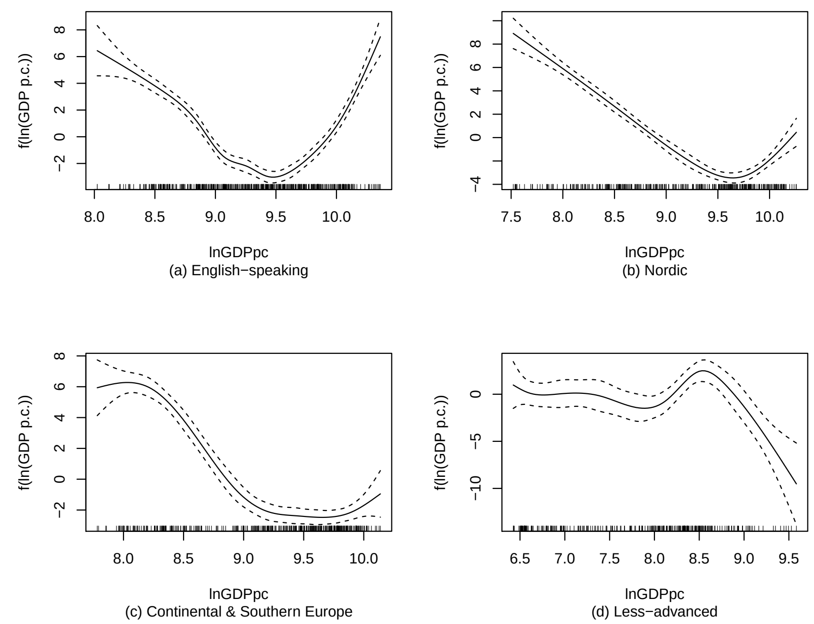
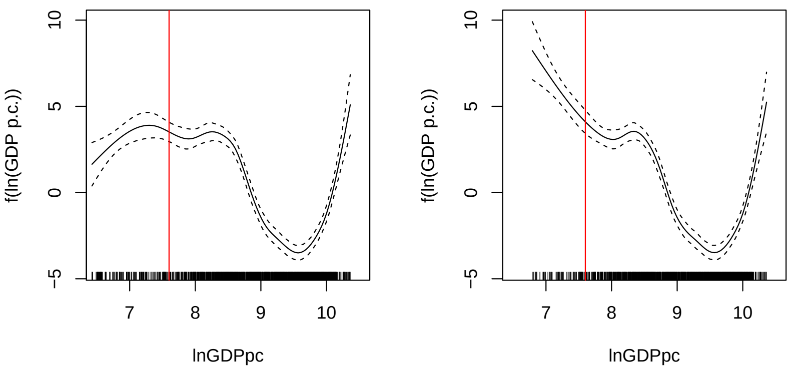


Figure 3: Illustration of the *top1*–development relation for four different subsets of the data (annual data 1900–2010). See Table 2 for details. The figures present only the smooth function f(ln(GDP per capita)) and not the constant term, so the reader should focus on the shape of the prediction – not the level. The solid line represents the estimates (smooth function f). The figures also show the 95% Bayesian credible intervals (dashed), and the covariate values as a rug plot along the horizontal axes.

were checked.[14] In most cases, leaving one country out of the sample had very little effect on our prediction. Five countries (out of 26) had some effect: First, dropping the USA out means dropping the highest GDP values out of the sample. However, this did not have a qualitative effect on the shape of the curve. Then, dropping Japan or South Africa from the sample resulted in the credible interval becoming wider at 'lower' levels of log GDP per capita (specifically, ln(GDP per capita) < 8) but the change in fitted curve was not drastic. Also dropping Argentina from the sample changed the prediction slightly around $7.5 <$ ln(GDP per capita) < 8.5 but its overall shape was still much the same as with the whole data. However, one country had a more evident effect on the overall shape of the curve, namely India. India (N=71) affects the prediction strongly at lowest levels of economic development, and Figure 4 below illustrates this. These results show that predictions for ln(GDP per capita) < 7.6 are very much dependent on few countries or observations.



(a) All 26 countries included
(b) India (IND) dropped from data

Figure 4: The effect of leaving one country out of the sample at a time (annual data 1900–2010): Case of India. The figures present the smooth function f(ln(GDP per capita)) in two cases. The solid line represents the estimates (smooth function f). The figure also shows the 95% Bayesian credible intervals (dashed), and covariate values as a rug plot along the horizontal axes. For comparison, plot (a) illustrates the model that is given in column (1) of Table 1 (illustrated also in Figure 2(a)).

[14] The models were estimated using smooth functions where the basis dimension of the smooth (before imposing identifiability constraints) was $k = 15$.

Each country was also separately studied to check how well its development process would fit the overall curve based on 26 countries.[15] Naturally, the development paths are not identical for all countries. For example, Switzerland (N=74), Portugal (N=24), Ireland (N=37) and South Africa (N=62) have downward peaks but they are not around ln(GDP per capita) $\approx$ 9.5.[16] Finland (N=90) and New Zealand (N=83) show a pattern that resembles an asymmetrical W-shape, but the 'last' downward peak is still around ln(GDP per capita) $\approx$ 9.5 for both countries. Moreover, Singapore (N=59) and Mauritius (N=52) show a U-shaped pattern with a 'plateau' (instead of a single 'peak') at the 'bottom'. The data from 'less-advanced' countries such as China (N=18) and Colombia (N=18) are not in line with the combined results, but these two *top1* series are notably shorter than the series of most other countries. In summary, one should be careful in making conclusions related to less-advanced economies since this group of countries is very heterogeneous and the sample does not include that many 'less-advanced' countries.

5.3. Controlling for urbanization

Does controlling for urbanization affect the results above? This type of investigation was also presented in Frazer (2006), who did not find a drastic change in the shape of the Gini–development relationship. Because the urbanization variable (share of urban population) is available every 5 years from 1950, the analysis is implemented using 5-year-averaged data. The averaged data are constructed using consecutive periods. For example, consider a particular country for which we have 5-year-averaged data. The 5-year periods are determined as follows: 1950–1954, 1955–1959, 1960–1964, ..., 2000–2004, 2005–2009.[17]

---

[15] The models were estimated using smooth functions where the basis dimension of the smooth (before imposing identifiability constraints) was $k = 5$ or $k = 10$.

[16] The suggested downward peaks were the following: Switzerland around ln(GDP per capita) $\approx$ 9.9; Portugal around ln(GDP per capita) $\approx$ 9.1; Ireland around ln(GDP per capita) $\approx$ 9.0; and South Africa around ln(GDP per capita) $\approx$ 8.2. But note that for South Africa, the ln(GDP per capita) values are in general lower than in 'advanced' countries.

[17] The annual, baseline results of subsection 5.1 were also checked for the 5-year-averaged data. In the case of 5-year data, the dependent variable is the average top 1% income share level in the 5-year period t, i.e. average of the 5-year period's *top1* values. This is explained by using 5-year-average of log GDP per capita values in period t. The averaged

First, one-dimensional smooths for the GDP and urbanization variables are estimated. The structure of the model becomes

$$top1_{it} = \alpha + f_1(ln(GDP\ per\ capita)_{it}) + f_2(urban\ population_{it}) + u_i + \epsilon_{it},$$

where α is a constant and smooth functions f_i $(i = 1, 2)$ can be approximated using penalized cubic regression splines. Here i refers to the country and t refers to the 5-year period, u_i is a country effect and $\epsilon_{it} \sim N(0, \sigma^2)$ is the error term. The country effects can be fixed (dummy for each country) or random ($u_i \sim N(0, \sigma_u^2)$). This type of model allows flexible functional forms but it does not allow interaction between our two variables. To allow for a highly flexible structure that also allows for complex interaction we can allow for a two-dimensional smooth. Then the specification becomes

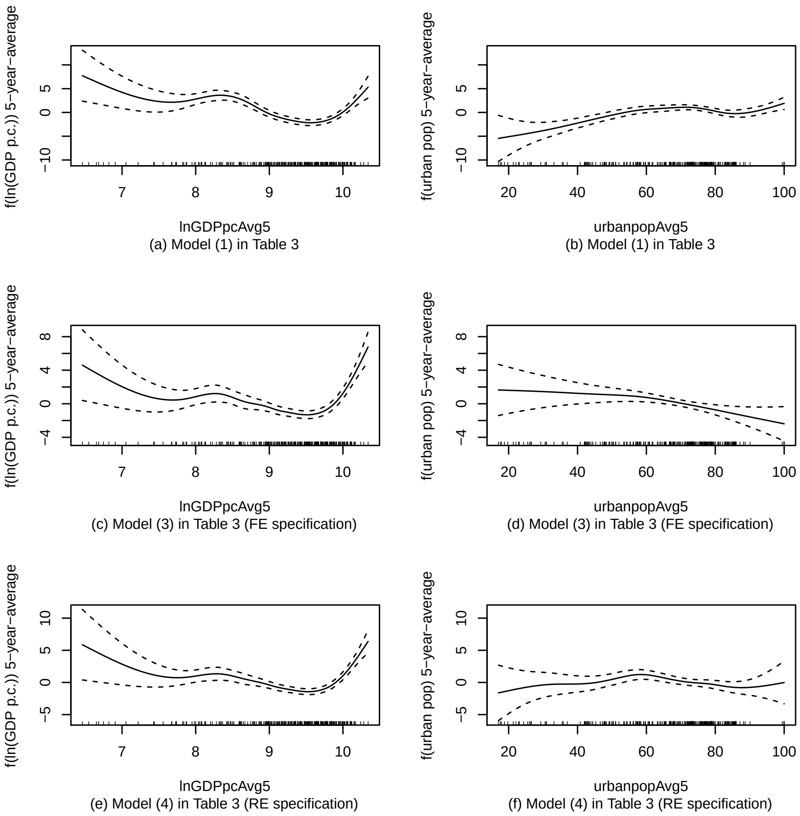
$$top1_{it} = \alpha + f_3(ln(GDP\ per\ capita)_{it}, urban\ population_{it}) + u_i + \epsilon_{it},$$

where f_3 is a bivariate (i.e., two-dimensional) smooth. Table 3 summarizes information on the results of these model types. It shows that the two-dimensional-smooth model in column (2) is preferred to the model given in column (1) if one compares model selection criteria. However, to properly compare the models in columns (1) and (2) we also need graphical illustrations. The qualitative results of model (2) are not in conflict with those of model (1) and, thus, we focus on model (1) in the main text. Illustrations of the two-dimensional smooth of model (2) can be found in Appendix F.

The models (1), (3) and (4) of Table 3 are illustrated in Figure 5. Plots (a), (c) and (e) reveal that including urbanization does not influence the downward peak at ln(GDP per capita) ≈ 9.5. The credible interval is still wide for values ln(GDP per capita) < 8. Plots (b), (d) and (f) suggest that an inverse-U-shaped association between *top1* and urbanization can be reasonable. Adding country effects in models (3) and (4) has fairly little effect on the overall shape of the smooths although urban population becomes non-significant in the random-effects model.

As a sensitivity check, different country groups were studied (compare to section 5.2; same grouping of countries as in Table 2). The results supported

data are constructed using consecutive periods: 1900–1904, 1905–1909, 1910–1914, ..., 2000–2004, 2005–2009. The baseline results for the 5-year data do not differ from the results discussed in subsection 5.1. To avoid repetition, these results are not discussed here in detail. Appendix E presents some results graphically.



(a) Model (1) in Table 3
(b) Model (1) in Table 3
(c) Model (3) in Table 3 (FE specification)
(d) Model (3) in Table 3 (FE specification)
(e) Model (4) in Table 3 (RE specification)
(f) Model (4) in Table 3 (RE specification)

Figure 5: Illustration of the smooths using 5-year-averaged data (years 1950–2009): the models (1), (3) and (4) in Table 3. The figures present only the smooth function and not the constant term, so the reader should focus on the shape of the prediction – not the level. The solid line represents the estimates (smooth function f). The figures also show the 95% Bayesian credible intervals (dashed), and the covariate values as a rug plot along the horizontal axes.

Table 3: Results on models with urbanization, using 5-year-averaged data (years 1950–2009): estimated degrees of freedom (edf) for each smooth $f_{\bullet}$. Constant terms are not reported. See also Figure 5 and Appendix F.

	depending variable: $top1_t$ (N=256)			
variable in period t	(1)	(2)	(3)	(4)
$f_1(log\ GDP\ pc_t)$	[edf=6.80^{a}]***	-	[edf=7.42^{a}]***	[edf=7.20^{a}]***
$f_2(urban\ pop_t)$	[edf=5.24^{a}]***	-	[edf=1.97^{a}]**	[edf=5.17^{a}]
$f_3(log\ GDP\ pc_t, urban\ pop_t)$	-	[edf=12.78^{b}]***	-	-
country effects	no	no	fixed	random
adjusted r^2	0.408	0.435	0.762	0.768
AIC	1169.6	1158.2	956.1	951.0

Approximate significance levels: (***) < 0.01 , (**) < 0.05 , (*) < 0.10.
These are calculated using the Bayesian estimated covariance matrix of the parameter estimators. The p-values are based on an F statistic.
[a]The basis dimension of the smooth before imposing identifiability constraints is $k = 15$.
[b]The basis dimension of the smooth before imposing identifiability constraints is $k = 5^2 = 25$ (tensor product smooth with rank 5 marginals).

the main findings above. For the English-speaking countries, the U shape between development and *top1* was significant even though urbanization was added into the model. In this U-shaped relation, the positive slope part of the U dominated. The urbanization–*top1* association support the inverse-U shape, which is also in line with the results based on all data. Similarly, the Nordic countries showed a U shape between *top1* and development. Moreover, the suggested urbanization–*top1* relation was inverse-U shaped but the shape was asymmetrical with respect to the peak. In the cases of Continental/Southern Europe or the 'less-advanced' countries, the results were not as clear as in the groups of Nordic or English-speaking countries.

Dropping one country at a time from the sample did not change the previous results on the *top1*–development relation. As before, the results on lower levels of GDP per capita were sensitive. For example, dropping India away from the sample had again some effect on the shape of the curve at lower levels of development.

5.4. Controlling for two sectors

Finally, two sectors were included in the 5-year-data models to check if the overall shape in the *top1*–development relation holds. The two sectors were described using urban population and employment in services (in percentages; as described in subsection 3.2 and Appendix C).[18] First, one-dimensional smooths were used, and then complex interactions between different variables were allowed (i.e., multi-dimensional smooths were estimated). Adding complex interactions did not improve the models: see the model structures in Table 4 and compare the goodness of fit of models (1), (2) and (3). Thus, specification (1) is discussed here.

Table 4: Results on models with two sectors, using 5-year-averaged data (years 1980–2009): estimated degrees of freedom (edf) for each smooth $f_\bullet$ and coefficients for linear terms. Constant terms are not reported. See also Figure 6.

	depending variable: $top1_t$ (N=129)				
variable in period t	(1)	(2)	(3)	(4)	(5)
$f_1(\textit{log GDP pc}_t)$	[edf=4.68[a]]***	[edf= 4.84[a]]***	-	[edf=4.96[a]]***	[edf=5.40[a]]***
$f_2(\textit{urban pop}_t)$	[edf=5.69[a]]***	-	-	[linear] 0.04	[edf=4.06[a]]**
$f_3(\textit{empl. in services}_t)$	[edf=2.79[a]]***	-	-	[linear] 0.11***	[linear] 0.15***
$f_4(\textit{empl. in services}_t, \textit{urban pop}_t)$	-	[edf=6.76[b]]***	-	-	-
$f_5(\textit{log GDP pc}_t, \textit{urban pop}_t, \textit{empl. in services}_t)$	-	-	[edf=16.60[c]]***	-	-
country effects	no	no	no	fixed	random
adjusted r^2	0.698	0.663	0.692	0.931	0.934
AIC	552.2	564.8	557.4	376.7	371.7

Approximate significance levels: (***) < 0.01 , (**) < 0.05 , (*) < 0.10.
These are calculated using the Bayesian estimated covariance matrix of the parameter estimators. The p-values are based on an F statistic.
[a]The basis dimension of the smooth before imposing identifiability constraints is $k = 10$.
[b]The basis dimension of the smooth before imposing identifiability constraints is $k = 5^2 = 25$ (tensor product smooth with rank 5 marginals).
[c]The basis dimension of the smooth before imposing identifiability constraints is $k = 5 \times 5 \times 3 = 75$ (tensor product smooth with rank 5 marginals for *log GDP pc* and *urban pop*, and rank 3 marginal for *empl. in services*).

In Figure 6, plots (a)–(c) describe the smooth functions in the model (1) of Table 4. Plot (a) shows the smooth of log GDP per capita, and the downward peak at approximately 9.5 still holds. Also the shape of the relation between *top1* and urbanization in plot (b) is quite similar to the correspond-

[18]Also another measure for service sector was tested to check for sensitivity with respect to service sector measure. The pooled data results using *services, etc., value added (% of GDP)* from World Bank (2014b) were tested. These data start already from the 1960s for some countries, but Swiss data are not available. Results were not in conflict with the results reported for models (1)–(3) in Table 4.

ing prediction in Figure 5(b). Moreover, plot (c) describes a positive and significant relationship between *top1* and employment in the service sector.

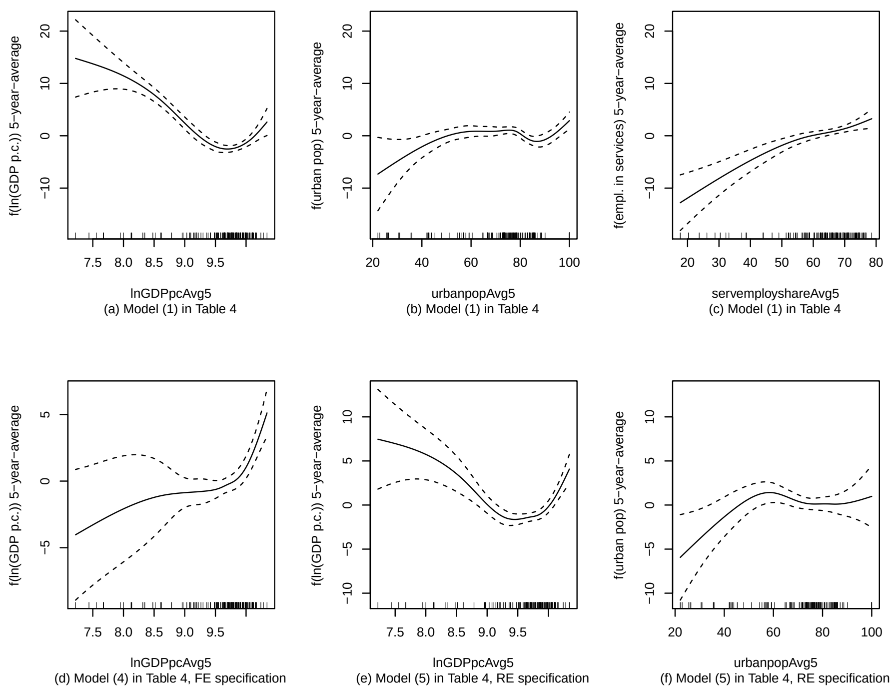


Figure 6: Illustration of the smooths, using 5-year-averaged data (years 1980–2009): the models (1), (4) and (5) in Table 4. The figures present only the smooth function and not the constant term, so the reader should focus on the shape of the prediction – not the level. The solid line represents the estimates (smooth function f). The figures also show the 95% Bayesian credible intervals (dashed), and covariate values as a rug plot along the horizontal axes.

In the previous subsections, the main findings regarding the 'later' phases of the development process did not change after including simple country-specific fixed effects (FE) or random effects (RE). The country effects were tested also here, and the results are provided in the columns (4) and (5) of Table 4. First, all variables were allowed to enter nonlinearly, but the estimation results guided toward linear terms for some sector variables: thus, some linear terms are reported in models (4) and (5). The results are easily interpreted with respect to linear terms and graphical illustrations are not needed. The coefficient for the service sector is positive and significant in both FE and RE specifications, whereas the coefficient for urbanization is non-significant in the FE specification. In Figure 6, plots (d)–(f) illustrate the smooth functions in the models with country effects. In the FE specification, a U-shape between *top1* and log GDP per capita cannot be confirmed, as the credible interval is very wide. But the positive slope at the very highest levels of GDP still holds (see plot (d)). In contrast, the RE specification maintains the U-shape in *top1*–development relation (see plot (e)). An inverse-U relation between inequality and urbanization is supported in the RE specification (see plot (f)).

In summary, the main result regarding the *top1*–development relation did not change after including controls for two sectoral measures. This is assuring because including these controls resulted in a much smaller number of observations to be used in estimation. Especially including the service sector variable shortened the time span considerably. The findings related to the sectoral measures are also intuitive in the spirit of Kuznets. The final section of this study provides discussion on the results.

6. Discussion

Kuznets (1955) proposed that inequality first increases during modernization, but later in the development process it starts to decrease. In actual data, he observed a plateau following a decline in inequality. Whereas Kuznets used income distribution data on three countries covering the first half of the 20th century, the results presented here use data covering 1900–2010. In addition, data from 26 countries are available. However, the data used here are not able to capture the relation between top 1% income shares and development (log of GDP per capita) at 'low' levels of economic development. But at somewhat higher levels of economic development various specifications suggest that the top 1% income share decreases. In addition, at even higher levels of development it increases again. This suggests a possibility for a reversal of the famous Kuznets curve at later stages of development. Can the result of increasing inequality be interpreted in the Kuznets (1955) framework?

Assume that the shift away from agriculture has already continued for a lengthy period, and that the proportion of agriculture is small. Then the overall inequality is dominated by the income distribution of the non-agricultural sector. Kuznets (1955) pointed out that the top income shares (in his text: the 5th quintile) would not decline if the per capita incomes would grow faster in the urban sector than in the agricultural sector, or if inequality in the per capita incomes of the urban sector would grow. These characteristics regarding 'urban' sector have actually been observed in many of the 'advanced' countries during the last decades. Overall inequality has increased. During the last decades, we have also observed the rapid growth of information technology and the expansion of the financial sector. One can speculate whether another type of shift in the economy has started even though this is not likely to be the sole explanation behind the evolution of top income shares. The results in this paper also support the idea that inequality first increases when the economy experiences urbanization, and then the relation becomes weaker. It is also found that as the service sector has grown inequality has increased.

Atkinson et al. (2011) and Roine and Waldenström (2014) discuss that there are problems in fitting changes in top 1% income shares into the story of shifts in the economy. Firstly, the decline in top income shares until the 1970s has been connected to capital income. In addition, the recent rise in top 1% income shares is hard to explain solely by skill-biased technological change. Thus, empirical results presented in this paper should not be taken as an exhaustive explanation of the trends in top 1% income shares. But this paper provides empirical support for a broad interpretation of the Kuznets process as new data have become available. Kuznets (1955) himself emphasized that there are various forces at play. This paper only augments the explanations suggested by, e.g., Atkinson et al. (2011) and Roine and Waldenström (2014). For example, the negative relation between development and *top1* could be (at least partly) explained by the introduction of progressive taxation. The negative relation between top marginal tax rates and top income shares has been shown empirically by Roine et al. (2009).

This paper emphasizes the results for 'advanced' countries. Investigating specific country groups revealed interesting features. The results show that the English-speaking and Nordic countries have surpassed a phase of negative relation between *top1* and development, and they now show a significant positive relation. Continental and Southern European countries show a partly similar pattern up to the point of negative relation. Due to data availability, very heterogeneous 'less-advanced' countries were pooled to discuss the development process at earlier stages of development – however, these development processes were hard to explain with the current data. Future will show if 'less-advanced' or Continental/Southern European countries will follow the same path as the Nordic and English-speaking countries seem to have taken. Specifically, more research on 'less-advanced' countries is needed when more data become available.

The results presented here are in line with the main results by Frazer (2006). The estimated curves of the *top1*–develoment relation are strikingly similar to the shape of Gini–development relation presented by Frazer. Even the downward peak around ln(GDP per capita) ≈ 9.5 that has been found in various specifications of this paper seems to be reasonable compared to Frazer's results. The similarity of these results also suggests that top 1% income shares provide a reasonable way to measure income inequality instead of the Gini coefficient, which is one of the most popular inequality indices.

Acknowledgements

This study has benefitted greatly from discussions with Ravi Kanbur, Matti Tuomala and Arto Luoma. The author also wants to thank the participants of the FDPE Econometrics Workshop II/2013, FDPE Labour and Public Economics workshop I/2014 and Annual Meeting of the Finnish Economic Association 2014. All the remaining errors are author's own.

Appendix A. Information on *top1* data

Table A.5: Sources for top 1% income share series used in this study. Series excluding capital gains have been selected whenever possible.

Country (abbreviation)	**N**	**Source**
Argentina (ARG)	39	Alvaredo et al. (2013b)
Australia (AUS)	90	Alvaredo et al. (2013b)
Canada (CAN)	91	Alvaredo et al. (2013b)[a]
China (CHN)	18	Alvaredo et al. (2013b)[b]
Colombia (COL)	18	Alvaredo et al. (2013b)
Denmark (DNK)	95	Alvaredo et al. (2013b)
Finland (FIN)	90	Alvaredo et al. (2013b)[c]
France (FRA)	96	Alvaredo et al. (2013b)[d]
Germany (DEU)	47	Alvaredo et al. (2013b)
India (IND)	71	Alvaredo et al. (2013b)
Indonesia (IDN)	28	Alvaredo et al. (2013b)
Ireland (IRL)	37	Alvaredo et al. (2013b)
Italy (ITA)	34	Alvaredo et al. (2013b)
Japan (JPN)	110	Alvaredo et al. (2013b)
Mauritius (MUS)	57	Alvaredo et al. (2013b)
Netherlands (NLD)	55	Alvaredo et al. (2013b)
New Zealand (NZL)	83	Alvaredo et al. (2013b)
Norway (NOR)	69	Alvaredo et al. (2013b)
Portugal (PRT)	24	Alvaredo et al. (2013b)
Singapore (SGP)	62	Alvaredo et al. (2013b)
South Africa (ZAF)	71	Alvaredo et al. (2013b)
Spain (ESP)	30	Alvaredo et al. (2013b)
Sweden (SWE)	80	Alvaredo et al. (2013b)
Switzerland (CHE)	74	Alvaredo et al. (2013b)[e]
United Kingdom (UK)	60	Alvaredo et al. (2013b)
United States (USA)	100	Alvaredo et al. (2013b)
	total: 1629	

Additional notes:
[a]Two overlapping series available. Here: series up to 1981 is based on tax data, and series from 1982 is based on Longitudinal Administrative Database.
[b]Urban China, information not from tax data.
[c]Two overlapping series available. Here: series up to 1989 is based on tax data, and the series from 1990 is based on the Income Distribution Survey.
[d]The figure for 1905 is for 1900–1910 averaged.
[e]For all years except 1933, the estimates relate to income averaged over the year shown and the following year. Thus, repeated value for two consecutive years is used in this study.

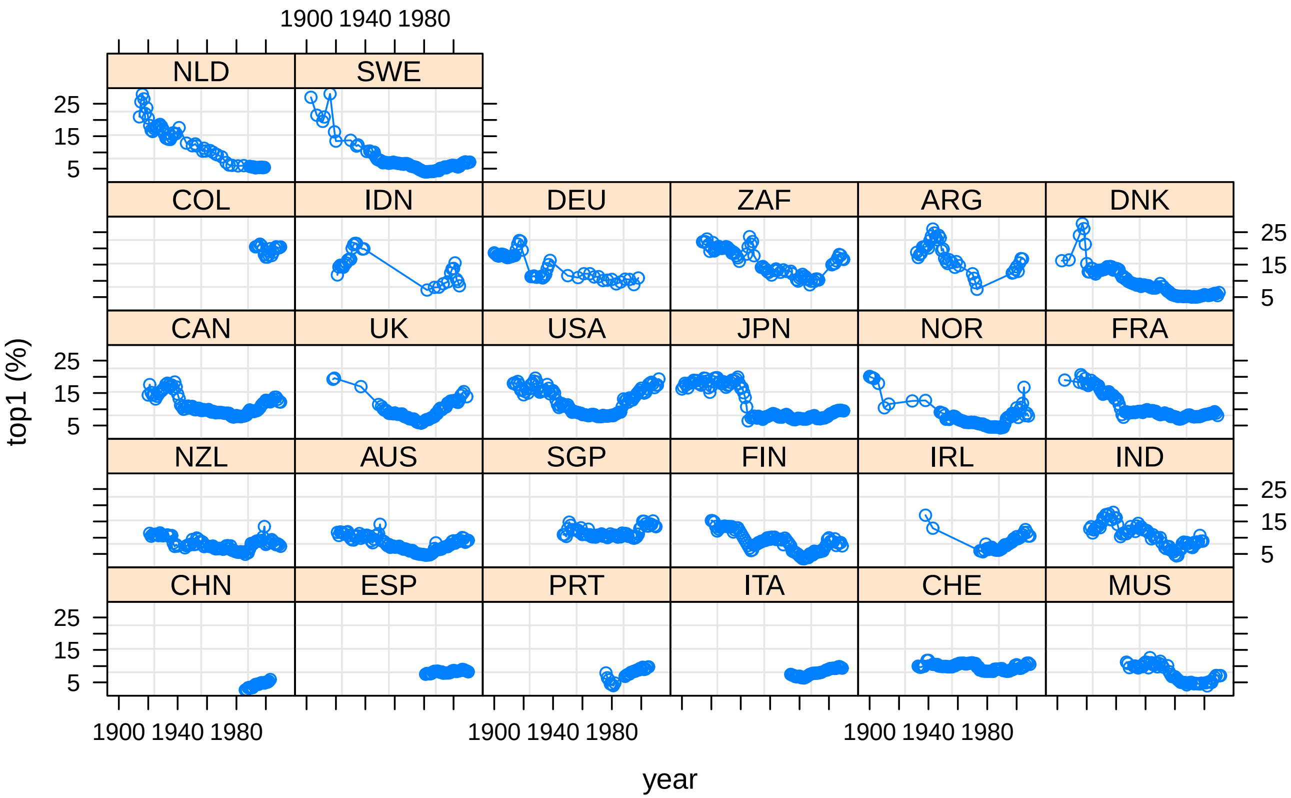


Figure A.7: Top 1% income share series for each country. See Table A.5 for country abbreviations.

Appendix B. Information on GDP variable

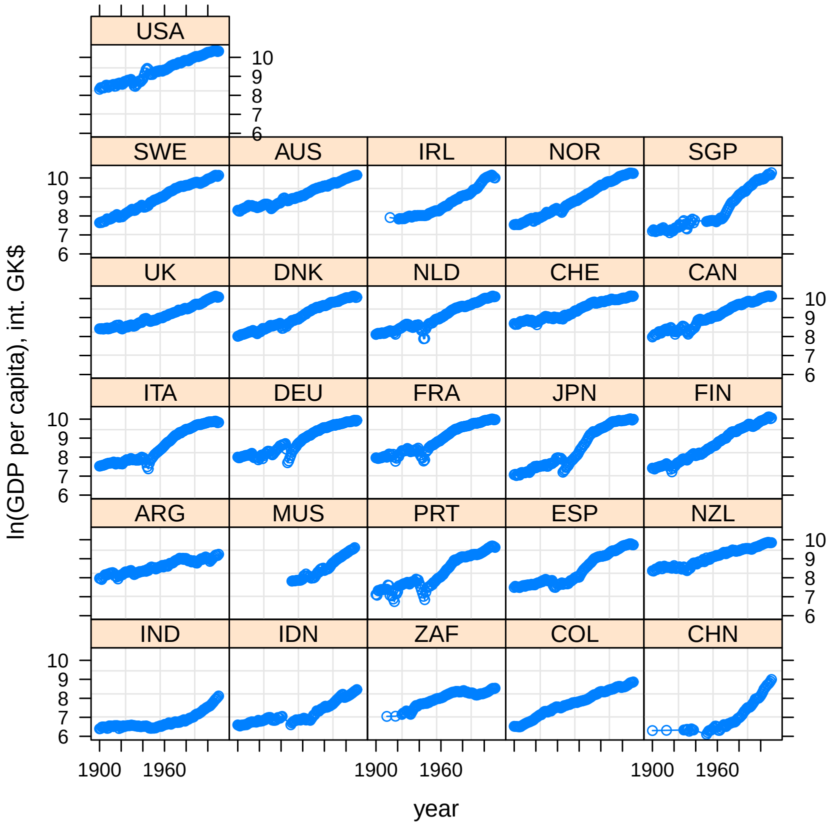


Figure B.8: Logarithm of GDP per capita in international 1990 Geary-Khamis dollars (I$) for each country. See Table A.5 for country abbreviations. Data source: update of Maddison Project (Bolt & van Zanden, 2013).

Appendix C. Information on sector variables

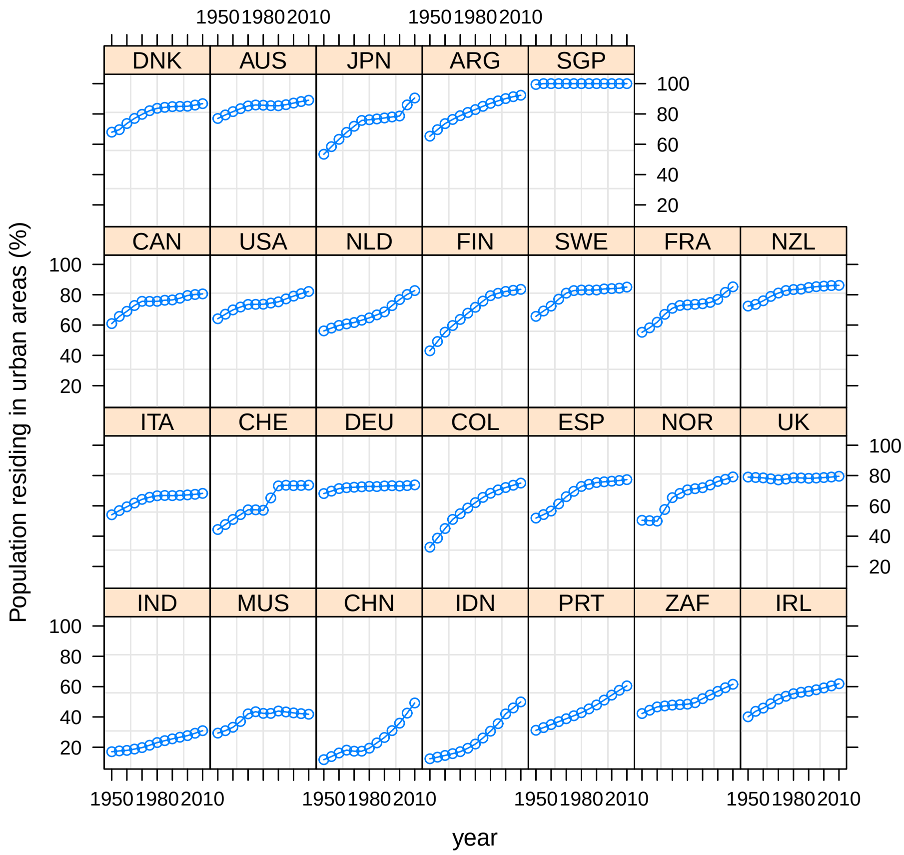


Figure C.9: Population residing in urban areas (%) is used to measure the level of urbanization. See Table A.5 for country abbreviations. Data source: United Nations (2012).

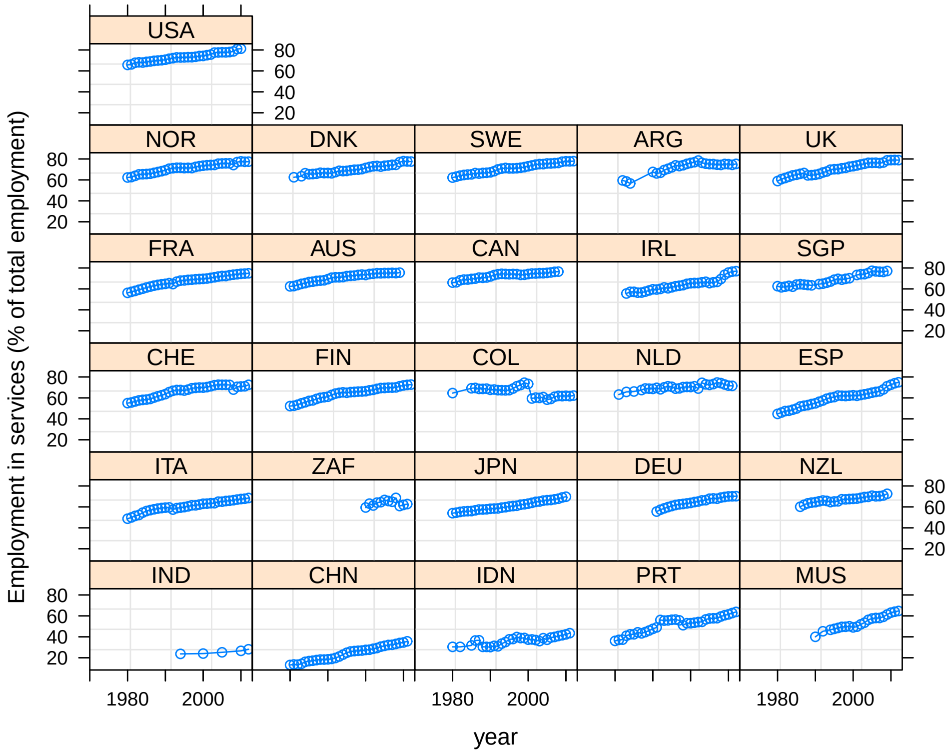


Figure C.10: Employment in services (% of total employment) is used to measure the service sector. See Table A.5 for country abbreviations. Data source: World Bank (2014a).

Appendix D. Tensor product smooths

This appendix provides additional information to section 4. The presentation here follows closely Wood (2006), and more detailed information can be found there. Tensor product smooths can be recommended if one uses a smooth that contains more than one variable but the scales of these variables are fundamentally different (i.e. measured in different units). A tensor product smooth can be presented in terms of the values of the function at a set of *knots*. Wood (2006, pp. 162–166) describes the case of a three-dimensional smooth but here the case of a two-dimensional smooth is described as an example. Tensor product smooths can be generalized to several dimensions.

Consider a smooth $f(x, z)$ comprised of two covariates, x and z. Assume that we have low-rank bases to represent smooth functions f_x and f_z of the covariates. Then, we can write

$$f_x(x) = \sum_{i=1}^{I} \alpha_i a_i(x) \quad \text{and} \quad f_z(z) = \sum_{l=1}^{L} \delta_l d_l(z),$$

where α_i and δ_l are parameters, and the $a_i(x)$ and $d_l(z)$ are known (chosen) basis functions such as cubic regression spline basis.

Consider then the smooth function f_x, and convert it to a smooth function of both x and z. This can be done by allowing α_i to vary smoothly with z. We can write:

$$\alpha_i(z) = \sum_{l=1}^{L} \delta_{il} d_l(z), \quad \text{and we get} \quad f_{xz}(x, z) = \sum_{i=1}^{I} \sum_{l=1}^{L} \delta_{il} d_l(z) a_i(x).$$

The relationship between the model matrix $\boldsymbol{X}$ for the whole model and the model matrices $\boldsymbol{X}_x$ and $\boldsymbol{X}_z$ for marginal smooths can be represented using the Kronecker product (denoted by $\otimes$). The ith row of $\boldsymbol{X}$ is $\boldsymbol{X}_i = \boldsymbol{X}_{xi} \otimes \boldsymbol{X}_{zi}$.

The penalty is a way to measure departure from smoothness. Assuming that each marginal smooth has its own penalty we can write:

$$J_x(f_x) = \boldsymbol{\alpha}^T \boldsymbol{S}_x \boldsymbol{\alpha} \quad \text{and} \quad J_z(f_z) = \boldsymbol{\delta}^T \boldsymbol{S}_z \boldsymbol{\delta}.$$

The $\boldsymbol{S}_\bullet$ matrices contain known coefficients, and $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ are the vectors of coefficients of the marginal smooths. When a penalty functional is the cubic spline penalty, then $J_x(f_x) = \int (\partial^2 f_x / \partial x^2)^2 dx$. Then consider that the 'wiggliness' (departure from smoothness) of f_{xz} can be measured by

$$J(f_{xz}) = \lambda_x \int_z J_x(f_{x|z}) dz + \lambda_z \int_x J_z(f_{z|x}) dx,$$

where, e.g., $f_{x|z}(x)$ is $f(x, z)$ considered as a function of x only, with z held constant (and $f_{z|x}(z)$ defined in a similar manner), and $\lambda_\bullet$ are the smoothing parameters that control the tradeoff between 'wiggliness' in different directions, and allowing the penalty to be invariant to the relative scaling of the covariates. When cubic spline penalties are used as the marginal penalties, then

$$J(f) = \int_{x,z} \lambda_x (\frac{\partial^2 f}{\partial x^2})^2 + \lambda_z (\frac{\partial^2 f}{\partial z^2})^2 dx dz.$$

This can be evaluated numerically.

Appendix E. Illustration: 5-year-average results for long series

The 5-year-averaged data covering 1900–2009 revealed the same type of relation that was described for annual data on these years (see subsection 5.1). The baseline model can be stated as

$$top1_{it} = \alpha + f(\ln(\text{GDP per capita})_{it}) + u_i + \epsilon_{it},$$

where α is constant, f is a smooth function, u_i is a country effect (fixed or random) and ϵ_{it} is the error term. Here i refers to country, t refers to each 5-year period, and we discuss averages.[19] Figure E.11 below describes 5-year models whose specifications resemble those described in Table 1 and Figure 2. The obtained shapes are very close to the corresponding ones in Figure 2.

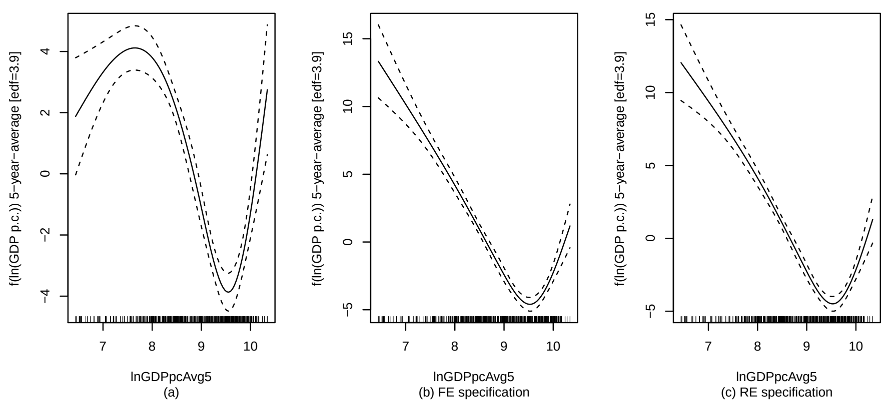
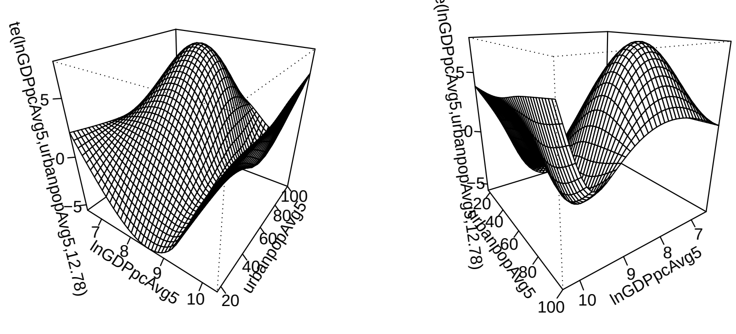


Figure E.11: Illustration of the *top1*–development relation, using 5-year-averaged data (years 1900–2009, here N=376). The figures present only the smooth function f(ln(GDP per capita)) and not the constant term, so the reader should focus on the shape of the prediction – not the level. The solid line represents the estimates (smooth function f). The figures also show the 95% Bayesian credible intervals (dashed), and covariate values as a rug plot along the horizontal axes. Plot (a) presents a model without country effects, plot (b) presents a model with country-specific fixed effects (FE), and plot (c) presents a model with country-specific random effects (RE). Compare the shapes in these plots to Figure 2.

[19]In estimation, the basis dimension of the smooth before imposing identifiability constraints was $k = 5$.

Appendix F. Additional plots: 5-year results with urbanization

The model (2) of Table 3 is illustrated from two views in Figure F.12. When one studies these kinds of plots it is useful to think of 'taking slices' with respect to one variable (i.e., holding one variable constant) to see how the surface changes as the other variable changes. The relation between the level of economic development and *top1* varies depending on the level of urbanization: at low levels of urbanization, there is a U-shape between GDP and *top1*; at highest levels of urbanization, there is a shape resembling a 'mirror image of a slanted S' (i.e., at low–middle GDP levels an inverse-U shape is proposed but at middle–high GDP levels a U shape is proposed). A U-shape between economic development and *top1* is suggested, and there is a downward peak at $9 < \ln(\text{GDP per capita}) < 10$ again. An inverted-U relation with respect to urbanization–development seems also reasonable.

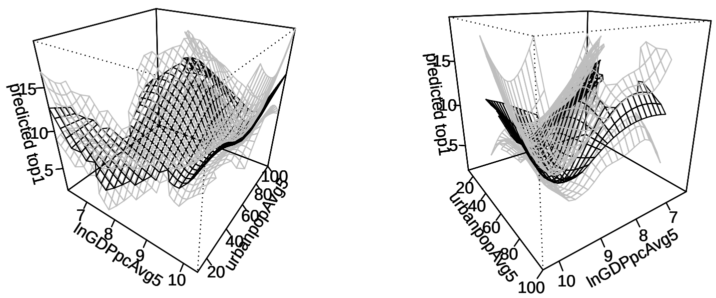


(a) Model (2) in Table 3

(b) Model (2) in Table 3

Figure F.12: Illustration of the 2-d smooth in the case of 5-year-averaged data (years 1950–2009): the model (2) in Table 3. The tensor product smooth is denoted by te(log GDP p.c., urban population, edf) on the vertical axis. The two covariates (GDP and urbanization) are on the horizontal axes. The same smooth is illustrated from two different views to get a better understanding on the shape. The figures present only the smooth function and not the constant term, so the reader should focus on the shape of the prediction – not the level. See also Figure F.13.

Figure F.13 can be used as help to investigate the 'problematic' areas: the plot grid nodes that are far from observed data points are dropped, and it is possible to study where the prediction goes far from data. This figure includes also credible intervals. As discussed in the main text, less-advanced countries' processes are harder to explain.



(a) Model (2) in Table 3 (b) Model (2) in Table 3

Figure F.13: Illustration of the prediction in the case of 5-year-averaged data (years 1950–2009): the model (2) in Table 3. The vertical axis shows the predicted top1. The two covariates (GDP and urbanization) are on the horizontal axes. The prediction is illustrated from two different views to get a better understanding on the shape. There are also two added surfaces to describe the 95% Bayesian credible intervals. Here, plot grid nodes that are too far from the points of urbanization and log GDP per capita are excluded: the grid has been scaled into the unit square along with urban population and log-GDP-per-capita variables and then grid nodes more than 0.1 from the predictor variables are excluded.

References

Aghion, P., Bolton, P., 1997. A Theory of Trickle-Down Growth and Development. Review of Economic Studies 64(2), 151-172.

Ahluwalia, M.S., 1976. Inequality, Poverty and Development. Journal of Development Economics 3(4), 307-342.

Ahluwalia, M.S., Carter, N.G., Chenery, H.B., 1979. Growth and poverty in developing countries. Journal of Development Economics 6(3), 299-341.

Alvaredo, F., 2011. A Note on the Relationship between Top Income Shares and the Gini Coefficient. Economics Letters 110(3), 274-277.

Alvaredo, F., Atkinson, A.B., Piketty, T., Saez, E., 2013a. The Top 1 Percent in International and Historical Perspective. Journal of Economic Perspectives 27(3), 3-20.

Alvaredo, F., Atkinson, A.B., Piketty, T., Saez, E., 2013b. The World Top Incomes Database. `http://g-mond.parisschoolofeconomics.eu/topincomes` (Sep 12, 2013).

Anand, S., Kanbur, S.M.R., 1993. Inequality and development: A critique. Journal of Development Economics 41(1), 19-43.

Atkinson, A.B., 1995. Incomes and the Welfare State: Essays on Britain and Europe. Cambridge University Press, Cambridge.

Atkinson, A.B. (2004a). ”Growth, Taxation and the Long-Run Evolution of Top Incomes”, mimeo, revised version of paper prepared for XVI Riunione Scientifica SIEP, Pavia, October 2004. `http://www-3.unipv.it/websiep/wp/386.pdf` (Oct 19, 2012).

Atkinson, A.B., 2004b. Income Tax and Top Incomes over the Twentieth Century. Hacienda Pública Española / Revista de Economía Pública 168(1), 123-141.

Atkinson, A.B., 2005. Top Incomes in the UK over the 20th Century. Journal of the Royal Statistical Society - Series A 168(2), 325-343.

Atkinson, A.B., 2007. Measuring Top Incomes: Methodological Issues, in: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century - A Contrast between Continental European and English-Speaking Countries. Oxford University Press, Oxford and New York, pp. 18-42.

Atkinson, A.B., Brandolini, A., 2001. Promise and Pitfalls in the Use of ”Secondary” Data Sets: Income Inequality in OECD Countries as a Case Study. Journal of Economic Literature 39(3), 771-799.

Atkinson, A.B., Piketty, T. (Eds.), 2007. Top Incomes over the Twentieth Century - A Contrast between Continental European and English-Speaking Countries. Oxford University Press, Oxford and New York.

Atkinson, A.B., Piketty, T. (Eds.), 2010. Top Incomes - A Global Perspective. Oxford University Press, Oxford and New York.

Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the Long Run of History. Journal of Economic Literature 49(1), 3-71.

Barro, R.J., 2000. Inequality and Growth in a Panel of Countries. Journal of Economic Growth 5(1), 5-32.

Bolt, J., van Zanden, J.L., 2013. The First Update of the Maddison Project; Re-Estimating Growth Before 1820. Maddison Project Working Paper 4. Data downloaded from website: `http://www.ggdc.net/maddison/maddison-project/data/mpd_2013-01.xlsx` (Sep 12, 2013).

Dahan, M., Tsiddon, D., 1998. Demographic Transition, Income Distribution, and Economic Growth. Journal of Economic Growth 3(1), 29-52.

Deininger, K., Squire, L., 1996. A New Data Set Measuring Income Inequality. World Bank Economic Review 10(3), 565-591.

Deininger, K., Squire, L., 1998. New ways of looking at old issues: Inequality and growth. Journal of Development Economics 57(2), 259-287.

Frazer, G., 2006. Inequality and Development Across and Within Countries. World Development 34(9), 1459-1481.

Galor, O., Tsiddon, D., 1996. Income Distribution and Growth: the Kuznets Hypothesis Revisited. Economica 63(250), S103-S117.

Hastie, T., Tibshirani, R., 1986. Generalized additive models (with discussion). Statistical Science 1(3), 297-318.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall/CRC, New York.

Kanbur, R., Zhuang, J., 2013. Urbanization and Inequality in Asia. Asian Development Review 30(1), 131-147.

Kuznets, S., 1953. Shares of Upper Income Groups in Income and Saving. NBER Publication No. 55, New York.

Kuznets, S., 1955. Economic Growth and Income Inequality. American Economic Review 45(1), 1-28.

Leigh, A., 2007. How Closely Do Top Income Shares Track Other Measures of Inequality? Economic Journal 117(524), F589-F603.

Maddison, A., 2010. Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD. `http://www.ggdc.net/maddison/maddison-project/`.

Piketty, T., 2001. Income Inequality in France 1901-98. CEPR Discussion Paper No. 2876.

Piketty, T., 2005. Top Income Shares in the Long Run: An Overview. Journal of the European Economic Association, 3(2-3), 1-11.

Piketty, T., Saez, E., 2003. Income Inequality in the United States 1913-1998. The Quarterly Journal of Economics 118(1), 1-39.

Piketty, T., Saez, E., 2006. The Evolution of Top Incomes: A Historical and International Perspective. American Economic Review 96(2), 200-205.

Robinson, P.M., 1988. Root-N-Consistent Semiparametric Regression. Econometrica 56(4), 931-954.

Robinson, S., 1976. A Note on the U Hypothesis Relating Income Inequality and Economic Development. American Economic Review 66(3), 437-440.

Roine, J., Vlachos, J., Waldenström, D., 2009. The long-run determinants of inequality: What can we learn from top income data? Journal of Public Economics 93(7-8), 974-988.

Roine, J., Waldenström, D., 2014. Long run trends in the distribution of income and wealth, in: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 2, North-Holland, Amsterdam, forthcoming.

United Nations, Department of Economic and Social Affairs, Population Division, 2012. World Urbanization Prospects: The 2011 Revision (POP/DB/WUP/Rev.2011/1/F2). Website: `http://esa.un.org/unpd/wup/CD-ROM/WUP2011-F02-Proportion_Urban.xls` (Nov 28, 2013).

Wood, S.N., 2006. Generalized Additive Models - An Introduction with R. Chapman & Hall/CRC, Boca Raton FL.

World Bank, 2014a. World Development Indicators, Employment in services (% of total employment). Website: `http://data.worldbank.org/indicator/SL.SRV.EMPL.ZS/` (Feb 25, 2014).

World Bank, 2014b. World Development Indicators, Services, etc., value added (% of GDP). Website: `http://data.worldbank.org/indicator/NV.SRV.TETC.ZS/` (Feb 25, 2014).