CESifo Working Papers
www.cesifo.org/wp

Spatial Inequality and Development – Is there an Inverted-U Relationship?

Christian Lessmann

CESifo Working Paper No. 3622
Category 6: Fiscal Policy, Macroeconomics and Growth
October 2011

CESifo *Working Paper No. 3622*

Spatial Inequality and Development – Is there an Inverted-U Relationship?

Abstract

This paper studies the hypothesis of an inverted-U-shaped relationship between spatial inequality and economic development. The theory of Kuznets (1955) and Williamson (1965) suggests that (spatial) inequality first increases in the process of development, then peaks, and then decreases. To test this hypothesis I have used a unique panel data set of spatial inequalities in 55 countries at different stages of economic development, covering the period 1980-2009. Parametric and semiparametric regressions are carried out using cross-section and (unbalanced) panel data. The results provide strong support for the existence of an inverted U, but importantly I also find spatial inequalities to increase again at very high levels of economic development. Although many factors may be contributing to this rise, one explanation rests on the process of tertiarization, i.e., the structural shift from industrial production towards a service base in the highest-developed economies.

JEL-Code: D310, R100, O150.

Keywords: regional inequality, Kuznets curve, panel data, semiparametric estimates.

Christian Lessmann
Technical University Dresden
Faculty of Business and Economics
01062 Dresden
Germany
christian.lessmann@tu-dresden.de

This version: October 2011
I would like to thank Susan Kircheis for helping me with the data collection and preparation.

1 Introduction

The growing spatial income inequalities around the world have begun to attract considerable interest among academics and politicians. Spatial inequalities are important for at least two reasons. The first reason is that interregional inequality increases undesirable interpersonal (or overall) inequality. The second reason is that interregional inequality often goes hand in hand with political and ethnic tensions which undermine social cohesion and political stability [Kanbur and Venables (2005a)].

Despite the importance of spatial inequality for policy concerns, little is known about the its determinants.[1] One of the most important theories on the determinants of regional inequality dates back to Kuznets (1955) and Williamson (1965). In his seminal paper, Kuznets conjectured that as countries develop from farm-based economies to industrial economies, income inequality first increases, then peaks, and then decreases. Thus, the trajectory of this relationship is inverted-U-shaped – what we call the *Kuznets curve* today. The reason is that in an early stage of development, very few people benefit from the increasing investment in physical capital, and income inequality increases. At a later stage of development, more and more workers shift from the agricultural sector to the industrial sector, and income redistribution takes place, so that inequality falls. Williamson adopted this idea for the case of spatial inequality. He argues that the industrialization was driven by the discovery and utilization of natural resources such as coal and iron. Those natural resources are often not equally distributed within countries (think of the western regions of France or the Ruhr area in Germany). The economic prosperity in the industrialization process is therefore also unequally distributed within countries, so that regional inequalities rise in this process. At a later stage of development, the more attractive employment opportunities in the booming regions attract workers from abroad, depressing wages in destination regions but increasing wages in home regions. Thus, a natural convergence process starts, possibly encouraged by government policies; therefore regional inequality falls, creating again an inverted-U-shaped relationship.

Surprisingly, only a few empirical studies – such as Williamson (1965), Amos (1988), Ezcurra and Rapun (2006), or Barrios and Strobl (2009) – have tried to provide evidence for the Kuznets curve in spatial inequality. A major reason for the scarcity of research in this field is the poor availability of regional economic accounts, which are necessary to calculate inequality measures. In the case of OECD countries, data collection is quite easy, since the OECD Regional Statistics, EUROSTAT, and Cambridge Econometrics (CAMECON) provide many useful regional data. Unfortunately, data collection is more difficult for other countries than OECD member states, since no international database contains the relevant information. The major problem for this kind of research is that it is essential either to have historical data for single countries or to include poorer countries in a cross-country data set, since the theories of Kuznets and Williamson point at the deep structural changes associated with the industrialization process. In this paper I use a unique panel data set on spatial inequality, which covers 55 countries at very different levels of economic development

[1] See Barrios and Strobl (2009) for an overview of theoretical studies concerning the relationship between spatial inequality and development.

for the period 1980–2009 to investigate this hypothesis. I have collected a new data set on spatial inequality around the world, where much of the regional data was provided by national statistical offices or central banks on individual request. I show, based on cross-country as well as panel data, that the relationship between spatial inequality and economic development has a nonlinear, inverted-U-shaped trajectory implying that economic growth first increases spatial inequality, and later – at higher stages of economic development – inequalities fall. But importantly, regional inequalities increase again at very high levels of economic development (25,000 US$), which may be related to tertiarization, i.e., a shift from manufacturing industries to service ones. This result is obtained in standard parametric regressions using polynomial functions of the income variable as well as in regressions which use a more flexible semiparametric approach. The inverted U with increasing spatial inequalities at very high levels of development was detected in a cross-section of countries as well as in panel fixed effect regressions. Thus, the inverted-U-shaped relationship holds for differences between countries and for changes within countries over time. Importantly, the general finding is robust to the inclusion of a number of covariates the literature has shown to affect spatial inequality.

The remainder of the paper is organized as follows: Section 2 surveys the existing theoretical and empirical literature on the link between spatial inequality and economic development. Section 3 presents the unique data set on regional inequality and discusses measurement issues. Section 4 provides the econometric analysis. Finally, Section 5 sums up the results and concludes.

2 Spatial inequality and development: Existing theory and evidence

Williamson (1965) was the first who suggested an inverted-U-shaped relationship between spatial inequality and economic development. Based on the ideas of Kuznets (1955), who studied personal inequality, he argues that regional inequalities are affected in quite a similar manner. The spatial concentration of wealth- and income-generating resources results in increasing regional inequalities in early stages of economic development, followed by a more widespread dispersion of income in later stages. Following Williamson, four reasons are decisive for the evolution of spatial inequalities: natural resources, migration, capital mobility, and government policies. He argues that most natural resources are point resources and thus are unequally distributed among the different regions of a country. A discovery of new resources will then increase unbalanced development of regions, and a selective influx of labor and capital, perhaps encouraged by government policies, will lead to a further increase in spatial inequality. At later stages of economic development, new resources will be discovered in less developed regions (or the demand for existing resources will increase), and government policies will concentrate on lagging regions, so that the process is reversed. Based on these ideas, he formulates the hypothesis

> that the early stages of national development generate increasingly large North-South income differentials. Somewhere during the course of development, some or all of the

> disequilibrating tendencies diminish, causing a reversal in the pattern of interregional inequality. Instead of divergence in interregional levels of development, convergence becomes the rule, with the backward regions closing the development gap between themselves and the already industrialized areas. [Williamson (1965), p. 9]

Thus, the relationship between spatial inequality and economic development is expected to be inverted-U-shaped. Williamson was able to find support for this hypothesis in cross-country data.

A more formal foundation of the inverted-U hypothesis is provided by Barrios and Strobl (2009), based on Lucas (2000). Their model analyzes the dynamics of regional growth after a technological shock (innovation) takes place which initially benefits one region. Growth is accelerated in this leading region, implying that regional inequalities increase. The other regions will follow the leading region with a lag, whose magnitude depends on differing technological capabilities. Those lagging regions which adopt the new technology will grow at the rate of the leading region plus an additional growth effect determined by the natural rate of convergence. Thus, regional inequalities increase, peak, and decrease. It is important to note that while the argumentation of Kuznets and Williamson applies to long-lasting structural changes, this framework suggests an inverted-U-shaped relationship even in a shorter time horizon.

Amos (1988) discusses the inverted-U hypothesis for U.S. intra-state inequalities. Since the U.S. has already reached a very high level of economic development, he argues that the inverted-U pattern in the scheme of Kuznets and Williamson must have been completed. Thus, he is primarily interested in what happens after the inverted U: stabilization or increase of regional inequalities (a further decrease would just imply that the inverted-U pattern has not been completed yet). The neoclassical growth theory suggests stabilization. Amos argues that neoclassical factor market adjustment mechanisms had more than 100 years to compensate for the disequilibrating technological shocks caused by the industrial revolution, and therefore regional inequality should have stabilized. In contrast, increasing inequalities may reflect other aspects of regional development not covered by the neoclassical theory: urban decay, suburbanization, rural decline, etc. Increasing inequalities may, however, also be the result of disequlibrating shocks and the beginning of a new inverted-U process which follows the initial one. The empirical findings of Amos point at increasing inequalities within U.S. states. Interestingly, this finding is consistent with studies of personal inequality such as List and Gallet (1999).

As mentioned in the introduction, existing empirical evidence on the relationship between spatial inequality and economic development is scarce. The highly influential study by Williamson (1965) was the first to explore a possible inverted-U-shaped relationship between spatial inequality and development. Williamson analyzed cross-country and time series data of 24 countries, including a number of developing countries such as Indonesia, India, and several South American countries. The evidence supports the hypothesis of an inverted U. A more recent cross-country study is provided by Ezcurra and Rapun (2006), who consider 14 Western European countries for the period 1980–2002. Using a semiparametric approach, the authors do not find an inverted-U-shaped relationship between spatial inequality and development. As the authors state, this is not surprising, since all considered countries have reached a high level of development. But there is some evidence that increases in GDP coincide with a decrease in spatial inequality at the beginning of the observation period, indicating that the inverted-U pattern had not been completed at that time. At later stages of economic development, spatial inequalities tend to stabilize. A related study has been carried out by Barrios and Strobl (2009), who focus on 12 EU countries for the period 1975–2000. Although only highly developed countries are considered, they find evidence of an inverted U, based on a semiparametric regression approach.

Besides the cross-country studies, there also exist some studies on single countries. Amos (1988) analyzed spatial inequality within U.S. states, finding that inequalities increased with development. The major finding is supported by Fan and Casetti (1994). Another case study is provided by Terrasi (1999) for Italian regions (1953–1993). Similarly to the U.S. case, her parametric estimates point at a U-shaped relationship between spatial inequality and development. Terrasi's interpretation of the empirical finding is that a "new era of divergence has begun in connection with the emergence of high technology industries and producer services as the new leading sectors" [Terrasi (1999), p. 508]. Altogether, the studies of highly developed countries point at increasing spatial inequalities at very high levels of economic development.

The discussion of the existing literature shows that no study has been carried out since Williamson (1965), which looks at countries at different levels of economic development. This is surprising insofar that the original theory of Kuznets and Williamson aims to explain the effect of deep structural changes, which are difficult to isolate in high-income economies such as those of western European countries or the U.S. without having historical regional accounts. The aim of this paper is to fill this gap in the literature. For this purpose, a unique data set on spatial inequality was collected, as described in the following section in detail. My reexamination of Williamson's work is also interesting in that several studies point at increasing spatial inequalities after the inverted-U pattern has completed. Based on the new cross-country data, this hypothesis can be tested for a wide range of countries. Of course, today's available econometric methodologies have several advantages over those of the 1960s, so that one might have more trust in the new findings. However, to make the results comparable to the initial study of Williamson, I conduct parametric regressions as well as semiparametric regressions, which are commonly used in the recent literature.

3 Spatial inequality around the world – a new data set

Spatial inequality matters because it is an important determinant of interpersonal income inequality. Yemtsov (2005) and Elbers et al. (2005) estimate that regional inequality explains about one-third of interpersonal income inequality. But spatial inequality also matters because it might be a consequence of ethnic discrimination or market failures such as excessive migration [Mills and Ferranti (1971), Boadway and Flatters (1982)].

The discussion in the introductory section mentions the problems related to data availability, but even if one has access to suitable regional data, the measurement of spatial inequality is difficult.

I resort to the weighted coefficient of variation (WCV) of regional GDP per capita (p.c.), which is widely used in the literature on spatial inequality [see, e.g., Williamson (1965), Ezcurra and Rapun (2006), Rodríguez-Pose and Ezcurra (2010)]:[2]

$$WCV: \quad = \quad \frac{1}{\bar{y}} \left[\sum_{i=1}^{n} p_i \left(\bar{y} - y_i \right)^2 \right]^{1/2}, \tag{1}$$

where $\bar{y}$ is the country's average GDP p.c., y_i is the GDP p.c. of region i, p_i is the share of the country's total population in region i, and n is the number of spatial units.[3] The advantages of this measure are that it is mean-independent, independent of the sizes and the number of spatial units, and robust against single extreme observations, and that it satisfies the Pigou-Dalton transfer principle [Dalton (1920), Pigou (1912)], which states that a transfer from rich to poor regions should reduce the inequality measure [see Sen (1973) and Mehran (1976) for details]. Other commonly used inequality measures such as the (log of the) standard deviation of regional GDP p.c., which are commonly used in the literature on growth and convergence [see, e.g., Barro and Sala-i-Martin (1992), Sala-i-Martin (1996)], are less appropriate in our context, since they cannot account for the heterogeneity of regions with respect to population size. This is a very important issue here, due to the lack of a uniform territorial classification for all countries in the data set. In countries with large economic differences and a very unequally distributed population, an unweighted inequality measure might be difficult to interpret. An example should illustrate the problem. The northern Canadian Territories are much poorer than the provinces to the south, so that an inequality measure might indicate large economic differences, although very few people are actually poor (note that the Territories are inhabited by only 100,000 people in total). Therefore, it is necessary to calculate a concentration measure such as WCV, which incorporates the different sizes of spatial units within a country. Another way to attack this problem is using a homogeneous territorial classification. Therefore, I refer to regional data based on OECD TL2 or NUTS2 level for OECD member countries.[4] Note that for non-OECD countries, where only state- or province-level data is available, predetermining the territorial level becomes increasingly important.

I have calculated the weighted coefficient of variation based on the regional GDP p.c. for 55 countries covering the period 1980–2009. Note that the frequency of the data varies by country: for the OECD countries the underlying panel is almost balanced, but for developing countries there are quite large gaps in the data. Table 1 presents the means of these calculations for the most recent years, 2000–2009, subdivided by the different regions of the world following the standard World Bank classification. A first observation from the summary statistics is the link between spatial inequality and development. High-income countries in the core of Europe, Scandinavia, and North America have much lower spatial inequalities than low- and middle-income countries

[2] See Bendel et al. (1989) for a comparison of standard inequality measures.

[3] Note that the Theil index is not applicable for cross-section analysis with large variations in the number of sub-national units of the countries considered, since its values range from 0 to $\ln n$ [Hale (2003)].

[4] A complete list of countries, territorial levels, periods covered, and sources is provided in Table A.1 in the appendix. NUTS refers to *Nomenclature of Territorial Units for Statistics*. Note that I have used the NUTS3 territorial level in the cases of Latvia, Lithuania, and Malta, since NUTS2 data is not provided. OECD territorial level 2 refers to macro-regions.

Table 1: Spatial inequality around the world

Country	WCV	Country	WCV
Europe & Central Asia		*North America*	
Austria	0.20	Canada	0.16
Belgium	0.35	United States of America	0.17
Bulgaria	0.29	Average	0.17
Croatia	0.21	*Latin America & Caribbean*	
Czech Republic	0.39	Bolivia	0.29
Denmark	0.11	Brazil	0.48
Finland	0.17	Chile	0.35
France	0.29	Colombia	0.46
Georgia	0.19	Mexico	0.59
Germany	0.20	Panama	0.46
Greece	0.13	Peru	0.42
Hungary	0.40	Average	0.44
Ireland	0.17	*East Asia & Pacific*	
Italy	0.27	Australia	0.09
Kazakhstan	0.75	China	0.51
Latvia	0.53	Indonesia	0.89
Lithuania	0.30	Japan	0.13
Netherlands	0.14	Korea, Rep. (South)	0.06
Norway	0.32	Mongolia	0.67
Poland	0.25	New Zealand	0.07
Portugal	0.25	Philippines	0.62
Romania	0.39	Thailand	0.88
Russian Federation	0.37	Average	0.44
Slovak Republic	0.46	*South Asia*	
Slovenia	0.18	India	0.42
Spain	0.21	*Sub-Sahara Africa*	
Sweden	0.21	South Africa	0.41
Switzerland	0.20	Tanzania	0.37
Turkey	0.43	Average	0.39
Ukraine	0.58	*Middle East & North Africa*	
United Kingdom	0.37	Iran, Islamic Rep.	0.56
Uzbekistan	0.51	Malta	0.07
Average	0.31	Average	0.31

in South America and Asia. But there are also interesting variations within the different country groups; for example, among the European countries the United Kingdom and Belgium have quite high spatial inequalities, while Denmark and the Netherlands are much more homogeneous.

Not only inequality levels are relevant for this kind of analysis, but also the development over time. Figure 1 shows the weighted coefficient of variation for China, India, Bolivia, Latvia, Russia, and the United States. Importantly, spatial inequalities vary quite a lot within countries over time,

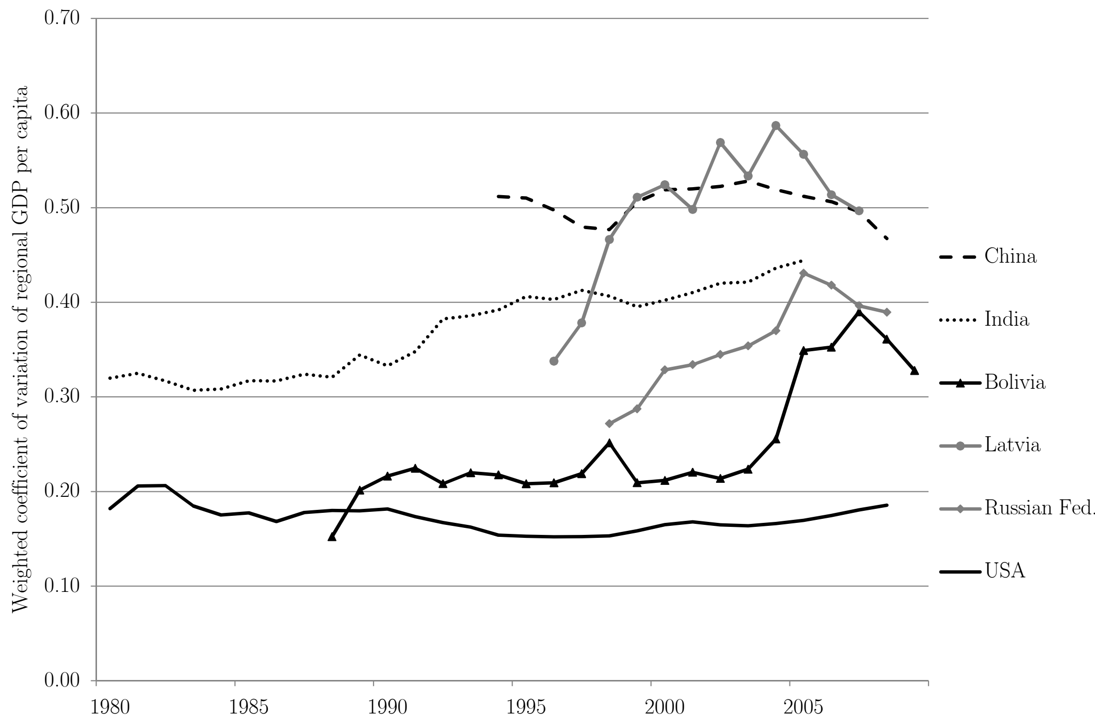

Figure 1: Trends in regional inequality 1980-2009

which is important for the panel data analysis conducted below. Concerning the individual time series, an interesting case is China, where the development of spatial inequality since the end of the 1990s has some resemblance to an inverted-U-shape curve. Jian et al. (1996) find, based on long time series data on Chinese provinces, that regions converged before the Cultural Revolution, diverged during it, and subsequently converged again. China has experienced rapidly increasing spatial inequality since the opening of the country to the world market in the 1990s, which was accompanied by fast economic growth [see Chen and Fleisher (1996), Wei et al. (2009), and Kanbur and Zhang (2005)]. Since then, spatial inequality has been on the decrease again, perhaps caused by the Western Development Program, which aims to restore a more balanced regional development [Lessmann (2011)]. This finding supports the ideas of Kuznets and Williamson.

We can also learn from the experience of Bolivia, which faced rapidly increasing spatial inequalities because the regions of the country have benefited differently from increasing resource revenues. But since the central government gained control of the natural gas resource revenues in 2006, spatial inequality has started to decrease again. Also, the data of Latvia and the Russian Federation, which both faced a rapid structural change after the breakup of the Soviet Union, support the hypothesis of a nonlinear relationship between spatial inequality and development. The case of India shows that the strong growth period, which started at the beginning of the 1990s, has increased spatial inequality significantly, and no turning point of this trend has been reached yet. The examples imply that there is quite a lot of variation in my measure of spatial inequality. However, as the example of the United States illustrates, this does not apply to all countries in the data set. This is one of the reasons why I explore cross-country data, which focuses on the variation of spatial inequality *between* countries, as well as panel data, which focuses on the variation *within* countries [Wooldridge (2002)].

4 Econometric analysis

4.1 Methodology

This study uses two different approaches to test for the pattern of an inverted-U-shaped relationship between spatial inequality and development. First, I examine cross-section data as presented in section 3, covering the period 2000–2009. The theory of Kuznets and Williamson suggests that less-developed (more-developed) countries tend to fall along the positively (negatively) sloped region of the Kuznets curve, which can be tested in a cross-country framework focusing on between-country differences. Here I follow Williamson (1965). Second, I analyze the (unbalanced) panel data set covering the period 1980–2009. Using panel data has the advantage that I can eliminate unobserved heterogeneity between countries by including country fixed effects [Baltagi (2005)]. Since there exist numerous unobservable factors driving spatial inequality, which might cause an omitted variable bias, this is very important for maintaining the quality of the analysis. Examples include geographic factors such as fragmentation, mountains, coasts, deserts, etc., which are determinants of spatial inequality, but difficult to consider in an econometric analysis which focuses on the variation in time. The country dummies capture all of these country-specific determinants of spatial inequality. In contrast to the cross-section estimations, panel regressions concentrate on within-country variations, which are very important here because they consider the dynamics of structural changes.

Concerning the estimation procedure, I consider two different econometric methods: a parametric ordinary least squares approach and a semiparametric partially linear model. The econometric representation of the Kuznets curve in the parametric regression approach is given by

$$WCV_{i,t} = \alpha_i + \sum_{j=1}^{k} \beta_j Y_{i,t}^j + \sum_{m=1}^{q} \gamma_m X_{m,i,t} + \mu_t + \epsilon_{i,t}, \qquad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \quad (2)$$

where $WCV_{i,t}$ is the weighted coefficient of variation of the regional GDP per capita for country i at time t; $Y_{i,t}$ is the log of the GDP p.c. at the country level, which enters the regression in a polynomial form of degree j; $X_{m,i,t}$ represents q different control variables; α_i are the estimated country fixed effects; μ_t are time fixed effects; and $\epsilon_{i,t}$ is random error term. The coefficients of interest are the β_j. For $k = 2$ the polynomial is quadratic, and I expect β_1 to be positive and β_2 to be negative, implying an inverted-U-shaped relationship between spatial inequality and development. But as Amos (1988) shows, spatial inequality might increase at high levels of development after the inverted-U pattern has been completed; therefore I also consider polynomials of higher degree. The estimation equation of the cross-sectional model looks similar to the panel data model represented by equation (2), but it has no time dimension t and no country and time fixed effects.

The parametric regression model described above has the advantage that it directly tests for the existence of an inverted U as suggested by the theory. The functional form of the effect of economic development Y on spatial inequality WCV is given by the polynomial of degree j. But by simply using higher order terms to estimate a possibly nonlinear relationship we place a fairly strong restriction on the possible link between the variables of interest that may not reflect the true underlying relationship. As suggested by Durlauf (2001), semiparametric methods are the more appropriate approach for studies of growth and convergence, because of parameter heterogeneity, which means that the effect of one variable on another cannot be captured by a constant coefficient, since the marginal effect varies by country or with other variables. By using a semiparametric approach one can investigate the possible nonlinear effect of economic development on spatial inequality in a flexible way, while simultaneously allowing for linear effects of other conditioning variables.[5] The equation to be estimated has the following form (omitting subscripts for reasons of clarity):

$$WCV = \alpha + f(Y) + \gamma X + \epsilon, \tag{3}$$

where X is a set of explanatory variables that are assumed to have a linear effect on WCV, $f(\cdot)$ is an unknown smooth function of Y, which I expect to be nonlinear, and ϵ is a random error term. Thus, γX represents the parametric and $f(Y)$ the nonparametric part of the model. I use the approach proposed by Yatchew (1997) to fit the partial linear model, which consists of four steps: (1) The data is sorted by ascending values of Y, and first differences of all the sorted data are calculated. (2) The γ parameters are computed with OLS, using the differences of X and WCV variables. (3) The original dependent variable WCV is adjusted for the linear effects by calculating $WCV - \hat{\gamma}X$. (4) The resulting "purged" dependent variable is used for a local linear regression on the independent variable Y to obtain an estimate of $f(\cdot)$.[6] Note that the use of higher order differences increases the efficiency of the estimator.

Both approaches, the parametric and the semiparametric one, allow for considering additional control variables that might affect spatial inequality. I control for the number of spatial units which has been used to calculate the inequality measures, since the territorial level is not always comparable over the whole set of countries considered. Controlling for the number of regions should reduce any bias caused by a measurement error related to this problem.[7] Also related to spatial issues is the log of area in square kilometers, which is used as an additional explanatory variable. A further determinant of spatial inequality may be the heterogeneity of the population living in the different parts of a country, since the different regions are often inhabited by different ethnic groups. Think of Belgium, with the Dutch-speaking Flemings living in the north and the French-speaking Walloons in the south, or India, with the Indo-Aryans in the north and Dravidians in the

[5] Yatchew (1998) and DiNardo and Tobias (2001) provide a very helpful discussion of semiparametric methods.

[6] For Stata I have used the plreg program by Lokshin (2006), as in other studies such as Araujo et al. (2008) and Lambert et al. (2010). See, e.g., Robinson (1988) for an alternative estimation approach for semiparametric models.

[7] I have also experimented with the average size of regions (number of units divided by area) and other indicators for fragmentation, but the number of units itself turned out to be the most important determinant.

south. As discussed in Kanbur and Venables (2005b), ethnic fractionalization may result in ethnic discrimination or conflict, promoting the divergence of regions. Thus, I control for the degree of ethnolinguistic fractionalization as calculated by Alesina et al. (2003). The works of Krugman and Elizondo (1996), Gianetti (2002), and Rodriguez-Pose and Gill (2006) suggest that trade openness affects spatial inequality. Therefore, I control for the sum of imports and exports as a share of the GDP. To capture agglomeration effects, I control for the urban share of the population.[8] Last but not least, I draw on the literature on decentralization and spatial inequality and include a dummy variable for federal countries [see, e.g., Lessmann (2009) and Rodríguez-Pose and Ezcurra (2010) for recent contributions].

4.2 Parametric regressions

4.2.1 Cross-section data

The results of different specifications of the parametric cross-section estimates as given by equation (2) are presented in Table 2. Note that I have used period averages of all considered variables for the period 2000 to 2009. I have refrained from using a longer period for averaging because the data of middle- and low-income countries is largely confined to recent years, so that this period is the most complete one.

Column (1) reports results of a bivariate model without any control variables, showing that more highly developed countries (in terms of the log of the GDP p.c.) have lower spatial inequalities. In column (2) the GDP enters in a quadratic form [$k = 2$ in equation (2)]. The coefficients have the expected signs (β_1 is positive and β_2 is negative), but neither of these coefficients is statistically significant, nor is their joint effect [see Brambor et al. (2006) on how to calculate marginal effects in interaction models]. In column (3) the two control variables which control for spatial effects (log of the number of spatial units and log of the total area in square kilometers) are added as explanatory variables. These control variables seem to be very important, since the explanatory power of the regression model increases significantly, as indicated by the adjusted R^2. Also, the coefficients β_1 and β_2 reach significance or are close to it. In column (4) all discussed control variables enter the regression. Thereby both main coefficients of interest are significant and also the control variables show expected signs. This specification of the model supports the hypothesis of an inverted-U-shaped relationship between spatial inequality and economic development.

However, as the discussion in section 2 has shown, there might be increasing spatial inequalities at higher levels of economic development. I thus consider a third order polynomial in the estimations reported in column (5).[9] All main coefficients of interest are statistically significant in these regressions. The signs of the coefficients β_1 and β_2 remain unchanged, and the sign of β_3 is positive, implying that spatial inequality increases after the inverted-U pattern has been completed, and thus supporting earlier findings of Amos (1988) and others.

[8] Following Lessmann (2009), I have also used a concentration measure of the population within countries, which does not turn out to affect the regional inequality significantly.

[9] I was not able to find any significant effects using higher order polynomials.

Table 2: Cross-section data: parametric estimates

Dependent variable: Weighted coefficient of variation of regional GDP p.c. (WCV)	(1)	(2)	(3)	(4)	(5)
log(GDP p.c.)	−0.097***	0.145	0.282	0.338*	3.764**
	(−5.77)	(0.68)	(1.43)	(1.75)	(2.58)
(log(GDP p.c.))^2		−0.014	−0.022*	−0.021*	−0.439**
		(−1.18)	(−1.88)	(−1.90)	(−2.41)
(log(GDP p.c.))^3					0.017**
					(2.25)
log(spatial units)			0.067**	0.108***	0.108***
			(2.48)	(4.82)	(4.63)
log(total area)			0.005	0.038***	0.030**
			(0.46)	(2.84)	(2.23)
ethnic fractionalization				0.183*	0.167*
				(1.90)	(1.87)
trade/GDP ratio				0.003***	0.003***
				(4.92)	(4.73)
urbanization				−0.003*	−0.003**
				(−1.85)	(−2.17)
federal dummy				−0.092**	−0.073**
				(−2.46)	(−2.18)
constant	1.199	0.191	−0.671	−1.794**	−10.860***
	(7.52)	(0.21)	(−0.77)	(−2.06)	(−2.90)
Obs. (N)	55	55	55	55	55
Adj. R^2	0.428	0.432	0.498	0.669	0.692

Note: t-values are reported in parentheses; standard errors are calculated using White correction; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

4.2.2 Panel data

The cross-section evidence reported in the previous section supports the inverted-U hypothesis, but – in addition – suggests that regional inequalities rise again at very high levels of economic development. Note that this empirical approach has focused on between-country variations, while the theory of Kuznets and Williamson is ultimately related to the development of regional inequalities within a country in time. Thus, a panel data approach may be better suited to test this theory. Moreover, a panel data analysis allows us to consider country fixed effects, thereby eliminating unobserved country heterogeneity. Unobserved heterogeneity is very likely to occur in this framework, since there are a number of geographical, political, and social determinants which I cannot control for because several potential control variables are time invariant by nature. Another advantage of panel data is that there are econometric procedures available which reduce any endogeneity bias. In particular, the level of economic development, which is our main explanatory variable of interest, may also be affected by regional inequalities. Alesina and Rodrik (1994) have demonstrated, for the case of interpersonal income inequalities, that inequalities also affect growth and development through government policies. Such a situation may also occur in case of regional inequalities – for example, if the government increases tax rates in order to finance redistributive policies which aim at reducing regional inequalities. Then the increased taxation can harm overall economic growth and development, and thus an endogeneity problem occurs. To take this potential problem into account, I suggest two different approaches. The first is simply to use a lag of the independent variable in the OLS regressions. This approach is similar to that of Barrios and Strobl (2009). The second approach is to employ a dynamic panel estimation using a difference GMM estimator, which uses lagged levels of the endogenous regressor as instruments. Note that this estimation procedure removes the country fixed effects through the first-differencing of the regression equation, and it includes a lag of the dependent variable as explanatory variable. The results of different specifications of equation (3) using the different estimation procedures are provided in Table 3.

Table 3: Panel data: parametric estimates

	Dependent variable: Weighted coefficient of variation of regional GDP p.c.					
	OLS	OLS[a]	OLS	GMM	GMM[b]	GMM
	(1)	(2)	(3)	(4)	(5)	(6)
WCV_{t-1}	–	–	–	0.077	0.499***	0.648***
				(0.37)	(3.48)	(7.43)
log(GDP p.c.)	0.344**	0.328**	0.087	0.389**	0.279**	−0.424
	(2.41)	(2.31)	(0.09)	(2.11)	(2.10)	(−0.87)
$(\log(\text{GDP p.c.}))^2$	−0.019**	−0.019**	0.012	−0.021**	−0.015	0.065
	(−2.08)	(−2.10)	(0.09)	(−2.02)	(−1.63)	(1.00)
$(\log(\text{GDP p.c.}))^3$	–	–	−0.001	–	–	−0.003
			(−0.22)			(−1.18)
trade/GDP	0.001	0.001*	0.001*	−0.001	−0.001	−0.001
	(1.59)	(1.72)	(1.79)	(−1.01)	(−0.85)	(−1.50)
urbanization	−0.005	−0.004	−0.004	−0.007**	−0.004**	−0.002
	(−1.59)	(−1.59)	(−1.56)	(−2.38)	(−2.07)	(−1.56)
country fixed effects	yes	yes	yes	–	–	–
time fixed effects	yes	yes	yes	yes	yes	yes
Obs.	901	899	899	790	790	790
N	55	55	55	54	54	54
Adj. R^2	0.355	0.346	0.345	–	–	–
Hansen J (p-value)	–	–	–	0.778	1.000	1.000
AR2 test (p-value)	–	–	–	0.694	0.811	0.894

Note: t-values are reported in parentheses; standard errors are calculated using White correction; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively. (a): the regression uses lagged values of log(GDP p.c.) in order to reduce a potential endogeneity bias; (b): log(GDP p.c.) and its square are treated as endogenous variables in the difference-GMM estimations in addition to the lagged dependent variable.

Columns (1)–(3) report the results of OLS regressions; columns (4)–(6) report the results of dynamic panel estimations. The result reported in column (1) is based on an OLS regression with country and time fixed effects. Note that only the time-varying control variables (trade/GDP and urbanization) can be considered in the estimation. The coefficient of the income variable is positive and statistically significant, and the coefficient of the quadratic term is negative and significant, supporting the inverted-U hypothesis. In column (2) I have used one-year-lagged values of the GDP variables in order to reduce any possible endogeneity bias. The results are very similar. In column (3) I use a third degree polynomial to explain regional inequalities. Interestingly, and in contrast to the cross-section estimates, there is no evidence of increasing inequalities at higher stages of economic development after the inverted-U pattern has been completed. In column (4), which reports difference GMM results, only the lagged dependent variable has been treated as endogenous. The main coefficients of interest are statistically significant and have the expected signs. In the estimations reported in column (5) I add the GDP variables to list of endogenous regressors. The results still support the inverted-U hypothesis. Finally, in column (6) I present results with a higher order polynomial, which do not suggest increasing inequalities at very high levels of economic development.

To sum up, the panel estimations which focus on the within-country variation in the data support the inverted-U hypothesis, but inequalities do not seem to increase again at higher levels of economic development. The second result, which contrasts the findings of the cross-section estimations, may be an outcome of the specific (viz., polynomial) functional form I have posed for the development-inequality relationship. Thus, a more flexible semiparametric approach as used in the following section might be better suited to analyze the development-inequality nexus.

4.3 Semiparametric regressions

4.3.1 Cross-section data

This section presents estimation results using a semiparametric regression procedure. To stick with the structure of the previous section, I first focus on a cross-section of countries, and subsequently present the panel estimates. As discussed in section 4.1, I estimate the semiparametric equation (3) using the Yatchew procedure. This procedure is available as a Stata routine and was graciously provided by Lokshin (2006). The estimation output consists of two parts: (1) a table which reports the regression coefficients of the linear part of the model, and (2) a graph which illustrates the functional form of the nonlinear part, that is, the relationship between spatial inequality and development. Table 4 reports the corresponding results. I have specified two different models. Column (1) reports the linear part of the semiparametric regressions with a reduced number of control variables in order to increase the number of available degrees of freedom. Importantly, the significance test (**V**) on the GDP variable is passed here [see Lokshin (2006) for details]. Using the full set of controls as in the regressions reported in column (2) yields very similar results, but the significance test on the GDP variable failed. The nonparametric part of the first specification is illustrated in figure 2. Note that I have selected a bandwidth of 0.8 for smoothing, meaning that 80% of the observations are used for calculating smoothed values for each point in the data except

Table 4: Cross-section data: semiparametric estimates (linear part of the model)

Dependent variable: Weighted coefficient of variation of regional GDP p.c. (WCV)	(1)	(2)
log(spatial units)	0.112***	0.110***
	(3.93)	(4.12)
log(total area)	0.029**	0.029**
	(2.05)	(2.02)
ethnic fractionalization	−0.003**	−0.003**
	(−2.12)	(−2.19)
trade/GDP ratio	0.003***	0.002***
	(3.72)	(3.44)
urbanization		0.210**
		(2.29)
federal dummy		−0.073
		(−1.42)
Significance test on log(GDP p.c.)		
V (p-value)	0.01***	0.41
Obs. (N)	53	53
Adj. R^2	0.337	0.395

Note: t-values are reported in parentheses; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

for end points. The graph supports the main hypothesis of an inverted-U-shaped relationship between spatial inequality and development, and it also shows increasing inequalities at high levels of economic development. The corresponding graph for the second specification looks very similar; therefore I have relegated it to the appendix [see figure 7]. All in all, the semiparametric regressions based on the cross-section data set support the parametric results, in particular with regard to the relevance of a third order polynomial.

4.3.2 Panel data

The last step of this analysis is to use the full panel data set in the semiparametric approach. I present two different specifications. In both specifications, I linearly control for time-varying control variables and for country and time fixed effects. The difference is that in the first specification I use the contemporaneous value of the log of GDP p.c., while in the second specification I use a one-year-lagged value in order to reduce a possible endogeneity bias. The results of the linear part of the model are presented in Table 5. The sign and significance of the control variables are similar to those in the parametric panel regressions as reported in Table 3. The development variable, which enters the regression nonlinearly, has also a statistically significant effect. The functional form of the relationship between regional inequality and development is shown by figure 3, based on the first specification. The graph reveals an inverted-U-shaped relationship with increasing inequalities

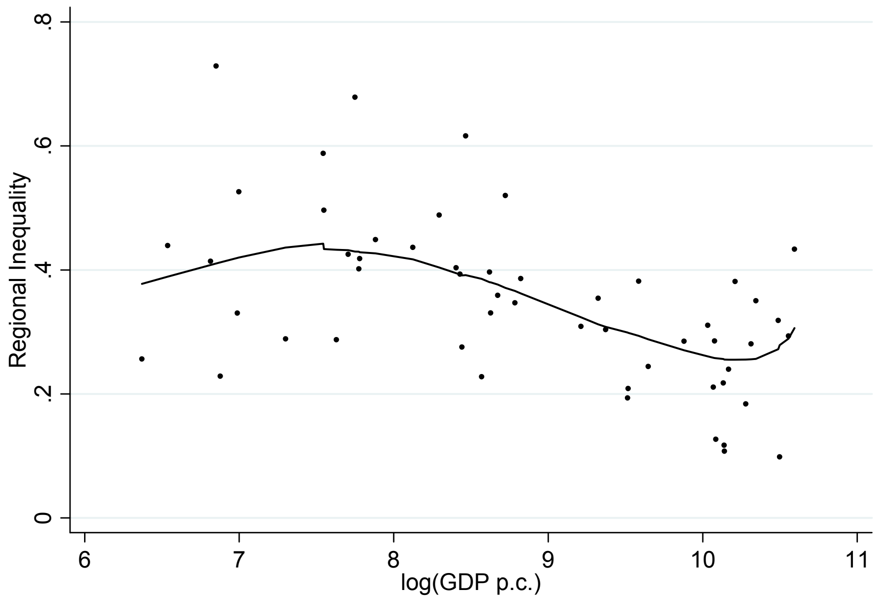

Figure 2: Cross-section: Nonlinear part of specification 1

Table 5: Panel data: semiparametric estimates (linear part of the model)

Dependent variable: Weighted coefficient of variation of regional GDP p.c. (WCV)	(1)	(2)[a]
trade/GDP	0.001	0.001
	(4.50)	(4.72)
urbanization	−0.003	−0.004
	(−3.20)	(−3.85)
country fixed effects	yes	yes
time fixed effects	yes	yes
Significance test on log(GDP p.c.)		
V (p-value)	0.000	0.000
N	899	897

Note: t-values are reported in parentheses; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively. (a): specification (2) uses a one-year-lagged value of the development variable.

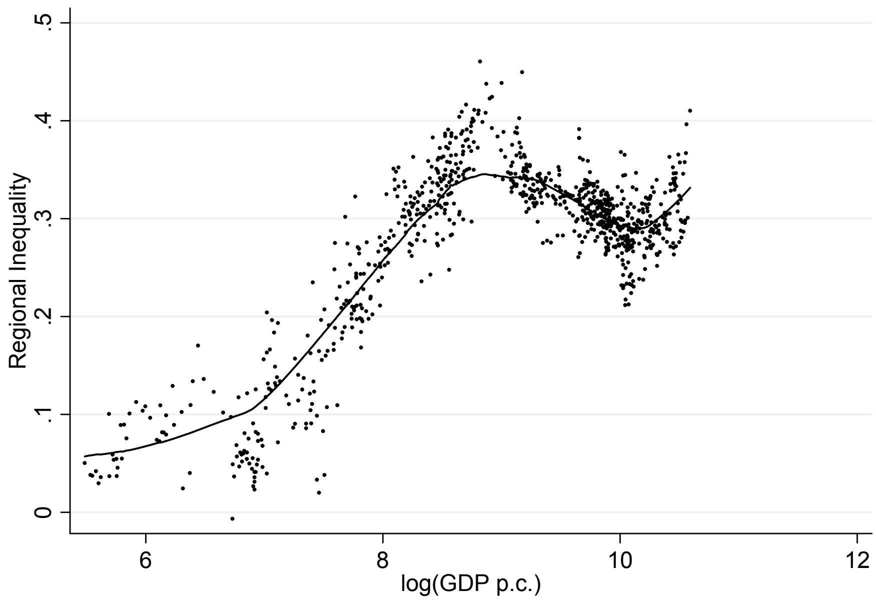

Figure 3: Panel: Nonlinear part of specification 1

at very high levels of economic development. Based on the full panel data set, it is also possible to quantify different thresholds of economic development at which the effect of regional inequalities changes. In the process of economic development, regional inequalities increase until log(GDP p.c.)= 8.88, which corresponds to approximately 7,000 US$ (in 2000 prices). Countries close to this threshold are, e.g., the Czech Republic and Mexico. Then regional inequalities fall again until log(GDP p.c.)= 10.16 (for Canada), which corresponds to approximately 26,000 US$. In richer countries (e.g., Japan), increased development is associated with increasing regional inequality. The process of industrialization leads first to greater regional inequalities; yet beyond a certain critical GDP level, further increases in economic development lead to less inequality. The analysis also shows that regional inequalities start to increase again at later stages of economic development. My result is consistent with previous studies of (income) inequality, which also tends to increase at very high levels of economic development [e.g., Amos (1988), Ram (1991), or List and Gallet (1999)]. As discussed by List and Gallet (1999), one explanation for the renewed positive relationship between inequality and development is the process of tertiarization, in which the economies shift from a manufacturing base towards a service base.

4.4 Robustness tests

In this subsection I provide a number of important robustness tests. First of all, one might be concerned about using the log of GDP p.c. as an indicator of economic development, since the logarithm itself is a nonlinear function. However, such a monotonic transformation of one variable does not affect the general result. If a logarithmic transformation is applied to the GDP p.c.

as explanatory variable, the distribution is less skewed, which has advantages for the parametric regression analysis. The semiparametric approach is flexible in any case. To show the robustness of my findings, I have employed a similar semiparametric panel estimation to that in the previous section, but without transforming the GDP p.c. Figure 4 presents the main result. Again, we can observe an inverted-U-shaped relationship, and regional inequalities increase again at high levels of economic development.

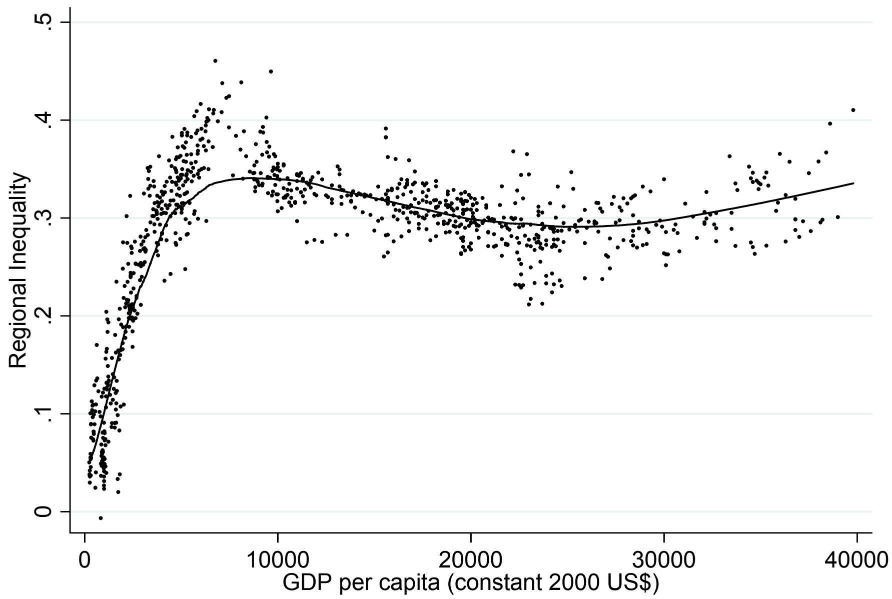

Figure 4: Panel robustness test without a logarithmic transformation

A next robustness test concerns a potential bias due to business cycle effects, which is particularly important for annual panels focusing on within-country variation. The following example should illustrate the problem. Consider a country consisting of two regions, a rich one and a poor one. Suppose further that the rich region has an industry which depends to a large extent on exports. Therefore, it is volatile over the business cycle, while the poor region has a less productive and less volatile economy. In such a situation, one might expect increasing regional inequalities in times of expansion, and decreasing inequalities in times of recession. This is for example the case in Germany, where the eastern part of the country has much less volatile growth rates than the western states, which are richer at the cost of higher volatility. To correct for a bias caused by the regular business cycle, I build 5-year period averages of the panel data set and I repeat the regressions. Table 6 presents the results. The country and time fixed effects OLS regressions (columns (1)–(3)) support the earlier findings. Interestingly, both coefficients are slightly smaller than in the annual panel, supporting the hypothesis of an upward bias caused by the business cycle. In the dynamic panel estimates I cannot find any significant effects. Note, however, that the lagged dependent variable is statistically insignificant in these regressions, implying that it is not necessary to estimate a “dynamic” panel. Moreover, the number of observations drops to 97

Table 6: Panel data: parametric estimates using 5-year period averages

	Dependent variable: Weighted coefficient of variation of regional GDP p.c.					
	OLS	OLS[a]	OLS	GMM	GMM[b]	GMM
	(1)	(2)	(3)	(4)	(5)	(6)
WCV_{t-1}	–	–	–	2.049	0.113	2.692
				(1.50)	(0.25)	(0.58)
log(GDP p.c.)	0.259**	0.169*	−0.740	−0.397	0.196	5.181
	(2.13)	(1.70)	(−0.90)	(−0.81)	(1.09)	(0.14)
$(\log(\text{GDP p.c.}))^2$	−0.014*	−0.017**	0.114	0.009	−0.001	−0.702
	(−1.77)	(−2.18)	(1.04)	(0.40)	(−0.12)	(−0.15)
$(\log(\text{GDP p.c.}))^3$	–	–	−0.005	–	–	0.029
			(−1.12)			(0.15)
trade/GDP	0.001*	0.001**	0.001*	0.001	−0.001	0.001
	(1.92)	(2.33)	(1.88)	(0.22)	(−0.24)	(0.18)
urbanization	−0.004	−0.002	−0.004	0.013	−0.004	0.017
	(−1.63)	(−1.27)	(−1.64)	(1.15)	(−1.28)	(0.59)
country fixed effects	yes	yes	yes	–	–	–
time fixed effects	yes	yes	yes	yes	yes	yes
Obs.	204	202	204	97	97	97
N	55	55	55	39	39	39
Adj. R^2	0.427	0.445	0.435	–	–	–
Hansen J (p-value)	–	–	–	0.937	0.161	0.794
AR2 test (p-value)	–	–	–	0.238	0.226	0.311

Note: t-values are reported in parentheses; standard errors are calculated using White correction; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively. (a): the regression uses lagged values of log(GDP p.c.) in order to reduce a potential endogeneity bias; (b): log(GDP p.c.) and its square are treated as endogenous variables in the difference GMM estimations in addition to the lagged dependent variable.

(39 countries), limiting the number of available degrees of freedom. The GMM estimates are thus not very meaningful. In addition, I have estimated the semiparametric model based on the 5-year period averages. The result of the nonlinear part of the model is illustrated in figure 5, supporting earlier findings.

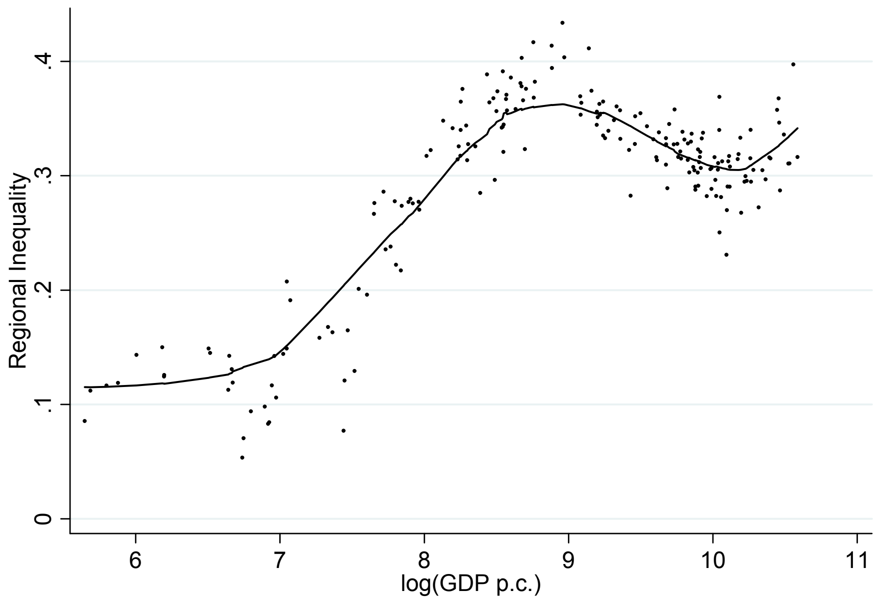

Figure 5: Robustness tests using 5-year averages

The next robustness test concerns the measure of regional inequality used as dependent variable. All presented estimations up to this point of the paper have used the population-weighted coefficient of variation as discussed in section 3. This measure is commonly used in the economic-geography literature, but studies on growth and convergence such as, e.g., Barro and Sala-i-Martin (1992) concentrate on unweighted measures. One might argue that weighting of observations by size of the regions distorts the inequality measure. Small regions, which may be extremely rich (e.g., capital regions) or poor (e.g., special zones of ethnic minorities), have only a minor effect on the overall indicator, although the deviations from the respective countries' means are very important in light of the risk of conflict and secession. To allow for this argument and to make my results comparable to the convergence literature, I also calculate the (unweighted) coefficient of variation (CV) of the regional GDP p.c. for each country:

$$CV := \frac{1}{\bar{y}} \left[\sum_{i=1}^{n} (\bar{y} - y_i)^2 \right]^{1/2}. \tag{4}$$

Figure 6 in the appendix presents the main findings of semiparametric panel estimates with the CV as dependent variable. Again, the main findings of the analysis can be confirmed.

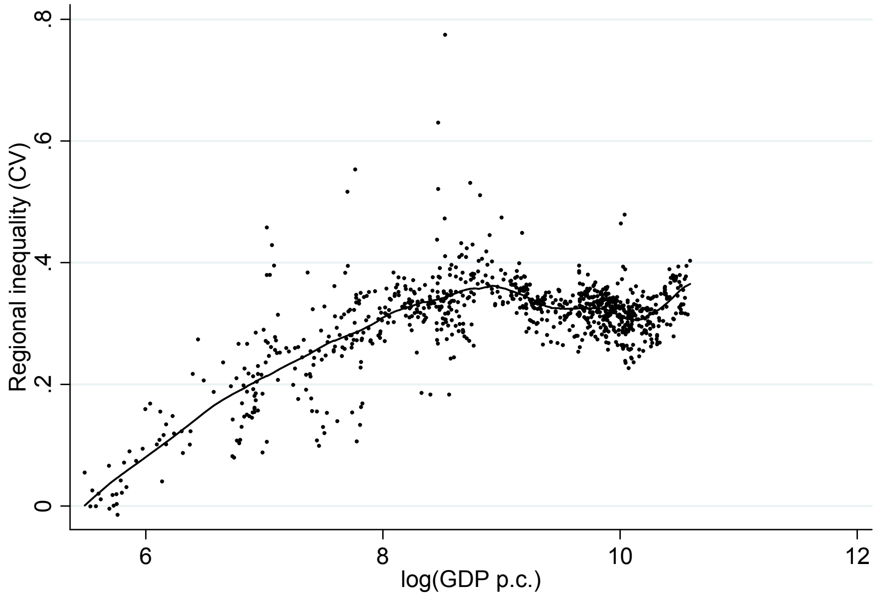

Figure 6: Robustness tests using 5-year period averages and the CV as inequality measure

5 Conclusion

This paper studies the hypothesis of an inverted-U-shaped relationship between spatial inequality and economic development. The theory of Kuznets (1955) and Williamson (1965) suggests that (spatial) inequality first increases in the process of development, then peaks, and then decreases. With the exception of the initial study of Williamson (1965) himself, empirical evidence for this hypothesis, which covers developing and developed economies, does not exist. This gives reason to reexamine the original work using a broader data set as well as recent econometric techniques. A further reason – and perhaps the most important one – is that something might have changed in the relationship between spatial inequality and development since the 1950s or 1960s, in that most countries have experienced very dynamic growth since then. For this purpose, a unique panel data set covering 55 countries at all stages of economic development was collected. The period covered by the unbalanced panel is 1980–2009. Cross-country and panel regressions have been carried out using a parametric as well as a semiparametric approach. I find evidence of an inverted U in models focusing on between- and within-country variations. In countries at late stages of economic development, spatial inequalities increase with increased GDP p.c. This result is in line with existing studies of high-income economies [see, e.g., Amos (1988), Fan and Casetti (1994), and Terrasi (1999)].

What do we learn from this study? When countries shift from agricultural to industrial economies, spatial inequalities increase. If a certain development level (in my study 7,000 US$) is reached, the relationship is reversed until high levels of economic development (26,000 US$) are reached, where inequalities start again to increase. The increase in spatial inequality at high levels of economic development may have been caused by an exogenous shock, or be a mark of the change from industrial to service-based economies. In the process of tertiarization one might expect a new inverted-U process where a few leading regions of a country adopt the new technology first, and other regions lag behind, so that spatial inequalities increase in the process of structural change. Unfortunately, this would imply that we should expect increasing regional imbalances within most developed countries for the next decades. This can be a breeding ground for conflicts and make precautionary political interventions necessary.

References

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth*, 8(2), pp. 155–194.

Alesina, A. and D. Rodrik (1994). Distributive Politics and Economic Growth. *The Quarterly Journal of Economics*, 109(2), pp. 465–90.

Amos, O. J. (1988). Unbalanced regional growth and regional income inequality in the latter stages of development. *Regional Science and Urban Economics*, 18(4), pp. 549–566.

Araujo, M. C., F. H. Ferreira, P. Lanjouw, and B. Oezler (2008). Local inequality and project choice: Theory and evidence from Ecuador. *Journal of Public Economics*, 92(5-6), pp. 1022–1046.

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data*. Chichester: Wiley.

Barrios, S. and E. Strobl (2009). The dynamics of regional inequalities. *Regional Science and Urban Economics*, 39(5), pp. 575 – 591.

Barro, R. J. and X. Sala-i-Martin (1992). Convergence. *Journal of Political Economy*, 100(2), pp. 223–251.

Bendel, R. B., S. S. Higgins, and D. A. Pyke (1989). Comparison of Skewness Coefficient, Coefficient of Variation, and Gini Coefficient as Inequality Measures within Populations. *Oecologia*, 78(3), pp. 394–400.

Boadway, R. and F. Flatters (1982). Efficiency and Equalization Payments in a Federal System of Government: A Synthesis and Extension of Recent Results. *Canadian Journal of Economics*, 15(4), pp. 613–633.

Brambor, T., W. R. Clark, and M. Golder (2006). Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*, 14(1), pp. 63–82.

Chen, J. and B. M. Fleisher (1996). Regional Income Inequality and Economic Growth in China. *Journal of Comparative Economics*, 22(2), pp. 141 – 164.

Dalton, H. (1920). The Measurement of the Inequality of Incomes. *Economic Journal*, 30(119), pp. 348–361.

DiNardo, J. and J. L. Tobias (2001). Nonparametric Density and Regression Estimation. *Journal of Economic Perspectives*, 15(4), pp. 11–28.

Durlauf, S. N. (2001). Manifesto for a growth econometrics. *Journal of Econometrics*, 100(1), pp. 65–69.

Elbers, C., P. Lanjouw, J. Mistiaen, B. Özler, and K. R. Simler (2005). Are Neighbours Equal? Estimating Local Inequality in Three Developing Countries. In R. Kanbur and A. Venables (editors), *Spatial Inequality and Development*, pp. 37–76. New York: Oxford University Press.

Ezcurra, R. and M. Rapun (2006). Regional Disparities and National Development Revisited: The Case of Western Europe. *European Urban and Regional Studies*, 13(4), pp. 355–369.

Fan, C. C. and E. Casetti (1994). The Spatial and Temporal Dynamics of U.S. Regional Income Inequality, 1950-1989. *The Annals of Regional Science*, 28(2), pp. 177–96.

Gianetti, M. (2002). The Effects of Integration on Regional Disparities: Convergence, Divergence or Both? *European Economic Review*, 46(3), pp. 539–567.

Hale, T. (2003). The Theoretical Basics of Popular Inequality Measures; Online Computation of Examples. University of Texas Inequality Project.

Jian, T., J. D. Sachs, and A. M. Warner (1996). Trends in regional inequality in China. *China Economic Review*, 7(1), pp. 1 – 21.

Kanbur, R. and A. Venables (2005a). *Rising spatial disparities and development*. Number 3 in United Nations University Policy Brief. Helsinki.

Kanbur, R. and A. Venables (2005b). Spatial Inequality and Development. In R. Kanbur and A. Venables (editors), *Spatial Inequality and Development*, pp. 3–11. New York: Oxford University Press.

Kanbur, R. and X. Zhang (2005). Fifty Years of Regional Inequality in China: a Journey Through Central Planning, Reform, and Openness. *Review of Development Economics*, 9(1), pp. 87–106.

Krugman, P. and R. L. Elizondo (1996). Trade policy and the Third World metropolis. *Journal of Development Economics*, 49(1), pp. 137–150.

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

Lambert, S., M. Ravallion, and D. van de Walle (2010). A Micro-Decomposition Analysis of Aggregate Human Development Outcomes. *Oxford Bulletin of Economics and Statistics*, 72(2), pp. 119–145.

Lessmann, C. (2009). Fiscal Decentralization and Regional Disparity: Evidence from Cross-section and Panel Data. *Environment and Planning A*, 41(10), pp. 2455–2473.

Lessmann, C. (2011). Foreign Direct Investment and Regional Inequality: A Panel Data Analysis. Working paper, Technische Universitaet Dresden.

List, J. A. and C. A. Gallet (1999). The Kuznets Curve: What Happens after the Inverted-U? *Review of Development Economics*, 3(2), pp. 200–206.

Lokshin, M. (2006). Difference-based semiparametric estimation of partial linear regression models. *Stata Journal*, 6(3), pp. 377–383.

Lucas, R. E. (2000). Some Macroeconomics for the 21st Century. *Journal of Economic Perspectives*, 14(1), pp. 159–168.

Mehran, F. (1976). Linear Measures of Income Inequality. *Econometrica*, 44(4), pp. 805–809.

Mills, E. S. and D. M. d. Ferranti (1971). Market Choices and Optimum City Size. *American Economic Review*, 61(2), pp. 340–345.

Pigou, A. C. (1912). *Wealth and Welfare*. London: Macmillan.

Ram, R. (1991). Kuznets's inverted-U hypothesis: evidence from a highly developed country. *Southern Economic Journal*, 57(4), pp. 1112–1123.

Robinson, P. M. (1988). Root- N-Consistent Semiparametric Regression. *Econometrica*, 56(4), pp. 931–54.

Rodríguez-Pose, A. and R. Ezcurra (2010). Does decentralization matter for regional disparities? A cross-country analysis. *Journal of Economic Geography*, 10(5), pp. 619–644.

Rodriguez-Pose, A. and N. Gill (2006). How does trade affect regional disparities? *World Development*, 34(7), pp. 1201–1222.

Sala-i-Martin, X. (1996). Regional Cohesion: Evidence and Theories of Regional Growth and Convergence. *European Economic Review*, 40(6), pp. 1325–1352.

Sen, A. (1973). *On Economic Inequality*. New York: Norton.

Terrasi, M. (1999). Convergence and divergence across Italian regions. *The Annals of Regional Science*, 33(4), pp. 491–510.

Treisman, D. (2008). Decentralization dataset. http://www.sscnet.ucla.edu/polisci/faculty/ treisman/.

Wei, K., S. Yao, and A. Liu (2009). Foreign Direct Investment and Regional Inequality in China. *Review of Development Economics*, 13(4), pp. 778–791.

Weltbank (2011). World Development Indicators 2010. Washington D. C.

Williamson, J. G. (1965). Regional Inequality and the Process of National Development: A Description of Patterns. *Economic Development and Cultural Change*, 13(4), pp. 3–45.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yatchew, A. (1997). An elementary estimator of the partial linear model. *Economics Letters*, 57(2), pp. 135–143.

Yatchew, A. (1998). Nonparametric Regression Techniques in Economics. *Journal of Economic Literature*, 36(2), pp. 669–721.

Yemtsov, R. (2005). Quo Vadis? Inequality and Poverty Dynamics across Russian Regions. In R. Kanbur and A. Venables (editors), *Spatial Inequality and Development*, pp. 348–408. New York: Oxford University Press.

Table A.1: Sources of regional data by country

Country	Territorial level	Period	Source
Argentina	23 provinces; 1 capital region	1991–2002	Dirección Provincial de Estadística
Australia	8 TL2 regions	1990–2008	OECD Regional Statistics
Austria	9 NUTS2 regions	1980–2004	Cambridge Econometrics
Belgium	11 NUTS2 regions	1980–2004	Cambridge Econometrics
Bolivia	9 departments	1988–2009	Instituto Nacional de Estadistica
Brazil	26 states; 1 federal district	2002–2007	Instituto Brasileiro de Geografica e Estatistica
Bulgaria	6 TL2 regions	1995–2007	OECD Regional Statistics
Canada	12 provinces and territories (Northwest Territories including Nunavut)	1981–2004	Statistics Canada
Chile	13 regions	1996–2006	National Statistics Institute
China	30 provinces, autonomous regions, and cities	1994–2008	National Bureau of Statistics of China
Colombia	33 departments	1990–2007	Departamento Administrativo Nacional de Estadística
Croatia	3 TL2 regions	1990–2007	OECD Regional Statistics
Czech Republic	8 TL2 regions	1990–2007	Cambridge Econometrics and OECD Regional Statistics
Denmark	3 NUTS2 regions	1980–2004	Cambridge Econometrics
Finland	6 NUTS2 regions	1980–2004	Cambridge Econometrics
France	22 NUTS2 regions	1980–2004	Cambridge Econometrics
Georgia	9 provinces	2003–2009	National Statistics Office of Georgia
Germany	30 NUTS2 regions (West)	1980–2004	Cambridge Econometrics
Greece	13 NUTS2 regions	1980–2004	Cambridge Econometrics
Hungary	7 NUTS2 regions	1990–2007	Cambridge Econometrics and OECD Regional Statistics
India	28 states and union territories	1980–2005	Directorate of Economics & Statistics of respective State Governments, and Central Statistical Organisation
Indonesia	33 provinces	2004–2008	Badan Pusat Statistik
Iran, Islamic Rep.	28 provinces	2000–2003	Statistical Center of Iran
Ireland	2 NUTS2 regions	1980–2004	Cambridge Econometrics
Italy	20 NUTS2 regions	1980–2004	Cambridge Econometrics
Japan	10 TL2 regions	1990–2005	OECD Regional Statistics
Kazakhstan	16 regions and cities	1998–2009	Agency of Statistics of the Republic of Kazakhstan
Korea, South	7 TL2 regions	1990–2007	OECD Regional Statistics
Latvia	6 NUTS3 regions	1996–2007	EUROSTAT
Lithuania	10 NUTS3 regions	1995–2007	EUROSTAT
Malta	2 NUTS3 regions	2000–2007	EUROSTAT
Mexico	32 states; 1 capital region	1980–2006	Instituto Nacional de Estadística y Geografía
Mongolia	21 provinces; 1 capital region	2000–2006	National Statistical Office

Table A.1 countinued

Country	Territorial level	Period	Source
Netherlands	12 NUTS2 regions	1986–2004	Cambridge Econometrics
New Zealand	2 TL2 regions	2000–2003	OECD Regional Statistics
Norway	7 NUTS2 regions	1980–2004	Cambridge Econometrics
Panama	9 provinces	2002–2007	Instituto Nacional De Estadistica
Peru	24 departments	2001–2009	Instituto Nacional de Estadística e informática – Dirección Nacional de Cuentas Nacionales
Philippines	17 districts	2002–2008	National Statistics Office
Poland	16 NUTS2 regions	1990–2007	Cambridge Econometrics and OECD Regional Statistics
Portugal	7 NUTS2 regions	1980–2004	Cambridge Econometrics
Romania	8 NUTS2 regions	1995–2007	EUROSTAT
Russian Federation	7 federal regions	1998–2008	Federal State Statistics Office
Slovak Rep.	4 TL2 regions	1990–2007	Cambridge Econometrics and OECD Regional Statistics
Slovenia	2 NUTS2 regions	1995–2007	EUROSTAT
South Africa	9 provinces	2001–2008	Statistics South Africa
Spain	18 NUTS2 regions	1980–2004	Cambridge Econometrics
Sweden	8 NUTS2 regions	1980–2004	Cambridge Econometrics
Switzerland	7 NUTS2 regions	1980–2004	Cambridge Econometrics
Tanzania	21 administrative regions	2002–2007	National Bureau of Statistics
Thailand	7 geographic regions	2001–2009	National Statistics Office Thailand
Turkey	26 TL2 regions	1990–2006	OECD Regional Statistics
U.S. of America	51 states	1980–2008	U.S. Department of Commerce, OECD Regional Statistics
Ukraine	27 districts	2004–2008	State Statistics Committee of Ukraine
United Kingdom	37 NUTS2 regions	1980–2004	Cambridge Econometrics
Uzbekistan	12 provinces; 1 republic; 1 capital region	2008	Uzbekistan in Figures – UinF
Venezuela	23 states; 1 federal district	2007	Banco Central de Venezuela

Table A.2: Summary statistics

Variable	Mean	Maximum	Minimum	Std. dev.	Observations
WCV	901	0.28	0.17	0.04	0.95
CV	901	0.30	0.19	0.06	1.30
GDP p.c.	901	13,158.92	10,345.99	229.00	39,800.00
spatial units (No.)	901	15.17	11.57	2.00	51.00
total area (1,000 km^2)	871	1,623.46	3,159.63	0.32	16,400.00
ethnic fractionalization	901	0.30	0.22	0.00	0.75
trade/GDP	901	70.85	35.61	12.40	195.00
urbanization	901	67.74	14.71	23.10	97.30
federal dummy	901	0.30	0.46	0.00	1.00

Table A.3: Data sources & definitions

Variable	Definition	Source
WCV	Population-weighted coefficient of variation of regional GDP per capita	various sources
CV	Coefficient of variation of regional GDP per capita	various sources
GDP p.c.	Log of the GDP per capita in 2005 $ prices.	Weltbank (2011)
spatial units	log of the number of regions considered for the calculation of measures of regional inequality.	various sources
total area	Log of area in square kilometers.	Weltbank (2011)
trade/GDP	Sum of imports and exports (total trade) as a share of the GDP.	Weltbank (2011)
ethnic fractionalization	Ethnolinguistic fractionalization is computed as one minus Herfindahl index of ethnolinguistic group shares, and reflects the probability that two randomly selected individuals from a population belonged to different groups.	Alesina et al. (2003)
urbanization	Share of urban living population in total population.	Weltbank (2011)
federal dummy	Dummy for countries with a federal constitution.	Treisman (2008)

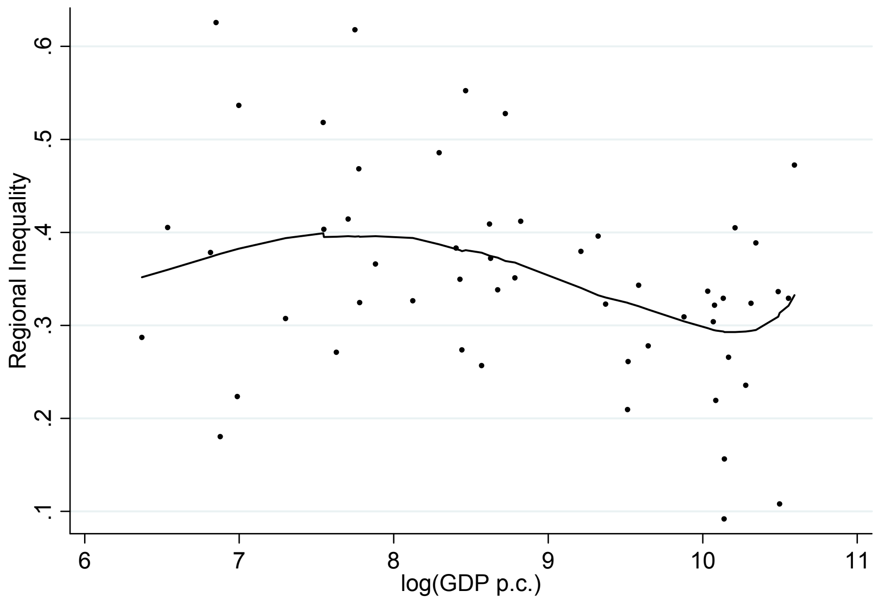

Figure 7: Nonlinear part of specification 2

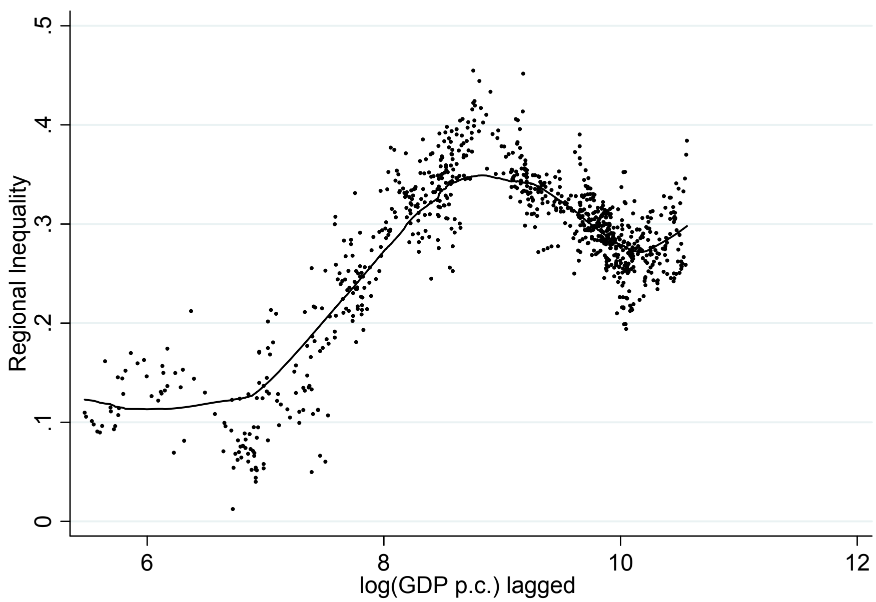

Figure 8: Panel: Nonlinear part of specification 2