Stages of Diversification

Jean Imbs
London Business School
and CEPR

Romain Wacziarg*
Stanford University
and NBER

September 2002

Abstract

This paper studies the evolution of sectoral concentration in relation to the level of per capita income. We show that various measures of sectoral concentration follow a U-shaped pattern across a wide variety of data sources: countries first diversify, in the sense that economic activity is spread more equally across sectors, but there exists, relatively late in the development process, a point at which they start specializing again. We discuss this finding in light of existing theories of trade and growth, which generally predict a monotonic relationship between income and diversification. (*JEL* F43, F15, O40)

This paper characterizes the pattern of sectoral diversification along the development path. Using data on sector level employment and value added, covering a wide cross-section of countries at various levels of disaggregation, we provide new and robust evidence that economies grow through two stages of diversification. At first, sectoral diversification increases, but there exists a level of per capita income beyond which the sectoral distribution of economic activity starts concentrating again. In other words, sectoral concentration follows a U-shaped pattern in relation to per capita income.

This new finding has potentially important implications for theories of trade and growth. Most existing theories predict a monotonic relationship between income and sectoral concentration. At early stages of development, countries are usually specialized in exploiting their natural resources endowments, and simple arguments based on the law of large numbers suggest that diversification should help dampen the aggregate effects of sector-specific shocks.[1] Similarly, in the context of high trading costs, an economy where consumers display a taste for diversity might find it efficient to open new sectors domestically rather than import desired goods at a high cost. If opening new sectors is costly, a relatively closed economy will tend to open new sectors, thus diversify, as factors are accumulated. This set of theories point to a negative relationship between income levels and the degree of sectoral concentration.

On the other hand, there is now a substantial body of literature emphasizing important reasons for sectoral specialization - or agglomeration - to occur over the development path. In Ricardian trade theory, open economies are predicted to specialize in producing a specific range of goods - so that specialization is expected to accompany any reduction in the impediments to trade, be they policy or technology driven.[2] More recently, an expanding body of theoretical and empirical literature, championed by Paul Krugman (1991) and recently surveyed in J. Peter Neary (2001), has taken interest in the geographic agglomeration of economic activity. In this literature, the presence of pecuniary externalities ("demand linkages") and costly trade can make it optimal for monopolistic competitors to cluster since their profits increase in local expenditure, itself a positive function of the number of local firms. The idea that transport costs fall with better technologies tends to reinforce this mechanism, as the demand linkages effect is decreasingly mitigated by the incentive to be located close to demand. Thus, economic activity in integrating economies tends to be increasingly agglomerated. Such geographic clustering would naturally translate into increasing observed degrees of concentration at the sector level within arbitrarily defined geographic zones, such as countries.

Thus, there is no theoretical consensus as to how measures of sectoral diversification should evolve as countries grow, although the force of diversification is probably more at play among low-income countries, and the force of concentration among richer economies. The findings in this paper show this to be the case empirically: each set of theories seems to be at play at different points in the development process. Sectoral diversification goes through two stages, at first one of increasing diversification, and later one of increasing concentration.

A data caveat is in order. Disaggregated data is arbitrary, in the sense that some economic activities are registered more coarsely than others, for no other reason than statistical conventions. Thus, an economy could display non-monotonic specialization patterns even though actual diversification stayed constant throughout. Suppose for instance there are three types of goods:

agriculture, manufacturing, and services, and structural transformation involves specializing first in agriculture, then in manufacturing and lastly in services. Suppose also that statisticians collect more disaggregated data on manufacturing than they do on either agriculture or services. Observed sectoral concentration would then artificially follow a U-shaped pattern.[3] While the reader should be cautioned about this kind of pitfall, our answer is to use various sources of data, various scopes for its coverage (economy-wide or manufacturing in isolation) and various levels of disaggregation (1, 2, 3 and 4 digits in the International Standard Industrial Classification). Such robustness analysis will lend support to the view that our result is not simply the outcome of the arbitrary sectoral classification of goods and services.

In fact, our result is an extremely robust feature of the data. The non-monotonicity holds above and beyond the well-known shift of factors of production from agriculture to manufacturing and on to services - in particular, the U-shaped pattern is present when focusing only on manufactured goods. It is valid whether a sector's size is measured by its share in total employment or whether it is measured by shares in value added. It holds within countries through time as well as in a pure cross-section, for a variety of levels of disaggregation and data sources. The estimated turnaround point occurs quite late in the development process and at a surprisingly robust level of income per capita. Thus, increased sectoral specialization, although a significant development, applies only to high income economies. Countries diversify over most of their development path.

The paper is structured as follows: Section 2 describes our benchmark non-parametric estimation, and establishes our result for various measures of dispersion and data coverage. Section 3 discusses extensions based on parametric estimation methods, drawing inferences from cross-country variation in the stages of diversification. We also briefly present robustness checks. Section 4 provides a roadmap for interpreting theoretically the paper's main empirical findings. The last section concludes.

1 The Evolution of Sectoral Concentration

This section documents the main empirical result of the paper: the relationship between various measures of sectoral concentration and the level of per capita income displays a U-shaped pattern. We also estimate the level of income at which minimum concentration level is attained.

1.1 Overview of the Data

We employ sectoral data from the International Labor Office (ILO, 1997) and United Nations Industrial Development Organization (UNIDO, 1997) to examine the evolution of several measures of sectoral concentration through time and in relation to the level of development.[4] For the ILO, the data pertain to employment shares across sectors, at the 1-digit level, covering all economic activities for the 1969-1997 time period. While the ILO data covers all economic activities, the UNIDO data covers only manufacturing, at the 3-digit level of disaggregation, extending from 1963 to 1996. One advantage of the UNIDO dataset is that measures of value added per sector are also available, allowing us to use an alternative measure of sector size.[5]

Both the ILO and UNIDO data sets span a wide range of industrial and developing countries.[6] In addition to ILO and UNIDO data, we also use data from the Organization for Economic Cooperation and Development (OECD, 1998) for 14 industrial countries, covering the period 1960-1993. For the OECD, data on both employment and value added per sector are available at the 2-digit level of disaggregation and cover all of national economic activity (although some non-manufacturing sectors appear at the 1-digit level of disaggregation only). The OECD data therefore provides a robustness check with respect to the level of disaggregation and allows a better focus on the upper end of the per capita income distribution.[7] As already suggested, by using various levels of disaggregation and various scopes for the coverage of the data, we hope to minimize the incidence of a somewhat arbitrary classification of goods and services across sectoral categories.

Our measures of sector size, used to construct various indicators of sectoral concentration, consist of employment shares and shares in value added. The use of employment shares as a measure of sector size is common in the empirical literature concerned with sectoral specialization.[8] However, results obtained using sectoral shares in value added can provide a generalization of the evidence based on sectoral labor inputs. All of our datasets consist of annual observations, so both within and between country variations are available.

Tables 1 and 2 contain summary statistics for the sectoral concentration data. We use a variety of measures for the concentration of employment across sectors. Throughout this paper we focus mostly on the Gini coefficient for the inequality of sector shares (GINI): the more equal the sector shares, the more diversified an economy. This is a common measure of sectoral and regional concentration in the economic geography literature.[9] However, since there are many different measures of dispersion and no particular reason to favor one or the other, we also sought to characterize the stages of diversification using a number of other measures:

- The Herfindahl index for the sectoral concentration of employment or value added (HERFIND)
- The coefficient of variation of sector shares (COEFVAR).
- The max-min spread (MAXMIN)
- The log-variance of sector shares (LOGVAR).

In general, as shown in Table 2, our five main measures of sectoral concentration are highly correlated among themselves. However, the magnitude of many of these correlations is not always sufficiently high to warrant using only one of these variables, particularly for value-added based measures in the OECD. To further assess the robustness of our results to alternative measures of dispersion, we also used the following indicators, for which results are available upon request:

- The share of the biggest sector in employment (BIGGEST)
- The mean-median spread (MEANMED)
- The interquartile range (IQR) of sector shares.

The interquartile range (IQR) was either weakly or even negatively correlated with all our other measures of diversification. Since the ILO data involves only 9 sectoral categories, the measures of dispersion obtained using the interquantile range were highly sensitive to which quantile range we chose. This, to a lesser extent, was also the case in the UNIDO 3-digit data, for which 28 sectors are available. Since sectors are coarsely defined, measures that rely on only two points in a given distribution, such as the max-min spread, the mean-median spread and the interquantile range, are likely to be poor measures of dispersion compared to measures that use data on the entire distribution (such as the Gini coefficient, the Herfindahl index and the coefficient of variation).[10]

1.2 Nonparametric Methodology

In order to investigate the shape of the relationship between our indices of sectoral diversification and income levels, we attempt to impose as little structure on the functional form as possible. This motivates the use of nonparametric methods to identify the shape of the relationship.[11] Moreover, we use a procedure that is locally robust: contrary to polynomial (or semi-parametric) methods, our method leads to estimating a relationship in which, for example, the shape of the curve linking income and diversification at high levels of income is not affected by observations at low levels of income.

The methodology we employ is derived from robust locally weighted scatterplot smoothing (*lowess*).[12] This consists, for each observation $n = 1...N$ on a dependent and an independent variable, say (y_n, x_n), in running a regression of variable y on variable x using a small amount of data around x_n. In our application y corresponds to a measure of sectoral concentration and x corresponds to income. The fitted value of this regression evaluated at x_n is used as the smoothed value in constructing the nonparametric curve linking y and x. The procedure is repeated for each observation (y_n, x_n) until the smoothed curve can be traced out - hence, the number of regressions is equal to the number of observations. The scheme involves two arbitrary choices: the choice of a bandwidth (the amount of data around x_n that is used in each regression) and a weighing scheme. The weighing scheme typically entails (weakly) decreasing weights on observations that are farther away from x_n. A "flat" or rectangular weighing scheme entails equal weighing of observations within the bandwidth.

The *lowess* procedure differs from other smoothing methods because it involves running a weighted regression of y on x for each subsample of data, and plotting the fitted value at x_n. Other smoothing methods simply plot the (weighted) average value of y for subsamples of the data centered at x_n. While both methods deliver similar smoothed curves, the advantage of *lowess* is that the estimated coefficients on variable x in a regression of y on x, for different subsamples centered at x_n, are being calculated and can be of independent interest. This is indeed the case in our application. We are not simply interested in the shape of the relationship linking sectoral concentration and income for a typical country, which is delivered by the smoothed curve. We are also interested in the sign and statistical significance of the coefficients on income in a within-country regression of sectoral concentration on income, for subsamples of the observations constructed at different income levels.

Since our nonparametric method differs slightly from the lowess procedure, we describe it in more detail. We start by partitioning the data into S subsamples according to overlapping income intervals of size J, with an overlap of size $J - \Delta$.[13] We use an income interval (or bandwidth) $J = \$5,000$. Each interval, given the range of income in the data, has a midpoint which is $\Delta = \$25$ away from the following interval. There are of course many ways to specify the bandwidth and increments, but our results were not sensitive to these choices.[14] For each subsample, we ran a simple fixed-effects linear regression of the measures of concentration on income. Thus, we employed a flat or rectangular weighing scheme.

For each of these regressions we computed a fitted value, evaluated at the midpoint of the income interval used to determine the sample, and plotted these fitted values against the income midpoint of each estimation interval. Hence, the ordinate of each point s on the fitted curve for the GINI coefficient (for example) was computed as:

$$\widehat{GINI}_s = \hat{\alpha}_s^{FE} + \hat{\beta}_s^{FE} \times x_s \tag{1}$$

where $s = 1...S$, and $\hat{\alpha}_s^{FE}$ and $\hat{\beta}_s^{FE}$ are the fixed effects estimates of the intercept and the slope on income, respectively, in a regression of the GINI coefficient on per capita income for the subsample $\left(x_s - \frac{J}{2},\ x_s + \frac{J}{2}\right)$, with $J = \$5000$. Adjacent points on the curve involve income increments $x_{s+1} - x_s = \Delta = \$25$. The resulting curve provides, for a typical country, the shape of the evolution of sectoral diversification throughout the development path (measured by per capita income). We also compute 5 percent confidence intervals by calculating the standard error of the predicted value $\widehat{GINI}_s$, and plot these confidence bands around the fitted curve.

Two additional issues concerning this methodology deserve clarification. The first one concerns the definition of the intercept term $\hat{\alpha}_s^{FE}$. Since we employ fixed-effects regression with country specific effects, each country is allowed to have its own intercept. For each subsample, these intercepts capture country-specific, time-invariant shifts in the level of sectoral diversification. The value of $\hat{\alpha}_s^{FE}$ we use to compute the fitted measures of diversification is the *average* of the estimated individual country fixed-effects for each subsample s, so the plotted non-parametric curve reflects the relationship between sectoral concentration and per capita income for an average country.

Secondly, equation (1) shows that changes in the fitted measures of sectoral diversification at different levels of income come from three sources: (i) changes in the estimated average intercept term $\hat{\alpha}_s^{FE}$ at each subsequent iteration of the non parametric procedure (ii) changes in the slope term $\hat{\beta}_s^{FE}$ and (iii) changes in the midpoint of the interval x_s. In fact, much of the variation in the smoothed curves comes from changes in the average degree of diversification ($\hat{\alpha}_s^{FE}$) across subsamples s.[15] Obviously, since different countries appear across subsamples, changes in $\hat{\alpha}_s^{FE}$ and in $\hat{\beta}_s^{FE}$, and hence variations in the curve, will be affected by between country variation in the data. Therefore, the resulting plots of $\widehat{GINI}_s$ are not reflective simply of within-country variation. On the other hand, *for each subsample,* the slope estimate on income purely reflects within-country variation in that subsample, since fixed-effects estimation is used.

1.3 Nonparametric Results

We first use the nonparametric procedure described above to obtain a flexible form for the relationship linking sectoral concentration and per capita income. Figures 1-3 display the fitted curves and the 95 percent confidence bands graphically, for the ILO data on employment, the UNIDO data on employment and the UNIDO data on value added, respectively.[16] Across datasets, the relationship between sectoral diversification and income per capita is highly non-monotonic: it appears to be U-shaped. This is the central stylized fact we establish in this paper.

Two observations are in order. Firstly, the U-shaped pattern is not symmetric: given the observed maximal level of income, the upward bending portion of the curve does not swing back up to the initial level of sectoral concentration. Secondly, the shift towards reconcentration seems to occur late in the development process: the non-parametric estimates can be used to compute the level of income at which the point of minimum concentration occurs. Specifically, we can take the value of income which minimizes the predicted value of sectoral concentration as the estimated minimum point. Table 3 reports these minima for all measures of sectoral concentration.

For example, for the ILO data, using the Gini measure of sectoral labor concentration, we find that the minimum point occurs when per capita income equals approximately $9,575 per year. The Summers-Heston data employed for the calculations is in constant 1985 US dollars, so this point occurs roughly at the level of income reached by Ireland in 1992.[17] In other words, according to these estimates the minimum point occurs quite late in the development process. It occurs somewhat earlier for the UNIDO data set ($8,675), suggesting that reconcentration occurs earlier within manufacturing than across a broader and wider set of sectors. However, the curve is flatter for the UNIDO data set, making the identification of the point of minimum concentration less precise. In all datasets, the precision in the estimation of the smoothed curves is very high, as the 5 percent confidence bands, obtained as ± 2 times the standard error of the prediction, track the point estimates very closely.

1.3.1 Significance of the upward-bending portion of the U-curve

The second use of our non parametric procedure is to allow us to assess the statistical significance of the within-country estimated coefficient on income for a variety of subsamples of the data. In particular, since the U-curve is asymmetric and the minimum point occurs late in the development process, the question of the statistical significance of its upward bending portion arises for high levels of income. A systematic way to ascertain the significance of the upward bending portion of the curve is to examine the significance of the fixed-effects slope coefficient on income for each of the sub-samples used in the non-parametric estimation of the relationship. This limits the incidence of outliers.

Figures 4 to 6 display in bold the range of coefficients for which the fixed-effects slope coefficient estimates are significantly different from zero at the 95 percent confidence level. The figures show, for the various data sets, that there exists a range of income per capita, strictly greater than the minimum point, at which the slope estimates become generally statistically significant and positive. In addition, Table 3 displays the range of interval midpoint income levels, around the point where the U-shaped curves reach their minimum, for which we cannot reject the null hypothesis that the estimated slope coefficient on income is equal to zero at the 95 percent confidence level.

To further assess the significance of the upward bending portion of the curve, we also use sectoral employment and value added data from the OECD. Indeed, focusing on data pertaining to a set of 14 industrial countries at a higher level of disaggregation than the ILO data and applying our non-parametric procedure demonstrates that the upward bending portion of the curve is much more pronounced for the OECD sample (Figures 7 and 8) than for the samples which included developing countries. This is particularly the case for value added data. By focusing on the higher end of the income distribution and extending the whole-economy data coverage to the 2-digit level of disaggregation, these OECD results provide further evidence that the upward bending portion of the U-curve is indeed a feature of the data. It is noteworthy that the estimated OECD minimum points (Table 3) be so similar to those obtained using ILO and UNIDO data (around $9,000 in 1985 PPP US dollars).

1.3.2 Differences across data sources

A striking feature of these results is the similarity in the shape of the estimated curves and in the level of income corresponding to minimal concentration, despite differences in data sources, sectoral coverage, level of disaggregation, and measures of concentration.

With respect to data sources, Figures 1-3 and Figures 7-8 reveal similar U shaped curves across the UNIDO, ILO and OECD datasets, whether the definition of sector size involves employment or value added.[18] All datasets suggest that the second stage of diversification starts on average when countries reach a level of per capita income roughly equal to $9,000 in 1985 constant US dollars, with the exception of the OECD value added data. For the latter, the phase of reconcentration seems to occur earlier, around $7,000, so that the smoothed curve is mostly upward sloping over observed OECD values of income. This level of income, however, is sufficiently high to be consistent with our earlier observation that the second stage of diversification starts late in the development process, as the OECD sample focuses by definition on economies that industrialized early. In Section 3 we discuss in more detail those differences.

A noteworthy discrepancy across datasets is the extent of reconcentration in the second phase. In the UNIDO dataset, which covers only manufacturing activity, the smoothed curves are relatively flatter in their reconcentration phase than in the economy-wide data from the ILO and OECD. It is plausible that such a difference is due to the coverage of the UNIDO dataset (manufacturing only rather than economy-wide), rather than the level of disaggregation or the country coverage. This is likely to be the case because the ILO and OECD datasets deliver similar curves, although they are obtained from different levels of disaggregation (1 and 2 digit data, respectively), and vastly different country coverages, but they both cover all economic activities. This finding implies that within-manufacturing specialization is slower in late stages of development than economy-wide specialization.

Our results are also very similar across alternative measures of concentration. Among the noteworthy differences are that U-shaped curves obtained using the log variance are typically much more pronounced than with alternative measures, and the minimum point obtained from this measure tends to occur slightly earlier.

2 Extensions and Robustness

2.1 Within versus Between Variation

As suggested above, while our nonparametric method allows us to obtain a flexible functional form linking sectoral concentration and income, it does not allow us to determine whether the resulting relationship is driven by between or within country variation in the data. This is because, by necessity and construction, different countries will appear in different subsamples used to draw the curves. Therefore, between country variation will be partly responsible for changes in the fitted measures of sectoral concentration. To examine whether the estimated U-shaped relationship is attributable to within or between country variation (or both), we turn to a quadratic specification.

Indeed, the nonparametric analysis presented above suggests that such a specification would adequately capture the stages of diversification. An added advantage of such a specification is that, contrary to our non-parametric methodology, it allows out-of-sample predictions on the stages of diversification, and it provides precise parameter estimates for the shape of the curve, above and beyond the graphical representation provided in Figures 1-3.

2.1.1 Within-country estimates

Table 4 displays the coefficients from fixed-effect regressions of sectoral concentration on per capita income and the square of per capita income for all measures of concentration. As before, within-country results demonstrate the existence of a statistically significant U-shaped relationship between sectoral concentration of employment and the level of per capita income, irrespective of the measure of dispersion. Figures 11-15 confirm graphically the existence of the non-linearity, by displaying the estimated within-country relationship (solid line) for the GINI coefficient, against the backdrop of the pooled-data scatter plot.[19] The economic significance of this relationship appears clearly for the ILO and OECD data sets, that is, for a broader level of disaggregation covering a wider range of sectors. Again, the relationship is flatter for the UNIDO data. However, the estimated fixed-effects coefficients on income (negative) and on income squared (positive) are highly significant statistically for all data combinations (Table 4).

The minimum point, after which countries begin to reconcentrate, can be calculated easily for each set of estimates, by setting the derivative of sectoral concentration with respect to income to zero. This leads to estimates that are very comparable to the minimum point estimates obtained using our non-parametric procedure, namely levels of 1985 constant US dollar per capita income centered around $9,000. The only difference worth noting is that the beginning of the second stage is estimated to occur later (around $11,000) for the ILO using the quadratic specification, relative to the nonparametric method. In short, a parametric analysis of the data focusing on the within-country dynamic variation in the measures of concentration leads to conclusions similar to those obtained in our non-parametric analysis.[20]

2.1.2 Between-country estimates

The evidence presented above pertains to within-country variation, which is perhaps most relevant as it describes what happens to measures of sectoral concentration in a typical country along the growth process. However, our result holds in a standard cross-country approach as well. To show this, we turn to between-country evidence, presented in Table 5. The table displays results based on a between estimator (OLS on country means) applied to all three data sources.[21] For the ILO and UNIDO samples, the U-shaped pattern is more pronounced than under fixed-effects, and it remains highly significant statistically despite the much reduced number of data points.[22] As expected from the small number of degrees of freedom, the curves obtained from OECD data are essentially flat. Table 5 also reports the implied minimum points, which are again in line with our previous findings.[23]

Within and between-country estimates are very similar. However, the magnitude of the coefficients on income squared (that is, the curvature of the U) is generally larger using the between estimator than using the within estimator. This is especially true for the UNIDO data set. This likely stems from measurement error in these data, leading to more errors-in-variables bias under fixed-effects relative to OLS on means.[24] The between-country estimates are therefore consistent with the hypothesis that the true curvature of the U-shapes might be more pronounced than the within-country dynamic variation in the data might alone suggest.

2.2 Country and Period Specific Analyses

2.2.1 Country examples

A straightforward way to work towards an empirical explanation for the stages of diversification is to examine specific country experiences. To the extent that cross-country variations in the dynamics of diversification can be related to other characteristics of these countries, we can hope to pinpoint factors, beyond income, that influence the location of a country on the U-curve.

Figures 16 to 18 report several representative country experiences chosen from our three data sources. As was already mentioned, most of the fourteen countries in the OECD sample are specializing. It is interesting to notice however that this applies to the large and relatively closed US economy as well as to small and open Belgium, at comparable levels of per capita income (Figure 16). Italy and France, on the other hand, are similarly sized and open to trade, but at very different stages of diversification, a fact that seems largely attributable to different per capita incomes over the available sample period.

This observation is confirmed in the UNIDO and ILO datasets: Figures 17 and 18 display the relationship between income and the Gini coefficient of sector shares for a set of highly different economies from the viewpoint of size and openness. China and Malaysia, two poor countries with very different sizes and exposures to international trade, are located at similar stages, while the same holds, among industrialized countries, for Denmark and the United States (Figure 17). In Figure 18, both stages of diversification appear clearly for Ireland and Spain, two fast growing countries with enough data to track the entire U-shaped curve over the sample period. One interesting feature of these examples is that there is some heterogeneity in the income corresponding to maximum diversification. In the case of Singapore this seems to have occurred around $2,500 of per capita income, in the case of Cyprus at $5,800 while in the case of Ireland it is estimated to have occurred around $7,000.[25] All three are small open economies and seem to have started specializing at levels of income lower than our benchmark estimate from the pooled sample ($9,000).

While they pertain only to a few illustrative data points, these country-specific examples suggest the possible tendency for open countries to start specializing at lower levels of income per capita. This does not alter the nature of the relationship between diversification and income, nor does it necessarily predict a straightforward link between the stages of diversification and openness, for even closed economies ultimately specialize once they reach a high enough level of income per capita. This anecdotal evidence suggests it is the *interaction* of income per capita and openness that determines the stages of diversification. The next section provides further evidence in favor of this possibility.

2.2.2 Dispersion of the minimum point

A more systematic way to get a sense of interesting deviations from the rule established earlier is to focus on those countries for which all stages of diversification are observed in-sample, and to examine when the minimum level was reached, and at what level of income per capita and trade openness. Table 6 presents this cross-section for our three employment datasets.[26] A number of observations are of interest. Firstly, there are 21 countries for which we observe all stages of diversification in the UNIDO sample, which has a longer time series of data available than the ILO. There were many fewer in the ILO and OECD samples, respectively 6 and 5. Hence, in all datasets a vast majority of countries are either monotonically diversifying, or monotonically specializing, making it difficult to precisely characterize the dispersion in the minimum point.

Secondly, the average income per capita at minimum specialization is remarkably close to the findings based on our two estimation methods and reported in Tables 3 and 4, with $9,161, $10,530 and $5,782 in the OECD, ILO and UNIDO employment samples, respectively. This is indeed striking, as the samples in Table 6 are very different from those used in earlier estimations, where purely diversifying or specializing countries were also included. There is, however, substantial dispersion of income levels around these means. For example, in the UNIDO sample, the standard deviation of income at the minimum point was $3,348.[27]

The main use of Table 6 is that it makes it possible to ask what determines deviations from cross-country means. The last two rows in each panel of Table 6 split the sample according to the median date of minimal specialization, in order to investigate whether there are identifiable differences in the characteristics of countries which reached the minimum early, and those that did so relatively late. We report averages for the Gini coefficient, income per capita, as well as the level of openness measured by the ratio of imports plus exports to GDP, for countries that reached their minimum before and after the sample median year. The split seems to be relevant for two reasons: countries that went through a minimum level of specialization relatively early tend to be substantially more open to trade on average, by 15 percentage points of trade to GDP in the UNIDO dataset (11.5 and 17 percentage points in the ILO and OECD datasets, respectively). Moreover, countries that tended to be late specializers were richer on average when they did so.[28] In Table 6, even closed economies end up specializing once they reach a high enough level of income. This suggests again that income per capita and openness are substitutes in determining the stages of diversification. Interestingly, this is consistent with the theories discussed in Section 4.

2.3 Robustness Tests

In this subsection we discuss a number of robustness checks.[29] In most cases, they failed to affect the main results on the stages of diversification.

4-Digit UNIDO data We examined the possibility that our pattern would hold at finer levels of disaggregation within the manufacturing sector. The UNIDO published annual data for manufacturing sectors at the 4-digit level of disaggregation (corresponding to a maximum of 81 sectors per country-year), from 1977 to 1997. One important caveat is that these data are quite spotty and likely to be of lesser quality than 1-digit and 3-digit employment data. Replicating the results for the quadratic specification presented above at the 4-digit level, however, led to interesting results: the U-curve pattern was preserved for all measures of sectoral labor concentration. Our estimates on income and income squared were close to statistically significant for the Herfindahl index of concentration, and consistently significant when using the between-country variation in the data.

They were not, however, statistically significant (although they were of the expected signs) when using a within estimator applied to the Gini coefficient of sectoral concentration. We ascribe the lower precision of our estimates at the 4-digit level to the lower quality of the data. In fact, given the poor quality of the 4-digit data it is remarkable that the stylized fact documented in this paper should show up at all.

Boundedness of the concentration measures Each measure of concentration used in this study is bounded above by 1 and below by 0. This may drive part of the non-linearity in the relationship between sectoral concentration and income levels. As the data stand, however, there are very few data points that lie near the bounds, where the boundedness of the measures could generate artificial nonlinearities. However, in order to examine this issue, we considered logistic transforms of each of the concentration measures. These provide a way of transforming the variables so they are not bounded above or below anymore. We then used these transformed measures as dependent variables in fixed effects regressions on income and income squared. The results provided further evidence concerning the robustness of the U-curve, as the coefficients were of the expected signs and statistically significant.

The Impact of Country Size One possible source of bias in our estimates is the equal treatment imposed to large and small countries in our samples. Not only is the level of sectoral diversification likely to be substantially higher in larger countries, but its dynamic relationship to income may also differ. To investigate the impact of country size, we constructed two subsamples, one excluding countries in the highest quartile of the distribution of country sizes, the other excluding countries in the smallest quartile. Size was measured by average population over the available sample period. We found no evidence that the estimated shape of the U-curves differed in the two subsamples, or that the estimated income at the minimum point differed significantly.[30]

Removing the agricultural sector Insofar as the results pertaining to the ILO dataset are driven by the structural shift away from agriculture, into manufacturing and services, and later from manufacturing to services, we might expect that the U-shaped pattern uncovered in the ILO sample would vanish when agriculture is removed from the sectoral coverage of this data set.[31] Indeed, if industrialization were the reason for the non-monotonicity, the curve would appear relatively flat at first, and increasing at higher levels of income, rather than U-shaped. However, this was not the case: the quadratic functional form appears supported by the data for the ILO dataset even when the agricultural sector (as well as mining and quarrying) is excluded from the data. Furthermore, we are confident that our stylized fact is not due only to movements of resources away from agriculture and into manufacturing and services, since the U-shaped pattern holds within the manufacturing sector alone (UNIDO data set).

Period-specific analysis We also examined whether the shape of the estimated curves changes according to whether the data originated from different time periods. To do so, we split the sample at 1980, and ran separate quadratic fixed-effects regressions for the two subsamples. We still found evidence for a U-shaped relationship between income and sectoral concentration in both cases, but the curvature of the U was always more pronounced before 1980 than after. This may be due to the fact that the second half of the sample period is, by virtue of the time trend in income, likely to be one where countries were observed closer to the bottom of their U-curve or in their second stage of diversification, where the magnitude of the slope on income tends to be smaller than in the first stage.

Region-specific analysis Finally, we performed sample split exercises according to geographic regions. We focused on subgroups of developing economies, namely Latin America, Sub-Saharan Africa and South-East Asia (as well as all non-OECD countries taken together). Our results hold in almost all cases. In particular, they hold for Latin America in all cases, for UNIDO value-added measures in Sub-Saharan Africa (we do not have enough ILO data for this subsample), and for ILO employment and UNIDO value-added data in Southeast Asia. We think the two exceptions largely stem from insufficient observations, as we have only 4 countries in Southeast Asia and 12 in Sub-Saharan Africa. Indeed, the non-monotonicity held very robustly among non-OECD countries taken as a whole.

3 Theoretical Interpretation

In this section, we review the main theories underpinning the incentives to diversify or specialize, and present a variety of theoretical hypotheses to account for the sequencing of diversification and specialization.

3.1 Diversification and Specialization

Theoretical reasons for countries to diversify are based on two types of arguments: one is related to the structure of preferences, the other is inspired from portfolio arguments. It is well-known that if agents have non-homothetic preferences, their consumption pattern will change as income grows. These Engel effects are generally understood as implying an expanding diversity of the goods consumed. Since in a closed economy production patterns respond to changes in the structure of demand, preference-based arguments are sufficient to generate increasing sectoral diversification throughout development.

An alternative is introduced in Acemoglu and Zilibotti (1997), where diversification occurs endogenously as a result of agents' decisions to invest in a range of imperfectly correlated risky projects, or "sectors". Each sector entails idiosyncratic risk, and diversification is imperfect because the number of sectors is limited by requirements on the minimal size of a given project. Due to these indivisibilities, sectoral diversification opportunities improve with the aggregate capital stock. Conversely, the more sectors are open, the easier it is to diversify idiosyncratic risk and thus to invest in risky projects. Therefore, "development goes hand in hand with the expansion of markets and with better diversification opportunities" (p.711). In this model, diversification is a process that accompanies economic growth through a portfolio motive.

Theoretical arguments for specialization are two-fold as well. Firstly, Ricardian theory relates specialization to the intensity of trade. For instance Dornbusch et al. (1977) present a model with a continuum of goods where the range of goods produced domestically and imported is endogenous. There, falling transport costs (or tariffs) result in a shrinking range of (non-traded) goods produced domestically, thus foster specialization.

The second class of explanations for sectoral concentration originates in the economic geography literature (Krugman, 1991). This literature has emphasized the importance of demand externalities in explaining the agglomeration of economic activities in specific regions or cities. Typically, these externalities make it optimal for monopolistic competitors to cluster since their profits increase in local expenditures, themselves a positive function of the number of local firms. The idea that transport costs fall with better technologies tends to reinforce the mechanism, as the demand linkages effect is decreasingly mitigated by the necessity to be located close to demand. Thus, increasingly integrated economies are expected to agglomerate regionally. Such regional agglomeration translates into increasing observed sectoral concentration once the world is arbitrarily partitioned into countries.[32]

3.2 Endogenizing the Stages of Diversification

Our evidence on the stages of diversification shows that specialization and diversification occur at different points in development. The aforementioned theories only predict a monotonic evolution of sectoral concentration. In the working version of this paper, we sought to reconcile various arguments concerning sectoral diversification and specialization, by proposing one of many possible models predicting a non-monotonic relationship between sectoral concentration and income levels.[33]

In our theory, the stages of diversification resulted from the interaction of productivity increases and trading costs. In a dynamic Ricardian model with a continuum of goods, an exogenous increase in a country's aggregate productivity level, *relative to the rest of the world,* translates into an increasing range of goods produced domestically. On the other hand, decreasing transport costs, as mentioned earlier, tend to be a force for increased concentration. We also show that the number of sectors is directly related to measures of diversification such as the Gini coefficient, even under the arbitrary truncation we impose on our data.[34] The observed stages of diversification then depend on which force dominates at any given point on a country's growth path. Under certain assumptions on the dynamic evolution of relative productivity and the fall in transport costs, countries are predicted to first diversify, and then reach a point at which the force of concentration dominates. This will occur, for example, if it is increasingly difficult to close the technological gap whereas transport costs decline linearly, or conversely, if the technological gap falls at a constant rate but the decline in transport costs accelerates as capital is accumulated.[35]

Gilles Saint-Paul (1992) presents a model where limited access to financial markets affects the pattern of domestic production, as sectoral diversification is the only available means to diversify away sector-specific income shocks and smooth consumption. Thus, in the context of incomplete markets, countries can be led to diversify for insurance purposes, and specialize again as financial markets deepen and the portfolio motive ceases to dominate comparative advantage considerations.

While these are by no means the only models consistent with the stages of diversification, both display the important feature that the stages of diversification are endogenous to both trade and economic growth.

In fact, this paper relates to an older line of research initiated by Hollis Chenery and Moshe Syrquin, who also took interest in the dynamics of economic structure, but focused on a much coarser approach, based on three sectors only. Furthermore, they did not document the U-shape we uncover in a much more general set of conditions. They also concluded, however, that structural change (during industrialization) responded endogenously to trade policy and economic growth. In particular, they write: "economies which pursued export-led growth [...] industrialized sooner, had higher rates of total factor productivity and tended to achieve the input-output structure of an advanced economy faster" (Chenery, Sherman Robinson and Syrquin, 1986, p. 358).[36] This is consistent with the suggestive evidence we present in the previous sections. In this sense, we show that the patterns of structural change during industrialization are but a part of a much more widespread phenomenon. In particular, it applies to economic activities within the secondary sector, and for a very large number of countries and data sources.

4 Conclusion

Using a wide panel of countries, this paper informs the theoretical debate about the evolution of sectoral diversification across time and across countries. The empirical verdict is clear. Poor countries tend to diversify, and it is not until they have grown to relatively high levels of per capita income that incentives to specialize take over as the dominant economic force. This non-monotonicity is a very robust feature of the data, and goes beyond the well-documented shift of resources from agricultural sectors to manufacturing and services.

Our stages of diversification reflect the reallocation of resources over a range of activities that is much less coarse than the well-know categorization into three sectors (primary, secondary, tertiary). This reallocation appears to be driven by the interaction of economic growth and openness to trade, in a way that is compatible with some combination of existing theories. Which of these is most relevant is a question we leave open for future research.

Appendix A. Sectoral Coverage

1. ILO 1-Digit Classification (9 sectors)

1. Agriculture, Hunting, Forestry and Fishing
2. Mining and Quarrying
3. Manufacturing
4. Electricity, Gas and Water
5. Construction
6. Wholesale and Retail Trade and Restaurants and Hotels
7. Transport, Storage and Communication
8. Financing, Insurance, Real Estate and Business Services
9. Community, Social and Personal Services

2. UNIDO 3-Digit Classification (28 sectors)

300 Total manufacturing

311 Food products

313 Beverages

314 Tobacco

321 Textiles

322 Wearing apparel, except footwear

323 Leather products

324 Footwear, except rubber or plastic

331 Wood products, except furniture

332 Furniture, except metal

341 Paper and products

342 Printing and publishing

351 Industrial chemicals

352 Other chemicals

353 Petroleum refineries

354 Miscellaneous petroleum and coal products

355 Rubber products

356 Plastic products

361 Pottery, china, earthenware

362 Glass and products

369 Other non-metallic mineral products

371 Iron and steel

372 Non-ferrous metals

381 Fabricated metal products

382 Machinery, except electrical

383 Machinery, electric

384 Transport equipment

385 Professional and scientific equipment

390 Other manufactured products

3. OECD 2-Digit Classification (20 sectors)

100. Agriculture, hunting, forestry and fishing

200. Mining and quarrying

310. Food, beverages and tobacco

320. Textiles, wearing apparel and leather industries

330. Wood and wood products, including furniture

340. Paper and paper products, printing and publishing

350. Chemicals, chemical petroleum, coal, rubber, plastic products

360. Non-metallic mineral products excl. products of petroleum & coal

370. Basic metal industries

380. Fabricated metal products, machinery and equipment

390. Other manufacturing industries

400. Electricity, gas and water

500. Construction

610+620. Wholesale trade and retail trade

630. Restaurants and hotels

700. Transport, storage and communication

810+820 . Financial institutions and insurance

830. Real estate and business services

910. Community and Social services

920. Personal services

Appendix B. Geographic Coverage

Algeria [ac]	Ethiopia [a]	Korea	Saudi Arabia [a]
Argentina [a]	Finland [e]	Kuwait [a]	Senegal [a]
Australia [e]	France [be]	Luxembourg [b]	Seychelles [b]
Austria	Gabon [a]	Madagascar [ac]	Singapore [b]
Bahamas [ac]	German Dem.[ad]	Malawi [ac]	South Africa [a]
Bangladesh	Germany, Fed.[e]	Malaysia	Spain
Barbados [b]	Ghana [a]	Mexico [b]	Sri Lanka
Belgium [be]	Greece	Morocco [b]	Sweden [e]
Benin [a]	Guatemala [a]	Mozambique [ac]	Switzerland [b]
Bolivia [b]	Guyana [ac]	Myanmar [b]	Syria [c]
Brazil	Haiti [ad]	Netherlands [be]	Taiwan [a]
Burkina Faso [a]	Honduras	New Zealand [d]	Thailand [b]
Burundi [a]	Hong Kong [b]	Nicaragua [a]	The Gambia [a]
Canada [e]	Hungary	Nigeria [a]	Trinidad/Tobago [b]
Chile	Iceland [d]	Norway [be]	Tunisia [d]
China	India [a]	Norway [ac]	Turkey
Colombia	Indonesia [b]	Pakistan	U.S.A [e]
Costa Rica [b]	Iran [a]	Panama [b]	U.S.S.R. [ad]
Cyprus	Iraq [ac]	Paraguay	U. A. Emirates [ac]
Czechoslovakia [a]	Ireland	Peru a	United Kingdom [e]
Denmark [e]	Israel [d]	Philippines	Uruguay
Dominican Rep.	Italy [e]	Poland	Venezuela
Ecuador	Jamaica [b]	Portugal [b]	Yugoslavia [a]
Egypt	Japan [e]	Puerto Rico [b]	Zaire [ad]
El Salvador	Jordan [ad]	Romania [b]	Zambia [ad]

a: Not in ILO dataset

b: not in UNIDO dataset

c: Not in UNIDO- Employment dataset

d: Not in UNIDO- Value Added dataset

e: In OECD dataset

References

Acemoglu, Daron and Zilibotti, Fabrizio. "Was Prometheus Unbound by Chance ? Risk, Diversification, and Growth." *Journal of Political Economy*, August 1997, 105(4), pp. 709-751.

Barro, Robert and Sala-i-Martin, Xavier. "Technological Diffusion, Convergence, and Growth." *Journal of Economic Growth*, March 1997, 2(1), pp. 1-26.

Chenery, Hollis; Robinson, Sherman and Syrquin, Moshe. *Industrialization and Growth: A Comparative Study*, Oxford: Oxford University Press for the World Bank, 1986.

Cleveland, William S. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association*, December 1979, 74(368), pp. 829-836.

Cuñat, Alejandro. "Sectoral Allocation in the Process of Growth." Mimeo, London School of Economics, 2000.

Dornbusch, Rudiger; Fischer, Stanley and Samuelson, Paul A.. "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods." American Economic Review, December 1977, 67(5), pp. 823-839.

Fagerberg, Jan. "A Technological Gap Approach to Why Growth Rates Differ," in Edward N. Wolff, ed., *The Economics of Productivity.* New York: Elgar, 1997, pp. 459-471.

Goodall, Colin. "A Survey of Smoothing Techniques." in J. Fox and J.S. Long, eds., *Modern Methods of Data Analysis.* Newbury Park, CA: Sage Publications, 1990, pp. 126-176.

Harrigan, James. "Specialization and the Volume of Trade: Do the Data Obey the Laws?", forthcoming in James Harrigan and Kwan Choi, eds., *The Handbook of International Trade*, Malden, MA: Basil Blackwell, 2002.

Hummels, David. "Have International Transportation Costs Declined?" Mimeo, University of Chicago, 1999.

Imbs, Jean and Wacziarg, Romain. "Stages of Diversification." Center for Economic Policy Research (London, UK) Working Paper no. 2642, December 2000.

International Labor Office. *Yearbook of Labor Statistics.* Geneva: International Labor Office Press, 1997.

Kalemli-Ozcan, Sebnem; Sorensen, Bent E. and Yosha, Oved. "Risk Sharing and Industrial Specialization: Regional and International Evidence.", forthcoming, *American Economic Review*, 2002.

Kim, Sukkoo. "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860-1987." *Quarterly Journal of Economics*, November 1995, 110(4), pp. 881-908.

Krugman, Paul. *Geography and Trade.* Cambridge, MA: MIT Press, 1991.

Krugman, Paul and Venables, Anthony. "Globalization and the Inequality of Nations." *Quarterly Journal of Economics*, November 1995, 110(4), p. 857-880.

Neary, J. Peter. "Of Hype and Hyperbolas: Introducing the New Economic Geography." *Journal of Economic Literature*, June 2001, 39(2), pp. 536-561.

Nelson, Richard R. and Phelps, Edmund. "Technological Diffusion and Economic Growth." *American Economic Review*, March 1966, 56(1), pp. 69-75.

Organization for Economic Cooperation and Development. *International Sectoral Database (ISDB).* Paris: Organization for Economic Cooperation and Development Press, 1998.

Royston, Patrick. "gr6: Lowess Smoothing." *Stata Technical Bulletin Reprints*, 1991, 1, pp. 41-44.

Saint-Paul, Gilles. "Technological Choice, Financial Markets and Economic Development." *European Economic Review*, April 1992, 36(4), pp. 763-781.

Heston, Alan and Summers, Robert. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988." *Quarterly Journal of Economics*, May 1991, 106(2), pp. 327–368.

United Nations Industrial Development Organization. *Industrial Statistics Database, 3-Digit Level of ISIC Code.* Vienna: United Nations Industrial Development Organization Press, 1997.

Ventura, Jaume. "Growth and Interdependence." *Quarterly Journal of Economics*, February 1997, 112(1), pp. 57-84.

* We thank Alberto Alesina, Lanier Benkard, John McMillan, Gian Maria Milesi-Ferretti, Paul Romer, Andrei Shleifer, an anonymous referee, seminar participants at the Universite de Lausanne, ECARES, Trinity College Dublin, ICRIER, CERGE-EI, the University of Chicago, the University of Manchester, the NBER Summer Institute 2000, the University of Maryland, Boston College, the IMF and Stanford University for useful comments. All remaining errors are ours. Jean Imbs gratefully acknowledges financial assistance from the Fonds National de la Recherche Scientifique (Switzerland), Grant 12-53584.98. Contact information: Imbs: London Business School, Dept. of Economics, Regent's Park, London NW1 4SA, UK; jimbs@london.edu, tel: (44-207) 706 6704. Wacziarg (corresponding author): Stanford Graduate School of Business, 518 Memorial Way, Stanford CA 94305-5015, USA; wacziarg@gsb.stanford.edu, tel: (650) 723 6069.

Notes

[1]This incentive has recently been related to the fact that aggregate volatility tends to fall with income per capita in Daron Acemoglu and Fabrizio Zilibotti (1997). In a similar vein, Sebnem Kalemli-Ozcan et al. (2002) relate specialization patterns to the extent of risk sharing between regions and between countries.

[2]See for instance Rudiger Dornbusch et al. (1977), or the survey by James Harrigan (2002).

[3]We thank an anonymous referee for pointing out this possibility. In fact, at the 1-digit level of disaggregation (International Labor Office data), services are covered by 6 sectors, while manufacturing is lumped into one category and agriculture and mining into two sectoral categories.

[4]Appendix 1 contains further details concerning data coverage.

[5]We also use employment data from the UNIDO at the 4-digit level of disaggregation (Section 2.4). These data cover the 1977-1996 period. While they are of relatively poor quality and their coverage is limited, they will further document the robustness of our results.

[6]The dataset was constructed so that the number of sectors available *through time* for each country was constant. This required abandoning observations on some sectors, when observations for a given country were not available for all years. On the other hand, the number of sectors available to compute the indices of concentration varied *across countries*. For fixed effects estimate we retained countries for which 27 or more sectors were available (UNIDO), or where 6 or more sectors were available (ILO). This variation is reflected purely in the country-specific effects. For regressions using between-country variation we restricted the sample to observations with the same number of available sectors (all 28 for UNIDO data and all 9 for the ILO).

[7]Since our per capita income data from Alan Summers and Robert Heston (1991) only extends to 1992, in most

cases some years of available data for both the ILO and the UNIDO datasets had to be discarded.

[8]See for instance Krugman (1991) and Sukkoo Kim (1995), among many others.

[9]See for instance Krugman (1991), p.55, who used "locational Ginis" to measure the geographic concentration of given sectors across locations. In contrast, we compute "sectoral Ginis", measuring the concentration of labor and value added across sectors in a given country.

[10]As a result of this, contrary to the other measures, our results were sensitive to the use of the interquartile range and, to a lesser extent, of the mean-median spread. If sector shares are measured with error, the error-to-truth ratio will be exacerbated in any measure of dispersion which uses differences in sector shares. This is not the case for the max-min spread, which by construction is highly correlated with the share of the biggest sector.

[11]In this application, a parametric approach could create problems. For example, fitting a quadratic relationship to the data could lead to erroneous conclusions. Suppose the true relationship between sectoral concentration and income were characterized by a semilog specification. A quadratic specification would lead us to erroneously infer an upward bending portion for the curve.

[12]See William S. Cleveland (1979) and Colin Godall (1990) for technical descriptions of this method of nonparametric analysis. See Patrick Royston (1991) for a computational implementation.

[13]The *values of S* across datasets were as follows: ILO dataset, $S = 529$. UNIDO employment dataset $S = 489$. UNIDO value added dataset: $S = 609$. OECD employment dataset $S = 409$. OECD value added dataset: $S = 409$. The average *number of observations* per subsample was as follows: ILO dataset, $N = 246$ (ranging from 88 to 394 observations per subsample); UNIDO employment dataset $N = 388$ (ranging from 108 to 963 observations); UNIDO value added dataset: $N = 358$ (ranging from 111 to 972 observations); OECD employment dataset $N = 155$ (ranging from 41 to 230 observations); OECD value added dataset: $N = 181$ (ranging from 65 to 256 observations).

[14]A typical *lowess* estimator would use 80% of the data in each regression, would weigh distant observations less, and would employ every ordered observation on income as a midpoint ($N = S$). See Royston (1991). Hence, if we had 1000 observations we would run 1000 regressions to obtain each curve. Here, by using increments of $25, instead, we reduce by about two thirds the number of regressions that need to be computed to plot each curve. Rectangular weights further reduce computational time. We checked that the curves obtained using different parametrizations are very similar.

[15]As suggested earlier, the term $\hat{\beta}_s^{FE} \times x_s$ is used solely for the purpose of evaluating the statistical significance of the relationship between income and sectoral diversification at various levels of income. Another way of saying this is that the curves would look very similar if we omitted the term $\hat{\beta}_s^{FE} \times x_s$ from equation (1) - but we would not

be able to ascertain precisely whether, say for incomes between \$12,500 and \$17,500, the relationship between GINI and income was indeed significantly positive statistically.

[16]The corresponding figures for the other measures of sectoral concentration are available upon request. Except for the interquartile range (IQR), and to a lesser extent for the mean-median spread, they look very similar to those obtained using the Gini coefficient.

[17]\$9,637 in Summers-Heston dollars.

[18]Interestingly, this suggests that employment-based measures of specialization, which are prevalent in the existing literature, adequately capture the distribution of economic activity across sectors.

[19]In other words, the scatter plot does not display deviations from the means. This is to allow comparability with pooled least squares plots, which reveals the source of the variation - cross-sectional (between) or time series (within).

[20]To further assess the significance of the upward bending portion of the curve, we restricted the sample to observations with income greater than the estimated point of minimum concentration and ran a simple fixed-effects regression of concentration measures on the level of income. For all specifications, the coefficient on the income variable was positive and highly significant statistically. These results are available upon request.

[21]The set of countries is reduced somewhat because the number of sectors used for each country has to be equal (countries with missing sector observations had to be deleted). OECD results are not likely to be very meaningful, since they involve only 14 countries. We include them for the sake of completeness.

[22]These results are confirmed when both the within and between variations are pooled. Random effects estimates which optimally weigh the within and between variations in the data are available upon request. As expected, compared to fixed-effects the statistical significance of the estimates is raised considerably, and their magnitude is larger.

[23]Graphs of the fitted curves against the backdrop of scatterplots of country averages, providing a graphical representations of the U-curves obtained from between estimates, are available upon request. They confirm our observations based on the parameter estimates.

[24]It is well known that under classical (white-noise) measurement error, errors-in-variables bias is exacerbated since we are differencing highly autocorrelated right-hand side variables (income and income squared). Hence, the error-to-truth ratio is raised by differencing. In any event, the U-shaped pattern remained statistically significant even when it was estimated to be relatively flat under fixed effects (in the UNIDO dataset).

[25]Belgium, in the OECD sample, experienced a turnaround point out of sample, at least below \$8,300.

[26]To construct Table 6, we excluded all countries where the minimal value of the Gini coefficient was observed at

one extreme level of income per capita. We then further dropped countries where no clear non-monotonicity was observable over the whole sample. The plots for the resulting countries are available upon request.

[27]The UNIDO sample contains three economies that were particularly poor when they started re-specializing: Madagascar, the Central African Republic and Congo. Data for these countries are very noisy, although it was not possible to rule out the presence of an observable non-monotonicity. Excluding them led to an average estimated income level at the minimum point equal to $6,540.

[28]When Madagascar, Central African Republic and Congo are excluded from the UNIDO sample, the contrasts are even more striking: average income per capita for early specializers becomes $5,381.3 vs. $7,700.6 for late specializers, and average levels of openness are 85.35% and 61.32%, respectively.

[29]The estimates corresponding to these robustness checks are available upon request.

[30]We also controlled for country size in the basic cross-sectional quadratic specification, confirming that larger countries are indeed more diversified in absolute terms. This did not affect the shape of the estimated relationship linking sectoral concentration and income.

[31]Doing so provides an additional robustness check with respect to the potentially arbitrary definition of sectors in the data.

[32]Krugman and Anthony Venables (1995) go further and introduce congestion costs, with immobile labor and wages that increase endogenously with specialization. Accordingly, there is a point at which deagglomeration becomes optimal. These dynamics have been documented by Kim (1993), who finds evidence of regional specialization in the US until the turn of the twentieth century, followed by a reversal of the trend between the 1930s and the 1980s. We were unable to uncover a systematic pattern of deagglomeration at very high levels of income. While the UNIDO data seems to display such a pattern in the upper tail of the income distribution, there were too few observations at those levels of income for this pattern to show up in a statistically significant way.

[33]See Jean Imbs and Romain Wacziarg (2000).

[34]As explained earlier, in our data the number of sectors in each country was kept constant over time.

[35]See for instance, Richard R. Nelson and Edmund Phelps (1966), Jan Fagerberg (1997) or Robert Barro and Xavier Sala-i-Martin (1997) for "technological gap" models providing justification for the assumption that the technological gap gets increasingly harder to close. See David Hummels (1999) for an empirical argument in favor of the assumption that transport costs only start falling late in development.

[36]For more recent research stressing the relationship between economic growth, trade openness and structural transformation, and combining insights from neoclassical growth theory and the Heckscher-Ohlin model of trade, see

Jaume Ventura (1997) and Alejandro Cuñat (2000).

Table 1 - Summary Statistics for the Sectoral Concentration Indices (pooled data)

	EMPLOYMENT		VALUE ADDED	
	Mean	**Std. Dev**	**Mean**	**Std. Dev**
ILO 1-digit	**885 obs.**			
GINI	0.479	0.098		
HERFIND	0.229	0.071		
COEFVAR	0.997	0.297		
MAXMIN	0.333	0.115		
LOGVAR	2.114	1.136		
UNIDO 3-digit	**1556 obs.**		**1493 obs.**	
GINI	0.573	0.102	0.570	0.112
HERFIND	0.118	0.098	0.107	0.068
COEFVAR	1.403	0.632	1.328	0.526
MAXMIN	0.226	0.134	0.207	0.110
LOGVAR	1.630	0.642	1.784	0.853
OECD 2-digit	**356 obs.**		**412 obs.**	
GINI	0.531	0.042	0.475	0.064
HERFIND	0.127	0.018	0.125	0.036
COEFVAR	1.176	0.157	0.987	0.154
MAXMIN	0.234	0.050	0.212	0.044
LOGVAR	1.073	0.150	1.032	0.370

Table 2 - Correlation Matrices for the Sectoral Concentration Indices (pooled data)

	EMPLOYMENT					VALUE ADDED				
	GINI	**HERFIND**	**COEFVAR**	**MAXMIN**	**LOGVAR**	**GINI**	**HERFIND**	**COEFVAR**	**MAXMIN**	**LOGVAR**
ILO 1-digit	**(885 Obs.)**									
GINI	1.000									
HERFIND	0.759	1.000								
COEFVAR	0.956	0.906	1.000							
MAXMIN	0.866	0.955	0.961	1.000						
LOGVAR	0.729	0.473	0.654	0.543	1.000					
UNIDO 3-digit			**(1556 Obs.)**					**(1493 Obs.)**		
GINI	1.000					1.000				
HERFIND	0.828	1.000				0.854	1.000			
COEFVAR	0.921	0.969	1.000			0.942	0.968	1.000		
MAXMIN	0.882	0.954	0.988	1.000		0.874	0.951	0.976	1.000	
LOGVAR	0.735	0.632	0.684	0.655	1.000	0.720	0.551	0.631	0.575	1.000
OECD 2-digit			**(356 Obs.)**					**(412 Obs.)**		
GINI	1.000					1.000				
HERFIND	0.708	1.000				-0.493	1.000			
COEFVAR	0.864	0.845	1.000			0.953	-0.295	1.000		
MAXMIN	0.760	0.904	0.934	1.000		-0.225	0.896	0.029	1.000	
LOGVAR	0.788	0.347	0.465	0.399	1.000	0.887	-0.390	0.818	-0.174	1.000

Table 3 – Non-Parametric Estimates of the Minimum Point (constant 1985 $)
and Range of Statistically Insignificant Slope Coefficients (95-percent confidence)

	GINI	HERFIND	COEF VAR	MAXMIN	LOGVAR
ILO – Employment					
Minimum Point	9,575	10,450	9,575	9,700	7,800
Y low*	5,875	5,725	5,800	9,675	4,850
Y high	10,950	10,825	10,800	10,875	8,175
UNIDO3 – Employment					
Minimum Point	8,675	8,825	8,800	9,000	7,300
Y low*	5,025	8,650	7,975	8,350	4,950
Y high	6,550	5,075	5,075	4,650	6,900
UNIDO3 – Value Added					
Minimum Point	8,725	7,225	9,825	9,925	9,475
Y low*	5,200	5,575	5,575	5,575	4,575
Y high	6,800	8,550	8,550	8,625	6,900

Table 3 - Non-Parametric Estimates of the Minimum Point (constant 1985 $)

and Range of Statistically Insignificant Slope Coefficients (95-percent confidence) (contd.)

	GINI	HERFIND	COEF VAR	MAXMIN	LOGVAR
OECD – Employment					
Minimum Point	9,250	9,175	8,375	8,650	9,250
Y low*	8,375	8,325	8,250	8,250	8,825
Y high	8,950	8,725	8,725	8,650	9,325
OECD – Value Added					
Minimum Point	6,975	7,625	6,975	6,975	6,050
Y low*	7,925	7,775	7,775	6,775	8,075
Y high	8,325	8,525	8,350	8,600	8,400

* Y low and Y high bound the range of statistically insignificant estimated slope coefficients β_s^{FE} (subsample midpoints with slopes statistically indistinguishable from zero).

Table 4a - Fixed-Effects Regressions of Sectoral Concentration on Income and Income Squared (unbalanced panel)[1]

ILO 1-Digit Data

	GINI	**HERFIND**	**COEFVAR**	**MAXMIN**	**LOGVAR**
ILO-Employment	885 obs, 64 countries				
Income	-0.027	-0.021	-0.0862	-0.0371	-0.0480
	(-12.11)	(-11.17)	(-11.35)	(-10.57)	(-1.29)
Income^2	0.001	0.001	0.0037	0.0016	0.0059
	(11.68)	(10.61)	(10.90)	(10.45)	(3.55)
Intercept	0.588	0.314	1.3419	0.4783	2.0336
	(63.23)	(40.48)	(43.05)	(33.22)	(13.36)
R-Squared	0.275	0.443	0.359	0.401	0.003
Minimum Point ($)	11,548	11,722	11,590	11,273	4,060

(t-statistics in parentheses)

[1] Throughout all the regressions that follow, the data on per capita PPP income is entered in thousands of 1985 constant US dollars to facilitate readability of the numbers.

Table 4b - Fixed-Effects Regressions of Sectoral Concentration on Income and Income Squared (unbalanced panel)

UNIDO 3-Digit Data

	GINI	HERFIND	COEFVAR	MAXMIN	LOGVAR
UNIDO3 - Employment	1556 obs, 67 countries				
Income	-0.0104	-0.0095	-0.0664	-0.0133	-0.0791
	(-7.44)	(-5.76)	(-7.05)	(-6.22)	(-5.03)
$Income^2$	0.0006	0.0005	0.0037	0.0007	0.0055
	(9.14)	(5.83)	(7.67)	(6.61)	(6.81)
Intercept	0.5975	0.1455	1.5837	0.2634	1.7984
	(134.97)	(27.63)	(52.87)	(38.64)	(36.00)
R-Squared	0.287	0.166	0.248	0.231	0.089
Minimum Point ($)	7,980	9,686	9,012	9,226	7,247
UNIDO3 – Value Added	1493 obs, 67 countries				
Income	-0.0161	-0.0073	-0.0706	-0.0125	-0.0826
	(-8.40)	(-4.52)	(-6.69)	(-5.42)	(-3.99)
$Income^2$	0.0009	0.0003	0.0033	0.0006	0.0059
	(9.57)	(3.79)	(6.55)	(5.17)	(5.95)
Intercept	0.6126	0.1308	1.5385	0.2446	1.9440
	(99.26)	(25.30)	(45.42)	(32.97)	(29.22)
R-Squared	0.388	0.210	0.328	0.286	0.074
Minimum ($)	9,185	12,500	10,701	10,997	7,025

(t-statistics in parentheses)

Table 4c - Fixed-Effects Regressions of Sectoral Concentration on Income and Income Squared (unbalanced panel)

OECD 2-Digit Data

	GINI	HERFIND	COEFVAR	MAXMIN	LOGVAR
OECD-Employment	356 obs, 14 countries				
Income	-0.030	-0.014	-0.1055	-0.0353	-0.0750
	(-11.67)	(-8.60)	(-8.36)	(-7.82)	(-13.98)
$Income^2$	0.002	0.0007	0.0056	0.0019	0.0037
	(13.44)	(10.00)	(9.80)	(9.17)	(15.49)
Intercept	0.659	0.184	1.6167	0.3816	1.4156
	(47.64)	(21.49)	(23.59)	(15.54)	(48.57)
R-Squared	0.137	0.121	0.174	0.148	0.014
Minimum Point ($)	9,639	9,542	9,465	9,464	10,021
OECD – Value Added	412 obs, 14 countries				
Income	-0.0101	-0.0030	-0.0298	-0.0041	-0.0006
	(-3.82)	(-3.05)	(-3.60)	(-1.56)	(-1.00)
$Income^2$	0.0007	0.0002	0.0020	0.0003	0.00001
	(5.46)	(4.52)	(5.30)	(2.88)	(0.48)
Intercept	0.5029	0.1326	1.0608	0.2131	0.0298
	(35.55)	(24.89)	(23.82)	(15.24)	(9.30)
R-Squared	0.075	0.007	0.126	0.029	0.005
Minimum Point ($)	7,730	7,437	7,484	5,976	22,952

(t-statistics in parentheses)

Table 5a - Between Regressions of Sectoral Concentration on Income and Income Squared

ILO 1-Digit Employment Data

	GINI	HERFIND	COEFVAR	MAXMIN	LOGVAR
ILO-Employment	51 countries				
Income	-0.0462	-0.0420	-0.1681	-0.0643	-0.3396
	(-5.61)	(-5.84)	(-5.99)	(-5.53)	(-2.52)
$Income^2$	0.0024	0.0022	0.0088	0.0034	0.0184
	(4.65)	(4.86)	(4.99)	(4.64)	(2.19)
Intercept	0.6506	0.3614	1.5820	0.5388	3.4626
	(27.45)	(17.48)	(19.59)	(16.10)	(8.93)
R-Squared	0.435	0.474	0.474	0.448	0.104
Minimum Point ($)	9,645	9,608	9,599	9,517	9,203

(t-statistics in parentheses)

Table 5b - Between Regressions of Sectoral Concentration on Income and Income Squared - UNIDO 3-Digit Data

	GINI	HERFIND	COEFVAR	MAXMIN	LOGVAR
UNIDO-Employment	51 countries				
Income	-0.0627	-0.0441	-0.3258	-0.0621	-0.2314
	(-4.99)	(-2.94)	(-3.75)	(-3.43)	(-3.29)
$Income^2$	0.0034	0.0023	0.0167	0.0030	0.0131
	(3.89)	(2.20)	(2.77)	(2.40)	(2.67)
Intercept	0.7547	0.2532	2.3991	0.4218	2.2702
	(25.09)	(7.05)	(11.54)	(9.71)	(13.48)
R-Squared	0.393	0.206	0.317	0.317	0.167
Minimum Point ($)	9,220	9,616	9,739	10,272	8,862
UNIDO-Value Added	50 countries				
Income	-0.0602	-0.0298	-0.2541	-0.0471	-0.2770
	(-4.38)	(-2.44)	(-3.25)	(-2.91)	(-2.92)
$Income^2$	0.0031	0.0015	0.0124	0.0022	0.0143
	(3.15)	(1.69)	(2.23)	(1.91)	(2.11)
Intercept	0.7479	0.2033	2.1310	0.3600	2.5340
	(22.83)	(6.98)	(11.43)	(9.31)	(11.18)
R-Squared	0.473	0.277	0.394	0.349	0.298
Minimum Point ($)	9,754	10,140	10,211	10,675	9,703

(t-statistics in parentheses)

Table 5c - Between Regressions of Sectoral Concentration on Income and Income Squared - OECD 2-Digit Data

	GINI	HERFIND	COEFVAR	MAXMIN	LOGVAR
OECD - Employment	14 countries				
Income	0.0203	0.0133	0.1626	0.0600	0.0620
	(0.38)	(0.72)	(1.04)	(1.34)	(0.29)
$Income^2$	-0.0007	-0.0005	-0.0058	-0.0022	-0.0029
	(-0.34)	(-0.67)	(-0.89)	(-1.19)	(-0.33)
Intercept	0.4000	0.0456	0.1145	-0.1445	0.7631
	(1.30)	(0.43)	(0.13)	(-0.56)	(0.62)
R-Squared	0.008	0.002	0.045	0.025	0.014
Minimum Point ($)	13,652	12,994	14,088	13,579	10,632
OECD - Value Added	14 countries				
Income	0.0413	0.0872	0.1677	0.1277	0.0243
	(0.38)	(1.38)	(0.72)	(2.05)	(2.03)
$Income^2$	-0.0014	-0.0037	-0.0060	-0.0054	-0.0010
	(-0.29)	(-1.32)	(-0.58)	(-1.98)	(-1.90)
Intercept	0.1999	-0.3600	-0.0910	-0.4966	-0.1130
	(0.34)	(-1.05)	(-0.07)	(-1.47)	(-1.74)
R-Squared	0.050	0.019	0.078	0.035	0.017
Minimum Point ($)	14,870	11,853	14,084	11,800	12,090

(t-statistics in parentheses) NOTE: The number of sectors varies across countries, but there are only 14 data points.

Table 6 - Characteristics of Observed Country-Specific Minimum Points

UNIDO – Employment Data				
Country	**Year**	**Income**	**Gini**	**Openness**
France	1968	8,228	0.440	34.47
Singapore	1968	2,430	0.431	311.45
Netherlands	1969	8,749	0.482	92.58
Puerto Rico	1971	6,006	0.526	118.91
Yugoslavia	1971	3,664	0.453	43.94
Spain	1972	6,490	0.422	26.53
Greece	1974	4,967	0.485	35.57
Czechoslovakia	1975	3,254	0.488	47.26
Portugal	1977	4,644	0.492	56.51
New Zealand	1978	10,036	0.534	54.21
Madagascar	1979	995	0.704	36.91
Cent. Afr. Rep.	1981	648	0.708	60.42
Congo	1981	2,044	0.599	97.64
Ireland	1981	6,985	0.486	90.74
Cyprus	1982	5,608	0.586	88.28
Canada	1983	14,105	0.451	44.88
USSR	1983	6,737	0.588	10.37
Austria	1985	11,131	0.465	58.95
Malaysia	1985	4,146	0.554	96.55
Mexico	1988	5,349	0.382	22.86
Bulgaria	1992	5,208	0.478	85.00
Standard Deviation	**6.81**	**3,348.3**	**0.086**	**62.15**
Average	**1978.2**	**5,782.1**	**0.512**	**72.10**
Average Before 1979		**5,405.7**	**0.496**	**78.03**
Average After 1979		**5,723.3**	**0.546**	**62.96**

Table 6 (continued) - Characteristics of Observed Country-Specific Minimum Points

ILO – Employment Data				
Country	**Year**	**Income**	**Gini**	**Openness**
Canada	1974	12,225	0.430	44.88
Finland	1975	9,609	0.283	48.58
Seychelles	1980	2,906	0.250	97.16
Luxembourg	1980	11,893	0.463	168.56
Paraguay	1983	2,075	0.511	37.07
Italy	1983	10,297	0.394	33.65
US	1984	16,255	0.487	13.91
Standard Deviation	**4.08**	**4,394.0**	**0.097**	**56.53**
Average	**1979.3**	**10,530.8**	**0.384**	**67.79**
Average Before 1980		**9,158.3**	**0.356**	**89.80**
Average After 1980		**10,337.8**	**0.398**	**78.32**

OECD – Employment Data				
Country	**Year**	**Income**	**Gini**	**Openness**
UK	1971	8,655	0.459	48.09
Norway	1972	8,746	0.485	84.88
Japan	1973	8,539	0.565	21.84
Italy	1977	9,016	0.515	33.65
Finland	1980	10,851	0.500	48.58
Standard Deviation	**3.78**	**960.8**	**0.039**	**23.71**
Average	**1974.6**	**9,161.4**	**0.504**	**47.41**
Average Before 1973		**8,646.7**	**0.503**	**51.60**
Average After 1973		**9,468.7**	**0.526**	**34.69**

Figure Titles

Figure 1

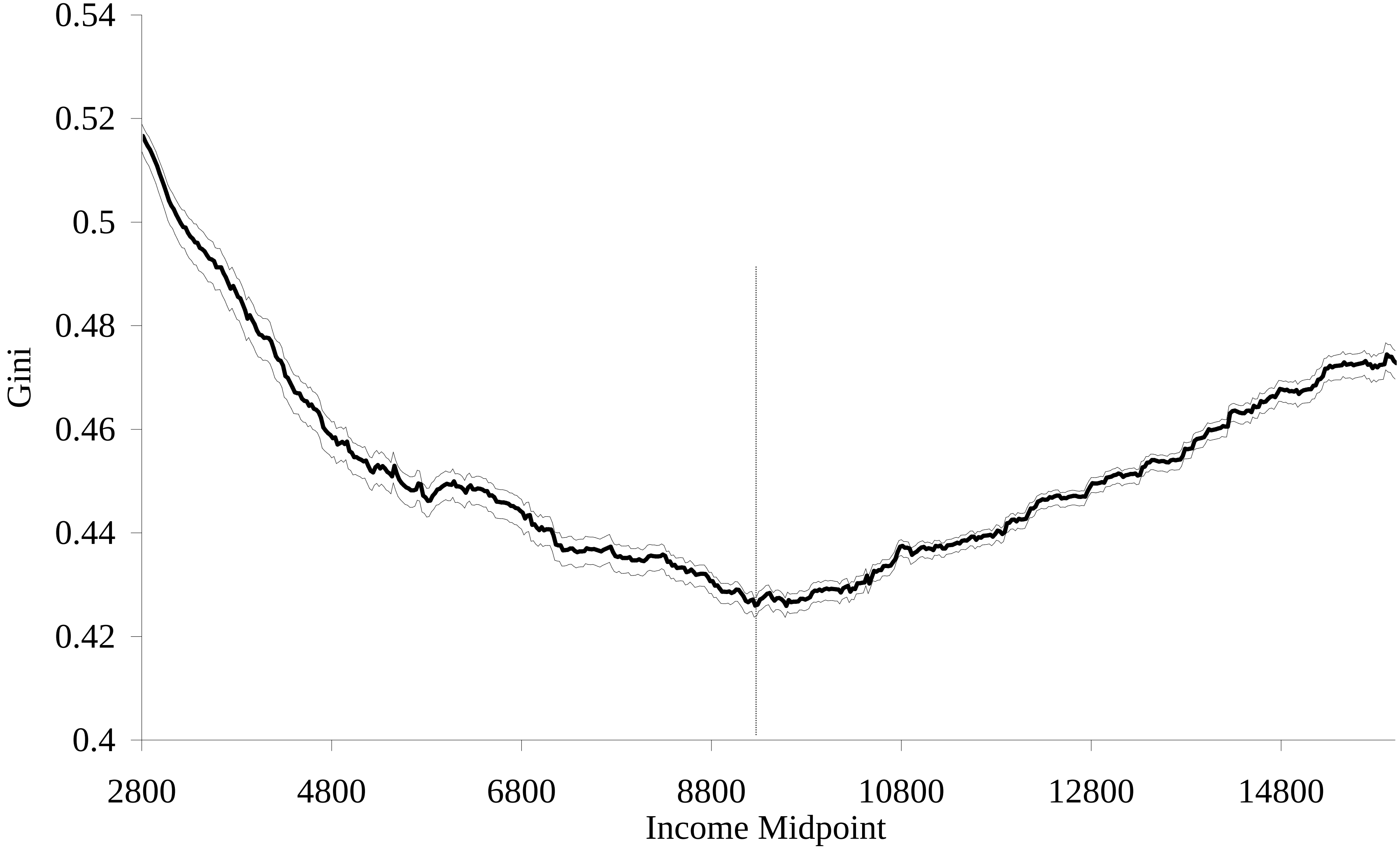

0.54
0.52
0.5
0.48
0.46
0.44
0.42
0.4
Gini
2800
4800
6800
8800
10800
12800
14800
Income Midpoint

Figure 2

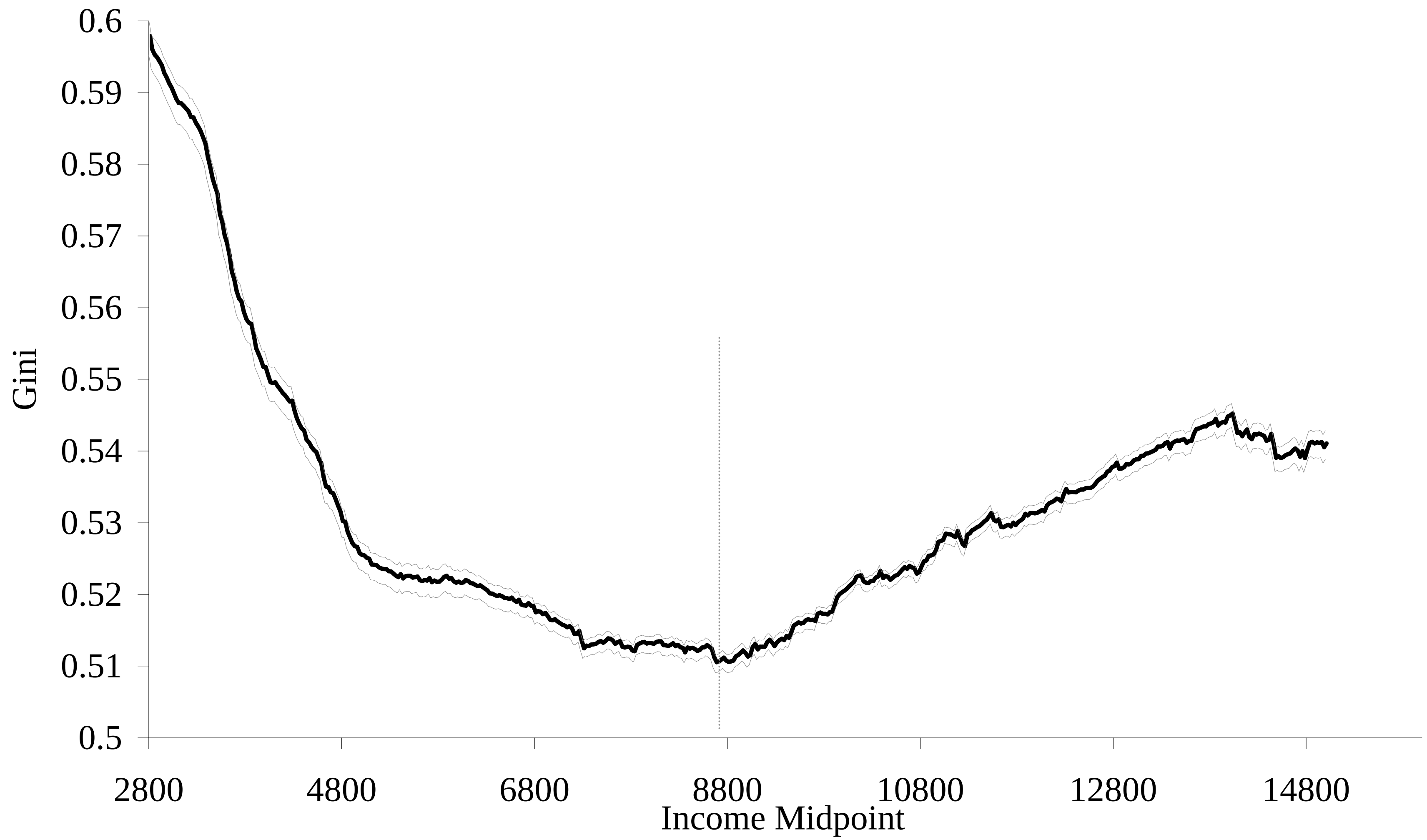

0.6
0.59
0.58
0.57
0.56
0.55
0.54
0.53
0.52
0.51
0.5
Gini
2800
4800
6800
8800
10800
12800
14800
Income Midpoint

Figure 3

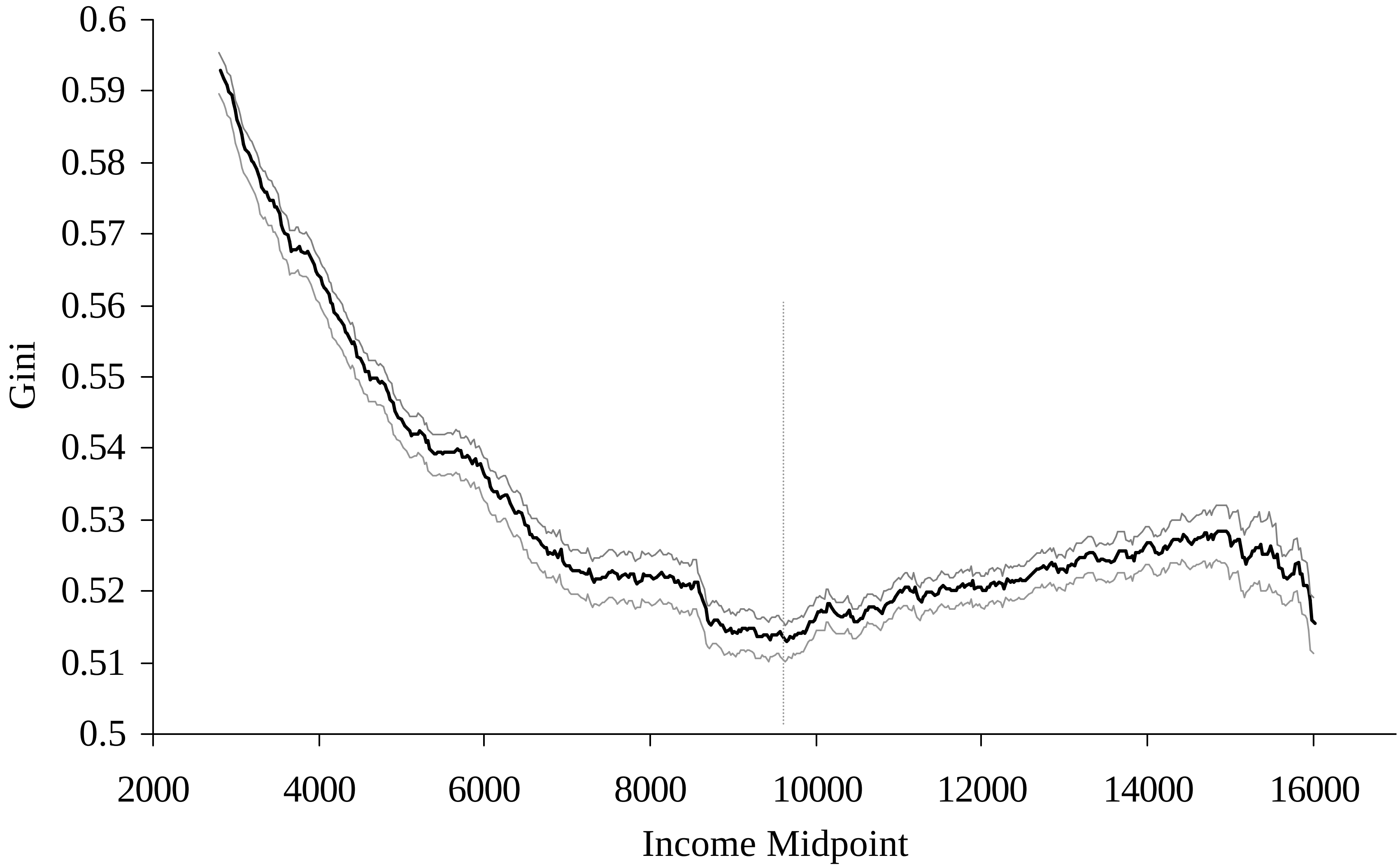

0.6
0.59
0.58
0.57
0.56
0.55
0.54
0.53
0.52
0.51
0.5
Gini
2000
4000
6000
8000
10000
12000
14000
16000
Income Midpoint

Figure 4

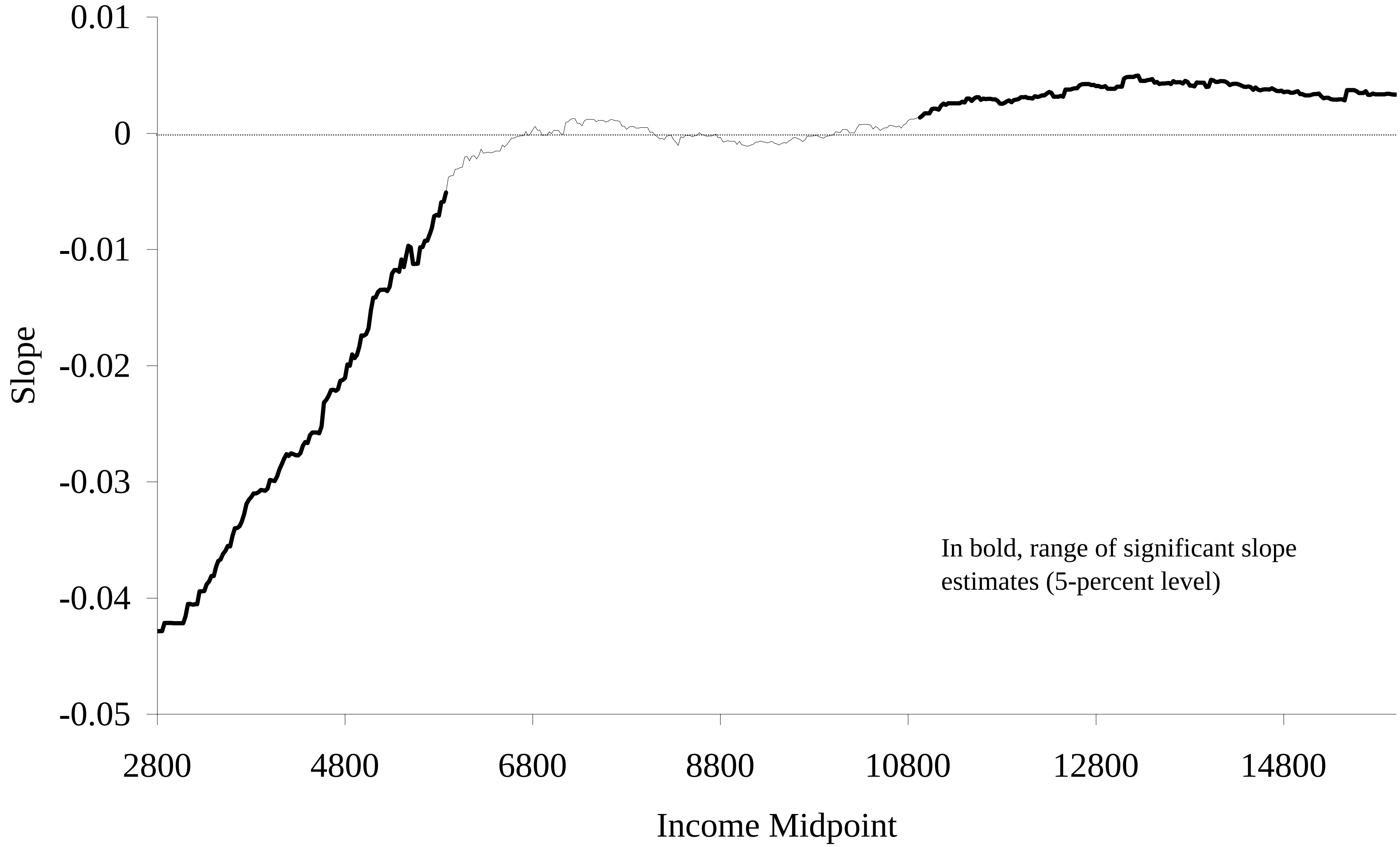

0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
Slope
2800
4800
6800
8800
10800
12800
14800
Income Midpoint
In bold, range of significant slope estimates (5-percent level)

Figure 5


0.01
0.005
0
-0.005
-0.01
-0.015
Slope
2800
4800
6800
8800
10800
12800
14800
Income Midpoint
In bold, range of significant slope estimates (5-percent level)

Figure 6

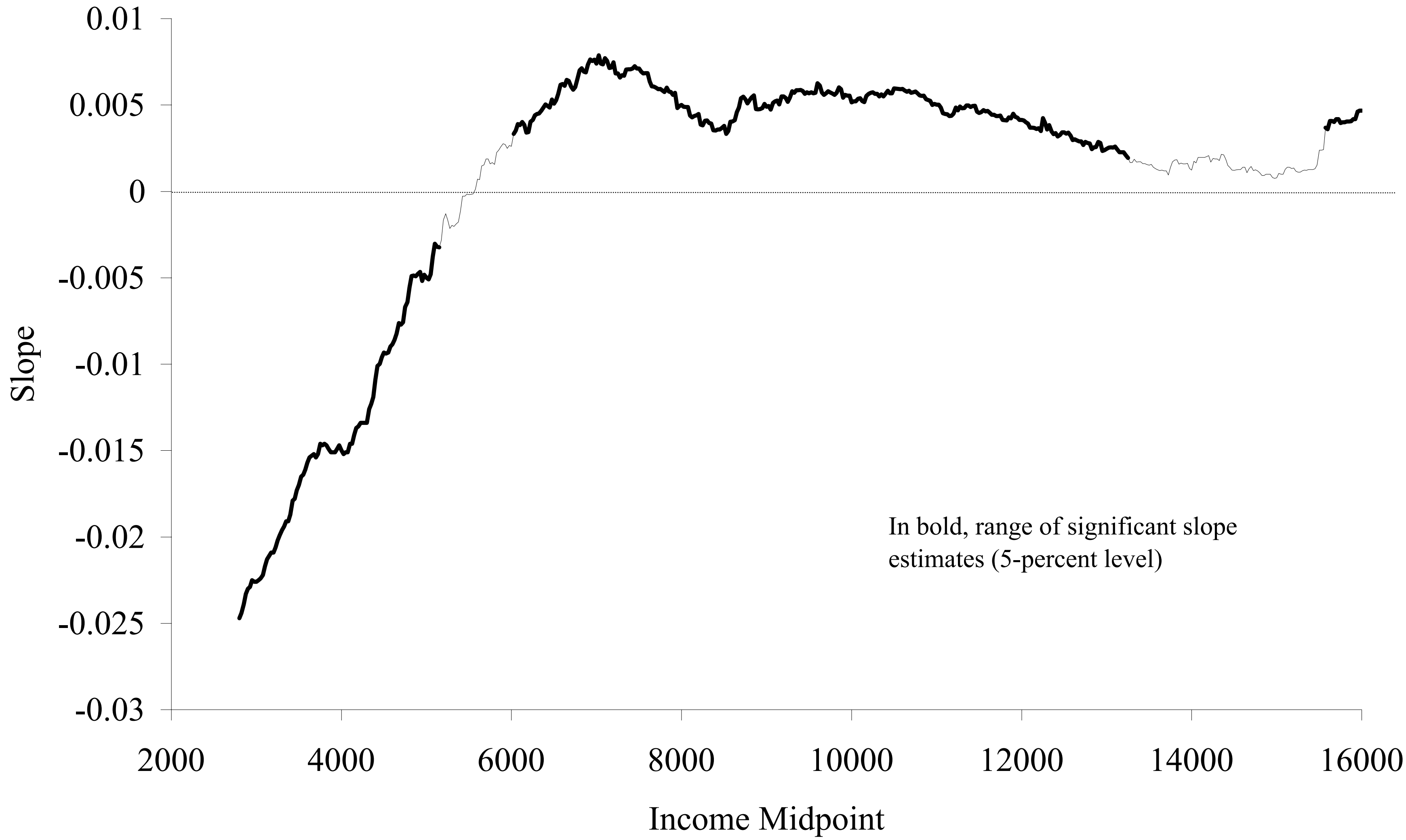

0.01
0.005
0
-0.005
-0.01
-0.015
-0.02
-0.025
-0.03
Slope
2000
4000
6000
8000
10000
12000
14000
16000
Income Midpoint
In bold, range of significant slope estimates (5-percent level)

Figure 7

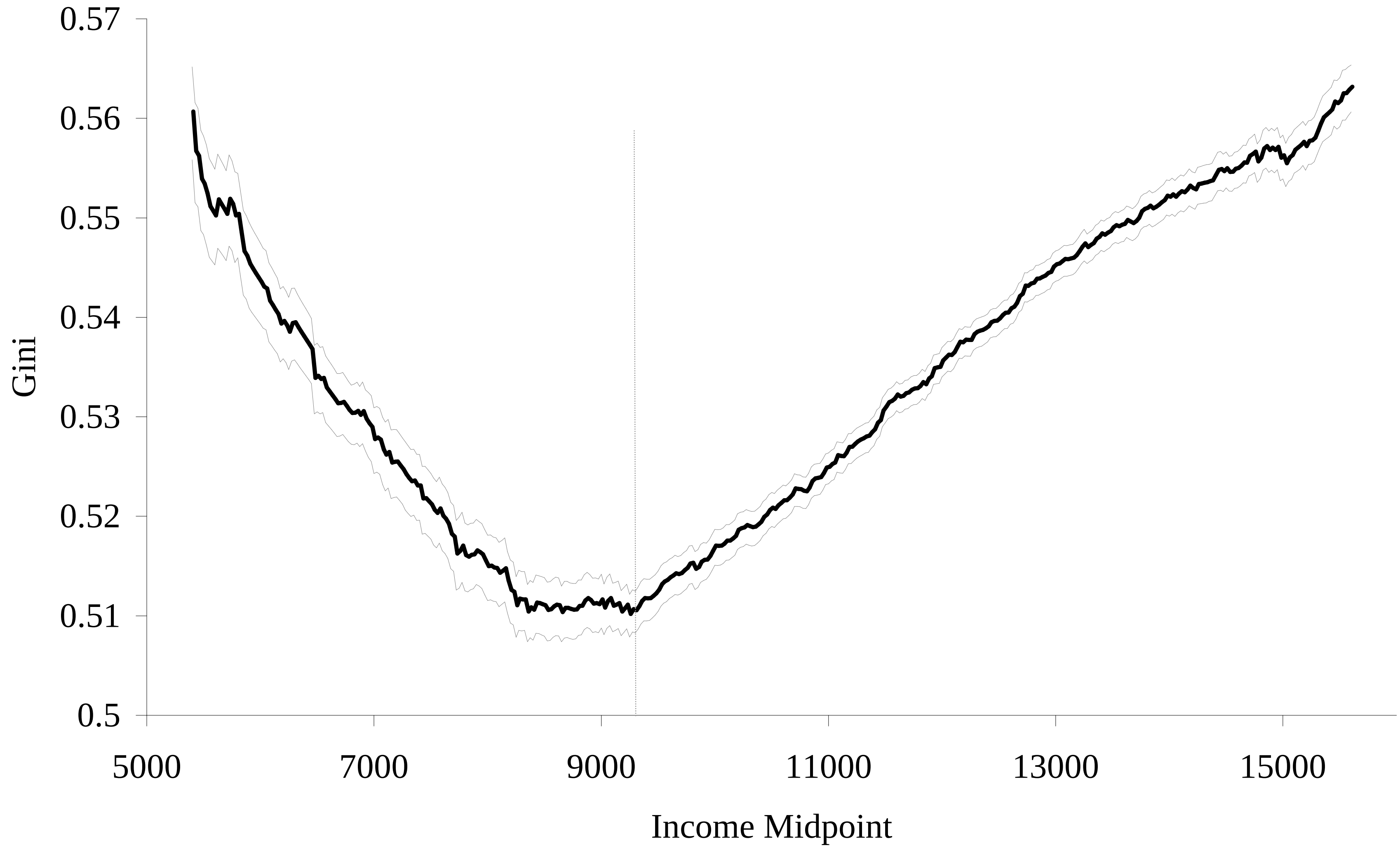

0.57
0.56
0.55
0.54
0.53
0.52
0.51
0.5
Gini
5000
7000
9000
11000
13000
15000
Income Midpoint

Figure 8

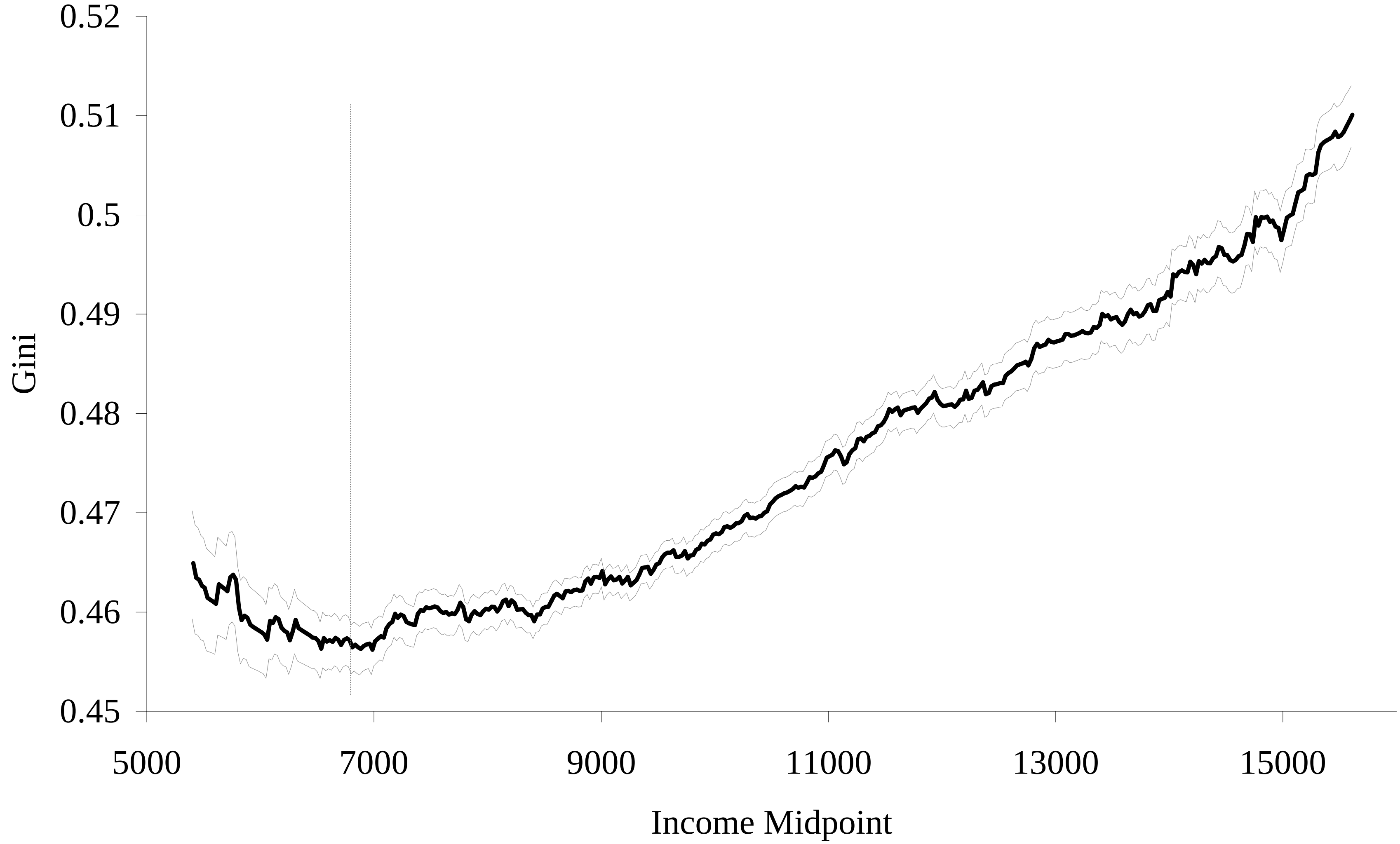

0.52
0.51
0.5
0.49
0.48
0.47
0.46
0.45
Gini
5000
7000
9000
11000
13000
15000
Income Midpoint

Figure 9

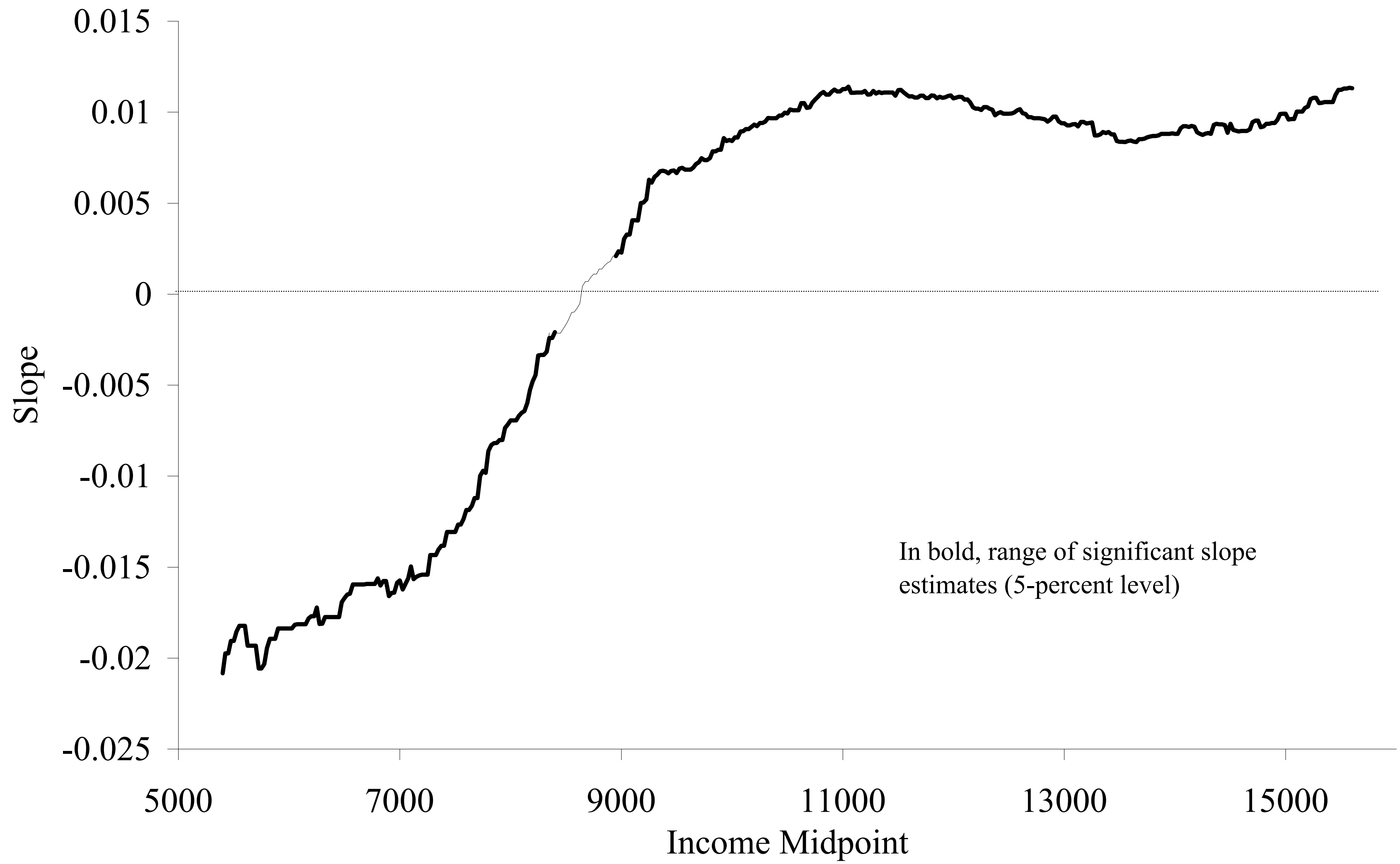

0.015
0.01
0.005
0
-0.005
-0.01
-0.015
-0.02
-0.025
Slope
5000
7000
9000
11000
13000
15000
Income Midpoint
In bold, range of significant slope estimates (5-percent level)

Figure 10

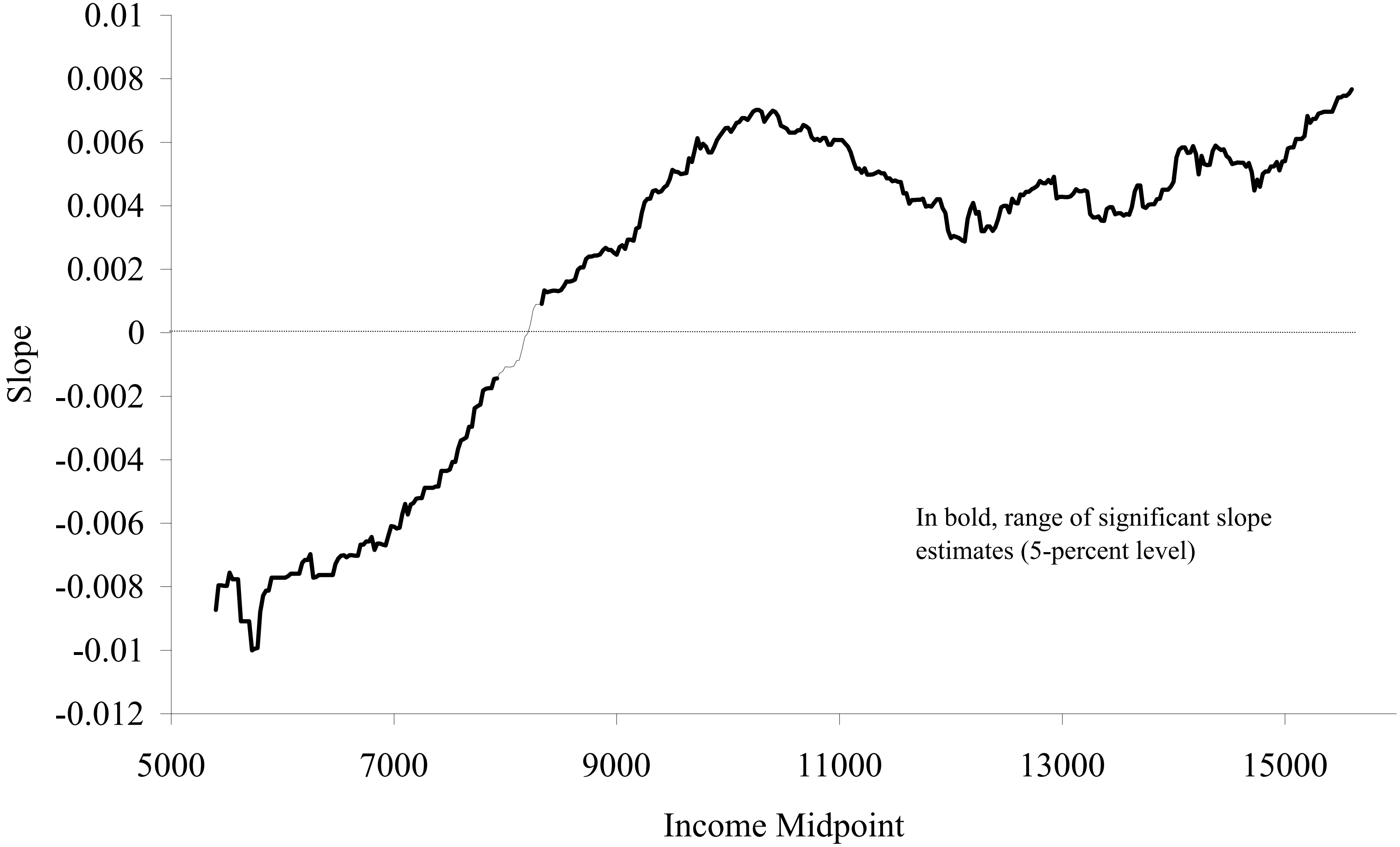
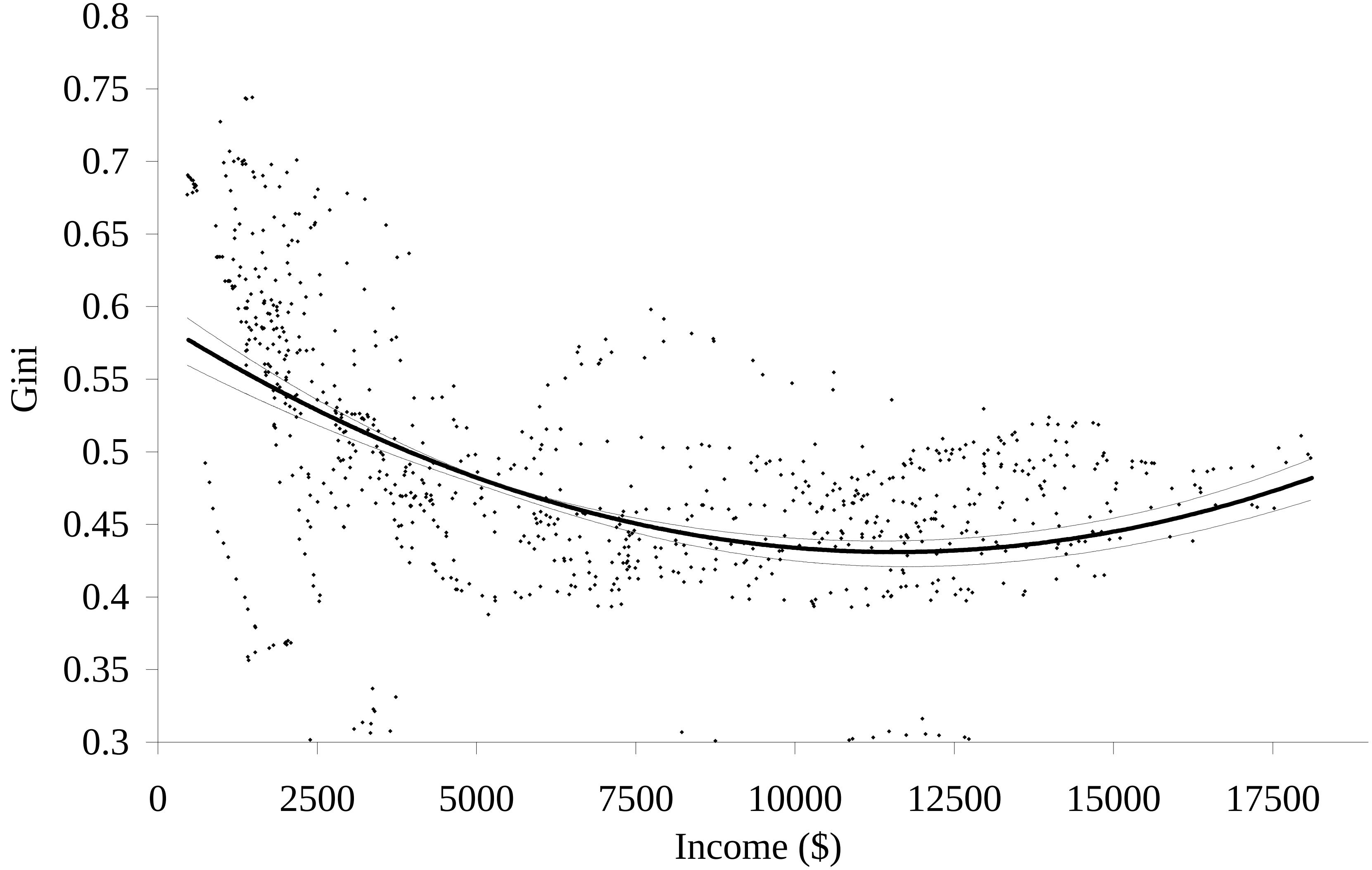

0.01
0.008
0.006
0.004
0.002
0
-0.002
-0.004
-0.006
-0.008
-0.01
-0.012
Slope
5000
7000
9000
11000
13000
15000
Income Midpoint
In bold, range of significant slope estimates (5-percent level)

Figure 11



0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
Gini
0
2500
5000
7500
10000
12500
15000
17500
Income ($)

Figure 12

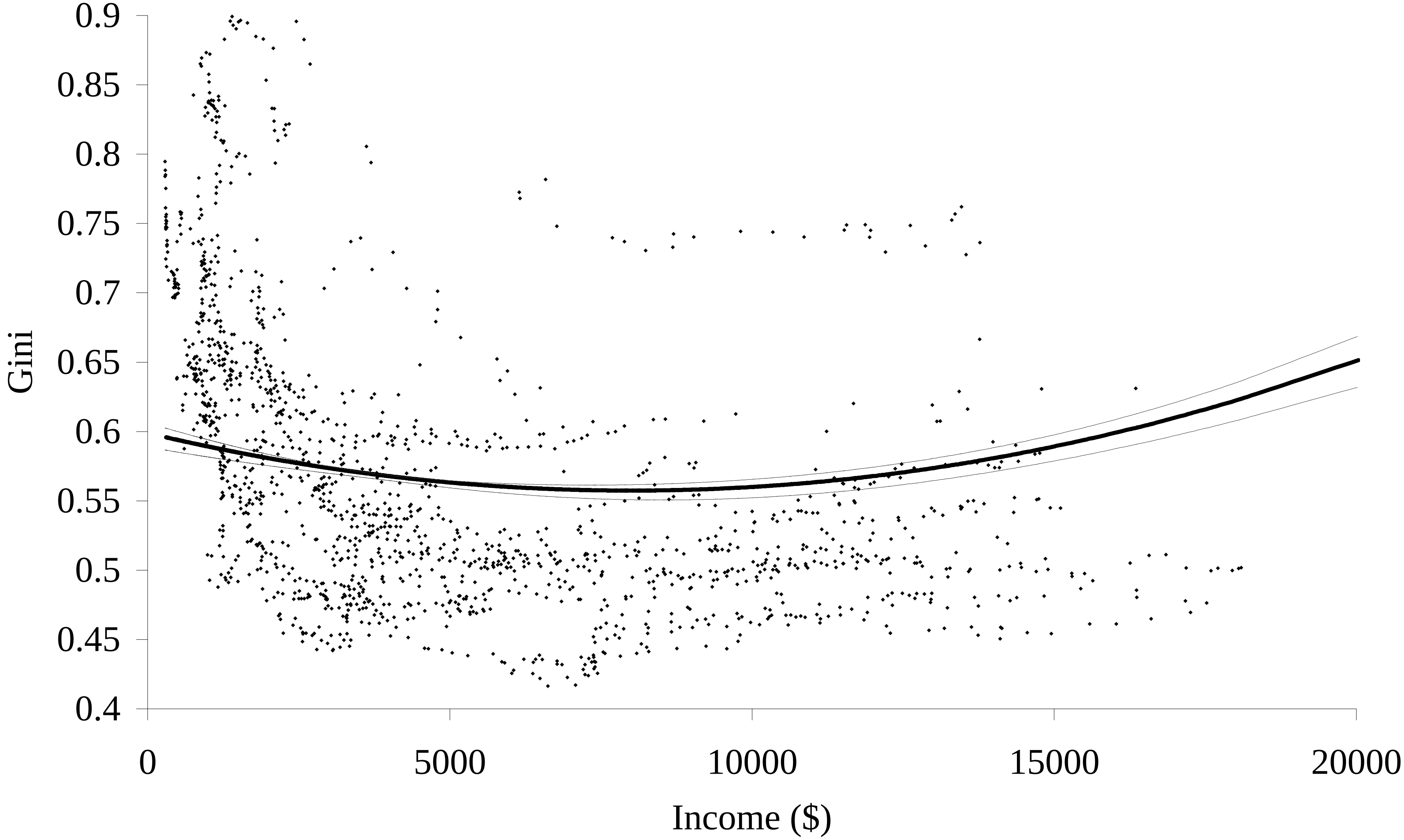

0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
Gini
0
5000
10000
15000
20000
Income ($)

Figure 13

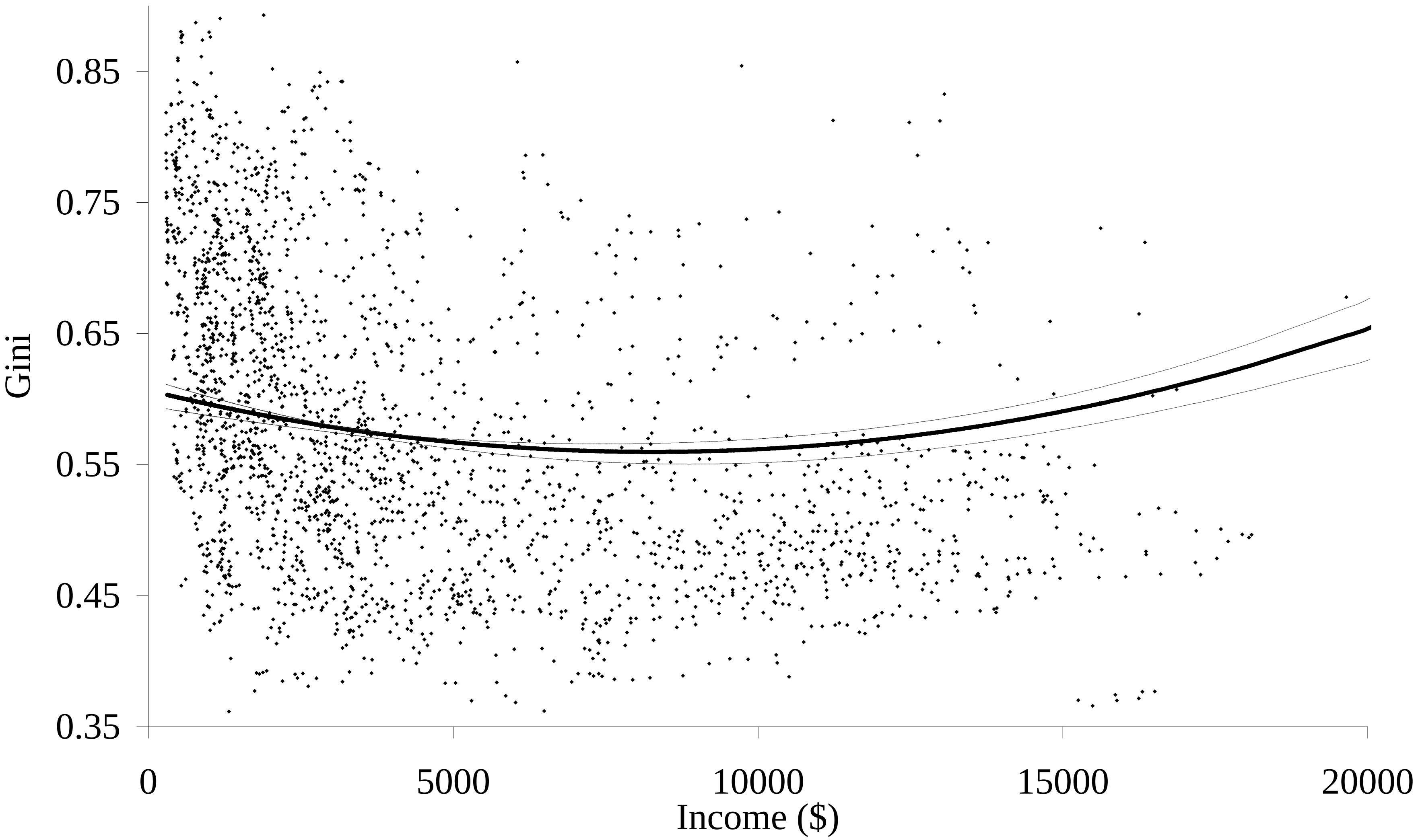
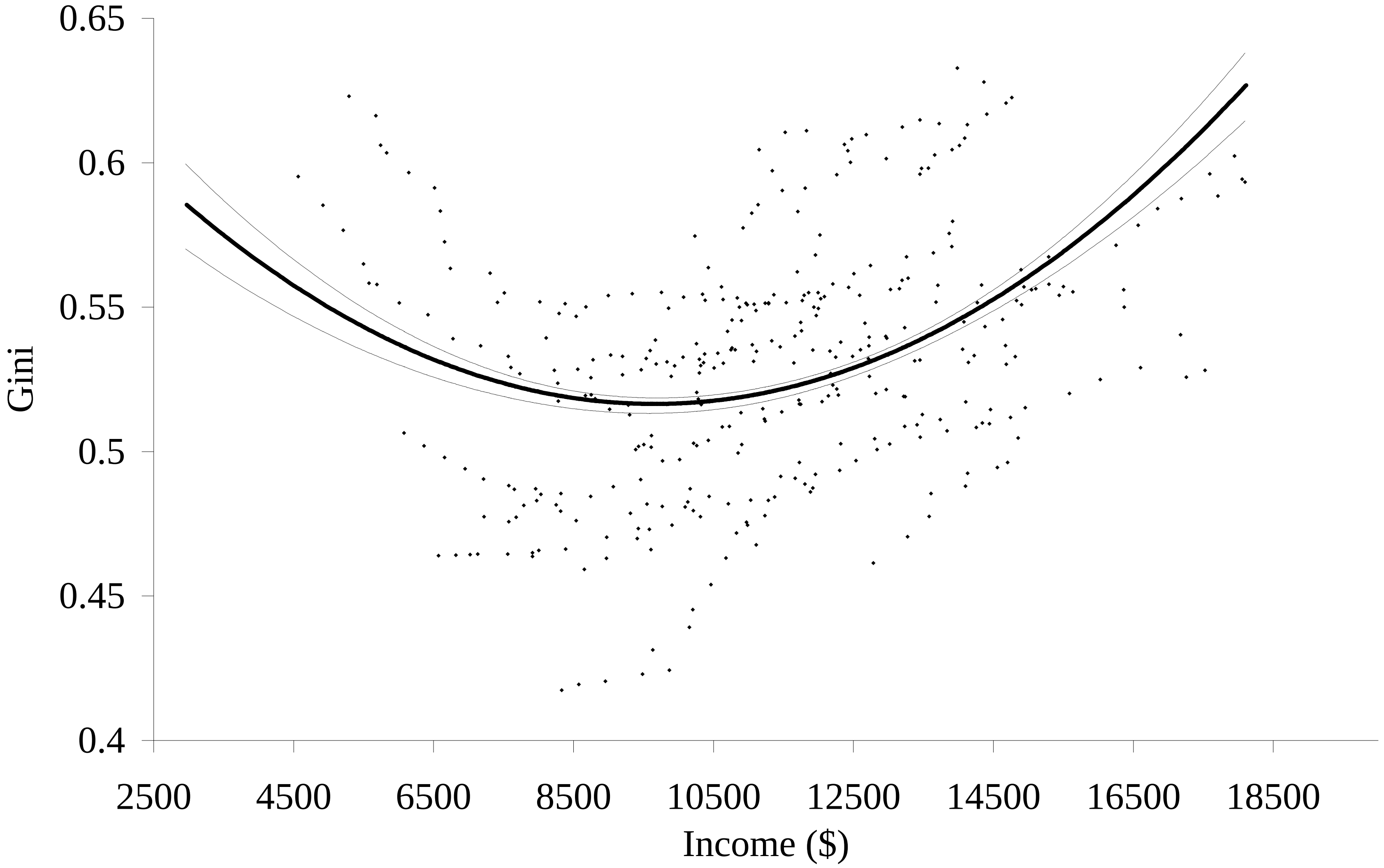
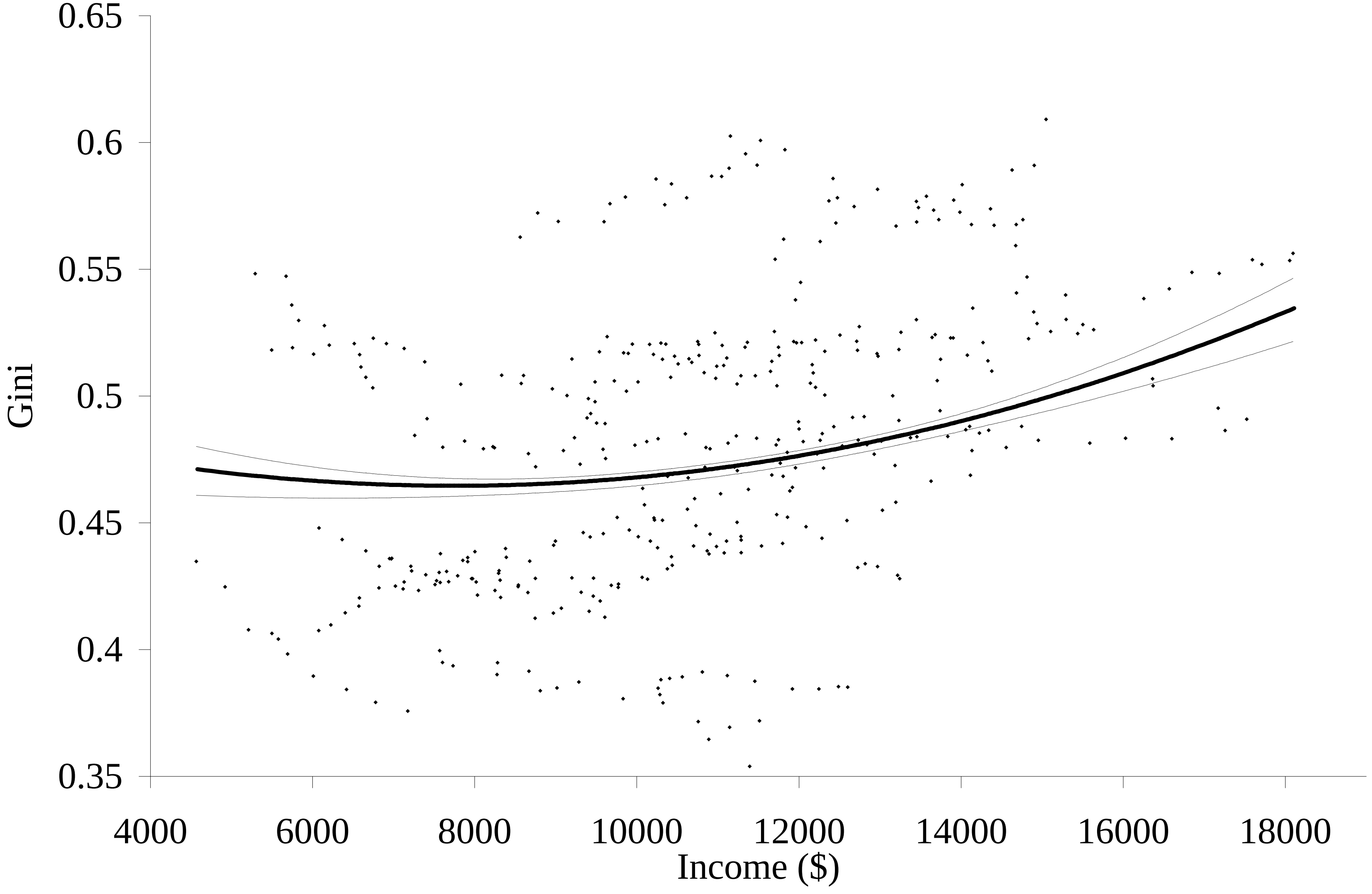


Figure 14



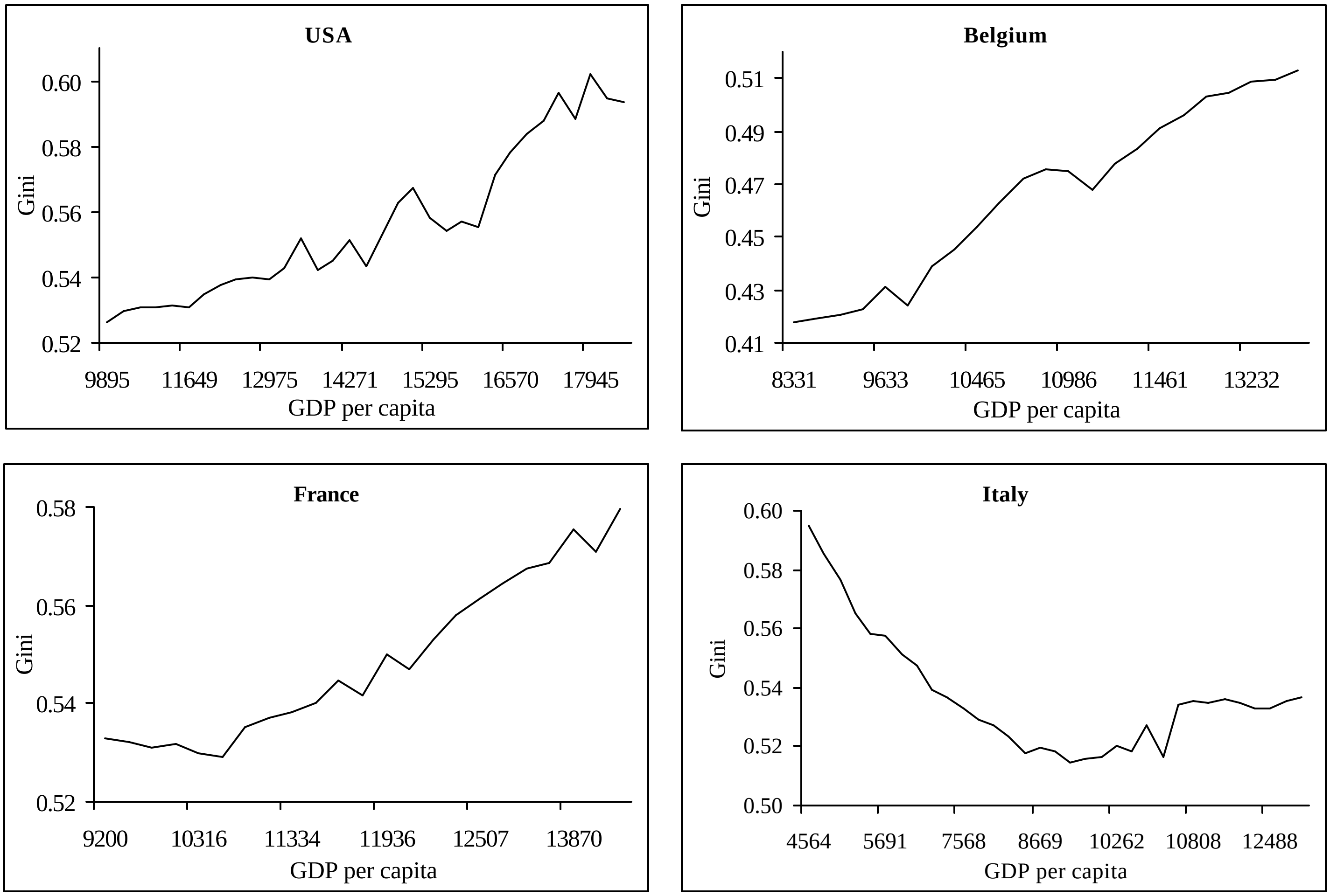
0.65
0.6
0.55
0.5
0.45
0.4
Gini
2500
4500
6500
8500
10500
12500
14500
16500
18500
Income ($)

Figure 15



0.65
0.6
0.55
0.5
0.45
0.4
0.35
Gini
4000
6000
8000
10000
12000
14000
16000
18000
Income ($)

Figure 16

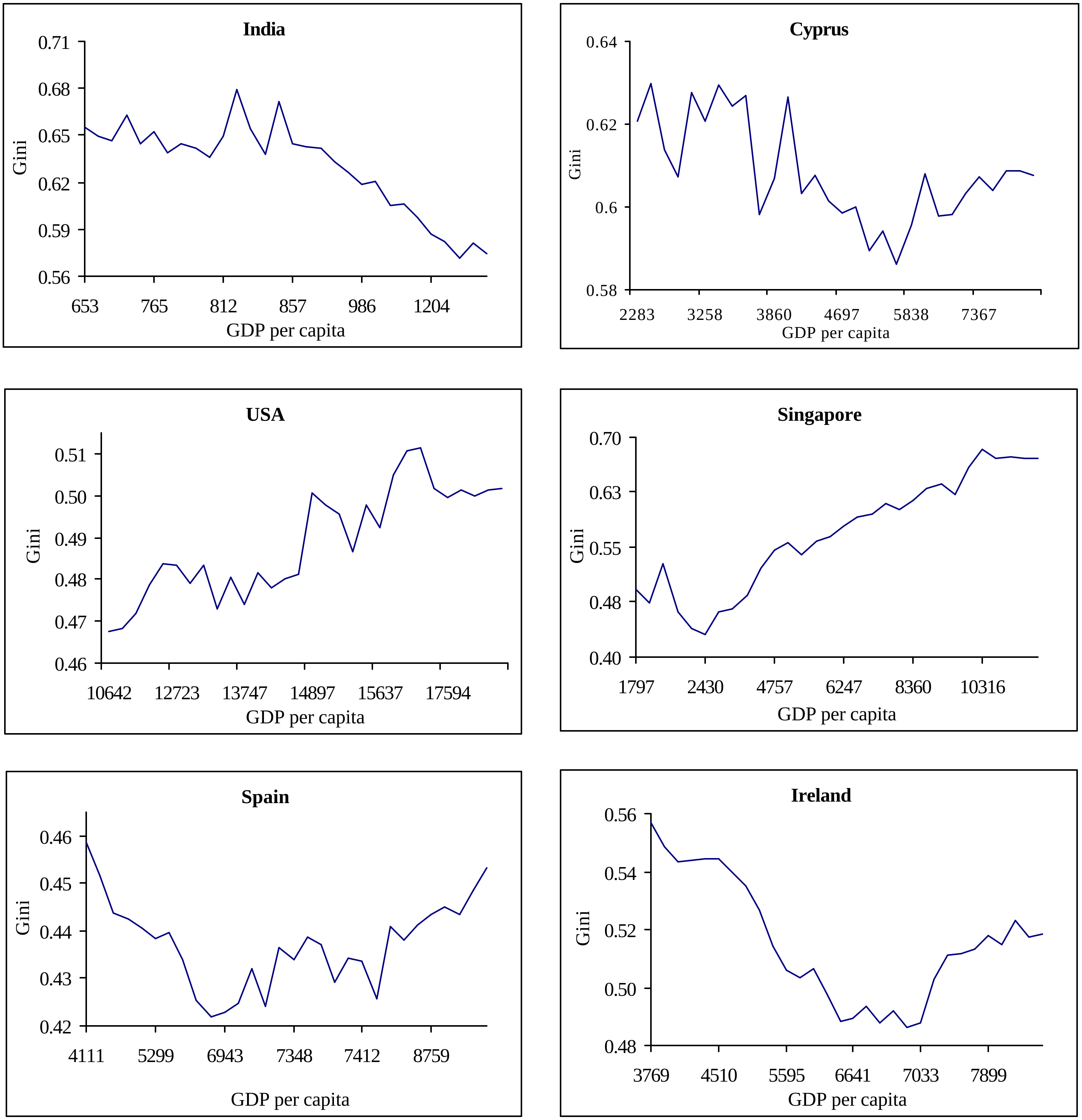


USA
Gini
0.60
0.58
0.56
0.54
0.52
9895
11649
12975
14271
15295
16570
17945
GDP per capita
Belgium
Gini
0.51
0.49
0.47
0.45
0.43
0.41
8331
9633
10465
10986
11461
13232
GDP per capita
France
Gini
0.58
0.56
0.54
0.52
9200
10316
11334
11936
12507
13870
GDP per capita
Italy
Gini
0.60
0.58
0.56
0.54
0.52
0.50
4564
5691
7568
8669
10262
10808
12488
GDP per capita

Figure 17


China
0.71
0.70
0.70
0.69
Gini
1261
1324
1326
1352
1378
1493
GDP per capita
Malaysia
0.49
0.48
0.47
0.46
0.45
0.44
Gini
3869
4146
4265
4420
5124
GDP per capita
Australia
0.44
0.43
0.42
0.41
0.4
0.39
Gini
10886
11514
12161
12784
14102
14704
GDP per capita
Denmark
0.51
0.50
0.49
0.48
0.47
0.46
0.45
Gini
10236
10921
11484
13449
13909
GDP per capita
USA
0.51
0.50
0.49
0.48
Gini
12963
13747
14832
15440
16570
17710
GDP per capita
Luxembourg
0.48
0.47
0.46
Gini
9782
10705
11546
11893
12900
14234
16280
GDP per capita

Figure 18


India
0.71
0.68
0.65
0.62
0.59
0.56
Gini
653 765 812 857 986 1204
GDP per capita
Cyprus
0.64
0.62
0.6
0.58
Gini
2283 3258 3860 4697 5838 7367
GDP per capita
USA
0.51
0.50
0.49
0.48
0.47
0.46
Gini
10642 12723 13747 14897 15637 17594
GDP per capita
Singapore
0.70
0.63
0.55
0.48
0.40
Gini
1797 2430 4757 6247 8360 10316
GDP per capita
Spain
0.46
0.45
0.44
0.43
0.42
Gini
4111 5299 6943 7348 7412 8759
GDP per capita
Ireland
0.56
0.54
0.52
0.50
0.48
Gini
3769 4510 5595 6641 7033 7899
GDP per capita