Econometrics 1 *Applied Econometrics with R*

Special Topic: U Shape Test

黄嘉平

中国经济特区研究中心 讲师

办公室：文科楼1726

E-mail: huangjp@szu.edu.cn

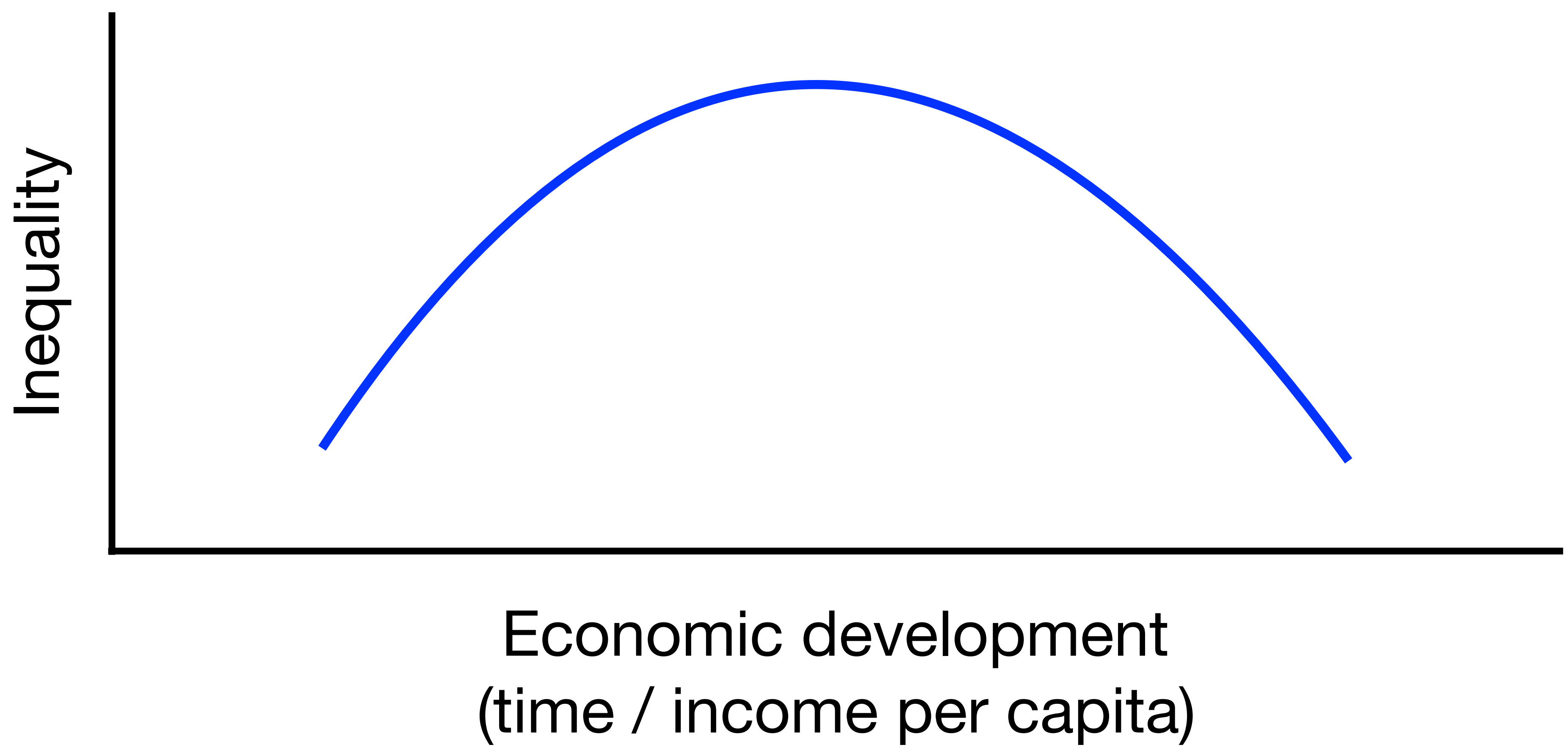
Tel: (0755) 2695 0548

Office hour: Mon./Tue. 13:00-14:00

The Kuznets Curve

- Simon Kuznets (1901–1985) claimed that

  Economic growth first increases and then reduces income inequality.



Robinson’s formulation

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

Assume the economy is divided into two sectors with different income distributions, and there is a monotonic increase in the relative population of one of the sectors over time.

Population shares: $W_1 + W_2 = 1$

Log mean of income: $Y = W_1 Y_1 + W_2 Y_2$

Log variance of income:

$$\begin{aligned}\sigma^2 =& W_1 \sigma_1^2 + W_2 \sigma_2^2 \\ &+ W_1 (Y_1 - Y)^2 + W_2 (Y_2 - Y)^2\end{aligned}$$

Robinson's formulation

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

Let W_1 be increasing over time. If $Y_1 > Y_2$, Y increases over time. One then has

$$\sigma^2 = AW_1^2 + BW_1 + C$$

where
$$A = -(Y_1 - Y_2)^2$$
$$B = (\sigma_1^2 - \sigma_2^2) + (Y_1 - Y_2)^2$$
$$C = \sigma_2^2$$

Claims of U and inverted U shaped relations in economics

- Laffer curve
  — tax rate vs. government revenue

- Environmental Kuznets curve
  — economic development vs. environmental quality

- In industrial organization
  — competition vs. innovation

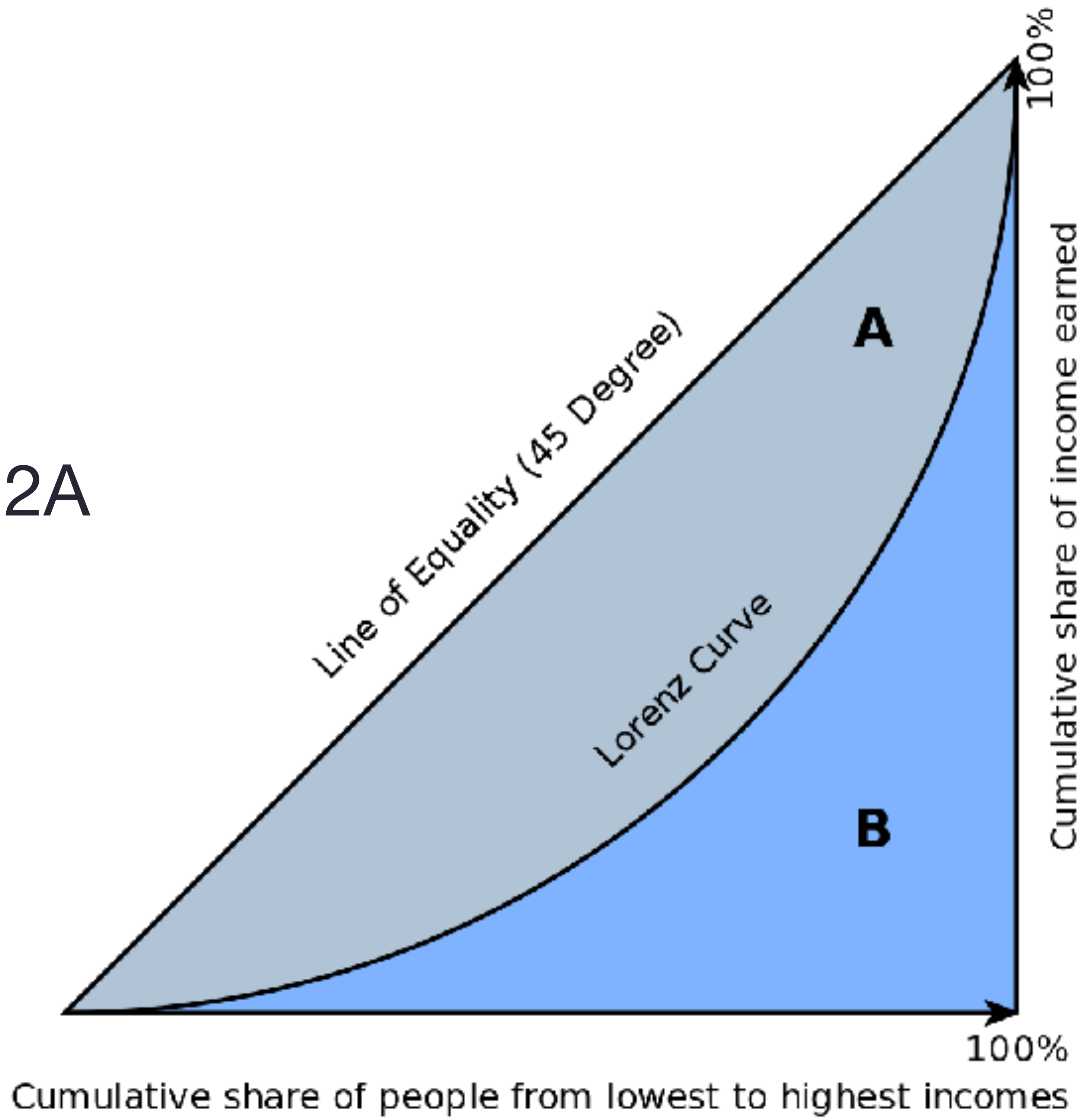
- In political science
  — level of democracy vs. tendency to war

The Gini coefficient: a measure of inequality

- The Lorenz curve

- The Gini coefficient

$$G = A / (A + B) = 2A$$



Source of the figure: wikipediaia/Gini_coefficient

Compare your country

Home English

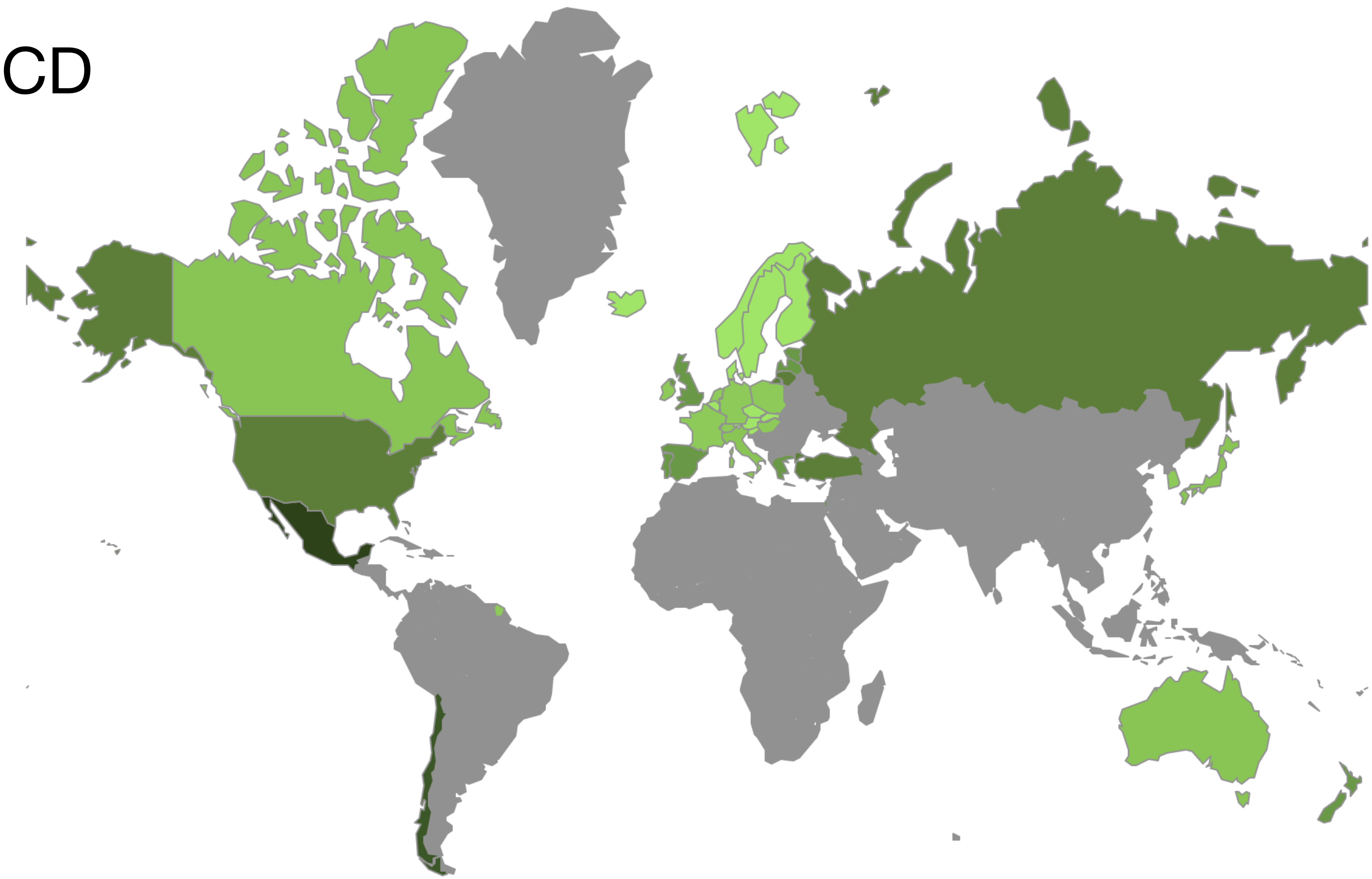
Income distribution and poverty

Change view

Overview Poverty by age

Gini Coefficient
Relative Income Poverty
Top 20% vs. bottom 20%

OECD



http://www.oecd.org/social/income-distribution-database.htm

Gini Coefficient, 2015

0.246 0.459

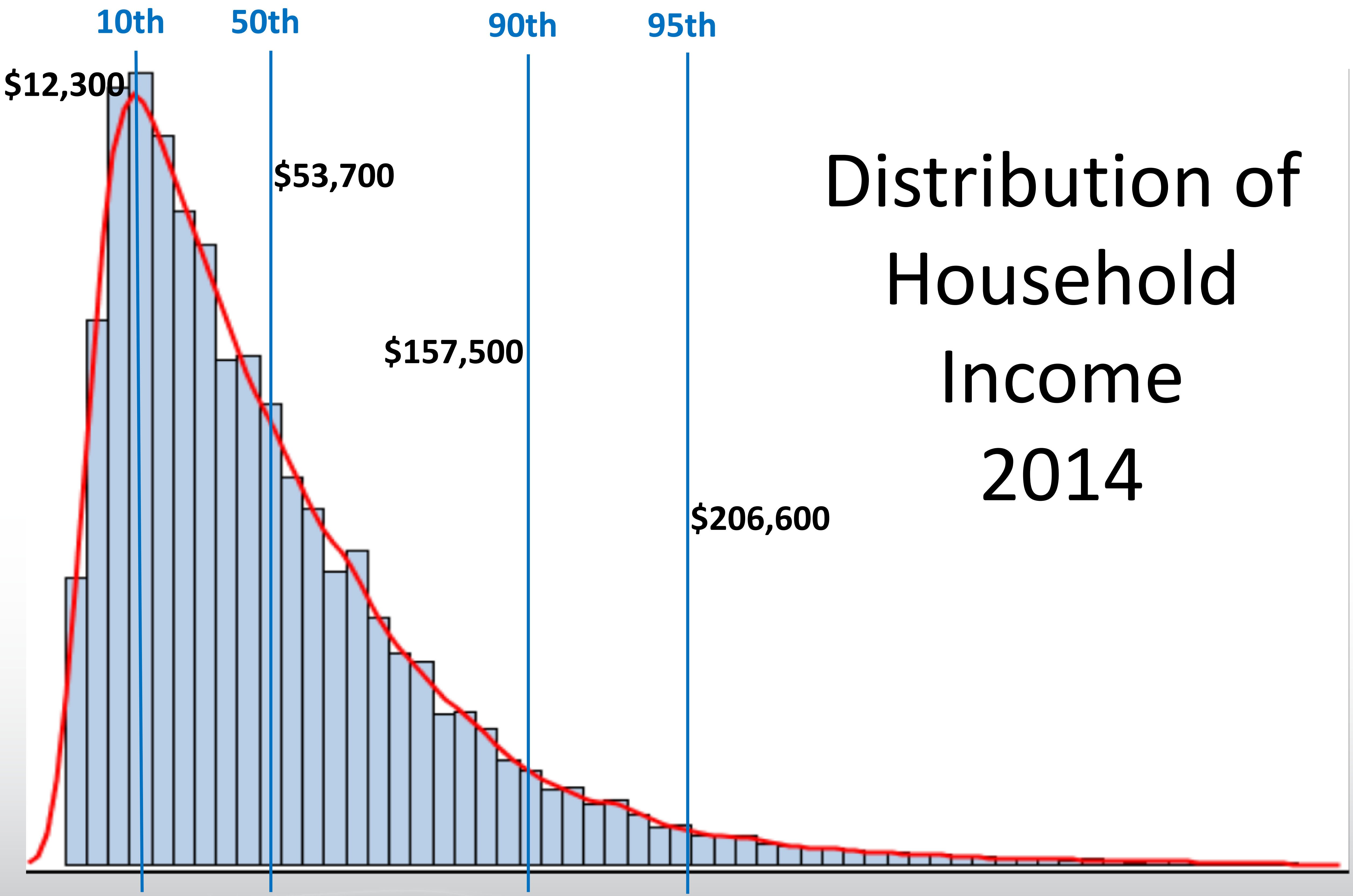
Share/embed Data source/about

OECD

The US

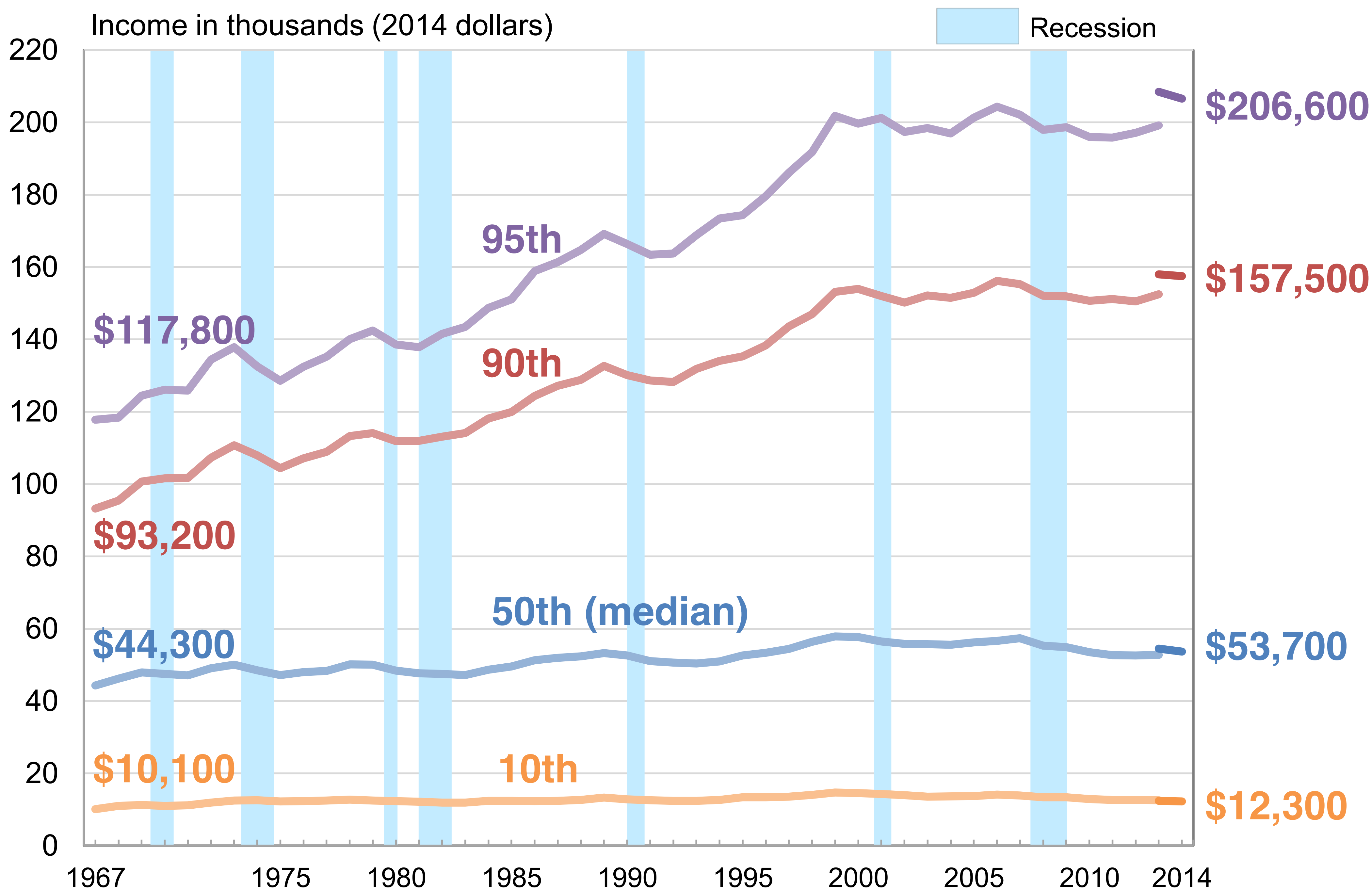
- The Current Population Survey
- Income and Poverty in the United States (yearly reports)
- The US Census Bureau https://www.census.gov/

Distribution of Household Income 2014



Source: U.S. Census Bureau, Current Population Survey, 2015 Annual Social and Economic Supplement.

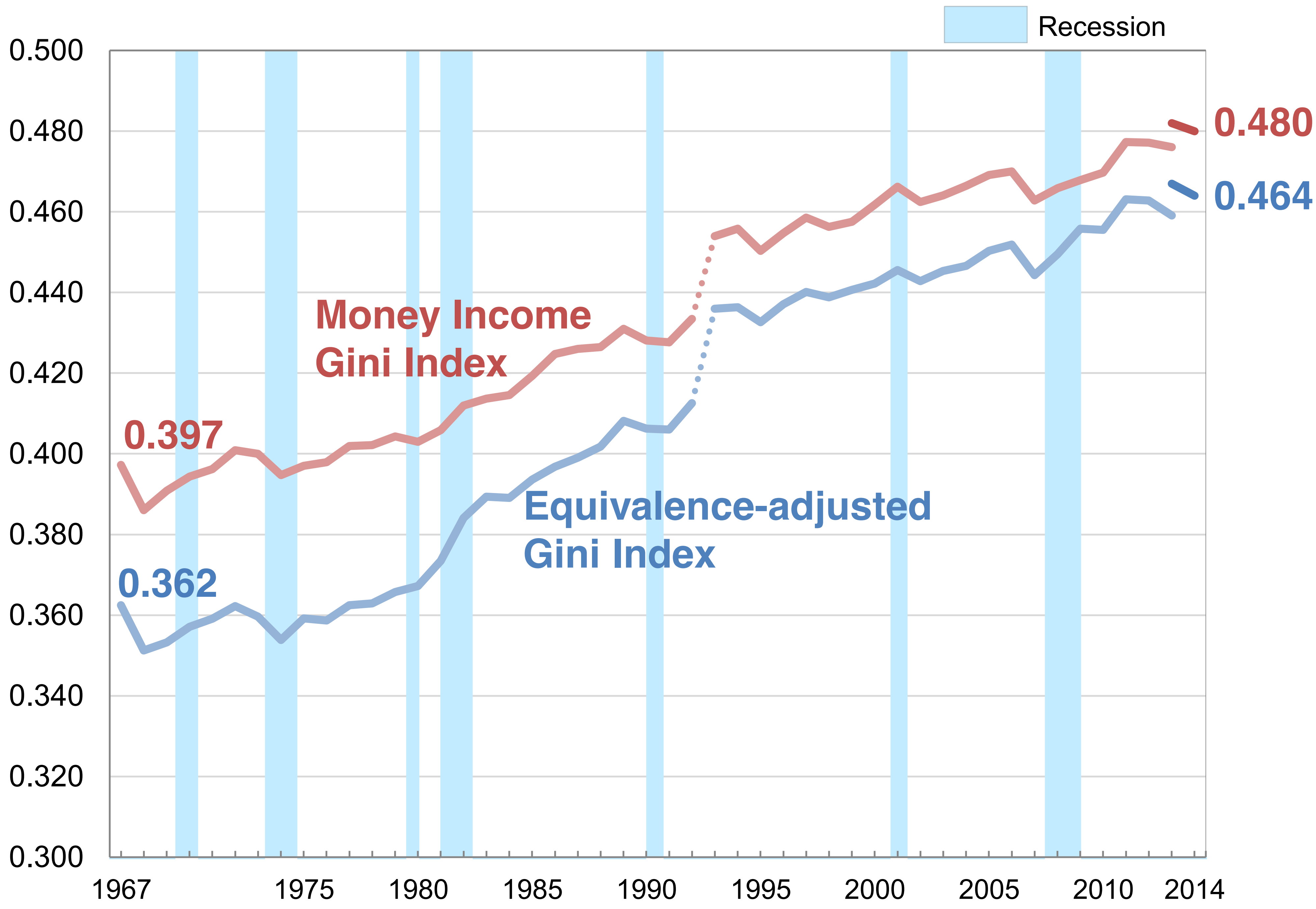
Real Household Income at Selected Percentiles: 1967 to 2014



Note: The 2013 data reflect the implementation of the redesigned income questions. See Appendix D of the P60 report, "Income and Poverty in the United States: 2014," for more information. Income rounded to nearest $100.

Source: U.S. Census Bureau, Current Population Survey, 1968 to 2015 Annual Social and Economic Supplements.

Gini Index of Money Income and Equivalence-Adjusted Income: 1967 to 2014



Note: The 2013 data reflect the implementation of the redesigned income questions. See Appendix D of the P60 report, "Income and Poverty in the United States: 2014," for more information. Change in data collection methodology in 1993.

Source: U.S. Census Bureau, Current Population Survey, 1968 to 2015 Annual Social and Economic Supplements.

A (questionable) case study

第 8 卷第 4 期 经 济 学（季 刊） Vol. 8，No. 4
2009 年 7 月 China Economic Quarterly July，2009

城市化、城乡差距以及全国居民总体收入差距的变动
——收入差距倒 U 形假说的实证检验

周云波*

摘 要 本文利用两部门模型，从理论和实证上分析了城市化、城乡差距与全国居民总体收入差距的关系。研究结果表明，改革开放以来，城市化是导致倒 U 现象出现的主要原因。首先，城市化的加快使得由城乡人口流动导致的收入差距于 2001 年开始缩小；其次，城市化意味着低收入农村人口比例的缩小和低收入城市人口比例的增加；最后，城市化促使我国农村内部收入差距拉大的速度放缓。由此，全国总体收入差距将在 2006—2009 年迎来倒 U 形曲线的拐点。

关键词 城市化，城乡差距，基尼系数

A (questionable) case study

- 假设农村和城市内部的收入分配是完全均等的。
- 假设城市人均收入和农村人均收入不同。在此基础上，探讨城镇化水平（城镇人口比重）与收入分配之间的关系。

假说 **1** 改革开放以后，虽然我国城乡内部居民收入差距在不断扩大，还没有出现转折的迹象，但是将城乡综合在一起的全国居民总体收入差距的变动很可能会出现倒 U 现象，并将很快迎来倒 U 曲线的拐点。

假说 **2** 城市化是导致全国居民总体收入差距变动出现倒 U 现象最主要的原因。

A (questionable) case study

- 文中使用的计量模型

$$G = \beta_0 + \beta_1 \mathrm{RJGDP} + \beta_2 \mathrm{RJGDP}^2 + \beta_3 \mathrm{UPSHARE} + \beta_4 \mathrm{UPSHARE}^2 + \beta_5 \mathrm{URGAP} + \beta_6 X + \mu,$$

- G: Gini coefficient
  RJGDP: 人均GDP
  UPSHARE: 城镇人口占总人口的比重
  UGGAP: 城乡收入比
  X: 其他控制变量

A (questionable) case study

表 3　模型(11)OLS 估计结果及检验

自变量	模型Ⅰ	模型Ⅱ	模型Ⅲ	模型Ⅳ
常数项	0.07439*** (11.51532)	0.07911*** (9.49418)	−0.11069*** (−8.09807)	−0.10297*** (−8.29165)
人均 GDP(元)	0.05609*** (9.22746)	0.05865*** (8.67874)	—	—
人均 GDP 的平方	−0.00953*** (−5.95325)	−0.008064*** (−6.29448)	—	—
城镇人口比重(%)	—	—	0.01250*** (12.04357)	0.01205*** (13.17788)
城镇人口比重的平方	—	—	−0.00015*** (−9.19740)	−0.00013*** (−9.75818)
城乡收入比率	0.08253*** (17.86316)	0.08218*** (16.84419)	0.09283*** (37.93077)	0.09117*** (44.12895)
外贸依存度(%)	0.00050** (2.43893)	—	0.000246** (2.20812)	—
通货膨胀率(%)	—	0.00037** (2.54978)	—	0.00036*** (4.70301)
时间截距变量	0.01407*** (3.16700)	0.01644*** (4.34234)	0.00412*** (3.23108)	0.00302*** (4.05027)

注：表3的完整信息请参照原论文。

A (questionable) case study

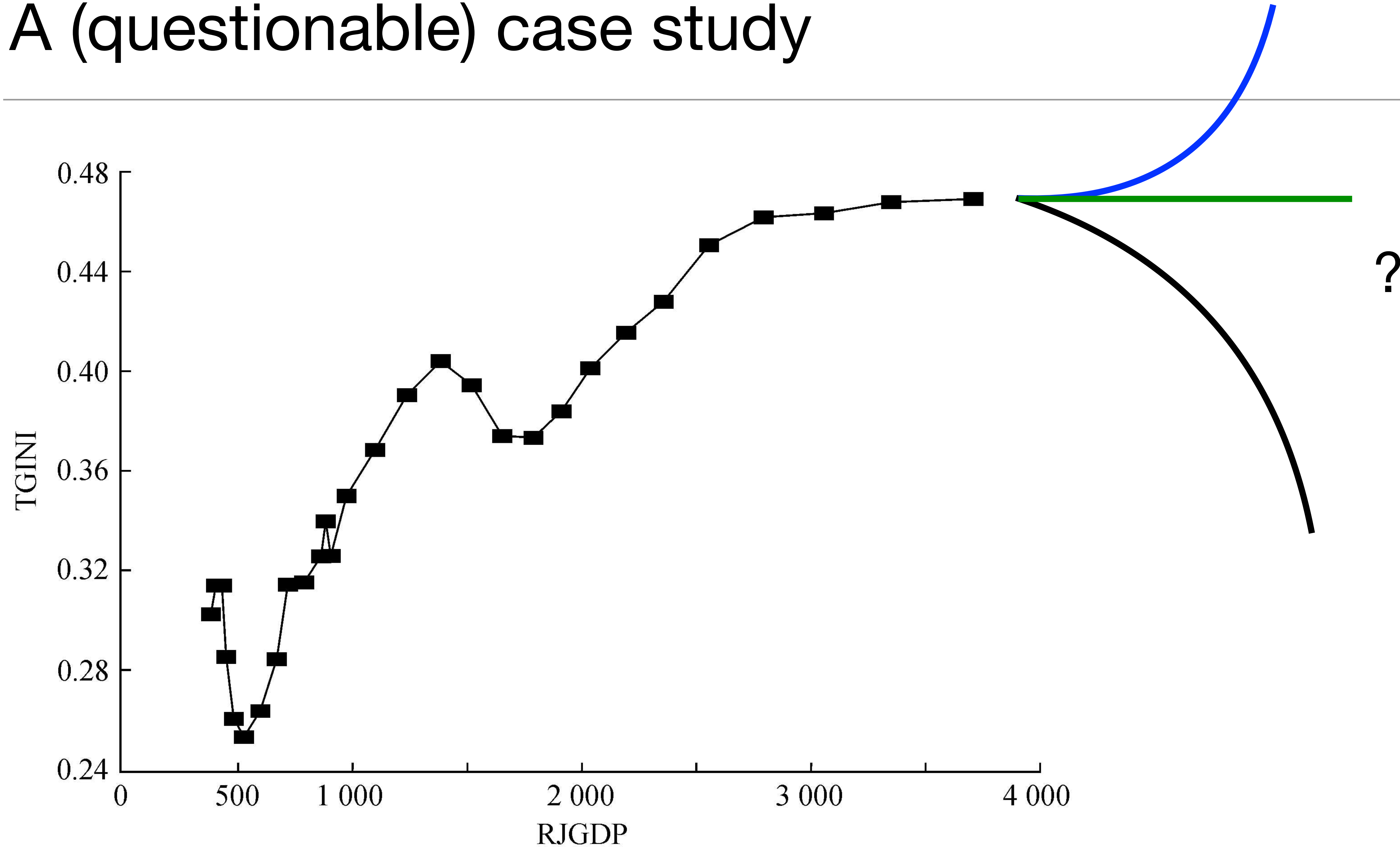
表 3 中的估计结果提供了如下一些信息：

（1）在模型 Ⅰ 和模型 Ⅱ 中，人均 GDP 及其平方项的估计系数是高度显著的，假说 1 得到证实。平方项为负值意味着全国居民总体收入差距的变动与人均 GDP 之间呈现出倒 U 形的关系，即收入差距倒 U 形假说得到了支持。

- Question:

  Is this test a valid way for testing U or inverted U shaped relations?

A (questionable) case study

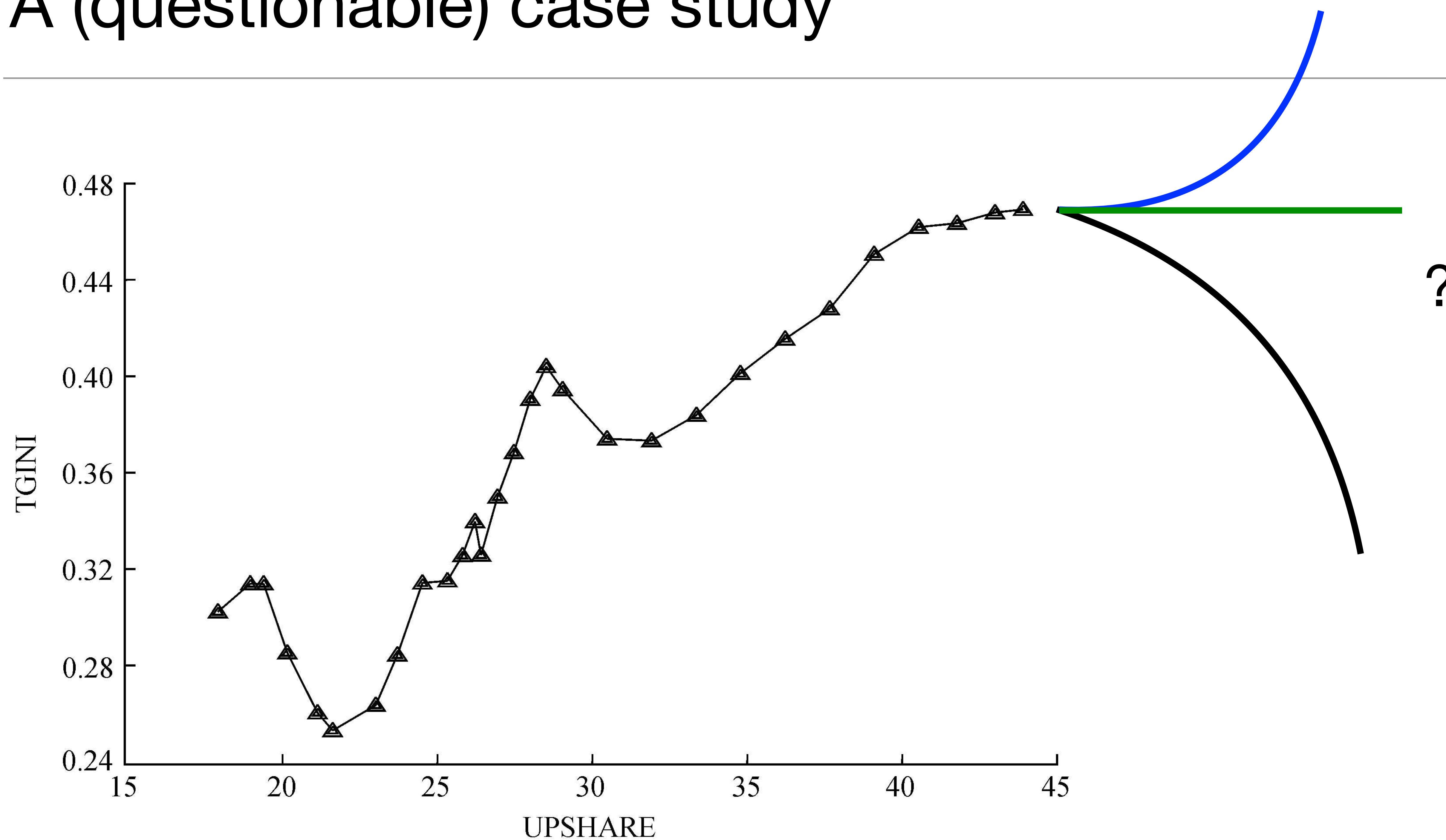


附图 2　人均 GDP 与全国居民收入基尼系数之间的散点图

资料来源：作者根据《中国统计年鉴》(1979—2007)年中的数据测算得到。

注：TGINI 为全国居民收入基尼系数，RJGDP 为按可比价格计算的人均 GDP。

A (questionable) case study



附图 3　城市人口比重与全国居民收入基尼系数之间的散点图

资料来源：作者根据《中国统计年鉴》(1979—2007)年中的数据测算得到。

注：TGINI 为全国居民收入基尼系数，UPSHARE 为城镇人口比重(%)。

A valid approach: Lind-Mehlum (2010)

OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 72, 1 (2010) 0305-9049
doi: 10.1111/j.1468-0084.2009.00569.x

With or Without U? The Appropriate Test for a U-Shaped Relationship*

Jo Thori Lind and Halvor Mehlum

Department of Economics, University of Oslo, PB 1095 Blindern, 0317 Oslo, Norway (e-mail: j.t.lind@econ.uio.no; halvor.mehlum@econ.uio)

A valid approach

- Research question:

  Given the estimates of a regression model, what is the test at some level of the presence of a U shape?

- A general model

$$y_i = \alpha + \beta x_i + \gamma f(x_i) + \xi' z_i + \varepsilon_i, \quad i = 1, \ldots, n.$$

- Assumption: $f(x)$ has a unique extreme point on the interval $[x_l, x_h]$ of the data. This is equivalent to require f' to be monotonic on this interval.

Hypotheses

- A U shape is implied by the condition

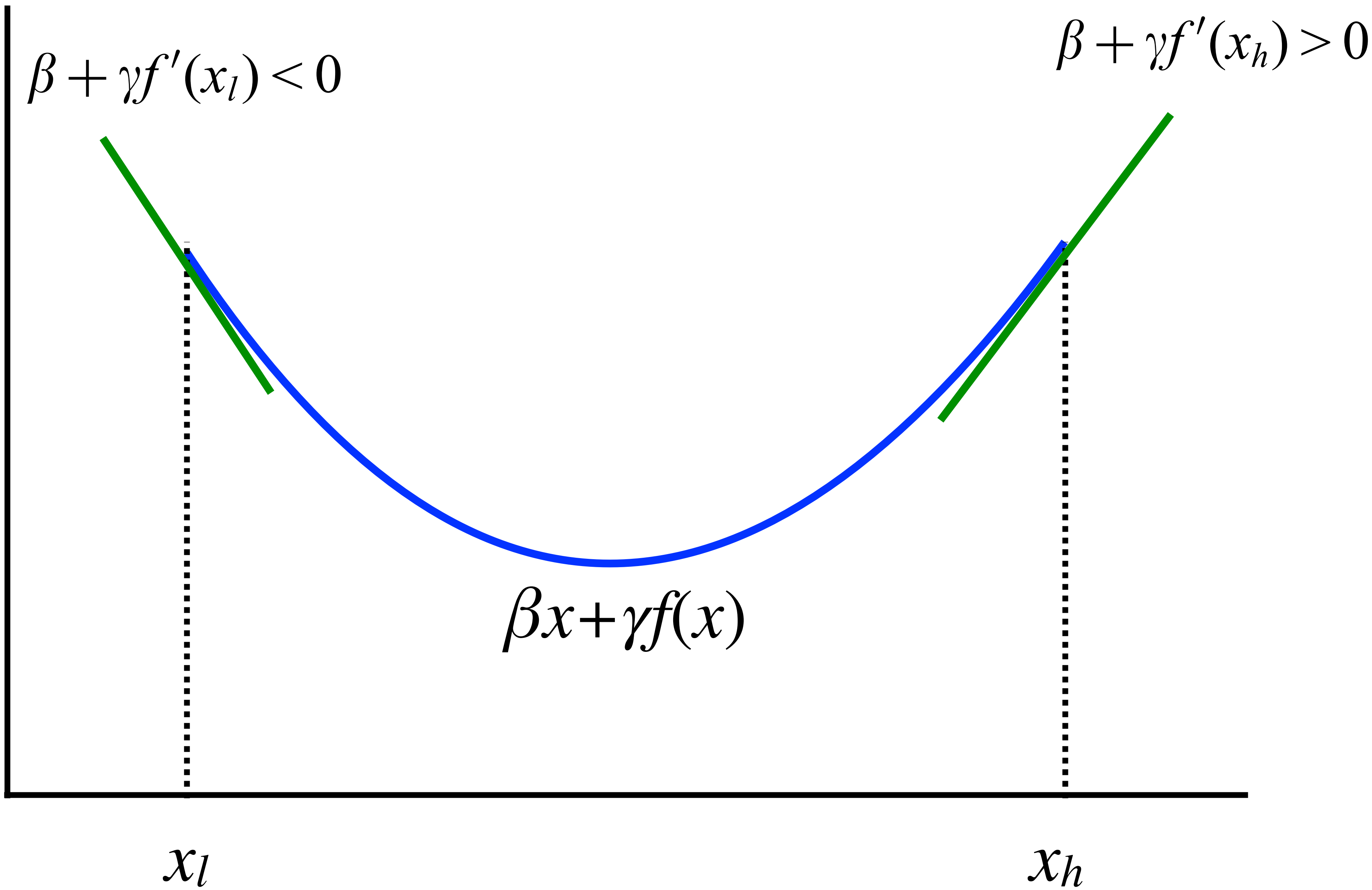
$$\beta + \gamma f'(x_l) < 0 < \beta + \gamma f'(x_h)$$

- Hypotheses

$$H_0 : \beta + \gamma f'(x_l) \geq 0 \ \text{ and/or } \ \beta + \gamma f'(x_h) \leq 0$$

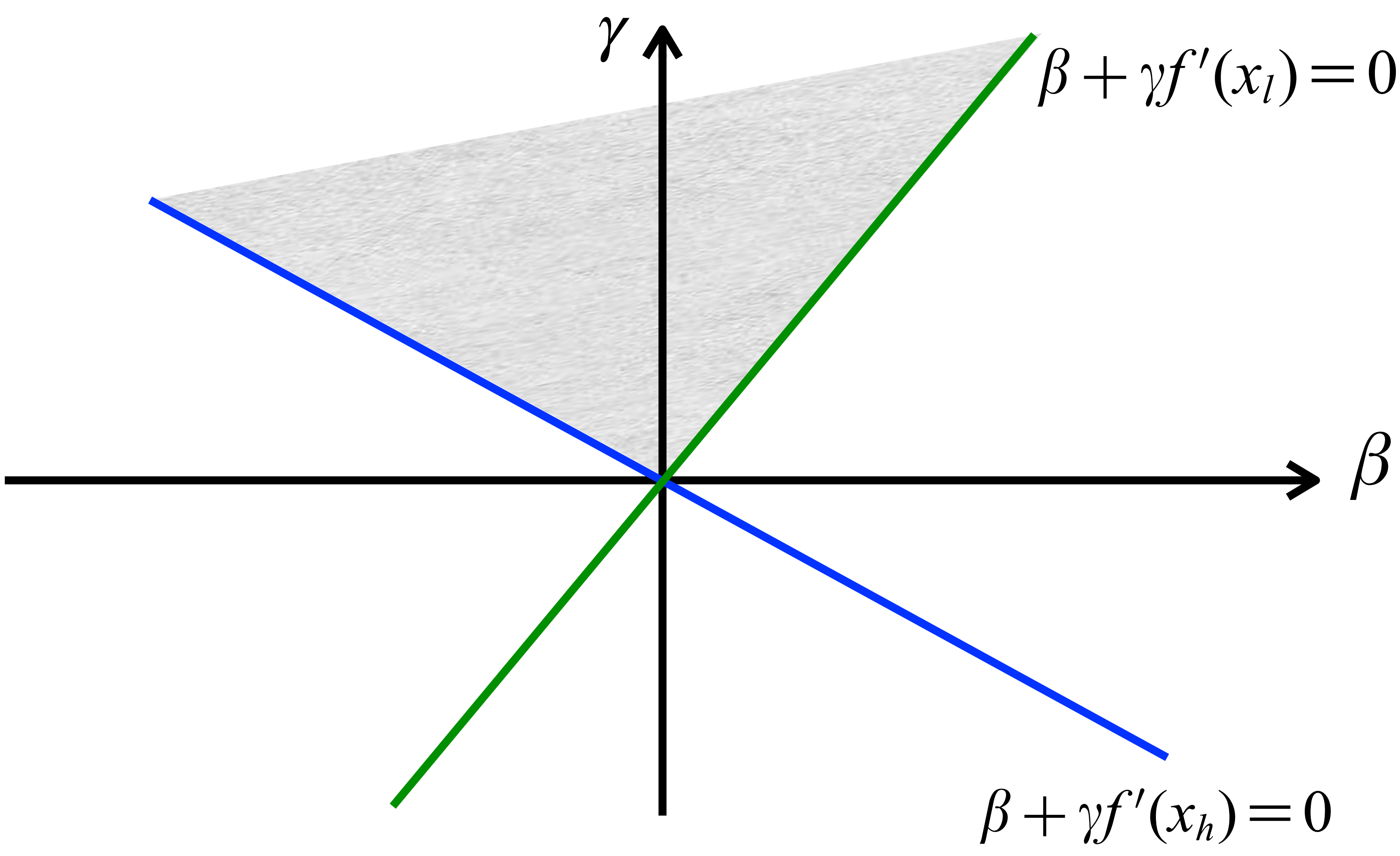
$$H_1 : \beta + \gamma f'(x_l) < 0 \ \text{ and } \ \beta + \gamma f'(x_h) > 0$$

A graphical description



The alternative hypothesis

$$H_1 : \beta + \gamma f'(x_l) < 0 \;\text{ and }\; \beta + \gamma f'(x_h) > 0$$



The Sasabuchi (1980) test

- A necessary condition for the existence of a U shape

Reject H_0 at the α level of confidence only if either,
H_0^L or H_0^H or both can be rejected at the α level of confidence,

where H_0^L and H_0^H are the null hypotheses in the two standard one-sided tests

$$H_0^L : \beta + \gamma f'(x_l) \geq 0 \text{ vs. } H_1^L : \beta + \gamma f'(x_l) < 0,$$
$$H_0^H : \beta + \gamma f'(x_h) \leq 0 \text{ vs. } H_1^H : \beta + \gamma f'(x_h) > 0.$$

The Sasabuchi (1980) test

The rejection area is the convex cone

$$R_\alpha = \left\{ (\beta, \gamma) : \frac{\beta + \gamma f'(x_l)}{\sqrt{s_{11} + 2f'(x_l)s_{12} + f'(x_l)^2 s_{22}}} < -t_\alpha \right.$$
$$\left. \text{and} \quad \frac{\beta + \gamma f'(x_h)}{\sqrt{s_{11} + 2f'(x_h)s_{12} + f'(x_h)^2 s_{22}}} > t_\alpha \right\}, \tag{5}$$

where s_{11}, s_{22} and s_{12} are the estimated variances of β and γ and the covariance between them, whereas t_α is the α-level tail probability of the t-distribution with the appropriate degrees of freedom.

Test for the quadratic form specification

- The quadratic form specification

$$y_i = \alpha + \beta x_i + \gamma x_i^2 + \xi' z_i + \varepsilon_i$$

- Criteria of rejection using two ordinary *t*-test

$$f'(x_l) < \hat{\theta}_l \equiv \frac{s_{12} t_\alpha^2 - \hat{\beta}\hat{\gamma} - t_\alpha \sqrt{(s_{12}^2 - s_{22} s_{11}) t_\alpha^2 + \hat{\gamma}^2 s_{11} + \hat{\beta}^2 s_{22} - 2 s_{12} \hat{\beta}\hat{\gamma}}}{\hat{\gamma}^2 - s_{22} t_\alpha^2},$$

$$f'(x_h) > \hat{\theta}_h \equiv \frac{s_{12} t_\alpha^2 - \hat{\beta}\hat{\gamma} + t_\alpha \sqrt{(s_{12}^2 - s_{22} s_{11}) t_\alpha^2 + \hat{\gamma}^2 s_{11} + \hat{\beta}^2 s_{22} - 2 s_{12} \hat{\beta}\hat{\gamma}}}{\hat{\gamma}^2 - s_{22} t_\alpha^2}.$$

The Fieller (1954) test for the extreme point

- The estimated extreme point satisfies

$$f'(\hat{x}^{\min}) = -\frac{\hat{\beta}}{\hat{\gamma}}$$

- A $(1-2\alpha)$ confidence interval for $\hat{x}^{\min}$ is given by $[\hat{\theta}_l, \hat{\theta}_h]$.

- The previous hypothesis test is equivalent to seeing whether the $(1-2\alpha)$ confidence interval for $\hat{x}^{\min}$ is inside the data range, i.e., $[\hat{\theta}_l, \hat{\theta}_h] \subset [x_l, x_h]$.

Performance of three tests using simulation

1. Whether the quadratic term is significant.
   → huge rate of type I errors: erroneously rejecting the absence of a U-shaped relationship.

2. Whether the quadratic term is significant and the estimated extreme point is in the data range (using Fieller's test).
   → stronger than (1), especially for large samples.

3. The suggested test (using Sasabuchi's test).
   → better than (1) and (2), but has somewhat lower test power than (2) when there is actually a U shape.

Fraction of type I errors

TABLE 1

Monte Carlo analysis of the performance of the test

		Number of observations									
		Correctly specified model					*Mis-specified model*				
Extreme point		10	50	100	500	1,000	10	50	100	500	1,000
0.6	(i)	0.11	0.18	0.32	0.91	1.00	0.12	0.26	0.45	0.98	1.00
	(ii)	0.11	0.18	0.32	0.91	1.00	0.12	0.26	0.45	0.98	1.00
	(iii)	0.10	0.18	0.30	0.82	0.97	0.13	0.30	0.49	0.98	1.00
0.8	(i)	0.11	0.18	0.31	0.91	1.00	0.12	0.26	0.45	0.98	1.00
	(ii)	0.11	0.18	0.31	0.89	0.97	0.12	0.26	0.45	0.98	1.00
	(iii)	0.09	0.11	0.14	0.36	0.57	0.11	0.20	0.31	0.83	0.98
1	(i)	0.11	0.19	0.31	0.91	1.00	0.12	0.26	0.44	0.98	1.00
	(ii)	0.11	0.19	0.31	0.51	0.50	0.12	0.26	0.44	0.93	0.98
	(iii)	0.07	0.06	0.05	0.05	0.05	0.09	0.13	0.17	0.45	0.68
1.2	(i)	0.11	0.19	0.32	0.91	1.00	0.12	0.25	0.45	0.98	1.00
	(ii)	0.11	0.18	0.26	0.10	0.03	0.12	0.25	0.45	0.81	0.90
	(iii)	0.05	0.02	0.01	0.00	0.00	0.08	0.09	0.11	0.23	0.36
1.4	(i)	0.11	0.18	0.31	0.91	1.00	0.13	0.25	0.45	0.98	1.00
	(ii)	0.11	0.16	0.13	0.00	0.00	0.13	0.25	0.44	0.71	0.78
	(iii)	0.04	0.01	0.00	0.00	0.00	0.08	0.07	0.08	0.13	0.19
1.6	(i)	0.12	0.18	0.32	0.91	1.00	0.13	0.25	0.45	0.99	1.00
	(ii)	0.11	0.11	0.05	0.00	0.00	0.13	0.25	0.44	0.63	0.68
	(iii)	0.04	0.00	0.00	0.00	0.00	0.08	0.06	0.07	0.09	0.12

Notes: Tests performed at the 5% level of significance. Results are based on 10,000 simulations. The tests studied are: (i) whether the quadratic term is significant; (ii) whether the quadratic term is significant and the estimated extremum point is in [0, 1]; and (iii) the test described previously.

An illustration using GDP and Gini coefficient

TABLE 2

Estimates of the Kuznets curve

Dependent variable: Gini index		
log per capita GDP (X_i)	$\hat{\beta} =$	32.27
		(11.63)***
log per capita GDP-squared (X_i^2)	$\hat{\gamma} =$	−1.88
		(0.65)***
Slope at X_l	$\hat{\beta} + 2\hat{\gamma}X_l =$	8.15
		(3.65)**
Slope at X_h	$\hat{\beta} + 2\hat{\gamma}X_h =$	−4.33
		(2.34)*
Appropriate U test		1.85
		[0.033]
Extremum point	$-\hat{\beta}/(2\hat{\gamma}) =$	8.60
90% confidence interval, Fieller method		[7.44, 9.58]
90% confidence interval, Delta method		[7.73, 9.47]

Notes: Robust standard errors in parentheses and *P*-values in square brackets. ***, ** and * denote significances at the 1, 5, and 10% levels.

GDP, gross domestic product.

An illustration using GDP and Gini coefficient

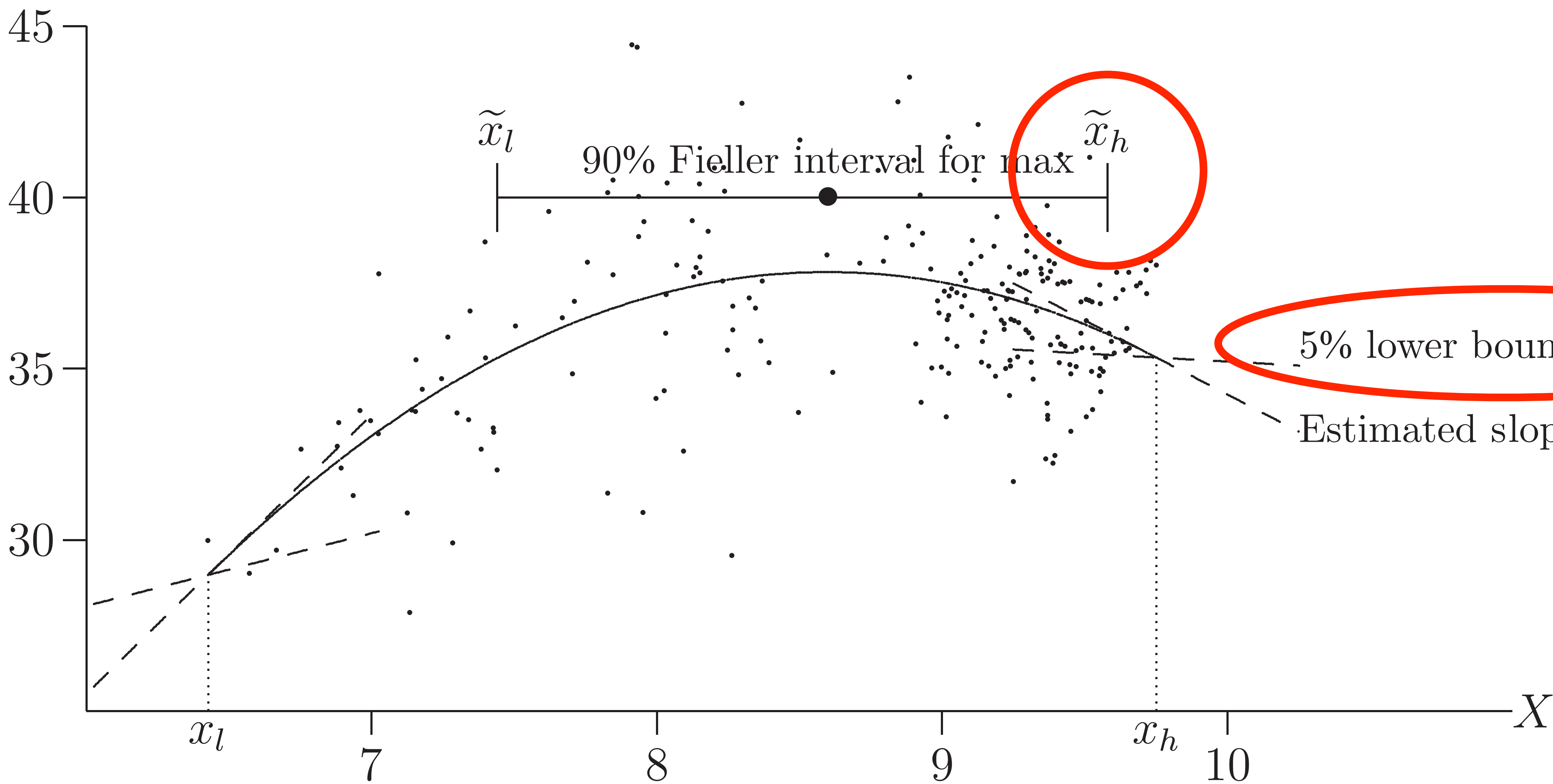


Figure 1. Illustrative Kuznets curve

The inverted U-shaped relationship exists over the range of the data, but the significance is weaker than traditional approaches.

A summary of Lind-Mehlum U test

1. Specify a model: $y_i = \alpha + \beta x_i + \gamma x_i^2 + \xi' z_i + \varepsilon_i$

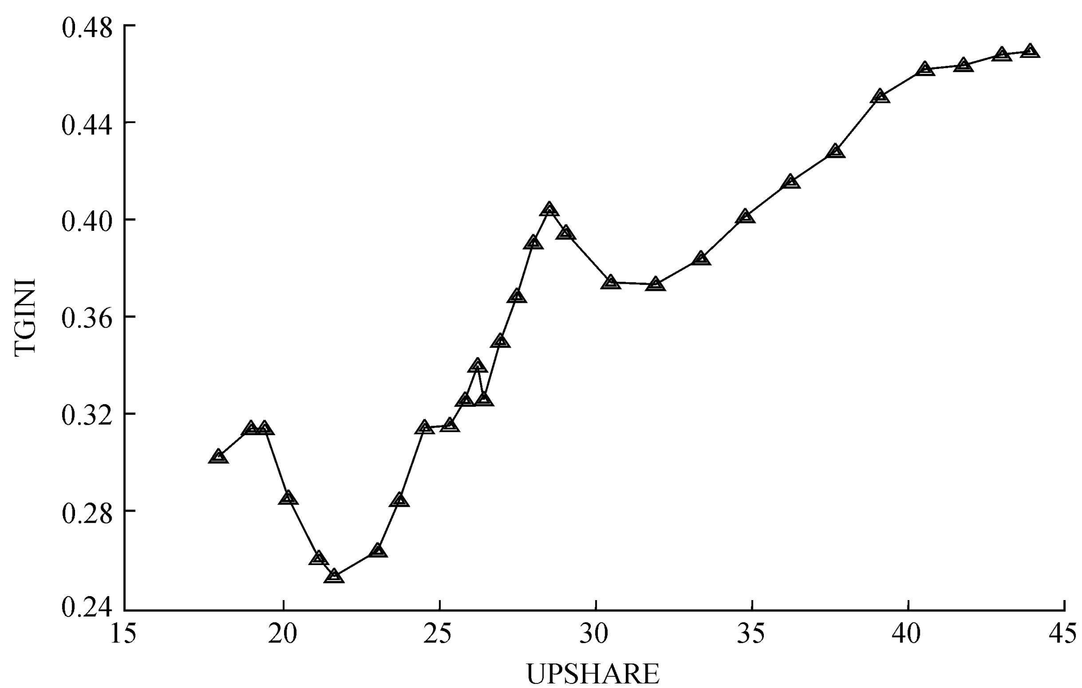
2. Estimate parameters.

3. Calculate the estimated slopes and standard errors at x_l and x_h.

4. Perform *t* tests for the slopes.

5. The weaker results of the above two *t* tests is the overall result.

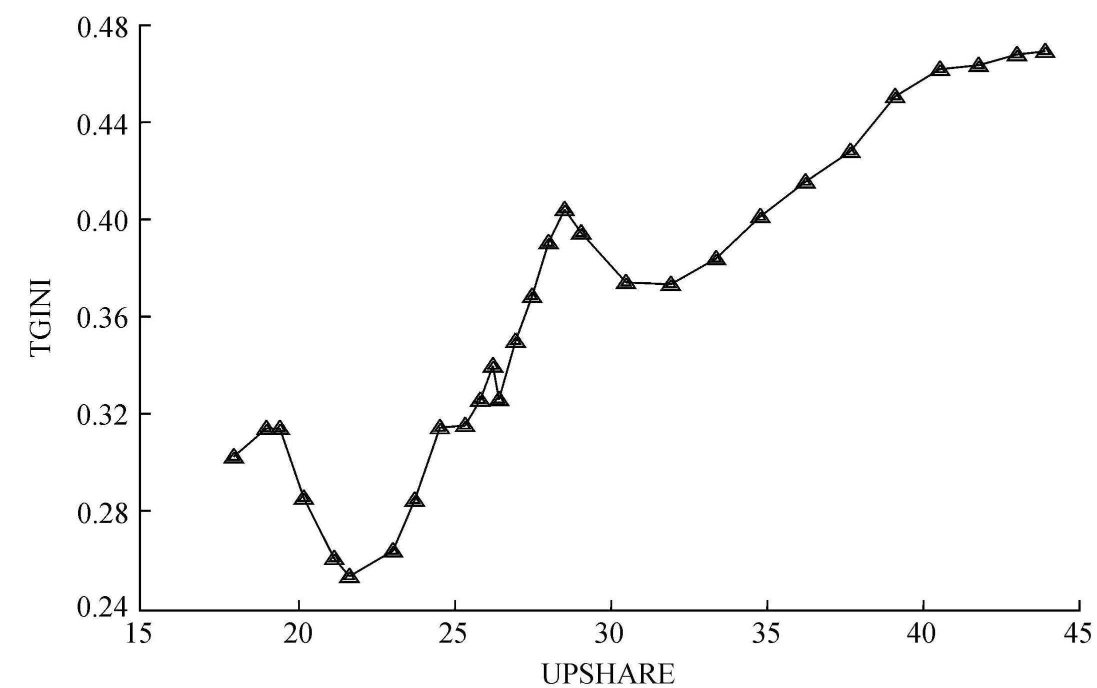
Back to the case study



附图 3　城市人口比重与全国居民收入基尼系数之间的散点图

资料来源：作者根据《中国统计年鉴》(1979—2007)年中的数据测算得到。

注：TGINI 为全国居民收入基尼系数，UPSHARE 为城镇人口比重(%)。



附图 3　城市人口比重与全国居民收入基尼系数之间的散点图

资料来源：作者根据《中国统计年鉴》(1979—2007)年中的数据测算得到。

注：TGINI 为全国居民收入基尼系数，UPSHARE 为城镇人口比重(%)。

- Take home practice: perform a Lind-Mehlum U test to check whether there is a inverted U-shaped relation as stated in this paper.