Taking Kuznets Seriously*

Jakob Madsen†

Holger Strulik‡

February 21, 2024

Abstract. Numerous studies have tested Kuznets' hypothesis of a hump-shaped relationship between inequality and industrialization by regressing the Gini coefficient on the level and square of per capita income. Here, we examine Kuznets' original idea that inequality first increases and then declines when workers move from agriculture to manufacturing. We collect sectoral wage and employment data for 17 advanced countries over the period 1800-2018 and compute the 'Kuznets Gini' as the Gini coefficient that results from changes in sectoral employment shares and the sectoral wage differential. We show the hump-shaped path of the Kuznets Gini and establish in panel regressions that the Kuznets Gini has been an important determinant of inequality.

Keywords: Kuznets curve, inequality, industrialization, economic development, Gini coefficient.

JEL: O14, O15, O40, N30, J10.

*We gratefully acknowledge financial support from the Australian Research Council (DP220102557).

† University of Western Australia, Department of Economics, 35 Stirling Hwy, Crawley, 6009 WA, Australia. email: jakob.madsen@uwa.edu.au

‡ University of Göttingen, Department of Economics, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany; email: holger.strulik@wiwi.uni-goettingen.de.

1. Introduction

In his seminal study, Kuznets (1955) speculated on a typical path of inequality in the course of economic development. While Kuznets (1955) gave various possible reasons for the evolution of inequality, the core of his study proposes the hypothesis that industrialization, understood as migration of labor from the agricultural to the industrial sector, results in increasing inequality during the initial stages of industrialization, as wages of industrial workers exceed wages of agricultural workers, and that inequality starts declining when the industrial labor share becomes sufficiently large.

In this paper, we examine Kuznets original idea by collecting data on wages and employment for 17 advanced countries over the period 1800-2018 and by computing the Gini coefficient that results from industrialization and the sectoral wage differential. We denote this Gini coefficient as the *Kuznets Gini* to distinguish it from the conventional Gini coefficient obtained from the overall income distribution. We then examine the ability of the Kuznets Gini, the capital share, and the top 1% income share to explain the evolution of overall income inequality, over the period 1800-2018. We limit the sample to developed countries due to unavailability of agricultural wage data, particularly, for a reasonable time horizon for developing countries.

Numerous studies have tested Kuznets' hypothesis of a hump-shaped relationship between inequality and industrialization by regressing the overall Gini coefficient on the level and square of per capita income, where the variables are mostly measured in logs. Earlier studies were based on cross-sectional data before the time-series data on income inequality constructed by Deininger and Squire (1996) made panel data studies feasible (see, for an overview of the literature and estimates, Thomas, 1991; Williamson, 1991; Deininger and Squire, 1998; Aghion et al., 1999; Kanbur, 2000; Acemoglu and Robinson, 2002; Helpman, 2004, Ch. 6; Banerjee et al., 2006; Barro, 2008; Chang et al., 2021; Batuo et al., 2022). Generally, while the early cross-sectional evidence was supportive of the Kuznets curve, the later time-series evidence has generally been unsupportive (see Helpman, 2004, Ch. 6). However, conventional tests of the Kuznets curve, where the overall income Gini coefficient is regressed on the level and square of per capita income, are fragile and sensitive to the data sample as shown by Anand and Kanpur (1993) and Chang et al. (2021).

The mixed evidence in studies based on the level and square of income is unsurprising for at least four reasons. First, increasing productivity is mostly driven by manufacturing and is often associated with a decreasing employment share in manufacturing. The income path is thus unlikely to capture a Kuznets curve for which the sectoral employment share is the main driving force. For example, the manufacturing employment share in total employment in our sample has declined significantly over the past 60 years despite substantial growth in per capita income over the same period. Second, the nexus between overall income inequality and the Kuznets curve is compounded by other dimensions of inequality, such as the functional income distribution and top income shares. Third, a significant portion of the results in the literature are based on cross-sectional data where unobserved heterogeneity is likely to create a bias in the estimates (see, for critical assessments, Anand and Kanbur, 1993; Deininger and Squire, 1998; Kanpur, 2000). Fourth, our analysis suggests that many developing countries are still on a path of increasing industrialization-driven inequality, thus preventing the empirical identification of a hump shape. See Williamson (1991) for a critical review of the literature on the conventional approach to the Kuznets curve.

While the Kuznets curve has achieved textbook status, the original Kuznets hypothesis of a hump-shaped path of inequality caused by the shift of labor from agriculture to manufacturing has, to our knowledge, never been tested in its own right. Kuznets (1955) notes that, "no adequate empirical evidence is available for checking this conjecture of long secular swing in income inequality" (p. 19), and he motivates his ideas with numerical illustrations. To the best of our knowledge, this lack of empirical support for the original Kuznets hypothesis has persisted to date, since consistent historical data on sectoral wages and employment for the entire industrialization period of the advanced countries were apparently not available.[1]

Here, we address this problem by constructing a new data set that enables us to compute the Kuznets Gini for the period 1800-2018. We show the hump shaped path of the Kuznets Gini with a peak between 1920 and 1950 in most countries. In panel regressions, we then show that the Kuznets Gini is a highly robust determinant of overall income inequality. For the core period of industrialization, 1880–1980, our regressions provide a partial correlation between the Kuznets Gini and the overall Gini coefficient of between 0.20 and 0.30 when we control for potential confounders as well as country and year fixed effects. Compared to the capital share and the top 1% income share, two other important determinants of income inequality proposed

[1]In later studies, Kuznets empirically examined differences in income inequality across countries and, more rudimentarily, changes in inequality over time (e.g., Kuznets, 1963). However, the original idea from his groundbreaking 1955 work, that industrialization, understood as the migration of labor from agriculture to a modern sector, was a driver of changing inequality, remained largely unexplored.

in the literature, the Kuznets Gini explains by far the greatest fraction of the variance in the overall Gini coefficient. Finally, we show that the conventional method used to test the Kuznets hypothesis yields misleading results and results that are sensitive to estimation period.

2. Analytical Framework

2.1. **The Kuznets Gini.** We measure income inequality by the Gini coefficient, which is defined as half of the relative mean absolute difference of all pairs of income in the population. Formally, the Gini coefficient at time t is given by

$$Gini_t \equiv \frac{\sum_{i=1}^{L_t} \sum_{j=1}^{L_t} |y_{it} - y_{jt}|}{2L_t^2 \bar{y}_t}, \tag{1}$$

in which L_t is population size, y_{it} is income of person i, and $\bar{y}_t$ is average income. In constructing the Kuznets-Gini, we follow Kuznets (1955) and assume that individuals earn an income in either agriculture or the modern (industrial) sector. We focus on low-skilled individuals and, since the dispersion of wage rates of low-skilled workers in agriculture and manufacturing is low, we assume that at a given time there are only two wages, $y_{it} \in \{w_t^A, w_t^I\}$, in which w_t^A is the average agricultural wage and w_t^I is the average industrial wage. Since there are only a few data points where the agricultural wage marginally exceeds the industrial wage, we additionally assume that $w_t^I \geq w_t^A$, such that (1) can be expressed as:[2]

$$Gini_t^{Kuz} = \frac{L_t^I L_t^A (w_t^I - w_t^A)}{(L_t^A + L_t^I)(w_t^I L_t^I + w_t^A L_t^A)} = \frac{w_t^I L_t^I}{w_t^I L_t^I + w_t^A L_t^A} - \frac{L_t^I}{L_t^I + L_t^A}, \tag{2}$$

in which L_t^A and L_t^I are (low-skilled) employment in agriculture and manufacturing, respectively. The Kuznets Gini is thus simply the income share of manufacturing in total wage income in agriculture and manufacturing minus the employment share in manufacturing. Note that the Kuznets Gini measures inequality within the low-skilled population, which was by far the majority of the population at the beginning of industrialization but later declined as the workforce became more educated. We therefore expect that the contribution of the Kuznets Gini to overall inequality decreases towards the end of the industrialization period.

For analytical convenience, we introduce the industrial wage premium $\phi_t \equiv w_t^I / w_t^A$ and, as a

[2] Due to data limitations, we cannot examine the role of within sector inequality as in some related studies that focus on more recent periods or on cross-sectional data, see e.g. Anand and Kanbur (1993).

measure of the level of industrialization, the employment ratio $\theta_t = L_t^I / L_t^A$ and rewrite (2) as:

$$Gini_t^{Kuz,V} = \frac{\phi_t \theta_t}{1 + \phi_t \theta_t} - \frac{\theta_t}{1 + \theta_t}. \tag{3}$$

We have added a V to the superscript to indicate that the wage ratio is allowed to vary over time. In the few cases where the agricultural wage exceeds the industrial wage, ϕ_t is defined as w_t^A / w_t^I and θ_t as L_t^A / L_t^I.

Additionally, we define $Gini^{Kuz,F}$ as the Kuznets Gini with fixed relative wages,

$$Gini_t^{Kuz,F} = \frac{\bar{\phi} \theta_t}{1 + \bar{\phi} \theta_t} - \frac{\theta_t}{1 + \theta_t}, \tag{4}$$

in which $\bar{\phi}$ is the average industrial wage premium over the sample period for each individual country.

Whether $Gini^{Kuz,V}$ or $Gini^{Kuz,F}$ is what Kuznets had in mind is not clear from his 1955 article. If the industrial wage premium is an endogenous response to industrialization, then $Gini^{Kuz,V}$ is the relevant measure of the Kuznets Gini. If not, then $Gini^{Kuz,F}$ is the relevant proxy. As shown in the empirical section, the principal results are independent of whether the focus is on $Gini^{Kuz,V}$ or $Gini^{Kuz,F}$.

We use wages for industrial unskilled labor as they are likely the closest substitutes for unskilled agricultural labor among the industrial workers. Some of the workers migrating from the rural to the urban areas may have entered an apprenticeship and eventually ended up as skilled workers, consequently leading to an even higher industrial wage premium and Kuznets Gini relative to our baseline case. Since the skill-premium between skilled and unskilled blue collar workers have not changed much over time, the principal results will not be affected by this consideration (Clark, 2005). We have not accounted for service sector jobs taken up by migrating agricultural labor because the occupational classification changes constantly over time and the data are highly fragmented. Furthermore, our focus is squarely on employed migrant workers; not rural individuals who transit to the urban centers through the educational system and, per definition, are not migrating across sectors. Finally, we focus on the industrial labor force where the lion's share of workers are either blue-collar workers or low brow white-collar workers.

2.2. **Peak Inequality.** It is straightforward to infer from (3) and (4) that during industrialization (conceptualized as rising θ_t) inequality first increases and then declines. The $Gini^{Kuz}$ is zero for $\theta = 0$ (nobody works in manufacturing) and for $\theta \to \infty$ (everybody works in manufacturing) and thus it is first increasing (when only a few work in manufacturing) and then declining (when many work in manufacturing). More precisely, the $Gini^{Kuz,F}$ assumes a maximum where

$$\frac{\partial Gini_t^{Kuz,F}}{\partial \theta_t} = \frac{\bar{\phi}}{(1+\bar{\phi}\theta_t)^2} - \frac{1}{(1+\theta_t)^2} = 0,$$

that is where

$$\theta_t = \theta^* \equiv \sqrt{1/\bar{\phi}}, \qquad Gini^{Kuz,F} = G^* \equiv 1 - \frac{2}{1+\sqrt{\bar{\phi}}}. \tag{5}$$

Inspection of (5) reveals that peak inequality is reached earlier (i.e., at a lower level of industrialization θ^*) and peak inequality G^* is higher in countries where the industrial wage premium $\bar{\phi}$ is larger. Since $\bar{\phi} > 1$, we conclude from (5) that $Gini_t^{Kuz,F}$ starts declining before employment in the industrial sector overtakes agricultural employment (before $\theta_t = 1$). For our sample average, we computed $\bar{\phi} = 1.71$ and thus $\theta^* = 0.76$.

While the industrial wage premium is mostly moving within a range between one and two in our sample, it differs markedly across countries at the world level: In the year 2018, the premium varied between 0.01 (Brunei) and 9.2 (Zambia) with a worldwide median of 1.6 (ILOSTAT, 2023). For the median country, the implied peak of the Kuznets Gini is thus at an industrialization level of $\theta^* = 0.79$. The world average industrialization level θ_t was 0.51 in the year 2003 (WDI, 2023), which suggests that many developing countries have been on the upward trajectory of the Kuznets curve, at least up until recently. This feature explains the difficulty in identifying a hump shaped Kuznets curve for developing countries. More generally, since the industrial skill premium differs markedly across developing countries, it is no surprise that the conventional approach to the Kuznets curve has had difficulties establishing a robust relationship between inequality and economic development, proxied by the level and the square of per capita income.

3. Data and Graphical Analysis

3.1. **Data for the Kuznets Gini.** To construct $Gini_t^{Kuz,V}$ and $Gini_t^{Kuz,F}$, we collected data for the following 17 OECD countries for which long data are available over the period 1800-2018 (henceforth G17): Australia, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, the UK, and the US. Detailed data sources are relegated to the Data Appendix, which also contains an account of the other variables used in the analysis below.

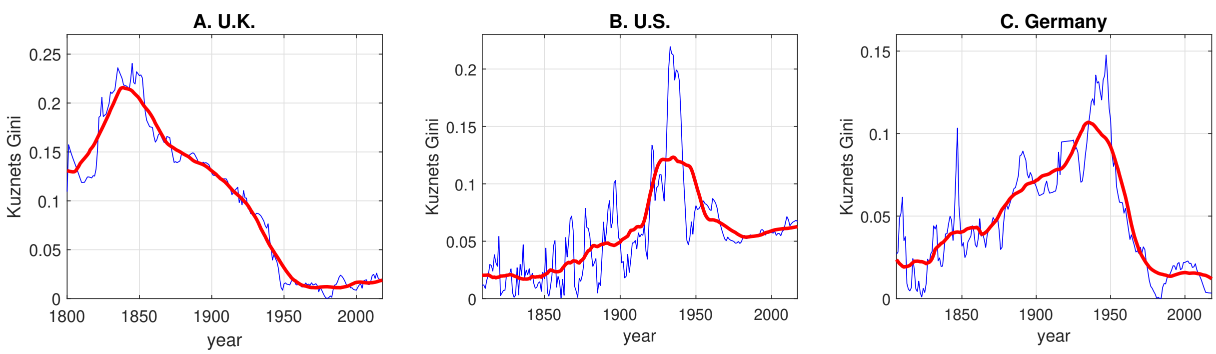
To estimate $Gini^{Kuz}$ from (3), we need data for wages of agricultural labor and unskilled blue-collar labor and employment in the agricultural and industrial sectors. The employment data are from various historical sources and censuses before regular labor market surveys were implemented in the first half of the 20th century. The construction of the wage data is a complex process because the data are fragmented, only available over short periods in most of the source material, often change in the statistical methodology, and need to be adjusted for hours worked per day, week, and year so that the payment periods in the two sectors are the same. We use primary source material whenever possible to ensure data consistency. Given the importance of wages for comparisons between trades, sectors, and regions and for the evolution of income inequality, wage data has almost always been collected throughout the history of the countries in our sample. The preindustrial data are mostly based on the records of big employers and employer organizations such as, guilds, manors, and some of the data is only available for a few locations within a country. Official statistical agencies, mostly established in the first half of the 19th century, took up the task of collecting wage data early on, often based on census surveys. Starting around 1870-1900, statistical agencies regularly conducted industrial surveys and, in some instances, relied on wage data from union members in the urban professions.

3.2. **Data for the Overall Gini, Top 1% Shares, and the Capital Share.** The data for the overall market Gini coefficient are mostly collected from national sources and spliced with the net Gini in the most recent decades, where the market Gini is constructed from pre-tax, pre-transfer income, and the net Gini is the post-tax post-transfer income. Since the data from national sources are relatively sparse in the 19th century we base our data construction on the Gini coefficients estimated by Moatsos et al. (2014). The data are mostly available at 10-year frequencies before the second half of the 20th century. The Top 1% income share is from the World Income Inequality Database (2022) and backdated using national sources and the ratio of per capita income and the average wage of agricultural and urban unskilled labor. The Top 1% income share is mostly based on annually available tax records. Finally, the capital share is estimated as one minus the labor share, where the labor share is the share of compensation to employees in total net national income (GDP minus depreciation of non-residential fixed capital). The labor compensation is adjusted for imputed labor income rent of self employed and family workers, and the net national income is purged from imputed house rent (see, for details, Kerspien and Madsen, 2023).

3.3. **Kuznets Curves.** Figure 1 shows the evolution of the $Gini^{Kuz,V}$ for the U.K., the U.S., and Germany, the three countries on which Kuznets (1955) mainly focused his comparative analysis. Blue lines show the raw data points and red lines show 15-year moving averages. The curves are hump-shaped as predicted by Kuznets. We can use them to re-examine Kuznets intuitive reasoning more thoroughly. Kuznets (1955, p. 19) writes:

"No adequate empirical evidence is available for checking this conjecture of a long secular swing in income inequality; nor can the phases be dated precisely. However, to make it more specific, I would place the early phase in which income inequality might have been widening, from about 1780 to 1850 in England; from about 1840 to 1890, and particularly from 1870 on in the United States; and, from the 1840's to the 1890's in Germany. I would put the phase of narrowing income inequality somewhat later in the United States and Germany than in England – perhaps beginning with the first world war in the former and in the last quarter of the 19th century in the latter."

Figure 1. The Kuznets Gini ($Gini^{Kuz,V}$) in the U.K., the U.S., and Germany



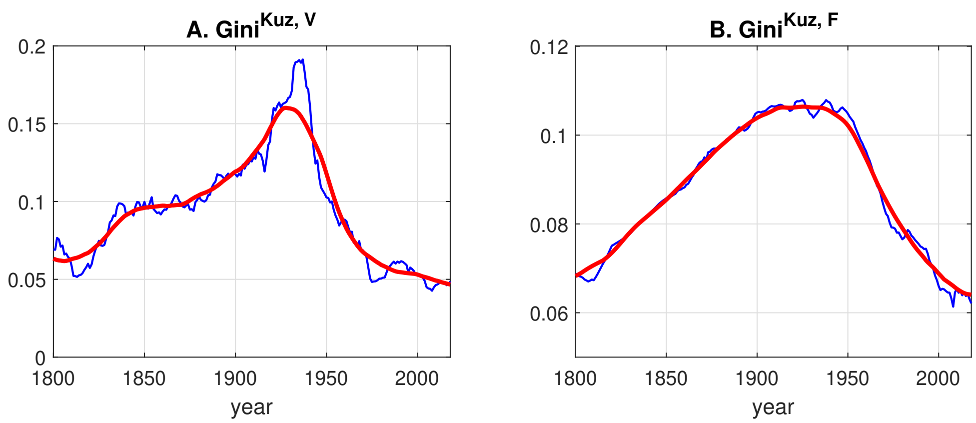
Blue thin lines: raw data; red thick lines: 15-year-moving-average.

Now, with adequate empirical evidence available, we conclude that Kuznets' intuition accurately assigned the period of increasing inequality in the U.K. and also correctly suspected that inequality in the US and Germany was lagging behind. However, he underestimated the timing of the peak of inequality in these countries by around 25 to 30 years.

Figure 2 shows the evolution of the average Kuznets Gini coefficient for the G17 countries. The left-hand side shows the $Gini^{Kuz,V}$ and the right-hand side shows the $Gini^{Kuz,F}$ with the industrial wage premium fixed at the average mean for each individual country. Both curves peak around 1930 and track each other reasonably well except for the spike of $Gini^{Kuz,V}$ over the period 1920-1940.

The spike can be explained by the agricultural crisis during the interwar period and the immediate post-WWI urban labor uprising that resulted in marked increases in wages of the urban blue-collar workers. The expansion in the world supply of agricultural products in the interwar period created an agricultural crisis that resulted in declining agricultural wages deflated by consumer prices (Madsen, 2001). This in turn, reduced the relative value of the marginal productivity of labor and hence the relative wages in agriculture. The elimination and reversal of the excess supply of agricultural produce on the international market throughout the 1940s reduced the industrial wage premium by 32%. Moreover, mechanization-induced productivity advances in the agricultural sector since approximately 1950, gradually reduced the industrial wage premium, leading to a steeper decline of $Gini^{Kuz,V}$ than $Gini^{Kuz,F}$.

Figure 2. The Kuznets Gini: G 17 – Average



Blue thin lines: raw data; red thick lines: 15-year-moving-average.

In general, however, Figure 2 shows that the migration of workers from agriculture to the industrial sector explains a large part of industrialization-driven inequality. Industrialization can motivate a Gini coefficient of 0.1 to 0.15 with maxima around 0.3 in the country-specific peaks. This suggests that industrialization may explain a significant part of overall inequality.

4. Regression Analysis

4.1. **Methodology.** For the Kuznets hypothesis to be a relevant partial explanation for the income inequality path, the Kuznets Gini must be significantly correlated with the overall income inequality over the course of industrialization. We would like to stress that we are not estimating a causal relationship but seek to find a mapping between the overall Gini and its components. To test for the relevance of the Kuznets Gini for overall income inequality, we estimate the following stochastic model:

$$Gini_{it} = \alpha_1 Gini_{it}^{Kuz,X} + \alpha_2 Top1_{it} + \alpha_3 S_{it}^{K} + \zeta Z_{it} + \gamma_t + \mu_i + \epsilon_{it}, \tag{6}$$

where $Gini$ is the overall income Gini coefficient; $Gini^{Kuz,X}$ is the Kuznets Gini, $X = V, F$; $Top1$ is the Top 1% income share; S^K is the income share of capital; Z is a vector of confounders; the γ_ts are time fixed effects; the μ_is are country fixed effects, and ϵ is a disturbance term. The model is estimated over various periods between 1800 and 2018.

Our stochastic model approximately follows the decomposition of the Gini coefficient derived by Checchi and Garcia-Penalosa (2010) who decompose the Gini into the functional income distribution, between-labor inequality, between-capital inequality, and the inequality between wage earners and government transfers to unemployed persons and pensioners. In our regression, the Kuznets Gini captures the inequality between low skilled workers; the capital share captures the distribution of income between capital and labor; and the top 1% income share includes the income distribution within the top-income earners regardless of whether their income originates from labor or capital.

While the top income share and the capital income share complement the Kuznets Gini in the overall Gini coefficient, they suffer from significant measurement errors. This is due to, for example, the large fraction of non-filers of income taxes before WWII, unreported labor income due to tax evasion after WWII, and the increasing share of employee fringe benefits in total wages, which render it difficult to precisely estimate historical top income shares (Piketty et al., 2018). The capital share is a much broader measure of capital income than that reported on tax returns on which most of the top income share data are based. Capital income in national accounts includes such items as imputed rents of homeowners, corporate retained earnings, property taxes, returns on pension funds, corporate taxes, and capital income earned by trusts and estates. These items do not show up in private accounts and are, therefore, not visible in the top 1% income share. The capital share data are not perfect either because of measurement problems, such as how to split the investment in intellectual property products between labor and capital (which is currently accounted for as capital income), zero profits recorded for the government sector, unavailability of wage statistics of non-manual professions before circa WWII, imputed house rent that never figure in earnings accounts, etc.

Confounders are included in some regressions to capture other sources of overall inequality that are not captured by the baseline regressors. As confounders we consider productivity growth and trade openness. Productivity growth is measured as the average geometric growth rate in per capita income over the previous 20 years to filter out the influence of erratic movements in the data and business cycle fluctuations. Trade openness is measured as imports plus exports divided by nominal GDP. Per capita income growth is included in the regression following the predictions of the model of Chu and Peretto (2019) in which increasing economic growth during industrialization leads to gradually increasing income inequality until the economy reaches a steady state. Similarly, Madsen et al. (2021) show that intangibles have been important drivers of the inequality path since 1860, and Grossman and Helpman (2018) show that the fundamental drivers of economic growth also determine the composition of workers with different skills and, therefore, affect wage inequality. Inequality may be affected by international trade due to knowledge spillovers that render innovations more productive, which in turn creates incentives for expansion of the idea-generating professions and, hence, the relative pay for the most able individuals. Furthermore, Adao et al. (2022) show that imports of cheaper foreign goods relative to domestic goods tends to raise relative demand for the factors owned by the rich because firms employing educated workers import a disproportionally high share of intermediate products.

Returning to the question of whether the wage premium should be fixed or variable in the estimates of the Kuznets Gini, a key issue is to what extent the evolution of the wage ratio is an outcome of industrialization. The upward trend in the wage ratio up until WWII and the downward trend thereafter were probably affected by industrialization, suggesting that the wage premium is somewhat endogenous to industrialization. However, it is beyond the scope of this paper to account for the endogenous part of the wage ratio due to the likely influence of fluctuations in real world food prices and tariffs on imports of food, union strength in the urban centers etc. Fortunately, as shown below, the principal results are the same regardless of whether the Kuznets Gini is based on a variable or fixed wage ratio.

4.2. **Estimation Results.** The results of estimating equation (6) are presented in Table 1. The coefficients of the Kuznets Gini are significantly positive in all cases regardless of the estimation period and whether the wage ratio is constant or varies over time. The exception is the period 1980-2018, in which the coefficient of $Gini^{Kuz,F}$ remains significantly positive while the coefficient of $Gini^{Kuz,V}$ becomes negative (the results are not shown). This degenerate result is obtained because in the post-1980 period the $Gini^{Kuz,V}$ is close to zero, reflecting the fact that wages in industry and agriculture had converged. After 1980, industrialization has been increasingly characterized by automation of low-skilled tasks, leading to increasing inequality and declining relative wages for low-skilled manufacturing workers (Prettner and Strulik, 2020). Thus, the post-1980 results have no bearing on the validity of the Kuznets curve.

The most important results are presented in columns (2) and (8), in which the estimation period is condensed to the years 1880–1980, during which the Kuznets curve has potentially the strongest effect on the overall Gini. Outside this period, the Kuznets Gini is close to zero and, therefore, should not have significant explanatory power for the overall Gini. In the following discussion, we focus on the 'core industrialization period' 1880-1980.

Table 1: Determinants of the Gini-Coefficient: 1800-2018

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
$Gini^{Kuz,V}$							0.21[0.11]	0.36[0.23]	0.35[0.22]	0.34[0.20]	0.33[0.20]
							(0.032)	(0.039)	(0.039)	(0.042)	(0.042)
$Gini^{Kuz,F}$	0.93[0.29]	0.97[0.30]	0.94[0.30]	0.84[0.27]	0.88[0.28]						
	(0.052)	(0.039)	(0.076)	(0.076)							
$Top1$	0.18[0.10]	0.13[0.08]	0.14[0.09]	0.13[0.08]			0.22[0.11]	0.10[0.06]	0.12[0.07]	0.11[0.07]	
	(0.033)	(0.040)	(0.040)	(0.041)			(0.034)	(0.040)	(0.040)	(0.040)	
S^K	0.04[0.06]	-0.01[0.01]	-0.01[0.01]	0.02[0.02]			0.01[0.01]	-0.06[0.06]	-0.03[0.05]	-0.06[0.14]	
	(0.009)	(0.039)	(0.025)	(0.023)			(0.010)	(0.025)	(0.024)	(0.024)	
$Growth$			0.004[0.01]						0.005[0.02]		
			(0.007)						(0.008)		
$Open$			0.02[0.09]						0.03[0.12]		
			(0.005)						(0.005)		
$\log y$				-0.03[0.08]	-0.03[0.08]	-0.03[0.09]				-0.06[0.14]	-0.05[0.14]
				(0.009)	(0.009)	(0.008)				(0.009)	(0.010)
$(\log y)^2$				-0.004[0.12]	-0.004[0.12]	-0.004[0.12]				-0.003[0.05]	-0.003[0.06]
				(0.001)	(0.001)	(0.001)				(0.001)	(0.001)
Est. Per	1800-2019	1880-1980	1880-1980	1880-1980	1880-1980	1880-1980	1800-2019	1880-1980	1880-1980	1880-1980	1880-1980
Obs.	3740	1717	1717	1717	1717	1717	3740	1717	1717	1717	1717

Notes. The dependent variable is the overall Gini. The figures in soft parentheses are standard errors robust to heteroscedasticity and serial correlation. The figures in square brackets are the absolute values of the partial correlation coefficients. *Growth* = 20-year growth rate in per capita GDP; *Open* = openness ((export + import)/GDP; S^K = the capital share; $Top1$ = top 1% income share; $Gini^{Kuz,F}$ = Kuznets Gini with the fixed industry wage premium; and $Gini^{Kuz,V}$ = Kuznets Gini with the time-varying industry wage premium; y = per capita income.

In terms of statistical and economic significance, the $Gini^{Kuz,F}$ has a larger impact on the overall Gini than the $Gini^{Kuz,V}$, which can also be seen from the partial correlation coefficients given in square brackets. In principle, since the model with the varying wage premium uses a more exact measure of income inequality, it should explain the overall Gini better than the model with the fixed wage premium. However, the coefficient of $Gini^{Kuz,V}$ is likely biased towards zero due to a negative correlation between $Gini^{Kuz,V}$ and the errors terms. As mentioned above, the overall Gini is, in many cases before 1950, interpolated between the world wars and between census years from which the historical data were derived. This means that the wage-induced fluctuations in $Gini^{Kuz,V}$ will not be echoed by the overall Gini but will be absorbed in the residuals. By contrast, the $Gini^{Kuz,F}$ is relatively smooth and has a within-country standard deviation that is 60% smaller than that of $Gini^{Kuz,V}$ in our focus period, 1880-1980.

The coefficients of the top 1% income share are significantly positive regardless of the estimation period and the partial correlation coefficients are sufficiently high to conclude that the top 1% income share is an important determinant of the overall Gini coefficient. The coefficients of the capital share are insignificant in almost all regressions, presumably because factor share effects are compounded by a negative correlation between the capital share and the ratio of unskilled and agricultural wages. From these results, however, it cannot be concluded that the capital share is an unimportant determinant of income inequality due to the measurement issues associated with the estimates of the capital share, as discussed in the previous section. Finally, the coefficients of per capita income growth are positive but not statistical significant. The coefficients of trade openness are significantly positive, in agreement with the prediction that the advanced countries' specialization in production of high-skill intensive goods causes inequality to increase along with increasing trade openness.

In terms of economic significance, the Kuznets Gini has been influential for the evolution of inequality over the period 1880-1980. Over this estimation period, the partial correlation coefficients are on average 0.30 ($Gini^{Kuz,F}$) and 0.21 ($Gini^{Kuz,V}$), suggesting that the Kuznets Gini explains a substantial part of the variance in the overall Gini. In comparison, the partial correlation coefficient of the Top 1% share is on average 0.08. A one standard deviation increase in $Gini^{Kuz,V}$ and $Gini^{Kuz,F}$ is associated with an 2.8 and 5.2 percentage point increase in the overall Gini, respectively. From these results it can be concluded that the Kuznets Gini has been influential for the evolution of inequality over the past two centuries.

4.3. **The Kuznets Gini and the Conventional Kuznets Curve.** The finding that the Kuznets Gini has been a important determinant of the overall Gini coefficient, begs the question of why the Kuznets hypothesis is so often rejected in the empirical literature. To check the results from the conventional analysis, we regress the overall Gini against the log and the squared log of per capita income, $\log y$ and $(\log y)^2$, with and without confounders in columns (4)-(6) and (10)-(11), over the period 1880-1980. The coefficients of $\log y$ are $(\log y)^2$ are both significantly negative at the 1% level, showing that the conventional Kuznets curve is not supported by the data. This result is unaltered for the estimation period 1800-2019 (the results are not shown). Thus, instead of a hump-shaped evolution, inequality declines monotonously with economic development measured by per capita income. Considering the results from the 'horse race regressions' in columns (5) and (11), we observe that the coefficients on income remain negative with small partial correlation while for the Kuznets Gini the size of the positive coefficient and the relatively large partial correlation with overall inequality remain robust to including the income terms.

5. Conclusion

In this paper we examine the original Kuznets hypothesis and show that the Kuznets Gini has been a significant determinant of overall income inequality in the advanced countries over the past two centuries. We find that the Kuznets Gini explains exactly what it is meant to explain: inequality increases in the first stage of industrialization and begins to decrease before industrialization reaches maturity. This conclusion is corroborated by the high partial correlation coefficients of the Kuznets Gini in our estimates where overall income inequality is regressed on the Kuznets Gini, the capital income share, and the top 1% income share.

Our analysis deviates from the literature that tends to interpret the validity of the Kuznets curve in terms of the shape of the overall Gini coefficient. Our interpretation of Kuznets' inequality hypothesis is that it is valid for any country that undergoes industrialization provided that industrial wages exceed agricultural wages. The absence of a clear relationship between the Kuznets curve and overall income inequality does not invalidate the Kuznets curve, but reflects omission of relevant confounders from the model. Using the conventional approach in which the overall Gini coefficient is regressed on the level and the squared log of per capita income, we find that the Kuznets hypothesis is rejected regardless of estimation period and the inclusion of confounders. The contribution of the Kuznets Gini to the overall inequality remains robust to the inclusion of the income terms. We therefore conclude that the hump of industrialization-driven inequality, in contrast to the conventional Kuznets curve, is a robust phenomenon of economic development.

While Kuznets' mechanism of industrialization-driven inequality became irrelevant for the group of advanced countries after the 1980s, it remains highly relevant for developing countries today. Assuming that the wage ratio between the industrial and agricultural sector is approximately equal to that of our sample, the Kuznets Gini is still on the upward trajectory for most developing countries. Kuznets' industrialization mechanism, therefore, remains important for the understanding of current and future income inequality for a large subset of countries where the ratio between industrial to agricultural employment is still low.

References

Acemoglu, D., and Robinson, J. A. (2002). The political economy of the Kuznets curve. *Review of Development Economics* 6(2), 183-203.

Acemoglu, D., and Restrepo, P. (2022). Tasks, automation, and the rise in US wage inequality. *Econometrica* 90(5), 1973-2016.

Adao, R., Carrillo, P., Costinot, A., Donaldson, D., and Pomeranz, D. (2022). Imports, exports, and earnings inequality: Measures of exposure and estimates of incidence. *Quarterly Journal of Economics* 137 (3), 1553-1614.

Aghion, P., Caroli, E., and Garcia-Penalosa, C. (1999). Inequality and economic growth: The perspective of the new growth theories. *Journal of Economic Literature* 37(4), 1615-1660.

Anand, S., and Kanbur, S. R. (1993). The Kuznets process and the inequality–development relationship. *Journal of Development Economics* 40(1), 25-52.

Banerjee, A. V., Benabou, R., and Mookherjee, D. (2006) Introduction and overview. In: Banerjee, A. V., Benabou, R., and Mookherjee, D. (eds.). *Understanding Poverty*, xiii-li. Oxford University Press.

Barro, R. J. (2008). Inequality and growth revisited. Working Paper Series on Regional Economic Integration No. 11. Asian Development Bank.

Batuo, M. E., Kararach, G., and Malki, I. (2022). The dynamics of income inequality in Africa: An empirical investigation on the role of macroeconomic and institutional forces. *World Development* 157, 105939.

Chang, S., Clance, M. W., Gupta, R., and Gozgor, G. (2021). A Reconsideration of Kuznets curve across countries: Evidence from the co-summability approach. *Journal of Developing Areas* 55(2).

Checchi, D., and Garcia-Penalosa, C. (2010). Labour market institutions and the personal distribution of income in the OECD. *Economica* 77(307), 413-450.

Chu, A.C., and Peretto, P. (2019). Innovation and inequality from stagnation to growth. University of Liverpool Working Papers in Economics 201910.

Clark, G. (2005). The condition of the working class in England, 1209–2004. *Journal of Political Economy* 113(6), 1307-1340.

Deininger, K., and Squire, L. (1996). A new data set measuring income inequality. *World Bank Economic Review* 10(3), 565-591.

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

Grossman, G. M., and Helpman, E. (2018). Growth, trade, and inequality. *Econometrica* 86(1), 37-83.

Helpman, E. (2004). *The Mystery of Economic Growth.* Cambridge, MA: Harvard University Press.

ILOSTAT (2023). https://ilostat.ilo.org/ Accessed 3.1.2023.

Kanbur, R. (2000). Income distribution and development. In: Atkinson, A.B. Bourguignon, F. (eds.), *Handbook of Income Distribution* Volume 1, 791-841. Amsterdam: North-Holland.

Kerspien, J. A. and J. B. Madsen (2023). Markups, Tobin's q and the declining labor share. *Journal of Money Credit and Banking*, forthcoming.

Kuznets, S. (1955). Economic growth and income inequality. *American Economic Review* 65, 1-28.

Kuznets, S. (1963). Quantitative aspects of the economic growth of nations: VIII. Distribution of income by size. *Economic Development and Cultural Change* 11(2, Part 2), 1-80.

Madsen, J. B. (2001). Agricultural crises and the international transmission of the great depression. *Journal of Economic History* 61(2), 327-365.

Madsen, J. B., Minniti, A. and Venturini, F. (2021). Wealth inequality in the long run: A Schumpeterian growth perspective. *Economic Journal* 131(633), 476-497.

Moatsos, M., Baten, J., Foldvari, P., van Leeuwen, B., and van Zanden, J. L. (2014). Income inequality since 1820. In Moatsos, M., Baten, J., Foldvari, P., van Leeuwen, B., and van Zanden, J. L. (eds.), *How Was Life?: Global Well-being since 1820*, 199-215. OECD Publishing, Paris.

Piketty, T., Saez, E., and Zucman, G. (2018). Distributional national accounts: methods and estimates for the United States. *Quarterly Journal of Economics* 133(2), 553-609.

Prettner, K., and Strulik, H. (2020). Innovation, automation, and inequality: Policy challenges in the race against the machine. *Journal of Monetary Economics* 116, 249-265.

Thomas, M. (1991). The evolution of inequality in Australia in the nineteenth century. In: Brenner, Y.S., Kaelble, H., Thomas, M. (eds.), *Income Distribution in Historical Perspective*, 149-173. Cambridge, UK: Cambridge University Press.

WDI (2023). https://databank.worldbank.org/source/world-development-indicators. Accessed on 5.1.2023.

Williamson, J. G. (1991). British inequality during the industrial revolution: Accounting for the Kuznets curve. In: Brenner, Y.S., Kaelble, H., Thomas, M. (eds.), *Income Distribution in Historical Perspective*, 57-75. Cambridge, UK: Cambridge University Press.

World Income Inequality Database (2022). WID https://wid.world/. Assessed on 22 November 2022.