Stages of Diversification Redux

Hoda Assem, Roberta Gatti, Daniel Lederman Office of the Chief Economist, MENA

3rd Annual Central Bank Conference on Development Economics in the MENA region Morocco, 14 September 2023

Outline

- 1. Motivation
- 2. Literature and Contribution
- 3. Data:
 - a) Dimensions and Measures of Diversification
 - b) Definition of Resource Endowments
- 4. Results:
 - a) Lowess Estimation
 - b) Sources of Variation
 - c) Quadratic Estimation
 - d) Revisiting Imbs and Wacziarg
 - e) Diversification in Resource-Poor and Resource-Rich
- 5. Conclusion

Diversification and Economic Volatility



Literature and Contribution

- Theoretical models describing the relationship between diversification and economic growth can be classified into two sets of models:
 - A negative relationship based on Neoclassical Ricardian or Heckscher-Ohlin trade theory.
 - A positive relationship (Acemoglu and Zilibotti, 1997).
- Empirical Studies show a U-shaped relationship
 - Imbs and Wacziarg (2003), Klinger and Lederman (2004; 2006), Hesse (2008) and Cadot et al. (2011).

Literature and Contribution

>The contribution of this paper to the literature is fourfold:

- 1. Updates our understanding of diversification patterns across countries and over time to 151 countries over 57 years.
- 2. Investigates whether observed patterns are driven by cross-country characteristics or by within-countries trajectories.
- 3. Expand the study of the relationship between diversification and economic development to the services sector.
- 4. Explores whether diversification patterns differ across resource-rich and resource-poor countries.

Dimensions of Economic Diversification: Employment and Value-added

UNIDO's Industrial Statistics Databases (INDSTAT) provides disaggregated data on the manufacturing sector.

>INDSTAT databases provide data on:

- Employment (Number of employees)
- Value added

> Data is provided at two levels of disaggregation:

- INDSTAT 2 comprises 23 sectors at the 2-digit level of ISIC Rev.3.
- INDSTAT 4 comprises 137 sectors at the 4-digit level of ISIC Rev.3.

Coverage of INDSTAT datasets varies considerably across countries.

Dimensions of Economic Diversification: Exports

> Merchandise Export Data from UN COMTRADE database:

- Dataset covers 215 countries for 58 years (from 1962 to 2019).
- 61 2-digit sectors and 625 4-digit sectors at SITC Rev1. classification.

Services Exports Data

- From WTO statistical portal.
- Two datasets are available:
 - 1. From 1980 2013: 22 sectors using EBOPS 2002 classification.
 - 2. From 2005 to 2021: 94 sectors using EBOPS 2010 classification.
 - 3. The two classifications concorded using Wettstein et al. (2019) and Brochert et al. (2020).
 - 4. The resulting dataset includes 158 countries and 13 sectors for a period of 40 years (1980 2019).

Measures of Diversification

Gini =
$$1 - \frac{1}{N} \sum_{k=1}^{n} (X_k + X_{k-1})$$

Where $x_k = \sum_{l=1}^k s_l$ represents the cumulative sector shares and $s_k = \frac{x_k}{\sum_{k=1}^N x_k}$ is the share of sector k in total employment/value-added and N is the total number of sectors (omitting country and time subscripts).

$$\text{HHI} = \frac{\sum_{k=1}^{n} s_k^2 - \frac{1}{N}}{1 - \frac{1}{N}}$$

Where $s_k = \frac{x_k}{\sum_{k=1}^n x_k}$ is the share of sector k in total employment / value added and n is the total number of sectors (omitting country and time subscripts).

Definition of Resource Endowments

- How to define natural resources?
 - Definition includes fuel and ores and metals.
- How to identify resource-rich countries (comparative advantage)?
 - Resource abundance is measured as the ratio between natural resources net exports and working age population.
 - This measure of resource abundance captures a country's comparative advantage in natural resources consistent with HO (Lederman and Maloney, 2003).
 - A resource-rich country is on average a net exporter of natural resources.
- Resource-rich countries divided based on their level of resource endowments into two groups:
 - Highly resource-rich countries are above median net exporters of natural resources per worker.
 - Moderately resource-rich countries are below median net exporters of natural resources per worker.



Non-Parametric "LOWESS" Estimation

The Gini index follows a U-shaped pattern in line with the literature on diversification and economic development (Imbs and Wacziarg, 2003; Klinger and Lederman, 2004 and 2006; and Cadot et al.,2011).

There are two stages of diversification: Countries first diversify then they start specializing at around PPP GDP per capita around 20,000 (constant 1985 US dollars).



This figure presents the estimated Non-Parametric "lowess" Curve for the Gini index for employment, value-added using UNIDO 2-digit ISIC Rev.3 data as well as exports using UN COMTRADE 2-digit SITC Rev.1 Data (Authors' calculations).



Non-Parametric "LOWESS" Estimation— Services Exports

Re-specialization in export services occurs at a higher level of GDP per capita compared with merchandise exports. However, the data is classified into only 13 sectors which is more aggregated than the other datasets used.



This figure presents the estimated Non-Parametric "lowess" Curve for the Gini index for exports in services using data from WTO statistical portal (Authors' calculations).

Breaking-down Sources of Variation

The change in the patterns of economic diversification (U-shaped pattern) across the development path is mainly driven by betweencountry rather than within-country variation.

Pooled, Within and Between

Value-added (2-digit)

Exports (2-digit)



This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UNIDO 2-digit Employment Data (Authors' calculations).

Quadratic Estimation: Within

• Fixed-Effects Regression of Sectoral Concentration on Income and Income Squared

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All Linear	All Quadratic	Resource-Poor Linear	Resource-Poor Quadratic	Resource-Rich Linear	Resource-Rich Quadratic
GDP per capita	0.00071***	0.00054***	0.00122***	-0.00584***	0.00037**	-0.00017
	(0.00013)	(0.00021)	(0.00020)	(0.00052)	(0.00017)	(0.00036)
GDP per capita square		0.00000		0.00022***		0.00000*
		(0.00000)		(0.00002)		(0.00000)
Constant	0.60639***	0.60740***	0.59386***	0.62422***	0.62756***	0.63086***
	(0.00125)	(0.00160)	(0.00176)	(0.00270)	(0.00195)	(0.00277)
Observations	4,970	4,970	3,393	3,393	1,577	1,577
R-squared	0.00648	0.00668	0.01165	0.07075	0.00306	0.00489
Number of iso3c1	140	140	93	93	47	47

Figure

Quadratic Estimation: Between

Between Regressions of Sectoral Concentration on Income and Income Squared (with year fixed effects)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All	All	Resource-Poor	Resource-Poor	Resource-Rich	Resource-Rich
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
GDP per capita	-0.00434***	-0.01185***	-0.00670***	-0.01521***	-0.00232	-0.01754
	(0.00115)	(0.00197)	(0.00178)	(0.00499)	(0.00294)	(0.01466)
GDP per capita square		0.00019***		0.00036*		0.00043
		(0.00004)		(0.00020)		(0.00041)
Constant	-3.66375*	-2.77362	-7.74183***	-6.66882**	-0.10016	0.05693
	(1.97675)	(1.79207)	(2.53702)	(2.53674)	(6.26328)	(6.21142)
Observations	4,970	4,970	3,393	3,393	1,577	1,577
R-squared	0.71510	0.77155	0.88303	0.89216	0.88934	0.90677
Number of iso3c1	140	140	93	93	47	47

Revisiting Imbs and Wacziarg (2003): S=796, J=10,000\$, Δ =50\$, N=4,929

Experimenting with different bandwidths doesn't qualitatively change the results.



This figure presents the estimated Non-Parametric Curve for the Gini index – UNIDO 2-digit Employment Data (authors' calculations).

Lowess vs Semiparametric

Revisiting Imbs and Wacziarg (2003): S=796, J=10,000\$, ∆=50\$, N=4,929

The first graph shows the number of observations within each subsample. The number of observations is lower for subsamples with higher income.



This figure presents slope estimates for Gini index – UNIDO 2digit Employment Data (authors' calculations).

Revisiting Imbs and Wacziarg (2003): S=796, J=10,000\$, Δ =50\$, N=4,929

Intercept estimates drive most of the U-shaped pattern in diversification.



This figure presents the intercept estimates for the Gini index – UNIDO 2-digit Employment Data (authors' calculations).

Patterns of diversification for resource-rich and resource-poor countries

Gini follows a similar pattern in resource-poor countries: Countries first diversify then they start specializing at the same turning point.

Resource-rich countries reach a plateau at around PPP GDP per capita of 10,000 (constant 1985 US dollars) and start specializing again later at a higher-turning point.



Average Gini

<u>HHI</u>

This figure presents the estimated Non-Parametric "lowess" Curve for the Gini index for employment, value-added using UNIDO 2-digit Rev3. data as well as exports using UN COMTRADE 2-digit SITC Rev1. Data (Authors' calculations).

Patterns of diversification for resource-rich and resource-poor countries

Three observations for resource-rich countries:

- 1. Gini of exports is always higher than Gini for value-added and employment.
- 2. After the turning point, re-concentration in resource-rich countries is higher in exports and value-added than in employment.
- 3. Economic concentration in resource-rich countries is higher than in resource-poor countries.

Conclusion

- This paper investigates the diversification of economies over the past 60 years with a special focus on trajectories of resource-rich and resource-poor countries across three dimensions: employment, value-added, and exports.
- We estimate non-parametrically a U-shaped curve between measures of economic concentration and per capita income levels.
- The evidence suggests that these patterns are driven by between-country rather than within-country variation.
- Diversification patterns also differ across resource-rich and resource-poor countries.

Thank You

Appendix

Coverage of UNIDO

- Coverage in terms of years, sectors, variables (employment or value-added) varies considerably across countries and levels of disaggregation (2-digit or 4-digit) depending on data availability.
- As a first step, we harmonize the number of sectors for all countries and years by adding the missing sectors and assigning them zero values.
- The dataset remains unbalanced in terms of countries and years.
- INDSTAT2 includes 172 countries for the period 1963-2019 (57 years) and 23 manufacturing sectors. The average number of years per country is 34 years with a minimum of 1 year and a maximum of 57 years.
 - <u>Figure 1</u> shows, for each country, the number of years for which employment data at the 2-digit level is available.
 - Figure 2 shows, for each country, the number of years for which value-added data at the 2-digit level is available.
- INDSTAT4 includes 143 countries for the period 1990-2018 (29 years) and 127 manufacturing sectors. The average number of years per country is 12 years with a minimum of 1 year and a maximum of 27 years.
 - Figure 3 shows, for each country, the number of years for which employment data at the 4-digit level is available.
 - Figure 4 shows, for each country, the number of years for which value-added data at the 4-digit level is available.

<u>Back</u>

Figure 1: Coverage of Employment at 2-digit level per country for the period (1963-2019)

Series1

Number of Years Per Country



Figure 2: Coverage for Value-added at 2-digit level per country for the period (1963-2019)

Number of Years by Country





Figure 3: Coverage for Employment at 4-digit level per country for the period (1990-2019)





Figure 4: Coverage for Value-added at 4-digit level per country for the period (1990-2019)

Number of Years by Country





Advantages and Limitations of INDSTAT

• Advantage of INDSTAT:

- UNIDO is the longest-running dataset with data on cross-country production.
- UNIDO allows comparability over time and across countries.
- Limitations of INDSTAT :
 - UNIDO covers only the manufacturing sectors: no information on services.
 - Considerable variation in data coverage across countries, years and sectors.
 - UNIDO doesn't include data on the informal sector:
 - Informality is higher in low-income countries and in labor intensive sectors.
 - Price variations: Value-added is presented in local currency current prices.

Correlation Between Income per capita and Data Availability

- Higher income countries have more observations in the dataset.
- The dataset includes 101 resource-poor countries and 50 resource-rich countries. Average number of years for a resourcepoor country is 37.6 with a minimum of 6 and a maximum of 57. Average number of years for a resource-rich country is 34.3 with a minimum of 5 and a maximum of 57.



Correlation Between Income per capita and Data Availability

Dependent Variable = Number of observations	
Log GDP per capita	6.358***
	(1.350)
Resource-rich	13.928
	(18.916)
Log GDP per capita* Resource-rich	-2.009
	(2.265)
Constant	-14.537
	(11.363)
Observations	141

Observations	141
R-squared	0.177

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Measures of Diversification: Gini Index

Gini index:

$$Gini = 1 - \frac{1}{N} \sum_{k=1}^{n} (X_k + X_{k-1})$$

- Where $x_k = \sum_{l=1}^k s_l$ represents the cumulative sector shares and $s_k = \frac{x_k}{\sum_{k=1}^N x_k}$ is the share of sector k in total employment / value added and N is the total number of sectors (omitting country and time subscripts).
- The Gini coefficient ranges between 0 and 1. The closer the Gini coefficient is to 1, the more unequal the distribution is.
- A Gini of zero points to a homogeneously diversified economy where all sectors have an equal share of total employment / value-added.
- A Gini of 1 means the economy is fully concentrated and all employment / valueadded is generated by a single sector.

Measures of Diversification: HHI Index

Herfindahl index:

$$HHI = \frac{\sum_{k=1}^{n} s_k^2 - \frac{1}{N}}{1 - \frac{1}{N}}$$

- Where $s_k = \frac{x_k}{\sum_{k=1}^n x_k}$ is the share of sector k in total employment / value added and n is the total number of sectors (omitting country and time subscripts).
- The Herfindahl index is normalized to range between zero and one. The closer the index is to 1, the more unequal the distribution.
- Other measures include coefficient of variation, max-min spread, log variance of sector shares and the Theil index.

Summary statistics

Measures of Diversification: Gini vs HHI

- Gini index is a measure of inequality of a distribution. It measures how the actual distribution differs from an equal distribution across sectors.
- Gini is affected by adding sectors with zero values and increases as the number of sectors increase.
- HHI is a measure of market concentration or market power.
- HHI is more affected by sizes of sectors than number of sectors and gives larger weight to larger sectors.
- HHI is unaffected by the number of sectors with zero values.

Classification of countries by level of natural resources

Classification of Resource-Rich Countries

highly resource-rich moderately resource-rich resource-poor





GDP per capita over time by level of natural resources

PPP GDP per Capita (constant 1985\$)	Mean	Median	Minimum	Maximum
Resource-Poor (N=100)	6838.003	6023.79	375.7068	27286.78
Resource-Rich (N=51)	8340.347	12773.2	564.1001	72442.78
Moderately Resource-Rich (N=26)	2699.976	2014.04	564.1001	7592.511
Highly Resource-Rich (N=24)	14471.19	16391.5	1801.431	72442.78



Nonparametric estimation using LOWESS

- LOWESS refers to locally weighted scatter plot smoothing. Smoothed values are
 obtained by running a regression of the dependent variable on the independent
 variable for each observation and a few of the data points near it.
- Differences between LOWESS and estimation methodology of Imbs and Wacziarg (2003):
 - Lowess perform a simple OLS regression of y on x. Imbs and Wacziarg (2003) perform a fixed effects regression of y on x using country fixed effects.
 - Lowess is more computationally intensive, performing N regressions for N observations. Imbs and Wacziarg (2003) reduce the number of regressions by using increments of 25\$.
 - Lowess typically uses 80% of observations for each regression. Imbs and Wacziarg (2003) use an overlap of 5,000\$ (30% of observations on average).
 - Lowess performs a weighted regression for each observation so that points further away from the observation receive less weight. Imbs and Wacziarg (2003) use equal weighing of observations in each subsample.


Nonparametric estimation using LOWESS

HHI index confirms the U-shaped pattern: Countries first diversify then they start specializing at around PPP GDP per capita 20,000 (constant 1985 US dollars).

This figure presents the estimated Non-Parametric "lowess" Curve for the HHI index for employment, value-added using UNIDO 2-digit ISIC Rev.3 data as well as exports using UN COMTRADE 2-digit SITC Rev1. Data (Authors' calculations).



Different Levels of Disaggregation



Different Levels of Disaggregation

- The level of disaggregation used does not affect the shape of the curve, but it affects the level of the curve.
- The effect is more important for the Gini index because the Gini index is affected by the number of sectors including those with zero values.
- The higher the level of disaggregation, the higher the Gini index is because there are more sectors with zero values and therefore more inequality in the distribution.
- HHI is also affected by the higher level of disaggregation but to a smaller extent. The higher the level of disaggregation, the lower the HHI because there is less market power/market concentration.



Breaking-down Sources of Variation

• The estimated curve for the pooled relationship corresponds to "smoother" nonparametric regressions (lowess) of Gini on GDP per capita.

$$y_{it} = \alpha + \beta x_{it} + \epsilon_{it}$$

 The estimated curve for the between-country relationship corresponds to "smoother" non-parametric regressions (lowess) of average Gini on average GDP per capita.

$$\bar{y}_i = \alpha^{BE} + \beta^{BE} \,\bar{x}_i + \bar{\varepsilon}_i$$

• The estimated curve for within-country relationship corresponds to "smoother" non-parametric regressions (lowess) of demeaned Gini on demeaned GDP per capita.

$$y_{it} - \bar{y}_i = \boldsymbol{\beta}^{FE} \left(x_{it} - \bar{x}_i \right) + \left(\varepsilon_{it} - \bar{\varepsilon}_i \right)$$

• Change in level of economic diversification across the development path is mainly driven by between-country rather than within-country variation.

Employment (2-digit)

Breaking-down Sources of Variation

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UNIDO 2-digit value-added Data (Authors' calculations).



Employment (2-digit)

Breaking-down Sources of Variation

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UN COMTRADE 2digit exports Data (Authors' calculations).



Employment (2-digit)

Within versus Between Variation: Within



Figure . Within-country relationship for Gini index – UNIDO 2-digit Employment Data (authors' calculations). The estimated curve corresponds to the OLS estimation of: $Y_{it} = \alpha + v_i + \beta GDP \ per \ capita + \gamma GDP \ per \ capita^2 + e_{it}$

Within: Resource-Poor



Figure . Within-country relationship for Gini index – UNIDO 2-digit Employment Data (authors' calculations). The estimated curve corresponds to the OLS estimation of: $Y_{it} = \alpha + v_i + \beta GDP \ per \ capita + \gamma \ GDP \ per \ capita^2 + e_{it}$

Within: Resource-Rich



Figure . Within-country relationship for Gini index – UNIDO 2-digit Employment Data (authors' calculations). The estimated curve corresponds to the OLS estimation of: $Y_{it} = \alpha + \nu_i + \beta GDP \ per \ capita + \gamma \ GDP \ per \ capita^2 + e_{it}$

Within versus Between Variation : Between



Figure . Between-country relationship for Gini index – UNIDO 2-digit Employment Data (authors' calculations). The estimated curve corresponds to the OLS estimation of: $\overline{Y_i} = \alpha + \beta \ \overline{GDP} \ per \ capita_i^2 + \gamma \ \overline{GDP} \ per \ capita_i^2 + e_i$

Between: Resource-Poor



Figure . Between-country relationship for Gini index – UNIDO 2-digit Employment Data (authors' calculations). The estimated curve corresponds to the OLS estimation of: $\overline{Y_i} = \alpha + \beta \ \overline{GDP} \ per \ capita_i^2 + \gamma \ \overline{GDP} \ per \ capita_i^2 + e_i$

Between: Resource-Rich



Figure . Between-country relationship for Gini index – UNIDO 2-digit Employment Data (authors' calculations). The estimated curve corresponds to the OLS estimation of: $\overline{Y}_i = \alpha + \beta \ \overline{GDP} \ per \ capita_i + \gamma \ \overline{GDP} \ per \ capita_i^2_i + e_i$

Comparison between Resource-Rich and Resource-Poor

Resource-rich countries are on average more specialized both in employment and value-added compared to resource-poor countries. The difference between the mean of the two groups is statistically significant.

		Mean	Std. dev.	Min	Max
Resource-Poor (N=101)	Number of years per country	37.60) 15.76	6	57
	Average Gini of employment	0.62	L 0.11	0.36	0.91
	Average Gini of value-added	0.63	L 0.12	0.36	0.96
Resource-Rich (N=50)	Number of years per country	34.30) 15.67	5	57
	Average Gini of employment	0.63	3 0.10	0.35	0.90
	Average Gini of value-added	0.67	0.11	0.41	0.93

Patterns of diversification for resource-rich and resource-poor countries

HHI follows a similar pattern in resource-poor countries: Countries first diversify then they start specializing at the same turning point.

Resource-rich countries reach a plateau at around PPP GDP per capita 10,000 (constant 1985 US dollars) and start specializing again later at a higher turning point.

Back



This figure presents the estimated Non-Parametric "lowess" Curve for the Gini index for employment, value-added using UNIDO 2-digit Rev3. data as well as exports using UN COMTRADE 2-digit SITC Rev1. Data (Authors' calculations).

Breaking-down dimensions of variation: Resource-Poor

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UNIDO 2-digit Employment Data (Authors' calculations).

Value-added (2-digit)





Breaking-down Sources of Variation: Resource-Poor

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UNIDO 2-digit Valueadded Data (Authors' calculations).



Breaking-down Sources of Variation: Resource-Poor

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UN COMTRADE 2digit Exports Data (Authors' calculations).



Breaking-down dimensions of variation: Resource-Rich

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UNIDO 2-digit Employment Data (Authors' calculations).

Value-added (2-digit)



Exports (2-digit)

Breaking-down Sources of Variation: Resource-Rich Value-added (2-digit) and GDP per capita (PPP) for Pooled (N=3.393) Between (N=93)

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UNIDO 2-digit Value-added Data (Authors' calculations).



Breaking-down Sources of Variation: Resource-Rich

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UN COMTRADE 2digit Exports Data (Authors' calculations).



Diversification Patterns in Highly Resource-Rich Countries

Diversification patterns in highly resource-rich countries follow a similar pattern to resource-rich countries.

Moderately resource-rich countries, on the other hand, are poorer than highly resource-rich countries specializing at a lower turning point.

This figure presents the estimated Non-Parametric "lowess" Curve for the HHI index for employment, value-added using UNIDO 2-digit Rev3. data as well as exports using UN COMTRADE 2-digit SITC Rev1. Data (Authors' calculations).



Break-downing dimensions of variation: Highly Resource-Rich

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UNIDO 2-digit Employment Data (Authors' calculations).

Employment (2-digit) and GDP per capita (PPP) for Resource-Rich countries Within (N=3,390) Pooled (N=3,393) Between (N=93) ∞ ω, 8 g G Gini 4 Gini с, 2 2 2 0 0 0 GDP per capita, PPP (constant 1985 GDP per capita, PPP (constant 1985) GDP per capita, PPP (constant 1985)

Value-added (2-digit)

Exports (2-digit)

Breaking-down Sources of Variation: Highly Resource-Rich

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UNIDO 2-digit Value-added Data (Authors' calculations).



Breaking-down Sources of Variation: Highly Resource-Rich

This figure presents the estimated Non-Parametric "lowess" Curve for Gini index – UN COMTRADE 2-digit Exports Data (Authors' calculations).

