# Unequal Global Convergence*

Shoumitro Chatterjee Elisa Giannone Tatjana Kleineberg
Kan Kuno Luca Looser

**July 2025**

**Abstract**

We revisit a classical question: what is the role of structural transformation in determining regional convergence? To do so, we construct a novel global dataset of regional GDPs and granular sectoral employment for more than 1500 regions and more than 90 countries, which starts in 1980 and covers a large range of income spectrum. We document three main facts. First, regional convergence within countries decreases over time around the globe and stalls in the most recent decade despite residual spatial inequality. Second, this decline in regional convergence is associated with structural transformation toward high-skill services. Third, high-skill service employment exhibits a higher regional concentration than other sectors, including other services, manufacturing, or agriculture. Through the lens of a spatial equilibrium model which embeds the standard drivers of structural change, we find a reinforcing interplay between structural change and spatial development. As an economy transforms toward high-skill services, regional convergence declines due to agglomeration effects in the high-skill service sector. Agglomeration effects increase economic growth which further accelerates structural change toward services and, in turn, widens unequal regional dynamics.

*Chatterjee: School of Advanced International Studies, The Johns Hopkins University, 1717 Massachusetts Ave NW, Room 743A, Washington, DC, USA 20036 (email: shoumitroc@jhu.edu). Giannone: Centre de Recerca en Economia Internacional, Ramon Trias Fargas, 25-27, Barcelona, Spain 08005 (email: elisa.giannonne@gmail.com). Kleineberg: World Bank. 1818 H Street, NW Washington, DC 20433 USA (email: tkleineberg@worldbank.org). Kuno: Department of Economics, The Pennsylvania State University, Kern Building, State College, PA, USA 16801 (email: kankuno@icloud.com). We thank Julieta Caunedo, Arnaud Costinot, David Cuberes, Dave Donaldson, Chris Papageorgiou, Marti Mestieri, Michael Peters, Tommaso Porzio, Richard Rogerson, Fabrizio Zilibotti and seminar participants at EIEF, EUI, Georgetown, Harvard/MIT, LSE, the Society of Economic Dynamics (UW Madison), Urban Economics Association (Washington DC), ABCDE (World Bank), EPED, STEG, CREI, U. Bocconi, U. Bologna, U. Toronto, Toronto Metropolitan, Warwick, Yale and York for their generous comments. We thank Elif Basaran and Xiaoqiao Chen for their excellent research assistance. We are grateful to Rafael Laporta, Nicola Gennaioli and Andrei Shleifer for guiding us through their original dataset and to the ECJRC and World Bank for providing access to European data and labor force survey data. Any errors are our own.

# 1 Introduction

The services sector is poised to be a central driver of economic growth in the 21st century. The twin forces of de-globalization in manufacturing and premature deindustrialization in many developing economies have strengthened the case for services-led development. Reflecting this shift, multilateral institutions such as the World Bank (Nayyar et al., 2021) and several economists (Rajan and Lamba, 2023) have advocated for services as a viable engine of growth in today's developing countries for which the traditional manufacturing-led growth does not seem possible.

This paper examines the spatial implications of service-led development. The link between structural transformation and regional income dynamics is a longstanding topic in macroeconomics, but the literature has largely focused on the historical shift from agriculture to manufacturing—the path which most of today's advanced economies followed. As today's developing countries chart a different course, a central question is whether services-led growth, especially if driven by high-skilled services, will spatially diffuse or concentrate in a few regions, and what this implies for the trade-off between aggregate growth and spatial inequality. These spatial patterns have implications not only for welfare but also for political economy, regional polarization and for growth itself.

Progress on this question has been constrained by the lack of harmonized, long-run data on subnational economic activity, especially in the developing world. We address this gap in the literature by assembling and validating a novel global dataset that tracks economic activity across more than 1,500 regions in 90 countries which span five continents and represent 80% of the world population. Time coverage spans from 1980 to 2019 for most countries and even further back for richer economies. Crucially, our final dataset includes data on regional GDP per capita and regional employment by narrow sub-sectors. This granularity of data allows us to decompose service employment into low-skill non-tradable services, high-skill public services, and high-skill private services—enabling a rich analysis of the spatial consequences of services-led development.

The paper has two main parts. In the first part, we construct and validate the data and, then, provide new empirical evidence on the structural transformation to services and regional convergence. *One of the main contributions of this paper is the development of a longitudinal dataset for regions within countries, which enables social scientists to analyze information on GDP, education, and granular sectoral employment and productivies at the regional level.* We start with the pioneer dataset of Gennaioli et al. (2014) which includes GDP and education data. We complement this data in important ways. First, we expand the coverage of regional GDP and education data of the initial data set for additional countries and the latest available years when possible by adding multiple data sources. Second, we supplement our dataset with city-level GDP data—particularly for Sub-Saharan Africa—purchased from The Economist Intelligence Unit. Third, we collect regional data on sectoral employment and wages across regions within countries and over time from national censuses, labor force surveys, statistical agencies and other sources, which complements the GDP and education information and allows us to include the evolution of regional productivity differences in the dataset. The core sample that we use in this paper covers fewer countries than some other papers because our main analysis requires time-consistent panel data between 1980 and 2019. Overall, our sample represents approximately 80% of world GDP and 66% of the world population. The data set has lower coverage of African countries, which is why we corroborate the findings with *The Economist* dataset. *We are actively planning to do field work in a selected set of African countries to extend the coverage of Sub-Saharian Africa.*

We establish the following three novel empirical facts. Regional income convergence within countries has slowed considerably between 1980 and the 2000s on average, but with substantial heterogeneity across countries (Fact 1). In some countries—such as the United States and India—regional convergence stalled altogether, while it still persists in others such as China. This fall in regional convergence is present in about half of the countries in our sample, which together represent approximately 70% of the sample population. We test for heterogeneity of our results in terms of size, continent, income level and OECD status of countries. We find evidence that low-income countries converge more than high-income ones after 1990. It is, however, evident that prior 1990, regional convergence was stronger for early developers. We find that these changes in spatial inequality are not systematically related to countries' income levels or growth rates. Instead, we find that a decline in regional convergence is closely associated with countries' structural transformation toward high-skill private services, even after controlling for country-specific heterogeneity and year fixed effects (Fact 2). In line with this finding, we document that employment in high-skill services is substantially more geographically concentrated than employment in agriculture, manufacturing or low-skill service sector (Fact 3). This pattern is particularly strong for less-developed countries, holds throughout the entire time period, and is robust to using different measures of regional concentration.

In the second part of this paper, we study these empirical facts through the lens of an economic geography and structural transformation's model. The model highlights a novel interplay between structural transformation and regional inequality.The model features three sectors: tradable high-skill services, non-tradable services, and non-services or goods that include agriculture and manufacturing. Firms in each sector use local labor as the only input for production. Local productivities in each sector evolve over time according to exogenous, sector-specific, and economy-wide growth rates. Productivity growth in the tradable high-skill service sector additionally depends on agglomeration effects which we model as a spillover on the sector's local employment. Individuals have non-homothetic preferences which allow them to shift expenditure shares towards high-skill and other services as their incomes rise. Individuals receive idiosyncratic preference shocks across regions and sectors and then choose their location and sector of employment. The model incorporates convergence and divergence forces. Productivity grows faster in the non-service-sector compared to low-skill services, which leads to regional convergence when combined with non-homothetic preferences (cf. Caselli and Coleman (2001)). Agglomeration effects in the tradable high-skill service sector instead lead to a spatial concentration of the sector's employment and act as a divergence force.

We calibrate our model and simulate it forward for 30 years to generate a path of structural transformation, growth, and spatial inequality. We then simulate counterfactuals to examine the mechanisms that drive these aggregate and spatial outcomes. Our current calibration and simulation results are only for illustrative purposes, as the more rigorous calibration to the data–which we describe below–is still in progress. We simulate three counterfactuals. The first eliminates agglomeration effects in the high-skill service sector and the second lowers exogenous productivity growth in high-skill services. The effects of these two counterfactuals go in the same direction but the effects are much more muted in the second one. Lower exogenous productivity growth in high-skill services reduces structural change toward the sector and lowers GDP per capita growth. However, agglomeration effects still increase the spatial concentration of high-skill service employment and spatial inequality overall, but to a smaller degree. In the third counterfactual, we eliminate income effects on the consumption side, which reduces overall structural change toward high-skill services due to constant sectoral expenditure shares but effects are smaller on regional income dynamics. These illustrative findings highlight that agglomeration effects, productivity growth in high-skill services, and income effects together generate a self-reinforcing interplay between regional inequality, structural transformation, and aggregate growth: As more workers move into high-skill services, the sector's productivity increases due to agglomeration effects, which increases spatial inequality and aggregate growth. Due to income effects, more growth again amplifies the employment shift towards high-skill services, generating a feed-back effect. Our results indicate a trade-off between regional disparities and faster aggregate structural transformation and growth.

As part of our ongoing work, we are improving the calibration and our quantitative analysis. We will use wage data by sector and region–which is available in many but not all labor force surveys–to calibrate initial productivity levels for each region and sector $A_{ij0}$. Given initial productivity levels, we can simulate the model forward to compute a path of sectoral employment, productivity, and GDP at the regional and aggregate level. We can then calibrate exogenous sectoral productivity growth $g_i$ and agglomeration effects in the high-skill service sector $\delta$ by targeting, first, data on country-wide sectoral productivity growth rates from the GGDC and ETD and, second, the spatial distribution of employment in the high-skill service sector from our data set. We can validate our calibration by ensuring that our model adequately fits regional convergence estimates over time. We plan to calibrate the model to data of the US, India, and Thailand, as these countries illustrate the main findings of our empirical results (cf. Figures 2 and 3 ). It will also allow us to ground our counterfactual analysis to these illustrative cases. For example, we can analyze how India's path of structural transformation, regional convergence, and aggregate growth would have changed if it had the same agglomeration effects or sectoral productivity rates as Thailand or the US.

**Related Literature** Our paper contributes to a growing literature on structural transformation and economic geography. In particular, recent work has studied the role of structural change in regional inequality. Caselli and Coleman (2001) and Eckert and Peters (2018) study how the structural transformation from agriculture to manufacturing increased regional convergence in the US from 1880 to 1920. Other papers study the implications of structural change for regional convergence in China (Hao et al., 2020), Spain (Budi-Ors and Pijoan-Mas, 2022) and France (Chen et al., 2023). Fan et al. (2022) show that service-led growth in consumer non-tradable service of India has created more inequality within the country and pushed for more growth. Bohr et al. (2024) introduces new tractable preferences for studying structural change and economic geography. In contrast to these papers, we center the analysis on high-skill high-productivity service growth. On top of that, this set of papers all focuses on a single country. Instead, we use subnational longitudinal data from many countries to document a systematic link between structural transformation towards high-skill private services and regional convergence. We further establish that this link holds independently of countries' development level which implies that regional convergence is declining for many of today's developing countries which are moving into services at relatively low development levels and without experiencing a strong industrialization phase Rodrik (2016). Overall, our findings add to the literature above by pointing out a new dichotomy in the role of structural transformation for spatial development.

Our paper also relates to a large macro development literature that studies the aggregate causes and implications of structural transformation as summarized by Herrendorf et al. (2014). Work by Buera and Kaboski (2012) has studied the role of services, and Huneeus and Rogerson (2020) studies the reasons behind premature deindustrialization. Our main contribution here is to add and study the spatial dimension and characterize the feedback effect of spatial differences on structural transformation and aggregate economic growth.

This paper also relates to a long empirical literature that studies regional convergence within and across countries, which was pioneered with the seminal work of Barro and Sala-i Martin 1992 and includes contributions from Sala-i-Martin 1996, Blanchard et al. 1992, Gennaioli et al. 2014, Ganong and Shoag 2017, Guriev and Vakulenko 2012. Gennaioli et al. (2014) study convergence between regions of the world in the cross-section. We augment and expand the dataset from Gennaioli et al. (2014) in coverage and by adding subnational detailed sectoral employment data and we then analyze the evolution of within-country convergence over time. These elements allow to a multitude of new studies relative to the new structural transformation to services and how this shapes geography.

Finally, this paper connects to a small number of studies that have produced remarkable collections of regional economic data. For example, Nordhaus (2006) assembled a dataset on a 1°×1° grid of per capita economic output, while Gennaioli et al. (2014) generated consistent subnational GDP and education data. Similarly, Smits and Permanyer (2019) compiled regional per capita for 161 countries spanning 1990 to 2017 and named the dataset DOSE. To date, to the best of our knowledge, DOSE represents the most extensive subnational dataset on economic output available, covering 1,660 regions in 83 countries with annual data from 1960 to 2020. It compiles gross regional product data from statistical agencies, the academic literature, and statistical yearbooks, and—distinctively—provides, in most cases, breakdowns of gross regional product by the three main sectors: agriculture, manufacturing, and services. Rossi-Hansberg and Zhang (2025) explores the most extensive geographic dataset covering GDP measures at a very fine geographic dataset for 2018-2020. Lagakos and Shu (2023) in their review article highlight the importance of micro data to make progress in understanding structural transformation. A central contribution of our study is the creation of a dataset with harmonized regional definitions over time, space, *and* economic variables. An extensive set of our data comes from individual and firm-level data, which allows us to capture broad heterogeneity. In this respect, this paper is related to Donovan et al. (2023) that construct a large dataset on labor markets spanning across many developing countries and providing an important resource for macro-development.

This paper is organized as follows. Section 2 reports the datasets used for the analysis. Section 3 reports the stylized facts we encounter in the data. Section 4 develops a model of structural transformation and economic geography to explain the patterns in the data. Section 5 concludes and highlights the work we are currently pursuing.

# 2 Data

We compile a unique dataset covering 1509 regions in 90 countries over an unbalanced panel from 1950 to 2019 for GDP and/or employment. Our preferred units of geography are the equivalent of states in the US or provinces in Italy. We make this choice of aggregation for two reasons. First, states or provinces are the finest spatial units for which data on GDP and sectoral employment are collected consistently across a broad range of countries. Second, states are crucial administrative borders and political decision-making units in most countries. Table 1 presents the geographic coverage of our core dataset that we use in this paper.[1] Compared to the full dataset, we restrict our core sample to countries which offer a balanced panel of regional GDP and sectoral employment data from at least 1990 to 2010. We require the long time series coverage to analyze regional income dynamics within countries. Our core sample comprises the 32 countries shown in the last column of Table 1. The core sample includes sizable samples in West Europe, East Europe, Asia, and North America and which represents 82% of world GDP. Despite our extensive data collection efforts, parts of Asia and Africa remain underrepresented.

We now briefly summarize the construction and validation of our data set, while further details are provided in the Data Appendix Section G.

**GDP Data.** We collect regional GDP data from several sources, which are shown in Table G.23 of Data Appendix Section G. For each country $i$, year $t$, and region $s$, we rescale the

[1]For a coverage of the full dataset, refer to the Data Appendix Section G.

Table 1: Summary table for GDP and employment data

| Region | GDP | | Employment | | Both | | |
|---|---|---|---|---|---|---|---|
| | Nb. Countries | 1990-2010 | Nb. Countries | 1990-2010 | Nb. Countries | Avg. Nb. Years | 1990-2010 |
| Africa | 3 | 3 | 17 | 8 | 2 | 44 | 0 |
| Asia | 12 | 9 | 14 | 9 | 11 | 45 | 8 |
| Australia and Oceania | 1 | 1 | 3 | 2 | 1 | 38 | 1 |
| East Europe | 16 | 5 | 13 | 4 | 13 | 30 | 1 |
| North America | 3 | 2 | 4 | 3 | 3 | 51 | 2 |
| South America | 6 | 5 | 18 | 11 | 6 | 43 | 4 |
| West Europe | 16 | 16 | 16 | 16 | 16 | 39 | 16 |
| **Total** | **57** | **41** | **85** | **53** | **52** | | **32** |

*Notes:* This table shows the number of countries that contain GDP and/or employment data as well as the number of average years per country in the unbalanced sample. The values are split by country groups. Author's calculation.

regional GDP per capita data to ensure that it aggregates to national measures of GDP per capita:

$$(\text{Regional GDP pc})_{ist}^{adjusted} = (\text{National GDP pc})_{it} \times \frac{(\text{Regional GDP share})_{ist}^{data}}{(\text{Regional population share})_{ist}}, \quad (1)$$

where national GDP and population data are taken from the Penn World Table version 10.0 (Feenstra et al., 2015). When regional population data is missing, we impute it with linear interpolation. When regional GDP per capita is missing, we interpolate it for each region with the following OLS regression:

$$(\text{Regional GDP per capita})_{ist} = \beta_0^s + \beta_1^s t + \beta_2^s (\text{National GDP pc})_{it} + u_{ist}, \quad (2)$$

where the predicted values are used to fill in the missing observations. We implement several data cleaning steps and consistency checks. For example, we exclude country-year observations where GDP per capita is missing for more than 10 consecutive years. Moreover, in cases where changes in data sources coincide with very high growth rates, we perform splicing to correct discontinuities. More details on the data cleaning process, adjustments for specific countries, regions, and time periods as well as data validation are reported in the Data Appendix.

Our dataset only includes three African countries, which we address in two ways. First, we use nightlight data to test the robustness of our results. Second, we purchased the dataset from *The Economist* , which has longitudinal data on GDP and population for 923 cities in 77 countries between 2004 and 2020.

**Sectoral Employment Data.** We collect sectoral employment data at the regional level from three main sources: Census micro-data from IPUMS (Ruggles et al., 2015, 2024), labor force survey micro-data from the World Bank Global Labor Database and i2d2 database, and regional data from the ARDECO database from the ECJRC (Auteri et al., 2024). To further increase data coverage, we collect data from national statistical agencies or other country-specific sources for Australia, China, Japan, South Korea, and the UK. Table G.24 in the Data Appendix provides the full list of data sources across all countries and time periods. To ensure comparability over time and across data sources, we standardize all geographic units at the state or province level. The census data from IPUMS provides regional identifiers that are harmonized over time (the "geolev1" variable). For the labor force surveys, we manually create regional crosswalks for each countries to map regions over survey-years and across different data sources. The ARDECO database provides standardized NUTS region identifiers for EU countries which are already harmonized over time and we choose the NUTS-level that corresponds most closely to the state-level. To merge sectoral employment and GDP data at the region level, we construct geographic crosswalks across these data sets., This harmonization adjusts, among others, for spelling variations and border changes and it might require the aggregation of several regions to ensure consistency over time and across data sources.

We classify sectoral employment into five sectors: agriculture, manufacturing, low-skill services, high-skill services with slow productivity growth, and high-skill services with high productivity growth. We choose the three categories within the service sector to account for the sector's large heterogeneity (Duarte and Restuccia, 2019). Low-skill services comprise, for example, wholesale and retail trade and transportation industries. High-skill services with low productivity growth include public administration, education and health, while financial and business services are classified as high productivity growth. In the rest of this paper, we therefore refer to high-skill services with low productivity growth as "high-skill public services" and those with high productivity growth as "high-skill private services". For each data source, we manually assign detailed industry codes to the five categories or we rely on previously harmonized and aggregated sub-categories when applicable. The detailed list of sectors in each category is listed in Appendix table G.26.

Our data cleaning procedures address irregularities such as abrupt, reversing changes and persistent shifts that deviate from national trends by removing problematic country-year observations and replacing them with interpolated values instead. For many countries, we combine multiple data sources to create the longest possible time series. In these cases, we choose a data source as the "primary" source if its sectoral employment share data has the smallest mean squared error relative to the WDI data at the national level. When combining data sources, we then adjust the levels of "non-primary" sources to avoid artificial discontinuities in the year where data sources change. The level-adjustment matches sectoral employment shares perfectly in an overlapping year and then uses sector-specific growth rates to adjust the rest of the time series from "non-primary" sources. After the data cleaning and merging, the final employment series is linearly interpolated over missing years and validated against the WDI data. Appendix Figure H.18 shows that out data set aligns closely with the WDI by plotting national employment shares from our data set against the counterparts from the WDI for agriculture, manufacturing and services. The close fit holds across all development levels. Deviations are largest for a couple of smaller island countries.

**Other Regional Indicators.** Additionally, our analysis incorporates a range of regional and country-level indicators to enrich our empirical investigation. To capture human capital, we employ data on years of schooling from Barro and Lee (2000). Measures of Free Trade Agreements and global market access from CEPII, along with road network information from the Global Roads Inventory Project (GRIP), serve as proxies for external and internal connectivity, respectively. To assess how political systems influence spatial patterns of economic growth, we use the democracy score from the Political-IV project. Recognizing that tropical countries have historically experienced poorer long-run economic performance for various reasons (Sachs, 2001, Acemoglu et al., 2001), we further incorporate long-run institutional and technological determinants—specifically, type of climate, distance to the coast, and ruggedness—from Nunn and Puga (2012). We add additional data from the GGDC Productivity Level Database (Inklaar and Timmer, 2008) and the Economic Transformation Database (Kruse et al., 2022) that have employment by sector at national level to complement and validate our regional sectoral indicators.

# 3 Novel Facts on Regional Convergence and Structural Transformation

In this section, we present a series of novel empirical findings on the evolution of regional income disparities within countries. A deeper understanding of these regional dynamics is essential for assessing not only the welfare implications for individuals but also the broader socio-political trajectories of nations. Our analysis focuses on two complementary measures of regional convergence, referred to as $\beta$- and $\sigma$-convergence.

To estimate the speed of $\beta$-convergence between regions within each country, we loosely follow the framework of Baumol (1986). To reduce the volatility of regional GDP over time, we implement an additional step that estimates the average GDP growth rate over a 10-year period starting in an initial period $t_0$ by estimating the following regression:

$$\log(GDP_{i,c,t}) = \alpha + \gamma_{i,c,t-t0}\,(t - t_0) + \varepsilon_t, \tag{3}$$

where $GDP_{i,c,t}$ denotes the per capita GDP of region $i$ in country $c$ at time $t$ and $t_0 = t - 10$ (i.e., the regression is performed over an 11-point—or equivalently, 10-year rolling—interval). The estimated coefficient $\hat{\gamma}_{i,c,t-t+10}$ therefore represents the average growth rate over the period from $t$ to $t + 10$. Subsequently, we estimate the following convergence regression for each country and initial period $t$:

$$\hat{\gamma}_{i,c,t+10-t} = \alpha + \beta_{c,t}\,\log(GDP_{i,c,t}) + \mathbf{X}'_{i,c,t}\gamma + \varepsilon_{i,c,t}, \tag{4}$$

where $\mathbf{X}_{i,c,t}$ is a vector of control variables (such as population and education), and the regression is weighted by the regional population at time $t$. A negative estimate of $\hat{\beta}_{c,t}$ indicates that poorer regions experienced faster growth than richer regions, implying convergence; conversely, a zero or positive coefficient suggests no convergence or divergence, respectively.

To assess $\sigma$-convergence, we use the Coefficient of Variation where the underlying dimensions are regions $i$ within a country $c$ at time $t$ and we study its change over time:

$$\Delta COV_{c,t} = \frac{\sigma_{c,t+10}}{\mu_{c,t+10}} - \frac{\sigma_{c,t}}{\mu_{c,t}}, \tag{5}$$

where $\sigma_{c,t}$ and $\mu_{c,t}$ denote the population-weighted standard deviation and mean of GDP per capita for country $c$ at time $t$, respectively. The $\Delta COV_{c,t}$ will be our alternative convergence measure that we will use in the text. As explained carefully by Sala-i-Martin (1996) and reiterated in Rodrik (2013), $\beta$-convergence is a necessary but not sufficient condition for $\sigma$-convergence. Even when $\beta$-convergence is present, countries may not converge in income levels if random shocks to growth are sufficiently large. In our sample of countries, as documented in Appendix Section F.2, we observe that $\beta$- and $\sigma$-convergence tend to mostly occur together and the main fact 2 is robust to either measure of convergence.

## 3.1 Fact #1: A Stall in Within-Country Convergence, 1981–2019

We begin by documenting a marked decline in within-country convergence over the period 1981–2019, a span during which we have a balanced panel data set for a substantial set of countries. Figure 1a presents the average within-country convergence coefficient, defined as $\beta_t = \frac{1}{C}\sum_c \beta_{c,t}$, along with 95% heteroskedasticity-robust confidence intervals. The figure reveals a pronounced secular decline in the convergence rate, from approximately 1% in the 1980s to values that are statistically indistinguishable from zero in the 1990s and 2000s.

Table 2 shows that 67% of countries in our sample experienced significant convergence spells in the 1980ies, which holds for only 54% of countries in the 2000s. These findings stand in stark contrast to the evidence on cross-country convergence over the same period, where both unconditional convergence and its rate have strengthened over time (Patel et al., 2018; Roy et al., 2016). Appendix Section D further shows that our results are robust to alternative specifications, including different population weighting schemes and the exclusion of China and India. As reported in Appendix Section D.2, despite the limitation that this data's coverage spans from 2004 to 2019, we find evidence of no regional convergence even there. We do robustness for several sample restrictions also excluding all small countries, that are below 10 million inhabitants, in Appendix Section F.1. Using our *The Economist* dataset, we run a robustness test for a sample of more than 900 cities where 19 Sub-Saharan African countries are included.

Table 2: The Decline in Within-Country Convergence

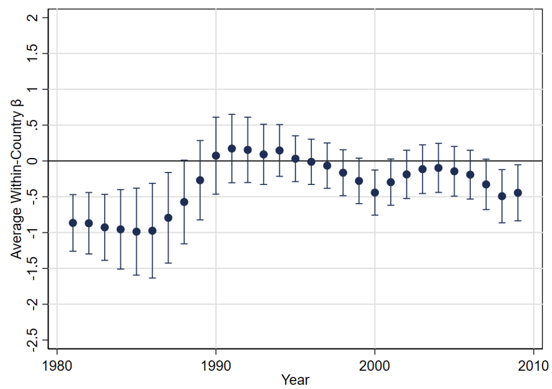
| | Ever converged | | |
|---|---|---|---|
| | 1981-1989 | 1991-1999 | 2001-2009 |
| Share of countries | 66.7% | 53.8% | 53.8% |
| Share of GDP | 74.9% | 50.0% | 53.6% |
| Share of population | 67.5% | 59.9% | 58.8% |

*Notes:* This table reports the share of our sample countries which has ever converged in a given decade, which is defined as having at least one $\beta$ estimate in this period that is negative and statistically significant at the 5% significance level. We further show the share of our sample's GDP and population that is represented by the respective countries. We use a balanced sample of 39 countries in this calculation.

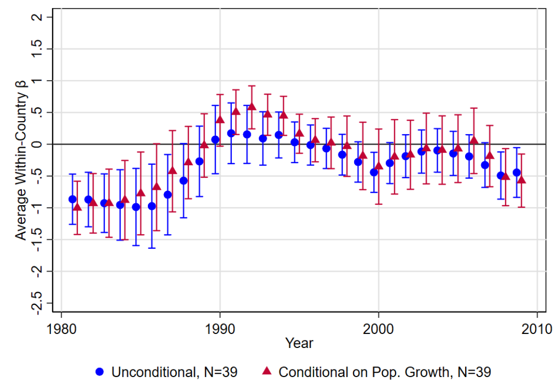
Our evidence extends to conditional convergence analyses following Solow (1956) and Mankiw et al. (1992). Controlling for population growth (while data limitations preclude conditioning on savings or investment at the regional level), Figure 1b confirms that the stall in convergence persists even when accounting for these covariates.

**Heterogeneity.** To explore further potential mechanisms, we analyze the heterogeneity in convergence patterns across countries. Most notably, we find a link between the timing of

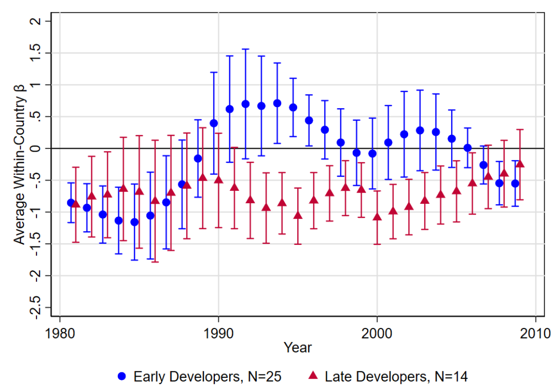
Figure 1: Within-Country $\beta$ Over Time




(a) Baseline: Population-weighted regression




(b) Conditional and unconditional convergence




(c) Heterogeneity: Early vs. late developers

*Notes:* This figure reports the average within-country $\beta$ convergence for a balanced sample of 39 countries between 1981 and 2019. In panel (a) the regressions for each country are weighted by population size, panel (b) compares unconditional and conditional convergence rates, where the latter control for population growth, panel (c) shows the heterogeneity in regional convergence between countries that are classified as early or late developers following Henderson et al. (2017)

.

countries' structural transformation and their convergence dynamics. Following Henderson et al. (2017), we classify countries as early or late developers.[2] Figure 1c illustrates that the decline in convergence emanates from early developers, while late developers exhibit less regional convergence patterns at the beginning of the sample while more starting around 1990 but slowly moving towards no regional convergence.

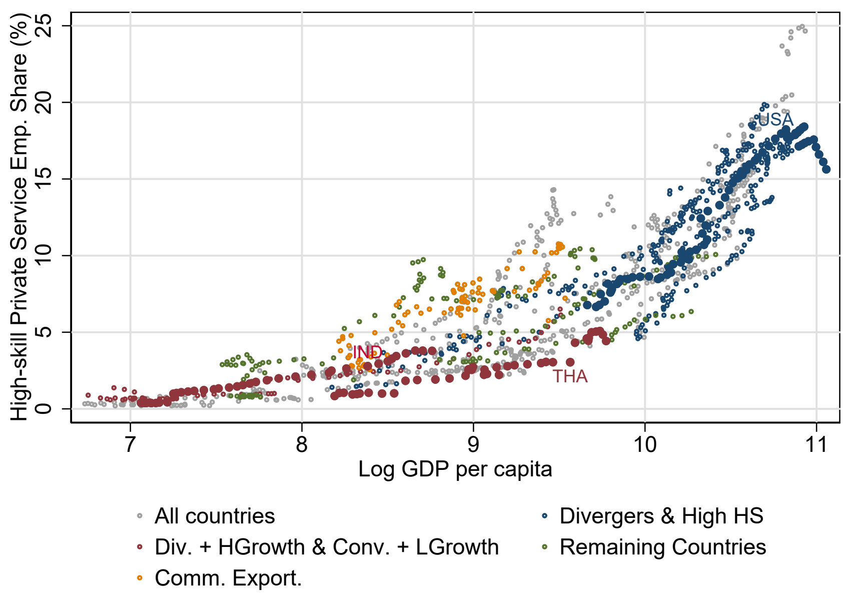
Appendix Section D groups countries based on additional layers of heterogeneity, based on OECD status, size, level of income, level of inequality and size of the service sector. All

[2]The categorization of each country is listed in Appendix Table D.5.

the heterogeneity is run based on the split of countries at the beginning of the sample to avoid compositional change.[3] Overall, the heterogeneity analysis points out a difference of convergence rates between high and low income countries, especially in the central years of our sample- between 1990 and 2000.

## 3.2 Fact #2: Structural Transformation and Regional Convergence

Figure 2: High-skill Service Employment and Regional Convergence




*Notes:* Employment data from own sources and GGDC/ETD for 27 countries. Convergers are defined as having no significant convergence spell after 1990, while divergers are those in the top quartile of the number of significant convergence spells after 1990. Countries with a high-skill private services employment share in 1991 above the median (6%) are defined as "High HS". "Growth" and "LGrowth" constitute the complement whereas the former are countries which grew more than the median (253%) of this remaining sample in number of employed people in this sector. "Comm. Export" refers to commodity exporters which are Peru and Colombia. Distinct countries are marked with their country code. The sample is unbalanced.

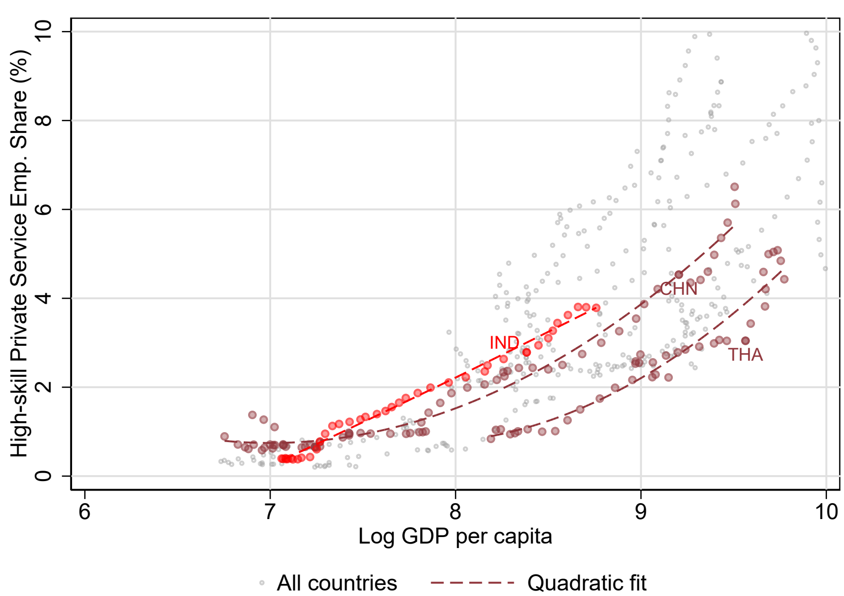
In this section, we show that the recent stagnation in within-country convergence coincides with structural shifts toward high-skill private services. Before turning to the formal econometric analysis, we present a graphical demonstration of the underlying mechanism. Figure 2 displays per-capita GDP against the employment share of high-skill private services for all

[3]The intersection of early/late developers and high/low-income countries is very high and hence, features a similar pattern.

country-year observations, distinguishing “high income” (blue) from “lower income” (red). Among high-income countries, most are divergers and most have high employment shares in high-skilled services. Instead, among low-income countries, there are strong differences. It stands out that India, which is a diverger or weak converger, has consistently higher levels of high-skill services for its development level compared to Thailand and China, which are strong convergers. These examples illustrate the broader patterns that we document: weak convergers consistently record higher shares in high-skill services than strong convergers, especially at lower income levels. Today’s advanced countries like the US, UK, and most European countries are weak convergers during the recent decades which are part of our sample, but at lower levels of GDP, they had high levels of industrialization and were strong convergers, as is well established in the literature.

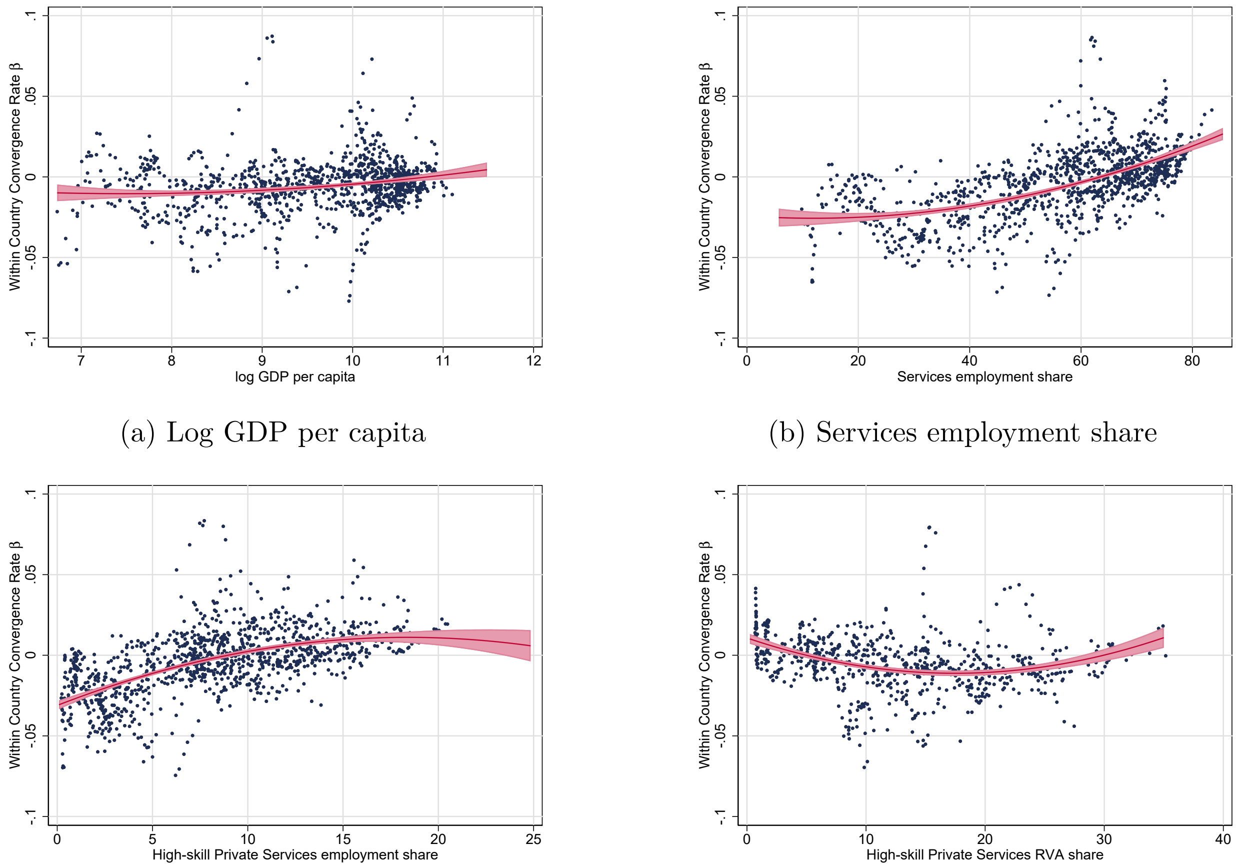
To zoom in even further on our mechanism, we isolate the three countries mentioned above–India, China and Thailand–in Figure 3. The clear separation between the path for India and the path for China and Thailand at early development stages visually encapsulates the mechanism at the heart of our analysis: economies that transition earlier and more heavily into high-skill services tend to experience weaker regional catch-up.

Figure 3: High-skill Service Employment and Regional Convergence: A Zoom




Together, Figures 2 and 3 show that the rise of high-skill services amplifies spatial concentration and diminishes within-country convergence. However, some countries are outliers which do not exhibit this relationship, in particular Brazil, France, and Scandinavian countries, which merit further investigation.

Figure 4: Structural Transformation and Regional Convergence




(a) Log GDP per capita

(b) Services employment share

(c) High-skill private services empl. share

(d) High-skill private services RVA share

*Notes:* Population weighted beta vs. log GDP per capita (a), vs. services employment share (b), the high-skill services employment share (c) and the the high-skill services real value added share (d) for the 39 countries which have employment. Estimates are residualized off year and country fixed effects. The red line shows the evolution of the average country. Confidence intervals are plotted around the estimated.

Figure 4 investigates the relationship between regional convergence and structural transformation more systematically by showing the correlation between countries' regional convergence and GDP per capita, service employment shares, high-skill service employment shares, and real value added in high-skill services. We residualize off country and year fixed effects and control for the Great recession by including a dummy variable that is equal to 1 if the time period of the 10-year-convergence regressions starts between the years 1997 and 2012. Recall that smaller (negative) $\beta$ values imply convergence while larger (positive) values imply divergence. Figures 4a - 4c show that regional $\beta_{ct}$-convergence has a slight negative correlation with GDP per capita and stronger negative correlations with employment shares in services and high-skill services. Figure 4d shows that regional convergence is only negatively correlated with real value added in high-skill services when a certain productivity threshold in high-skill services is reached.

Next, we turn to a multivariate analysis of the link between regional convergence and structural transformation by regressing the estimated $\beta_{ct}-$ and $\Delta\sigma_{ct}-$convergence-rates on log GDP per capita, employment shares in services or high-skill services, and real value added in high-skill services. Table 3 shows the results. Again, higher values of $\beta_{ct}$ imply *less* convergence, so that a positive coefficient in the regression implies a negative association with regional convergence. Columns 2 and 3 of Table 3 indicate that an increase in service employment shares is associated with lower convergence rates, but it is not statistically significant (i.e., higher $\beta$s). The association is stronger and statistically significant for employment in high-skill private services. Similarly, when we put on the y-axis $\Delta\sigma_{ct}-$convergence rates, we confirm the same results in column (4). Columns (5) and (6) further analyze the relationship between regional convergence and the high-skill private service sector, including productivity in this sector, which we obtain from the GGDC and ETD database, as well as time fixed effects. Higher private service sector productivity is associated with lower convergence rates, which remain significant even when controlling for the sector's employment shares. Table E.7 reports results for the balanced sample. Appendix Table E.8 reports results using ETD data. Appendix Table F.16 shows more specifications for $\beta-$ convergence. All the results are robust to our main findings. To isolate the role of the high-skill private services, we run similar regressions for the low-skill and high-skill public services, respectively. As Appendix Table E.9 and E.10 highlight, it is only the high-skill private services that drive the decrease in within-country convergence rates.

**Coefficient of Variation along Development** A potential concern is that, if spatial inequality naturally declines with overall development, then by the time a significant fraction of labor is employed in the services sector, little residual inequality remains to be closed. Figure 5 documents a negative relationship between spatial inequality—measured by the coefficient of variation of regional GDP per capita—and high-skill service employment shares from 5% to 20%. Nevertheless, spatial inequality appears to stagnate between 40%-50%, suggesting that significant regional disparities persist even as the services sector expands.

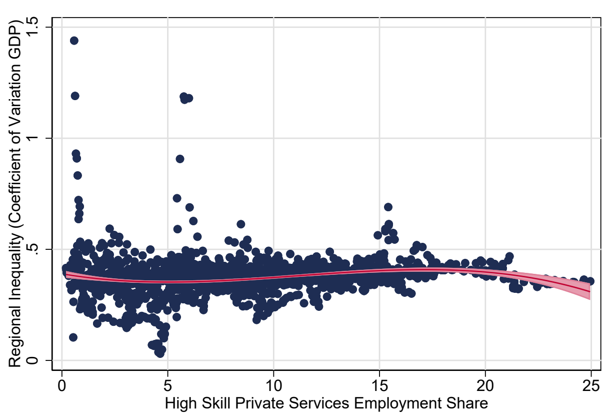
Moreover, our analysis indicates that the observed convergence patterns are not merely a

Table 3: Within-Country Convergence and Structural Transformation

| | (1) 10-year | (2) 10-year | (3) 10-year | (4) $\Delta CoV$ | (5) 10-year | (6) $\Delta CoV$ |
|---|---|---|---|---|---|---|
| Ln GDP pc. | -0.0075 | -0.0094 | -0.0135 | 0.0241 | -0.0237 | 0.0060 |
| | (0.0038)$^{*}$ | (0.0078) | (0.0027)$^{***}$ | (0.0151) | (0.0068)$^{***}$ | (0.0256) |
| Share serv. | | 0.0151 | | | | |
| | | (0.0740) | | | | |
| Share HS priv. serv. | | | 0.2621 | 0.6178 | 0.2689 | 0.4949 |
| | | | (0.0641)$^{***}$ | (0.3091)$^{*}$ | (0.0876)$^{***}$ | (0.2769)$^{*}$ |
| Great Recession | | | | -0.0446 | 0.0173 | -0.0382 |
| | | | | (0.0230)$^{*}$ | (0.0187) | (0.0723) |
| RVA per worker, HS priv. serv. | | | | | 0.0741 | 0.2145 |
| | | | | | (0.0328)$^{**}$ | (0.4075) |
| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | | | | | ✓ | ✓ |
| N | 1531 | 1531 | 1531 | 1520 | 951 | 940 |
| N country | 53 | 53 | 53 | 53 | 28 | 28 |
| $R^2$ | 0.5330 | 0.5335 | 0.5633 | 0.4118 | 0.7318 | 0.6169 |

*Notes*: This table presents the regression estimates where the dependent variable in each specification is the estimated $\beta-$convergence over a 10-year rolling window for each country in our unbalanced panel. "RVA per worker" measures the Real Value Added per workers in the high-skill private service sector that we obtain from the GGDC database. "Great Recession" is an indicator which equals one if the time period period of the 10-year-convergence regression start between the years 1997 and 2012. Specifications include country and/or year fixed effects. Robust standard errors clustered at the country level are reported in parenthesis. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

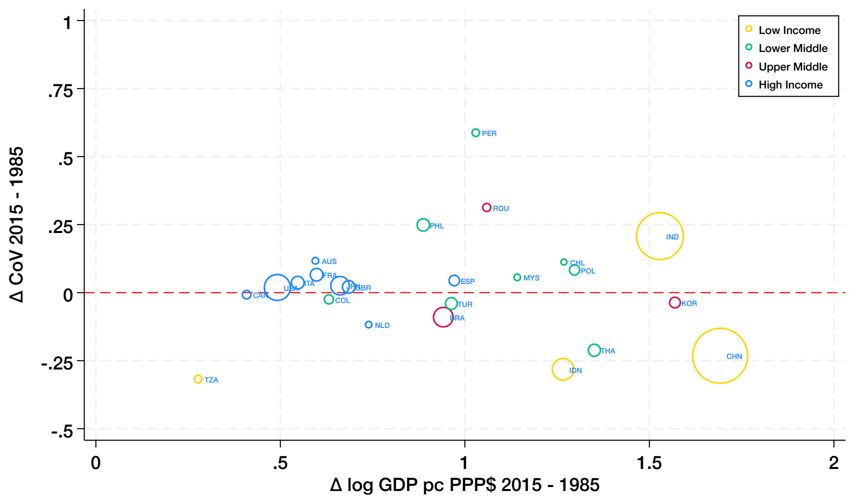
Figure 5: A Fall and Stagnation of Inequality with Structural Transformation




*Notes:* This figure plots the coefficient of variation of GDP per capita by country against the high-skill private service share in the economy. Estimates are residualized off country fixed effects and contain a recession dummy if the convergence measure is calculated in between 1997-2012. The red line shows the evolution of the average country. The sample is unbalanced.

function of the overall level of development and does not depend on the population of the country. Figure 6 displays the change in the Coefficient of Variation between 2015 and 1985 for each country with a population of less than 10 millions in relation to the change in national GDP per capita over the same time period. Several observations stand out. First, changes in income seem to be uncorrelated with changes in regional inequality, leading credence to the fact that inequality and development in our sample do not seem to be very correlated. Second, changes in regional inequality can be found almost among the whole development spectrum. Third, the correlation does not show a particular relationship with the size of the country.

Figure 6: Growth in regional inequality and economic development




*Notes:* This figure displays the correlation between the change in Coefficient of Variation and national GDP per capita between 1985-2015. The Coefficient of Variation is demeaned. The size of the circles represent the size of the national population in 1985. The classification into Low, Lower Middle, Upper Middle and High Income countries follows the World Bank definition and are recorded in 1987. Author's calculation.

The empirical evidence raises the question why the transition to high-skill private services is associated with lower regional convergence? To further investigate potential mechanisms behind this link, we next investigate the regional distribution of employment for all sectors.

## 3.3 Fact #3: Regional Concentration of Services

Given the strong relation between countries' stall in regional convergence and their structural transformation towards high-skill services, we next analyze the spatial concentration of employment across sectors. To do so, we use two measures of regional concentration: the

Herfindahl Index (HHI) and the Gini coefficient. The HHI is given by:

$$\text{HHI} = \sum_{i=1}^{N} \left( \frac{E_i}{E_{\text{total}}} \right)^2,$$

where $E_i$ is the employment in region $i$, $E_{\text{total}}$ is the total employment in all regions (i.e., the sum of employment across all regions), and $N$is the number of regions. The HHI values range from 0 to 1, where higher values indicate higher regional concentration and lower values indicate a more even distribution across regions.

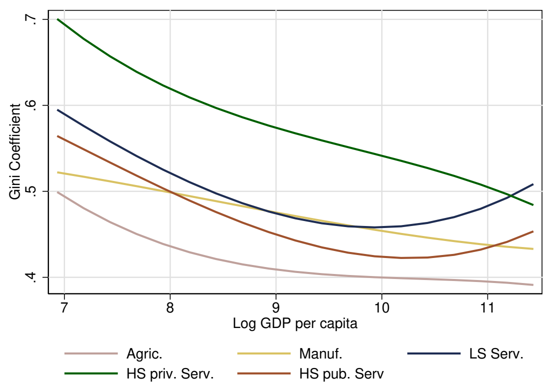
We compute the Gini and Herfindahl index for sectoral employment shares across all regions in each country-year. In Table 4, we show the average Gini and Herfindahl indices for each sector, which represent unweighted averages across all countries and years of the balanced sample of 39 countries. Column 3 shows that the Herfindahl index for agricultural employment indicates the lowest regional concentration, while high-skill private services are most spatially concentrated. Manufacturing, low-skill services, and high-skill public services lie in between and exhibit similar levels of regional concentration. More specifically, high-skill private services are two times more concentrated than agriculture, 53% more concentrated than manufacturing and high-skill public services and 50% more concentrated than low-skill services. This finding is robust to other measures of regional concentration: Column 1 of Table 4 shows the Gini coefficient of sectoral employment and Column 2 normalizes the Gini of sectoral employment by the Gini of regions' overall employment size to adjust for general differences in the size distribution of regions.

Table 4: Regional Concentration of Sectoral Employment

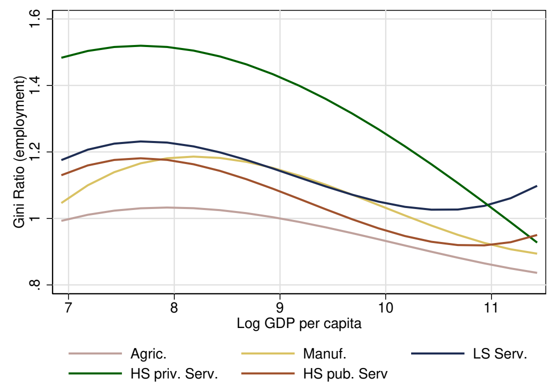
| | Gini<br>(1) | Gini Ratio<br>(2) | HHI<br>(3) |
|---|---|---|---|
| Agriculture | .39 | .94 | .13 |
| Manufacturing | .48 | 1.15 | .17 |
| Low-skill Services | .48 | 1.15 | .18 |
| High-skill public Services | .46 | 1.12 | .17 |
| High-skill private Services | .58 | 1.46 | .26 |

*Notes*: This table measures the regional concentration of sectoral employment in a balanced sample between 1980 and 2010. For each country-year, we compute the Gini and Herfindahl index of sectoral employment across all regions. "Gini-Ratio" divides the Ginis of sectoral employment by the Gini of overall employment to adjust for countries' heterogeneity in the overall size distribution of regions.

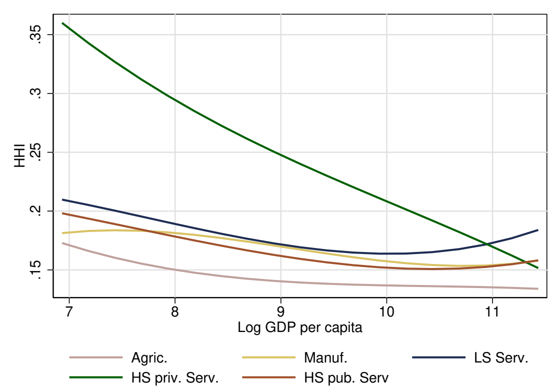
Figure 7: Sectoral Concentration By Development




(a) Gini Coefficient




(b) Gini-Ratio (employment)




(c) HHI

*Notes:* This figure shows sectoral concentration along log GDP per capita in the balanced data set from 1980-2010. The solid line represents a quadratic fit residualized off country fixed effects. For each country-year, the "Gini-Ratio (employment)" is defined as the Gini Coefficient of sectoral employment divided by the Gini Coefficient of overall employment.

Figure 7 further shows that this ranking of sectors' spatial concentration holds across all development levels. In particular, high-skill private service employment is more spatially concentrated than all other sectors, even if regional concentration overall tends to be lower in richer countries. Appendix Section F reports several robustness checks for these findings. For example, we corroborate our findings in an unbalanced sample which includes more countries and longer time periods.

### 3.4 Discussion of Empirical Facts and Model Implications

Overall, we document three novel facts about regional convergence and structural transformation for a large set of countries and across several decades. First, we show that regional convergence within-countries declines over time and stalls in the 2010s for most countries. Second, we show that the decline in convergence is present and more pronounced in countries that experience a stronger employment shift toward services, in particular, toward high-skill private services. Third, we show that employment in high-skill private services is more spatially concentrated than employment in other sectors.

The strong link between countries' regional convergence and changes in their sectoral structure suggests that innate differences in sectors' production functions could play a leading role in explaining countries' paths of regional inequality. In particular, the high spatial concentration of high-skill private service employment – which we demonstrate for countries at all development levels – indicates that agglomeration economies might be important in this sector. Agglomeration effects in the high-skill private service sector together with non-homothetic preferences can then generate feedback effects between spatial inequality and structural transformation. As economies become richer, consumers' demand shifts toward services, reinforcing the sector's local agglomeration effects which increase spatial inequality and economic growth. The increased growth then shifts consumers' demand even further towards services which starts the feedback cycle again.

These patterns are consistent with a large literature that demonstrates agglomeration economies, network externalities, and knowledge spillovers in the service sector. Among others, Davis and Dingel (2019) show that these forces can create self-reinforcing dynamics wherein urban centers attract a disproportionate share of skilled labor and innovative activities. This concentration can then impede the diffusion of economic gains to peripheral regions, thereby decreasing regional convergence. Giannone (2017) further shows that technological change reinforces spatial disparities by favoring regions with preexisting advantages in infrastructure and human capital. In particular, Moretti (2021) shows that high-tech sectors tend to concentrate in a few places and estimates strong agglomeration externalities. Kleineberg and Lebrand (2024) confirm our findings on the regional concentration of sectoral employment in a novel global dataset that uses internationally comparable measures of *cities* as the geographic unit of interest.

Next, we develop a stylized model framework that can rationalize our empirical findings and provide insights on the interplay between structural transformation and spatial growth.

# 4 A Model of Spatial Structural Change to Services

In this section, we present a stylized model of structural transformation and economic geography. The model incorporates the standard drivers of structural transformation: non-homothetic preferences and sector-specific productivity growth. Motivated by our empirical evidence, we disaggregate services into high-skill private and general services and we allow high-skill services to be tradable and to have local agglomeration effects. Despite its parsimonious structure, the model can generate the patterns of regional convergence and divergence across countries and links the shift toward high-skill services to spatial inequality and aggregate growth. We calibrate the model and conduct counterfactual simulations to quantify the contribution of agglomeration and sectoral productivity growth to changes in countries' regional convergence, sectoral employment shares, and aggregate growth.

## 4.1 Description of the Model

**Model Setup.** There are $J$ regions which we index by $j$ and $I$ sectors indexed by $i$. Workers decide where to locate in each period and receive idiosyncratic taste shocks $\mu_j$ for regions and $\mu_i$ for sectors, both originating from a Type-1 Extreme Value distribution. The parameters $\nu_J$ and $\nu_I$ scale the variances of the idiosyncratic shocks. Workers choose their location $j$ and sector $i$ to maximize their utility given sectoral and regional wages $w_{ij}$.

**Sectors.** We model three sectors: non-tradable services ($z$), tradable services ($s$) and tradable goods or non-services ($g$). Non-tradable services bundle the low-skill and high-skill public services from our empirical results. The goods sector bundles agriculture and manufacturing. Prices of the non-tradable services is determined by local market clearing $p_{zj}$, while the prices of the freely tradable sectors equalize across regions $p_s$ and $p_g$.

**Consumption.** Individuals have non-homothetic PIGL preferences over the three sectors. Under PIGL preferences, the indirect utility of consumption for an individual who earns wages $w_{ij}$ is given by:

$$V_{ij} = V(w_{ij}, P_j) = \frac{1}{\varepsilon}\left(\frac{w_{ij}}{(p_g)^{\phi^g}(p_s)^{\phi^S}(p_{zj})^{\phi^Z}}\right)^{\varepsilon} - \sum_{i\in\{g,s,z\}} v^i \ln P_j^i \qquad (6)$$

where each region's price index is equal to: $P_j \equiv p_g^{\phi^G} p_s^{\phi^S} p_{zj}^{\phi^Z}$ and $\sum_{i\in\{g,s,z\}} \phi^i = 1$. The income elasticity is given by $\varepsilon \in (0,1)$. Using Roy's identity, the consumption share of sector $s$ can be rewritten as[4]:

$$\psi^s(w_{ij}, P_j) = \phi^s + v^s \left(\frac{w_{ij}}{P_j}\right)^{-\varepsilon}, \tag{7}$$

where $s$ is the sector of consumption and $i$ the sector of individuals' employment which determines their income (but not their consumption choices conditional on income).

**Utility Maximization.** Households choose to live in a location $j$ and to work in a sector $i$ to maximize their utility by solving:

$$U_{ij} = \max_j \quad \max_i V(w_{ij}, P_j) + \nu_J \mu_j + +\nu_I \mu_i \tag{8}$$

where $i$ denotes the sector of employment, $j$ the region, and $\mu_j$ and $\mu_i$ individuals' idiosyncratic preference shocks. The price of tradable high-skill services is normalized to 1. Workers first learn their preference shock across regions $\mu_j$ and choose to move to a region $j$. After arriving in region $j$, they learn their sectoral preference shock $\mu_i$ and then choose their sector of employment $i$. We solve the problem backwards using the properties of the Type 1 extreme value distribution. First, conditional on living in region $j$, the share of individuals who choose to work in sector $i$ is equal to:

$$\pi_{i|j} = \frac{exp(V_{ij})^{1/\nu_I}}{\sum_i' exp(V_{i'j})^{1/\nu_I}}. \tag{9}$$

Using the properties of the Type 1 extreme value distribution again, we can express the expected utility of individuals in region $j$ before knowing their idiosyncratic sectoral preference shock as:

$$EV_j = \nu_I ln(\sum_{i'} exp(V_{i'j})^{1/\nu_I}). \tag{10}$$

When choosing their location $j$, individuals maximize this expected utility, so that we can express the share of individuals who choose to live in region $j$ as:

$$\xi_j = \frac{exp(EV_j)^{1/\nu_J}}{\sum_j' exp(EV_j')^{1/\nu_J}}. \tag{11}$$

These choices determine the distribution of workers across sectors and regions $L_{ij}$. There is no temporal persistence in this distribution since there are no costs or frictions when changing regions or sectors.

[4] Similar to Fan et al., (2023), we can calculate the aggregate consumption share of sector $s$ in $j$:$\bar{\psi}_j^s = \phi^s + v^s(\frac{E(w)}{P_j})^{-\varepsilon}(\frac{E(w^{1-\varepsilon})}{E(w)^{1-\varepsilon}})$

**Production.** In each region $j$, sectors $i = g, s, z$ produce goods or services with a location-specific productivity $A_{ij}$ and linear production functions that use local labor as the only input:

$$Y_{ij} = A_{ij} N_{ij}. \tag{12}$$

Individuals' location and sectoral choices determine the labor supply in each sector and region:

$$N_{ij} = \pi_{i|j} * \xi_j * N, \tag{13}$$

where $N$ is the total labor force in the economy.

The evolution of sectoral and regional productivity over time depends on initial local productivity levels $A_{ijt-1}$ and nationwide exogenous growth rates $g_i$ as follows:

$$A_{ijt} = e^{g_{it}} A_{ijt-1} \quad \text{for} \quad i = a, m \tag{14}$$

Motivated by our empirical findings, we additionally allow for agglomeration economies in the production function of high-skill private services, which we model as a spillover parameter $\delta$ on sectoral employment. Hence, productivity in the tradable service sector $s$ evolves according to:

$$A_{sjt} = e^{g_{st}} A_{ijt-1} N_{sjt}^{\delta}, \tag{15}$$

where $N_{sjt}$ is employment in high-skill services $s$ in region $j$.

Agglomeration spillover generate semi-endogenous growth. As workers move to regions that are productive in sector $s$, they further augment the sector's productivity in the next period. These local productivity spillovers accumulate over time and can have profound effects on regional income dynamics.

**Equilibrium.** Given a set of exogenous parameters that characterize individuals' preferences $\{\epsilon, \phi_i, \nu_i, \nu_J, \nu_I\}$, initial productivity levels $\{\{A_{ij0}\}_i^I\}_j^J$, and a set of normalizations such as $p_s = 1$ and $\sum_j N_j = N$, the competitive equilibrium in each period $t$ is characterized by a set of allocations and productivities $\{\{C_{ijt}, N_{ijt}, A_{ijt}\}_i^I\}_j^J$ and wages and prices $\{\{p_g, p_{zj}, w_{ij}\}_i^I\}_j^J$ such that the following conditions hold in each period:

(i) Given idiosyncratic preferences, workers choose their location, sector, and consumption to maximize their utility.

(ii) Local and sectoral productivities $A_{ijt}$ are given by the productivity processes presented in Equations 14 and 15.

(iii) Profit maximization of the firm and the zero-profit condition in each sector $i$ and region $j$ implies:

$$w_{ij} = p_{ij} A_{ij},$$

(iv) Labor market clear in each location and each sector so that local production in sector $i$ employs all local workers that choose to work in the sector so that:

$$N_{ij} = \pi_{i|j} * \xi_j * N \tag{16}$$

(v) For the non-tradable service sector, local prices $p_z j$ adjust to ensure that local production equals local consumer demand so that:

$$p_{zj} A_{zj} N_{zj} = p_{zj} C_{zj} = \Phi_{zj} * I_j, \tag{17}$$

where $\Phi_{zj}$ is the local expenditure share on non-tradable services derived from workers consumption choices and $I_j = \sum_i w_{ij} N_{ij}$ is total labor income of region $j$.

(vi) For the tradable sectors, supply and demand must clear at the country level so that:

$$p_i \sum_j A_{ij} N_{ij} = p_i \sum_j C_{ij} = \sum_j \Phi_{ij} * I_j, \tag{18}$$

where $\sum_j \Phi_{ij} * I_j$ is the country-wide expenditure on sector $i$.

**Key Model Mechanisms.** The model speaks to our empirical findings in the following way. As countries' productivities grow and they become richer, non-homothetic preferences shift consumers' demand and therefore countries' employment shares toward the service sector. Since the high-skill service sector $s$ is freely tradable and exhibits agglomeration effects, the sector's employment increases most in regions where its employment is initially high. These dynamics then increase the variance of service employment across regions, amplifying the sector's spatial concentration and reducing regional convergence.

The mechanisms of this simple model can therefore replicate the link between regional convergence and structural transformation that we documented above (Fact 2). More specifically, the agglomeration economies of the service sector attract workers into regions with high initial service employment as workers want to take advantage of the spillovers. In turn, this sorting generates further spillovers and increases productivity differences across regions. While agglomeration economies push for regional divergence, the model also has convergence forces such as workers' preference shocks and faster exogenous productivity growth in the non-service sector that eventually stabilize the spatial and sectoral distribution of workers.

## 4.2 Calibration - In Process

We are in the process of calibrating the model to the time series and geography of the US, India and Thailand. The cases of these three countries are representative for our broad empirical findings, as shown in Figures 2 and 3. The US is an advanced country with regional divergence and with a large employment share in high-skill services; India is a weak converger with a large employment share in high-skill service employment for its development level, and Thailand is a strong converger with low employment in high-skill services. Calibrating the model to these key cases allows us to illustrate the underlying mechanisms that drive the documented link between economies' transformation toward services and their stall in regional convergence.

Table 5 presents our chosen parameters for the current simulation. For the agglomeration effects, we first choose a value of $\delta = 0.05$ based on estimates from the literature. We then use US data on aggregate growth in sectoral real value added from the GGDC and BEA from 1980 to 2010 to calibrate the growth rates of exogenous productivity in each sector conditional on the value of $\delta$. In the next iteration, we plan to calibrate the agglomeration effects $\delta$ by targeting observed changes in regional income inequality $\sigma$ over time. For the dispersion of regional and sectoral preference shocks, we choose values from the literature. For regional preference shocks, we use $\nu_J = 5$ which is the mean of the estimates from Bryan and Morten (2019). For sectoral preference shocks, we set $\nu_I = 3.5$ as calibrated by Lee (2024). Finally, we set the income elasticity $\epsilon$ to 0.4 as estimated in Fan et al. (2022). The rest of the parameters–initial productivities $A_{ij0}$ and the PIGL parameters $\nu_i, \phi_i$–are currently set for illustration purposes.

Table 5: Simulation Parameters

| Parameter | Description | Targeted Moment | Literature | Value |
|---|---|---|---|---|
| *Production* | | | | |
| $g_a$ | Prod. Growth Goods | ✓ | | 0.034 |
| $g_m$ | Prod. Growth other Services | ✓ | | 0.033 |
| $g_s$ | Prod. Growth high-skill Services | ✓ | | 0.030 |
| $\delta$ | Agglomeration in Service | | ✓ | 0.05 |
| $A_{ij}$ | Initial Prod. by sector and location | | | – |
| *Preferences* | | | | |
| $\nu_I$ | T1-EV variance across regions | | Bryan and Morten (2019) | 5 |
| $\nu_J$ | T1-EV variance across sectors | | Lee (2024) | 3.5 |
| $\phi_g, \phi_s, \phi_z$ | PIGL Exp. Shares at limit | | | 0.6, 0.31, 0.09 |
| $\upsilon_g, \upsilon_s, \upsilon_z$ | PIGL Income Effects | | | 0.6, -0.5, -0.1 |
| $\epsilon$ | PIGL Income Elasticity | | Fan et al. (2022) | 0.4 |

## 4.3 Model Mechanisms and Counterfactuals Results

We use the model for simulations to illustrate how non-homothetic preferences, uneven sectoral productivity growth, and agglomeration effects in the high-skill service sector affect the link between structural transformation, regional convergence, and aggregate growth. To do so, we simulate the economy forward for 30 years (i.e., 1980-2010) for the baseline calibration and several counterfactual calibrations and we then compute changes over time for each outcome of interest.

Table 6 presents the results. Row (1) shows the results for the baseline calibration. Row (2) sets agglomeration effects to zero, i.e., $\delta = 0$. Row (3) lowers the exogenous growth rate in the high-skill service sector $g_s$ so that total productivity growth in the sector is half of the baseline growth. Last, row (4) eliminates income effects by setting $\nu_i$ to 0 in the PIGL preferences. Preferences then reduce to Cobb-Douglas with fixed expenditure shares and we adjust the expenditure share coefficients $\phi_i$ so that sectoral expenditure shares in the counterfactual are the same as in the baseline calibration for the first time period. The columns present how our outcomes of interest change over time in each simulation. To measure aggregate effects on structural transformation and development, we report the percent change in the aggregate high-skill service employment share (column 1) and real GDP per capita (column 2). To measure changes in spatial heterogeneity, we report the percent change in the standard deviation of high-skill employment (column 3) and in the coefficient of variation in GDP per capita across regions (column 4).

For the baseline economy, row (1) of Table 6 shows that employment in high-skill services increases by 13.2% and GDP per capita by 313% (or 4.9% per year) over the 30-year period. Spatial inequality rises substantially - the standard deviation of high-skill service employment increases by 290.6% and the coefficient of variation of real GDP per capita by 25% across regions.

Row 2 shows that these patterns would have been very different if we eliminate the agglomeration economies in the high-skill service sector. The employment share in high-skill services would have increased by only 1.6% (instead of 13.2%) and real GDP per capita would have grown by less than half with a percent growth of 87.2% over 20 years or 2.1% per year. However, without agglomeration economies, outcomes would have been much more equal across space. The standard deviation in high-skill service employment would have increased by on 13% (instead of 290%) and the coefficient of variation in real GDP per capita would have slightly decreased by 1.6%. Next, we implement a counterfactual that keeps agglomeration effects but lowers the exogenous productivity growth rate in high-skill services. This counterfactual lies in between the previous cases: Structural transformation, GDP per capita growth, and spatial inequality all increase but much less than in the baseline calibration. The aggregate and distributional effects are now smaller because the lower productivity growth in the high-skill service sector makes it relatively less important for determining aggregate outcomes. Instead, the sectors without agglomeration effects now contribute relatively more to aggregate outcomes, which mute the impact of the agglomeration effects on spatial inequality. Last, we show in row (4) that income effects are very important for the results on structural transformation and spatial inequality by making preferences homothetic. In this case, the employment share in high-skill services would have decreased by 4% since this sector has faster overall productivity growth than the other sectors in our calibration. Real GDP per capita growth is slightly slower with 4.5% per year (instead of 4.9% per year) because consumer demand for the fastest-growing sector does not increase as it does with non-homothetic preferences. Agglomeration effects in high-skill services still lead to a large increase in the concentration of high-skill employment with the standard deviation increasing by 122%, but the coefficient of variation in real GDP per capita decreases slightly by 1.3% because employment increases in the non-service and low-skill service sector, which acts as a convergence force.

These illustrative simulation results demonstrate that agglomeration forces and overall productivity growth in the service sector are important drivers that generate a self-reinforcing interplay between regional inequality and structural transformation. Eliminating agglomeration economies would increase spatial equality but would also reduce aggregate GDP growth, indicating a trade-off between regional dynamics and faster growth. Non-homothetic preferences further amplify the effects: once economies grow richer, consumers demand more high-skill services, which increases the sector's productivity through agglomeration effects, and further increases GDP growth but also spatial inequality.

## 4.4 Current Work and Next Steps

The model presented above highlights the interplay and self-reinforcing mechanisms between employment shifts into the high-skill service sector and regional convergence. To align closer with our empirical results, we are currently working on the following improvements to our quantitative exercise and calibration.

We are changing the preferences in the model to non-homothetic CES and we are working

Table 6: Agglomeration Economies in Service Sector and Regional Convergence

| Scenario | $\%\Delta Ls_{hs}$ | $\%\Delta GDPpc$ | $\%\Delta StdDevLs_{hs}$ | $\%\Delta CoefVarGDPpc$ |
|---|---|---|---|---|
| baseline | 13.2 | 313.6 | 290.6 | 24.9 |
| $\delta$=0 | 1.6 | 87.2 | 13.0 | -1.6 |
| Lower $g_s$ | 6.9 | 163.1 | 152.9 | 5.1 |
| No income effect | -3.9 | 273.8 | 121.5 | -1.3 |

*Notes:* This table shows the simulation results for the baseline calibration and various counterfactuals in each row. We simulate each model forward for 30 years and then compute the changes over time for each outcome of interest, which are listed in the columns. Column (1) represents the percent change in the aggregate high-skill service employment share, Column (2) presents the percent change in real GDP per capita over the 30 year period. Column (3) presents the percent change in the standard deviation of high-skill employment across regions. Column (4) presents the percent change in the coefficient of variation in GDP per capita across regions. Row (1) shows the baseline calibration. Row (2) sets agglomeration effects to 0. Row (3) lowers the exogenous growth rate in the high-skill service sector so that total productivity growth in the sector is half of the baseline growth. Last, row (4) eliminates income effects by setting $\nu_i$ to 0 in the PIGL preferences. Preferences then reduce to Cobb Douglas with fixed expenditure shares and we readjust the $\phi_i$ coefficients so that sectoral expenditure shares in this counterfactual are the same as in the baseline calibration for the first time period.

on estimating the parameters of these preferences using the same strategy and data as in Comin et al. (2021). We cannot take existing estimates from the literature since we use different sectoral classifications by distinguishing between high-skill private services and other services. We are furthermore working on using sectoral and regional wage data to calibrate initial productivity levels for each region and sector $A_{ij0}$. This wage data is available for many countries in the labor force surveys and in some censuses.

Given preferences and initial productivities, we can then simulate the model forward to calibrate agglomeration effects $\delta$ and the exogenous growth rates of sectoral productivity $g_i$ as described above. We will target changes in the spatial concentration of service sector employment from our dataset to calibrate agglomeration economies. Then, for a given value of $\delta$, we can solve for the exogenous sectoral productivity growth rates $g_i$ in our model, which ensure that aggregate sectoral productivity growth in our model exactly matches the growth rates in sectoral real value added that we obtain from the GGDC and ETD database. We are still determining whether we want to calibrate the model to the actual regions of different countries, which allows for an exact model inversion, or if we continue with a more stylized model calibration.

We can implement this calibration separately for the US, India, and Thailand to examine how initial sectoral and regional productivities $A_{ij0}$, exogenous sectoral productivity growth $g_i$ and agglomeration estimates $\delta$ vary across these three countries which each illustrate different paths of structural transformation, regional convergence, and development. This calibration approach will then allow us to implement more tangible counterfactuals and accounting exercises. In particular, we can then analyze how the heterogeneity in agglomeration forces and sectoral productivity growth across these countries relates to their paths of regional convergence, spatial inequality, and structural transformation. In addition, we can evaluate how different the path of regional inequality and structural transformation would have been in particular countries or country-groups if we assigned them the relevant structural parameters from another country or country-group. For example, we can evaluate how different outcomes would be if a weak converger such as India had the same spillover parameters as a strong converger such as Thailand.

# 5 Conclusion

In this paper, we assemble and validate a longitudinal dataset which provides detailed GDP and sectoral employment data for more than 1500 regions and more than 90 countries between 1980 and 2019. We use this novel dataset to revisit a classical question in macroeconomics: how is structural transformation toward services associated with regional convergence? We present new empirical facts. First, we find that regional convergence within countries is decreasing over time around the globe and stalls in the most recent decade. To understand which mechanisms drive the stall, second, we show that the decline in convergence is particularly pronounced in countries that experience a stronger employment shift toward services, in particular, toward high-skill private services. Third, we focus on the spatial sectoral employment and we show that employment in high-skill private services is more spatially concentrated than in any other sector. While the literature has shown that the transition from agriculture to manufacturing has led to regional convergence, our results indicate the opposite for the shift toward services, in particular, towards high-skill high-growth services.

To study the mechanisms that link service-led structural transformation and regional convergence, we therefore develop a stylized model framework that can rationalize our empirical findings. The model captures the key drivers of structural change, such as non-homothetic preferences and sector-specific productivity growth, and allows for agglomeration economies in the high-skill tradable service sector. We then calibrate the model and implement counterfactuals that quantify the contribution of specific mechanisms to countries' observed changes in regional convergence and sectoral employment. We find that eliminating agglomeration economies in the service sector would reduce regional inequality, but would decrease employment growth in the service sector, slowing countries' structural transformation and economic development. These findings demonstrate a trade-off between unequal regional dynamics and service-led growth.

# References


**Acemoglu, Daron, Simon Johnson, and James A Robinson**, "The colonial origins of comparative development: An empirical investigation," *American economic review*, 2001, *91* (5), 1369–1401.

**Auteri, D, C Attardo, M Berzi, C Dorati, F Albinola, L Baggio, G Bucciarelli, I Bussolari, and L Dijkstra**, "The Annual Regional Database of the European Commission (ARDECO) - Methodological Note," 2024.

**Barro, RJ and JW Lee**, "Barro-Lee data set," *International data on educational attainment: Updates and implications. Boston: Harvard University. Retrieved November*, 2000, *18*, 2004.

**Barro, Robert and Xavier Sala i Martin**, "Convergence," *Journal of Political Economy*, 1992, *100* (2), 223–251.

**Baumol, William**, "Productivity Growth, Convergence, and Welfare: What the Long-run Data Show," *American Economic Review*, 1986, *76* (5), 1072–85.

**Blanchard, Olivier Jean, Lawrence F Katz, Robert E Hall, and Barry Eichengreen**, "Regional evolutions," *Brookings papers on economic activity*, 1992, *1992* (1), 1–75.

**Bohr, Clement E, Marti Mestieri, and Frederic Robert-Nicoud**, "Economic Geography and Structural Change," *arXiv preprint arXiv:2412.03755*, 2024.

**Bryan, Gharad and Melanie Morten**, "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

**Budi-Ors, Tomas and Josep Pijoan-Mas**, "Macroeconomic Development, Rural Exodus, and Uneven Industrialization," *CEPR Discussion Paper No. DP17086*, 2022.

**Buera, Francisco J. and Joseph P. Kaboski**, "The Rise of the Service Economy," *American Economic Review*, October 2012, *102* (6), 2540–69.

**Caselli, Francesco and Wilbur John Coleman**, "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, June 2001, *109* (3), 584–616.

**Chen, Natalie, Dennis Novy, Carlo Perroni, and Horng Chern Wong**, "Urban-Biased Structural Change," *Unpublished Manuscript*, 2023.

**Comin, Diego, Danial Lashkari, and Martí Mestieri**, "Structural change with long-run income and price effects," *Econometrica*, 2021, *89* (1), 311–374.

**Davis, Donald R. and Jonathan I. Dingel**, "A Spatial Knowledge Economy," *American Economic Review*, January 2019, *109* (1), 153–70.

**Donovan, Kevin, Will Jianyu Lu, and Todd Schoellman**, "Labor market dynamics and development," *The Quarterly Journal of Economics*, 2023, *138* (4), 2287–2325.

**Duarte, Margarida and Diego Restuccia**, "Relative Prices and Sectoral Productivity," *Journal of the European Economic Association*, 05 2019, *18* (3), 1400–1443.

**Eckert, Fabian and Michael Peters**, "Spatial structural change," *Unpublished Manuscript*, 2018.

**Fan, Tianyu, Michael Peters, and Fabrizio Zilibotti**, "Growing Like India," 2022.

**Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer**, "The Next Generation of the Penn World Table," *American Economic Review*, October 2015, *105* (10), 3150–82.

**Ganong, Peter and Daniel Shoag**, "Why Has Regional Convergence in the U.S. Stopped?," *Journal of Urban Economics*, June 2017, *102*, 76–90.

**Gennaioli, Nicola, Rafael LaPorta, Florencio Lopez de Silanes, and Andrei Shleifer**, "Growth in Regions," *Journal of Economic Growth*, 2014, *19* (3), 259–309.

**Giannone, Elisa**, "Skill-Biased technical Change and Regional Convergence," 2017.

**Guriev, Sergei and Elena Vakulenko**, "Convergence among Russian regions," *CEFIR/NES Working Paper*, 2012, (180).

**Hao, Tongtong, Ruiqi Sun, Trevor Tombe, and Xiaodong Zhu**, "The effect of migration policy on growth, structural change, and regional inequality in China," *Journal of Monetary Economics*, 2020, *113*, 112–134.

**Henderson, J Vernon, Tim Squires, Adam Storeygard, and David Weil**, "The Global Distribution of Economic Activity: Nature, History, and the Role of Trade1," *The Quarterly Journal of Economics*, 09 2017, *133* (1), 357–406.

**Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi**, "Growth and structural transformation," in "Handbook of economic growth," Vol. 2, Elsevier, 2014, pp. 855–941.

**Huneeus, Federico and Richard Rogerson**, "Heterogeneous Paths of Industrialization," Technical Report, National Bureau of Economic Research 2020.

**Inklaar, Robert and Marcel P. Timmer**, "GGDC Productivity Level Database: International Comparisons of Output, Inputs and Productivity at the Industry Level," WorkingPaper, GGDC 2008.

**Kleineberg, Tatjana and Mathilde Lebrand**, "The Distribution of Economic Activity across Cities: A Global Perspective," 2024.

**Kruse, Hagen, Emmanuel Mensah, Kunal Sen, and Gaaitzen de Vries**, "A manufacturing renaissance? Industrialization trends in the developing world," *IMF Economic Review*, August 2022, *71* (2), 439–473.

**Lagakos, David and Martin Shu**, "The role of micro data in understanding structural transformation," *Oxford Development Studies*, 2023, *51* (4), 436–454.

**Lee, Munseob**, "Allocation of female talent and cross-country productivity differences," *The Economic Journal*, 2024, *134* (664), 3333–3359.

**Mankiw, N Gregory, David Romer, and David N Weil**, "A contribution to the empirics of economic growth," *The quarterly journal of economics*, 1992, *107* (2), 407–437.

**Moretti, Enrico**, "The effect of high-tech clusters on the productivity of top inventors," *American Economic Review*, 2021, *111* (10), 3328–75.

**Nayyar, Gaurav, Mary Hallward-Driemeier, and Elwyn Davies**, *At your service?: The promise of services-led development*, World Bank Publications, 2021.

**Nordhaus, William D.**, "Geography and macroeconomics: New data and new findings," *Proceedings of the National Academy of Sciences*, 2006, *103* (10), 3510–3517.

**Nunn, Nathan and Diego Puga**, "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

**Patel, Dev, Justin Sandefur, and Arvind Subramanian**, "Everything You Know about Cross-Country Convergence Is (Now) Wrong," Oct 2018.

**Rajan, Raghuram and Rohit Lamba**, *Breaking the Mould: Reimagining India's Economic Future*, Penguin Random House India Private Limited, 2023.

**Rodrik, Dani**, "Unconditional convergence in manufacturing," *The quarterly journal of economics*, 2013, *128* (1), 165–204.

_ , "Premature Deindustrialization," *Journal of Economic Growth*, 2016, *21*, 1–33.

**Rossi-Hansberg, Esteban and Jialing Zhang**, "Local GDP Estimates Around the World," Working Paper 33458, National Bureau of Economic Research February 2025.

**Roy, Sutirtha, Martin Kessler, and Arvind Subramanian**, "Glimpsing the End of Economic History? Unconditional Convergence and the Missing Middle Income Trap," *Center for Global Development Working Paper*, 2016, (438).

**Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, , and Matthew Sobek**, "Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]," 2015.

**_ , Lara Cleveland, Rodrigo Lovaton, Sula Sarkar, Matthew Sobek, Derek Burk, Dan Ehrlich, Quinn Heimann, and Jane Lee**, "Integrated Public Use Microdata Series, International: Version 7.5 ," 2024.

**Sachs, Jeffrey D**, "Tropical underdevelopment," Technical Report, National Bureau of Economic Research 2001.

**Sala-i-Martin, Xavier**, "Regional cohesion: evidence and theories of regional growth and convergence," *European Economic Review*, 1996, *40* (6), 1325–1352.

**Smits, Jeroen and Iñaki Permanyer**, "The Subnational Human Development Database," *Scientific Data*, 03 2019, *6*, 190038.

**Solow, Robert M**, "A contribution to the theory of economic growth," *The quarterly journal of economics*, 1956, *70* (1), 65–94.

# A Description of Main External Data Sources

**Dose:** The DOSE (MCC-PIK Database of Sub-National Economic Output) provides harmonized economic output data for over 1,660 regions across 83 countries, covering the period from 1960 to 2020. It includes total gross regional product (GRP) and, for most observations, sectoral GRP for agriculture, manufacturing, and services. The data are available in local currency units (LCU) and US dollars at both current and 2015 market prices, ensuring comparability across time and space. The database is available here: https://zenodo.org/records/7573249

**World Bank Global Labor Database:** The Global Labor Database (GLD) is a World Bank initiative designed to harmonize labor force and household surveys with labor-related modules. It aims to cover all labor force surveys worldwide, with a focus on lower income countries, though on occasion the GLD team may cover other household surveys with a sufficient labor module. As of April 2024, the GLD holds 345 surveys from 24 countries (1 high-income countries, 9 upper medium-income, 11 lower middle-income, and 9 low-income countries). The database can be accessed here. https://worldbank.github.io/gld/README.html

**I2D2:** The International Income Distribution Database (I2D2), is a database developed by the World Bank and contains more than 1,500 household surveys. It contains annual earnings, educational attainment, and employment rates.

**ARDECO Database from the ECJRC:** The Annual Regional Database of the European Commission's Joint Research Centre provides harmonized time-series data on demographic and socio-economic variables such as GDP, employment and wages at regional and sub-regional levels within Europe. Data can be retrieved from the ARDECO explorer: https://urban.jrc.ec.europa.eu/ardeco/explorer?lng=en.

**Global Roads Inventory Project (GRIP):** GRIP offers a harmonized global dataset compiling approximately 60 geospatial datasets on road infrastructure. Available for download at https://www.globio.info/download-grip-dataset.

**GGDC: Productivity Level Database:** Released by the Groningen Growth and Development Centre, this database presents data on relative prices and labor productivity across multiple countries and sectors, based on International Comparison Program benchmarks. As of 2023, it covers 84 countries and 12 sectors, aiding in the analysis of productivity differences and economic performance. Details are available at https://www.rug.nl/ggdc/productivity/pld/releases/pld-2023.

**Economic Transformation Database:** Developed by the Groningen Growth and Development Centre (GGDC) in collaboration with UNU-WIDER, the Economic Transformation Database (ETD) provides comprehensive, long-term, and internationally comparable sectoral data on employment and productivity. It covers 12 sectors from over 50 economies in Africa, Asia, and Latin America between 1990 and 2018. The database is accessible at https://www.rug.nl/ggdc/structuralchange/etd/?lang=en.

# B Coverage of the Sample

Table B.1: Data coverage

| | Nb. countries (1) | Avg. nb. years (2) | 1980-2010 (3) | 1990-2010 (4) | 1980-2019 (5) | 1990-2019 (6) | 2000-2019 (7) |
|---|---|---|---|---|---|---|---|
| *Panel A: GDP* | | | | | | | |
| Africa | 3 | 47 | 3 | 3 | 3 | 3 | 3 |
| Asia | 12 | 43 | 8 | 9 | 8 | 9 | 10 |
| Australia and Oceania | 1 | 38 | 0 | 1 | 0 | 1 | 1 |
| East Europe | 16 | 29 | 3 | 5 | 3 | 5 | 16 |
| North America | 3 | 51 | 2 | 2 | 2 | 2 | 3 |
| South America | 6 | 43 | 5 | 5 | 5 | 5 | 6 |
| West Europe | 16 | 39 | 16 | 16 | 16 | 16 | 16 |
| Total | 57 | | 37 | 41 | 37 | 41 | 55 |
| *Panel B: Employment* | | | | | | | |
| Africa | 17 | 22 | 2 | 8 | 0 | 1 | 3 |
| Asia | 14 | 30 | 4 | 9 | 3 | 5 | 6 |
| Australia and Oceania | 3 | 28 | 0 | 2 | 0 | 1 | 1 |
| East Europe | 13 | 27 | 0 | 4 | 0 | 4 | 13 |
| North America | 4 | 44 | 3 | 3 | 2 | 2 | 2 |
| South America | 18 | 38 | 9 | 11 | 3 | 3 | 4 |
| West Europe | 16 | 40 | 15 | 16 | 15 | 16 | 16 |
| Total | 85 | | 33 | 53 | 23 | 32 | 45 |
| *Panel C: GDP & Employment* | | | | | | | |
| Africa | 2 | 19 | 0 | 0 | 0 | 0 | 1 |
| Asia | 11 | 32 | 3 | 8 | 2 | 4 | 4 |
| Australia and Oceania | 1 | 35 | 0 | 1 | 0 | 1 | 1 |
| East Europe | 13 | 26 | 0 | 1 | 0 | 1 | 13 |
| North America | 3 | 42 | 2 | 2 | 1 | 1 | 2 |
| South America | 6 | 34 | 3 | 4 | 3 | 3 | 4 |
| West Europe | 16 | 39 | 15 | 16 | 15 | 16 | 16 |
| Total | 52 | | 23 | 32 | 21 | 26 | 41 |

**Notes**: This table displays the number of countries that are present in the data set and have GDP, employment data or both. Column 3-7 displays the number of countries that are present in this respective time period. Author's calculation.

# C Representativeness of the Sample

Table C.2: Representativeness of the Samples

| Period | Share of World Population | Share of World GDP | Avg Growth GDP p.c. | Growth relative to World Avg | # Countries | Avg years of education |
|---|---|---|---|---|---|---|
| 1980-1990 | 0.675 | 0.856 | 1.93% | 1.60 | 34 | 6.49 |
| 1990-2000 | 0.662 | 0.794 | 2.82% | 1.54 | 34 | 7.80 |
| 2000-2010 | 0.647 | 0.779 | 3.74% | 1.04 | 34 | 9.03 |
| 2010-2020 | 0.642 | 0.773 | 2.30% | 1.33 | 34 | 9.67 |
| All Years | 0.656 | 0.802 | 2.80% | 1.13 | 34 | 8.16 |

*Notes:* This table reports the main summary statistics such as share continent population, share of continent GDP, average GDP Growth per capita, GDP growth relative to the world average, # of countries and average years of education of our sample over the decades.

Table C.3: Representativeness of the Sample by Income Groups

| | Share of Continent Population | Share of Continent GDP | Avg. GDP Growth Per Capita | Growth relative to World Avg | # Countries | Avg years of education |
|---|---|---|---|---|---|---|
| | | | High Income | | | |
| 1980-1990 | 0.922 | 0.948 | 2.12% | 1.03 | 16 | 9.24 |
| 1990-2000 | 0.897 | 0.922 | 2.65% | 1.02 | 16 | 10.04 |
| 2000-2010 | 0.887 | 0.898 | 2.14% | 1.16 | 16 | 10.81 |
| 2010-2020 | 0.916 | 0.916 | 1.62% | 1.14 | 16 | 10.52 |
| All Years | 0.905 | 0.921 | 2.24% | 1.03 | 16 | 10.22 |
| | | | Middle Income | | | |
| 1980-1990 | 0.554 | 0.561 | 6.27% | 5.28 | 5 | 5.39 |
| 1990-2000 | 0.541 | 0.651 | 4.48% | 0.94 | 5 | 6.99 |
| 2000-2010 | 0.535 | 0.595 | 4.04% | 0.78 | 5 | 8.40 |
| 2010-2020 | 0.568 | 0.598 | 0.46% | -2.93 | 5 | 8.84 |
| All Years | 0.549 | 0.601 | 3.50% | 1.01 | 5 | 7.41 |
| | | | Low Income | | | |
| 1980-1990 | 0.707 | 0.732 | 0.70% | 0.58 | 13 | 4.24 |
| 1990-2000 | 0.693 | 0.752 | 2.63% | 0.81 | 13 | 5.34 |
| 2000-2010 | 0.675 | 0.762 | 5.51% | 0.89 | 13 | 6.48 |
| 2010-2020 | 0.663 | 0.778 | 3.50% | 1.05 | 13 | 7.38 |
| All Years | 0.686 | 0.755 | 3.29% | 0.86 | 13 | 5.49 |

*Notes:* This table reports the main summary statistics such as share continent population, share of continent GDP, average GDP Growth per capita, GDP growth relative to the world average, # of countries and average years of education of our sample by income group and over the decades in our sample. We divided countries in high income (more than 67th percentile), middle income (between 67th and 33th percentile) and low income (33th percentile and less).

Table C.4: Representativeness of the Sample by Continents

| | Share of Continent Population | Share of Continent GDP | Avg. GDP Growth Per Capita | Growth relative to World Avg | # Countries | Avg years of education |
|---|---|---|---|---|---|---|
| | | | Africa | | | |
| 1980-1990 | 0.148 | 0.253 | -3.83% | 0.95 | 3 | 3.66 |
| 1990-2000 | 0.144 | 0.270 | 0.20% | 0.29 | 3 | 4.21 |
| 2000-2010 | 0.139 | 0.225 | 4.68% | 0.84 | 3 | 5.82 |
| 2010-2020 | 0.135 | 0.179 | 2.58% | 13.08 | 3 | 5.67 |
| All Years | 0.142 | 0.235 | 1.74% | 1.16 | 3 | 4.61 |
| | | | Asia | | | |
| 1980-1990 | 0.795 | 0.743 | 4.02% | 2.09 | 6 | 5.94 |
| 1990-2000 | 0.772 | 0.757 | 3.74% | 1.14 | 6 | 7.01 |
| 2000-2010 | 0.759 | 0.737 | 4.84% | 0.87 | 6 | 8.88 |
| 2010-2020 | 0.756 | 0.742 | 3.16% | 1.10 | 6 | 8.36 |
| All Years | 0.771 | 0.745 | 4.00% | 1.01 | 6 | 7.48 |
| | | | Europe | | | |
| 1980-1990 | 0.833 | 0.955 | 2.11% | 1.20 | 16 | 7.61 |
| 1990-2000 | 0.522 | 0.733 | 2.52% | 2.18 | 16 | 8.67 |
| 2000-2010 | 0.544 | 0.735 | 3.34% | 0.85 | 16 | 9.71 |
| 2010-2020 | 0.559 | 0.678 | 2.32% | 1.24 | 16 | 10.06 |
| All Years | 0.617 | 0.780 | 2.59% | 1.26 | 16 | 9.08 |
| | | | North America | | | |
| 1980-1990 | 0.888 | 0.983 | 1.11% | 0.55 | 3 | 9.27 |
| 1990-2000 | 0.880 | 0.982 | 1.76% | 0.88 | 3 | 10.36 |
| 2000-2010 | 0.873 | 0.978 | 1.35% | 1.19 | 3 | 10.41 |
| 2010-2020 | 0.941 | 1.071 | 1.90% | 1.23 | 3 | 10.26 |
| All Years | 0.893 | 1.000 | 1.65% | 1.04 | 3 | 10.09 |
| | | | Oceania | | | |
| 1980-1990 | 0.807 | 0.867 | 2.23% | 0.97 | 1 | |
| 1990-2000 | 0.803 | 0.861 | 3.04% | 1.01 | 1 | 11.42 |
| 2000-2010 | 0.804 | 0.864 | 2.02% | 1.01 | 1 | 12.41 |
| 2010-2020 | 0.811 | 0.865 | 1.33% | 0.92 | 1 | |
| All Years | 0.806 | 0.864 | 2.19% | 0.98 | 1 | 11.92 |
| | | | South America | | | |
| 1980-1990 | 0.761 | 0.744 | 0.48% | 0.63 | 5 | 4.71 |
| 1990-2000 | 0.761 | 0.735 | 4.21% | 0.89 | 5 | 5.65 |
| 2000-2010 | 0.760 | 0.731 | 5.59% | 1.14 | 5 | 6.76 |
| 2010-2020 | 0.819 | 0.818 | 1.19% | -2.64 | 5 | 7.01 |
| All Years | 0.773 | 0.754 | 3.09% | 1.04 | 5 | 5.68 |

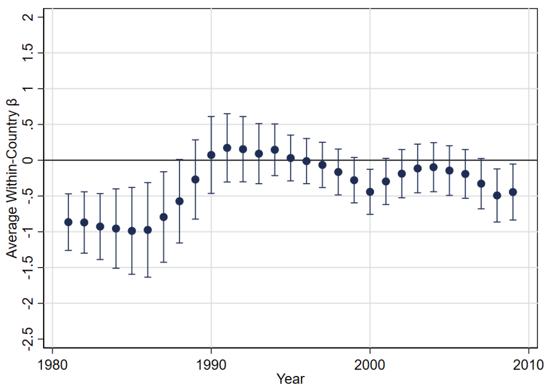
*Notes:* This table reports the main summary statistics such as share continent population, share of continent GDP, average GDP Growth per capita, GDP growth relative to the world average, # of countries and average years of education of our sample by continent and over the decades in our sample.

# D Robustness for Fact #1

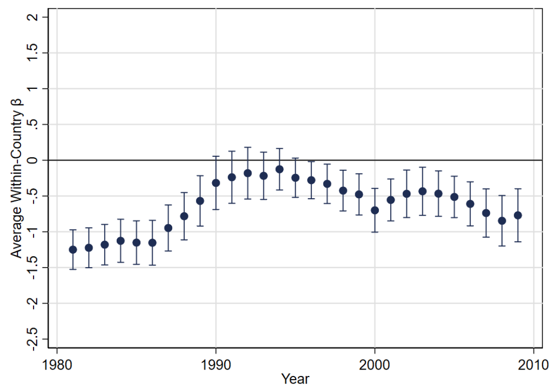
## D.1 Weighting and Heterogeneity

We report robustness exercises for fact #1 where we use the unweighted sample and control for population growth. We also keep the unbalanced panel of countries throughout the entire time periodin figure D.1. We confirm the robustness of fact 1 to different sample variations in figure D.2. We also report the different layers of heterogeneity as shown in figure D.3. Table D.5 reports which country belongs to each heterogeneity group.

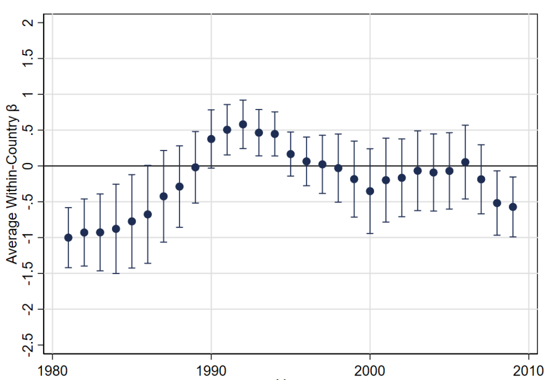
Figure D.1: Convergence over time, robustness to $\beta$ calculation



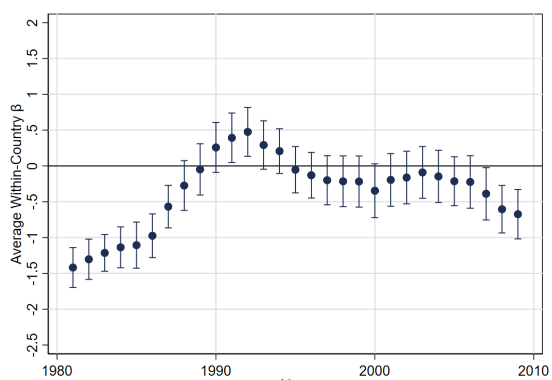

(a) Baseline: population weighted and not controlled for population growth




(b) Alternative: not-population weighted and not controlled for population growth



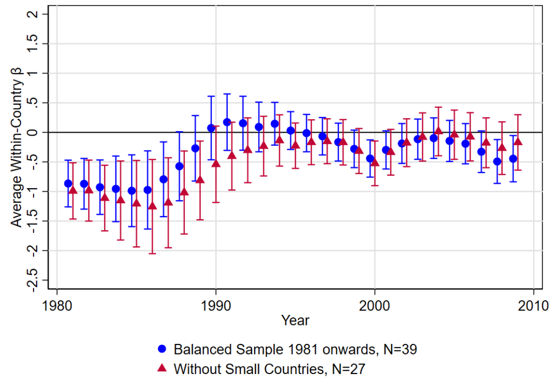

(c) Alternative: population weighted and controlled for population growth



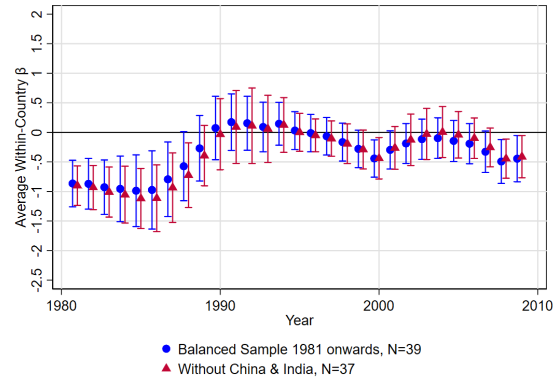

(d) Alternative: not-population weighted and controlled for population growth

*Notes:* This figure shows the the robustness to fact 1 for the 39 countries, where we vary the empirical specification in the way $\beta$ has been calculated.

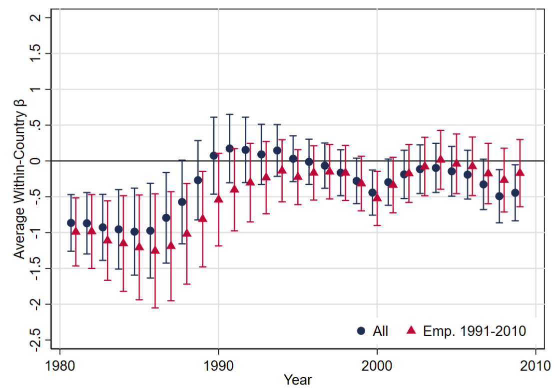
Figure D.2: Convergence over time, robustness to sample definition




(a) Baseline vs. without small countries




(b) Baseline vs. without China and India




(c) Baseline vs. without small countries with employment 1991-2010

*<u>Notes:</u>* This figure shows the the robustness to fact 1 for the 39 countries, where we vary the sample definition.

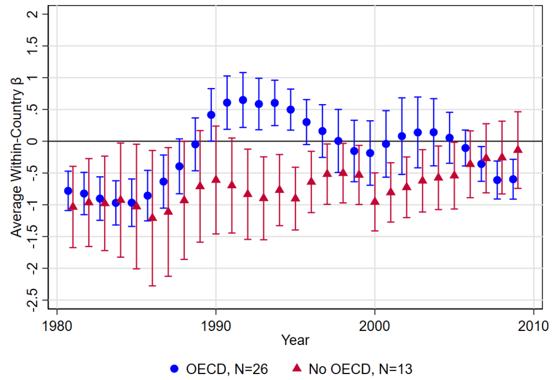
## D.2 Convergence in African countries

A potential concern with our main analysis is that the estimates may not fully capture the dynamics in African countries, where data availability is limited and many nations are still at an early stage of development.

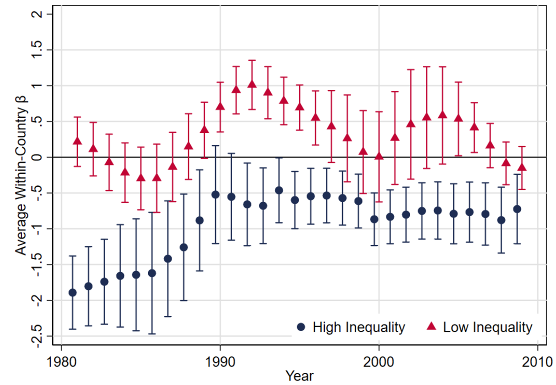
**Convergence between Cities:** Using GDP and population data from 923 cities in 77 countries, we examine the convergence dynamics at the urban level. Figure D.4 illustrates that, between 2004 and 2020, there is a noticeable lack of convergence even among cities within the same country. This dataset, provided by the Economist Intelligence Unit, also includes information for 19 African countries[5].

[5] Angola, Benin, Burkina Faso, Cameroon, Congo-Brazzaville, Congo-Kinshasa, Côte d'Ivoire, Ghana, Kenya, Malawi, Mozambique, Nigeria, Senegal, Somalia, South Africa, Sudan, Tanzania, Zambia, and

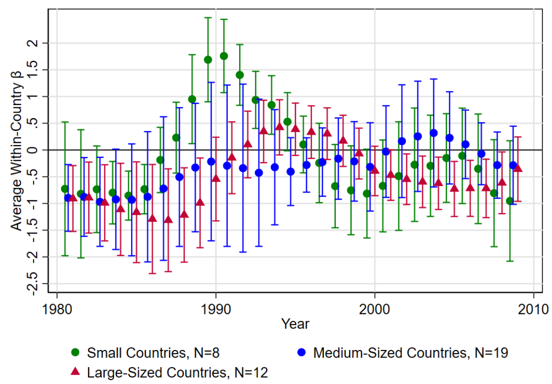
Figure D.3: Convergence heterogeneity over time



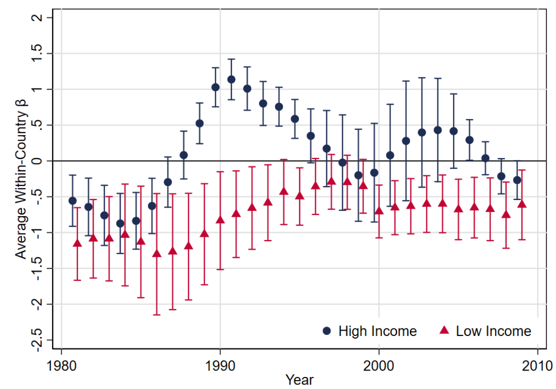

(a) OECD vs. No-OECD



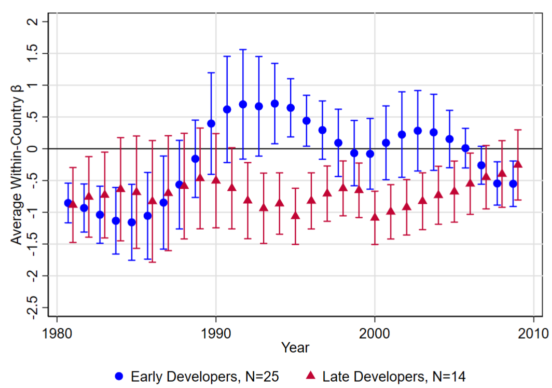

(b) High vs. Low Inequality



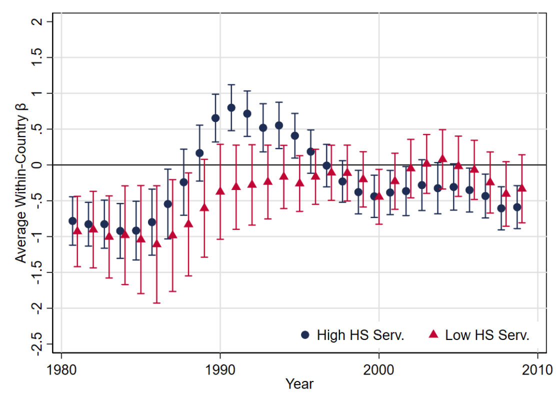

(c) Medium vs. Large Countries




(d) High vs. Low Income




(e) Early vs. Late Developers




(f) High vs. Low HS Priv. Services Share

*Notes:* This figure shows the heterogeneity of within country convergence across different subgroups. The definitions of countries can be found in table D.5

Zimbabwe.

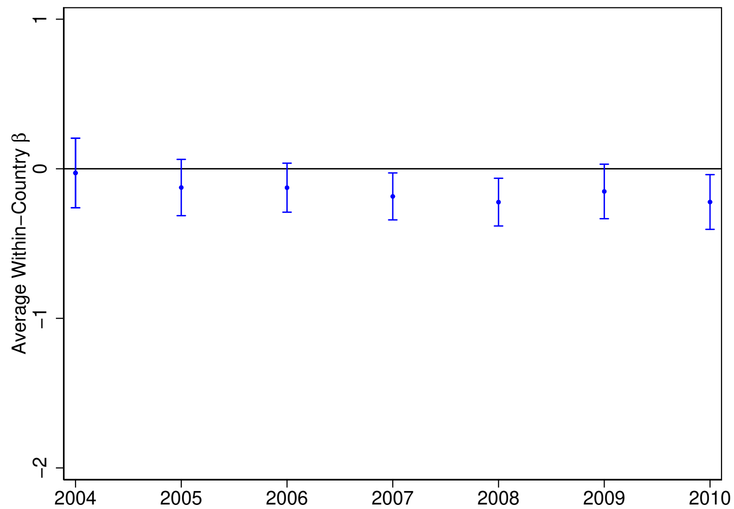



Figure D.4: Within-country $\beta$ convergence between cities (2004–2020).
*Notes:* The figure reports the estimates of within-country $\beta$ convergence using 10-year rolling windows, with the unit of analysis being a city.

**Convergence Using Nighttime Lights as a Proxy for GDP**: As an additional robustness exercise, we use nighttime light data from the Defense Meteorological Satellite Program (DMSP) as a proxy for GDP. This dataset spans from 1993 to 2018, though we limit our analysis to 2014 since in that year sensor changed and this might distort luminosity readings. Figure D.5 displays the evolution of within-country $\beta$ estimates, normalized to the initial year, for the global sample and isolating Africa. Our results indicate an overall decrease in $\beta$-convergence of approximately 0.7 percentage points globally. The main takeaway of this robustness analyis is that Africa is aligned wit the rest of the world in terms of $\beta$-convergence patterns.

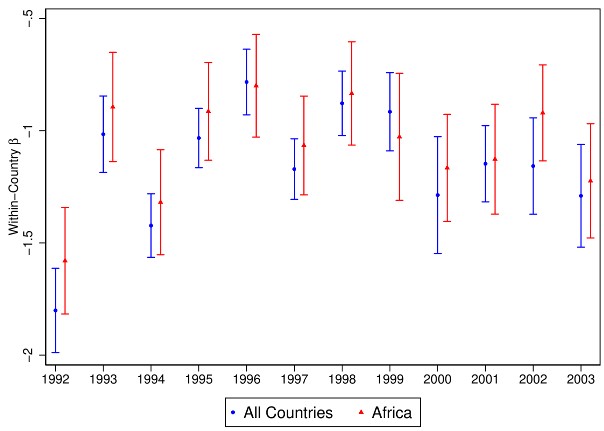



Figure D.5: Within-country $\beta$ convergence using nighttime lights data.

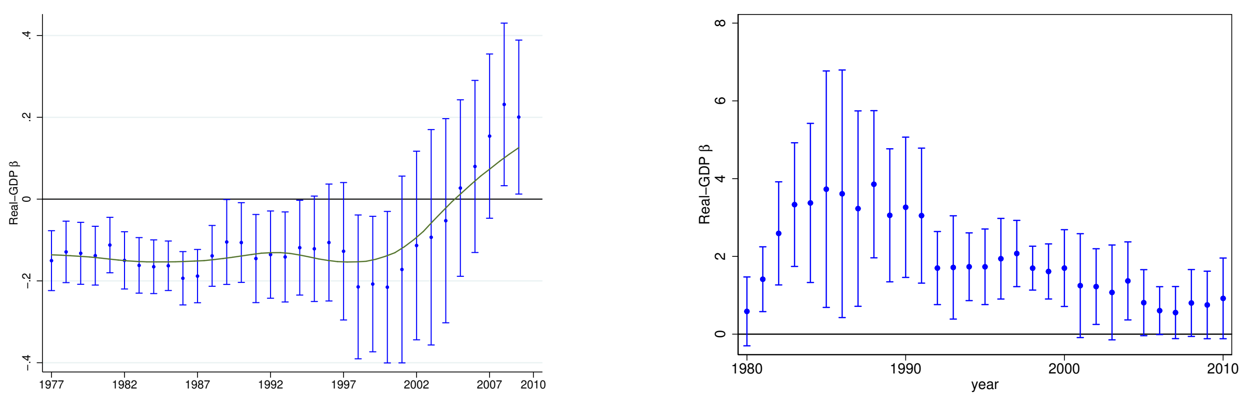
*Notes:* This figure reports the within-country $\beta$ convergence estimates based on 10-year rolling windows for the countries included in the nighttime lights dataset.

In summary, these robustness checks reinforce our headline results by demonstrating that our findings are robust to alternative regional definitions and data sources, particularly through the inclusion of additional African data in both the GDP and nighttime lights analyses.

## D.3 Convergence in Real GDP

A final concern could be that if prices are lower in poorer regions, an observed lack of convergence in nominal GDP may be misleading. Addressing this is much harder since regional price data or GDP deflators are hard to obtain for most countries. For now, we have obtained data on real GDP by states for the United States and India. In figure D.6, we show that in the United States and India, there hasn't been any regional convergence since the 2000s even in real GDP. We are currently working on obtaining similar data for other countries.

Figure D.6: Regional Convergence in Real GDP




(a) United States

(b) India

Table D.5: Country definition

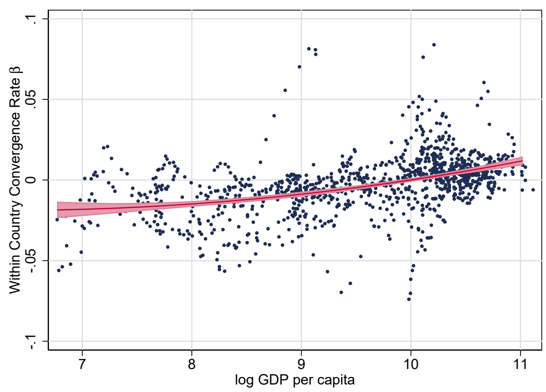
| | Country (1) | Size (2) | High Ineq. (3) | High Income (4) | High HS Services (5) | High HS Serv. Conc. (6) | High Rel. HS Serv. Conc. (7) | Early Developers (8) | OECD (9) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Australia | Medium | 0 | 1 | | | | 1 | 1 |
| 2 | Austria | Small | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 3 | Belgium | Medium | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 4 | Bolivia | Small | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| 5 | Brazil | Large | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| 6 | Canada | Medium | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| 7 | Chile | Medium | 1 | 0 | | | | 1 | 1 |
| 8 | China | Large | 1 | 0 | | | | 0 | 0 |
| 9 | Colombia | Medium | 1 | 0 | 0 | 1 | 1 | 0 | 1 |
| 10 | Denmark | Small | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| 11 | Finland | Small | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| 12 | France | Large | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| 13 | Greece | Medium | 0 | 1 | 0 | 1 | 1 | 1 | 1 |
| 14 | Hungary | Medium | 0 | 1 | | | | 1 | 1 |
| 15 | India | Large | 1 | 0 | | | | 0 | 0 |
| 16 | Indonesia | Large | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 17 | Ireland | Small | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 18 | Italy | Large | 0 | 1 | 0 | 1 | 0 | 1 | 1 |
| 19 | Japan | Large | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 20 | Kenya | Medium | 1 | 0 | | | | 0 | 0 |
| 21 | Malaysia | Medium | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 22 | Netherlands | Medium | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 23 | Norway | Small | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 24 | Peru | Medium | 1 | 0 | | | | 0 | 0 |
| 25 | Philippines | Medium | 1 | 0 | | | | 0 | 0 |
| 26 | Poland | Medium | 0 | 0 | | | | 1 | 1 |
| 27 | Portugal | Medium | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 28 | Republic Of Korea | Large | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| 29 | Romania | Medium | 0 | 0 | | | | 1 | 0 |
| 30 | South Africa | Medium | 0 | 0 | | | | 1 | 0 |
| 31 | Spain | Medium | 0 | 1 | 0 | 1 | 0 | 1 | 1 |
| 32 | Sweden | Small | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| 33 | Switzerland | Small | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 34 | Tanzania | Medium | 1 | 0 | | | | 0 | 0 |
| 35 | Thailand | Medium | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| 36 | Turkey | Large | 1 | 0 | | | | 0 | 1 |
| 37 | Uk | Large | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 38 | United States | Large | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| 39 | West Germany | Large | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| | Total | | 19 | 19 | 13 | 13 | 13 | 25 | 26 |

Note: This table reports the definitions for the heterogeneity analysis. These characteristics are fixed across the time periods and are collected in 1981. The definitions are as follows. High inequality: above median Gini coefficient of GDP per capita in the balanced group. High income: above median GDP per capita in the balanced group. Size: Population size. High HS Services: above median HS private services if exists. Otherwise in this country is not counted. High HS Serv. Conc.: above median HS private services Gini coefficient for which it exists. Otherwise, country is excluded. High Rel. HS. Serv. Conc.: above median HS private services Gini coefficient divided by population gini for which it exists. Otherwise, country is excluded. Early developers: Definition based on Henderson et al. (2017). Equals one if the country is defined as having ”high education”.

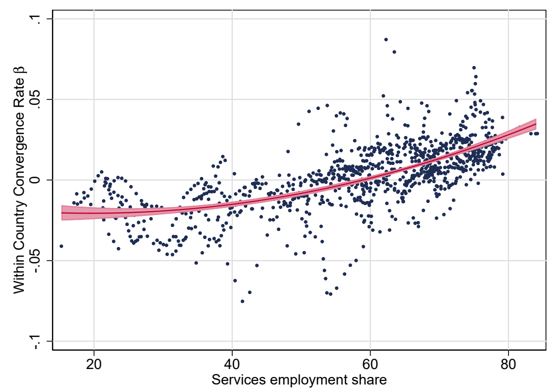
# E Robustness for Fact #2

In this section, we report robustness exercises for fact #2. Specifically, we change specifications to keep a balanced panel and without weights by population size as in figure 4. In all these different scenarios, we find that the results are very similar suggesting that changing specifications does not alter the results discussed in the main text.

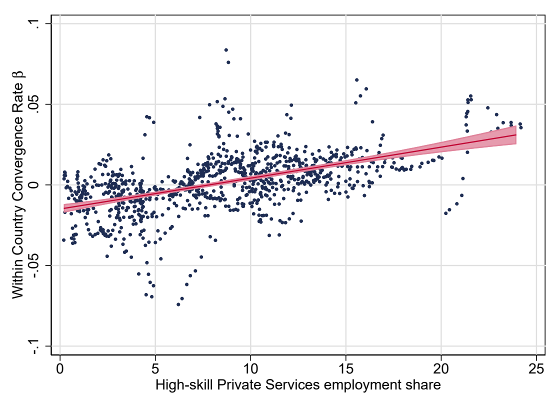
Figure E.7: Structural Transformation and Regional Convergence, balanced data set




(a) Log GDP per capita



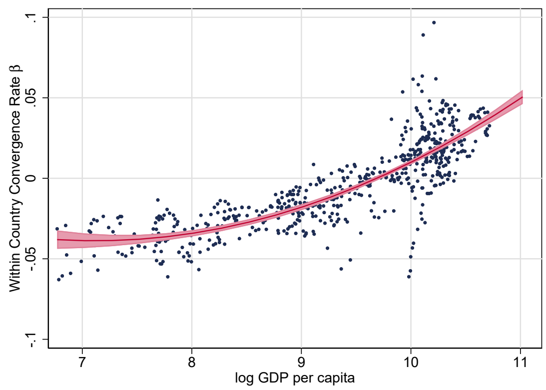

(b) Services employment share



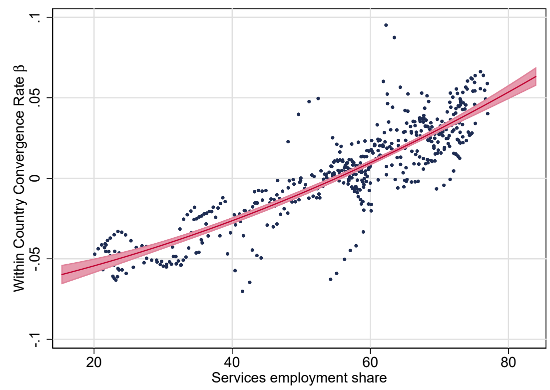

(c) High-skill private services employment share

*Notes:* Population weighted beta vs. log GDP per capita (a), vs. services employment share (b) and the high-skill services employoment share for the balanced panel for 1980-2019. Estimates are residualized off country fixed effects and contain a recession dummy if the convergence measure is calculated in between 1997-2012. The red line shows the evolution of the average country. Confidence intervals are plotted around the estimated.

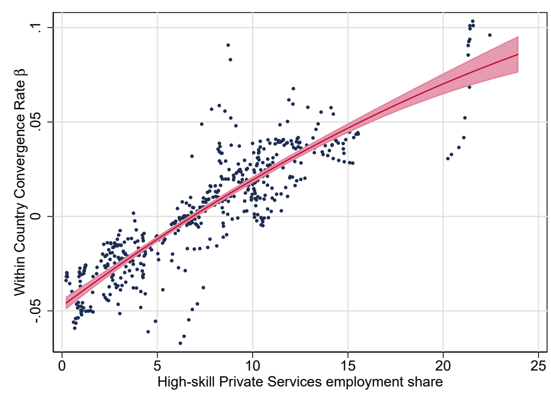
Figure E.8: Structural Transformation and Regional Convergence before 1997




(a) Log GDP per capita



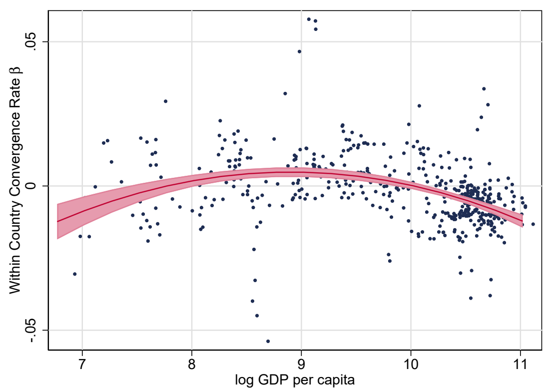

(b) Services employment share



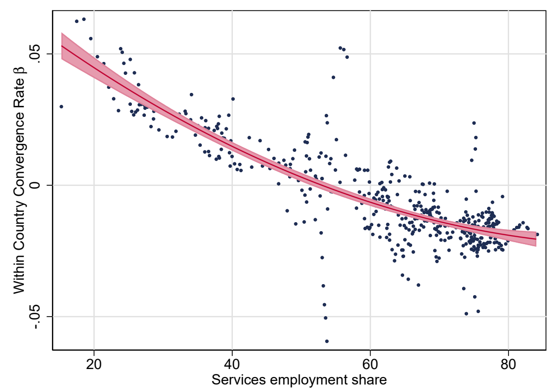

(c) High-skill private services employment share

*Notes:* Population weighted beta vs. log GDP per capita (a), vs. services employment share (b) and the high-skill services employoment share for the balanced panel before 1997. Estimates are residualized off country fixed effects and contain a recession dummy if the convergence measure is calculated in between 1997-2012. The red line shows the evolution of the average country. Confidence intervals are plotted around the estimated.

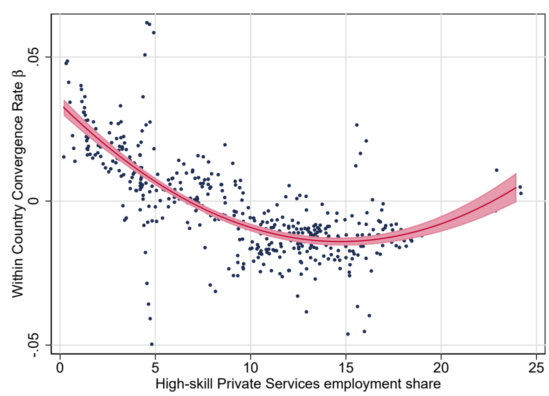
Figure E.9: Structural Transformation and Regional Convergence after 1997




(a) Log GDP per capita



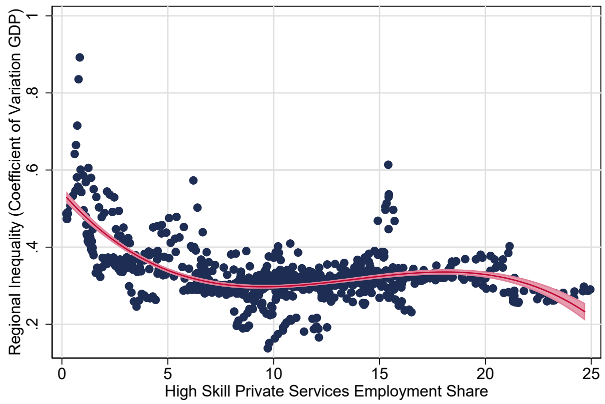

(b) Services employment share




(c) High-skill private services employment share

*Notes:* Population weighted beta vs. log GDP per capita (a), vs. services employment share (b) and the high-skill services employment share for the balanced panel after 1997. Estimates are residualized off country fixed effects and contain a recession dummy if the convergence measure is calculated in between 1997-2012. The red line shows the evolution of the average country. Confidence intervals are plotted around the estimated.

Figure E.10: A Fall and Stagnation of Inequality with Structural Transformation




*Notes:* This figure plots the coefficient of variation of GDP per capita by country against the high-skill private service share in the economy. Estimates are residualized off country fixed effects and contain a recession dummy if the convergence measure is calculated in between 1997-2012. The red line shows the evolution of the average country. The sample is balanced for 1980-2019.

Table E.6: Within-Country Convergence and Structural Transformation

| | (1)<br>10-year | (2)<br>10-year | (3)<br>10-year | (4)<br>10-year | (5)<br>10-year | (6)<br>10-year | (7)<br>10-year |
|---|---|---|---|---|---|---|---|
| Ln GDP pc. | -0.0195 | -0.0235 | -0.0292 | -0.0246 | -0.0220 | -0.0274 | -0.0369 |
| | (0.0050)$^{***}$ | (0.0073)$^{***}$ | (0.0051)$^{***}$ | (0.0035)$^{***}$ | (0.0037)$^{***}$ | (0.0027)$^{***}$ | (0.0067)$^{***}$ |
| Share serv. | | 0.0311 | | | | | |
| | | (0.0749) | | | | | |
| Share high-skill serv. | | | 0.3825 | 0.4356 | | 0.4532 | 0.3398 |
| | | | (0.1052)$^{***}$ | (0.1352)$^{***}$ | | (0.1460)$^{***}$ | (0.1546)$^{**}$ |
| Great Recession | | | | -0.0057 | | -0.0050 | 0.0257 |
| | | | | (0.0046) | | (0.0053) | (0.0136)$^{*}$ |
| RVA per worker | | | | | 0.1620 | 0.1557 | 0.1275 |
| | | | | | (0.0182)$^{***}$ | (0.0136)$^{***}$ | (0.0228)$^{***}$ |
| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | | | | | | | ✓ |
| N | 1094.0000 | 1094.0000 | 1094.0000 | 1094.0000 | 716.0000 | 716.0000 | 716.0000 |
| N country | 39 | 39 | 39 | 39 | 25 | 25 | 25 |
| $R^2$ | 0.6538 | 0.6551 | 0.6869 | 0.6941 | 0.6994 | 0.7370 | 0.7874 |

*Notes*: This table presents the regression estimates where the dependent variable in each specification is the estimated $\beta$-convergence over a 10-year rolling window for each country in our unbalanced panel. "RVA per worker" measures the Real Value Added per workers in the high-skill private service sector that we obtain from the GGDC database. "Great Recession" is an indicator which equals one if the time period period of the 10-year-convergence regression start between the years 1997 and 2012. Specifications include country and/or year fixed effects.

Table E.7: Within-country convergence, structural transformation and labor productivity, balanced sample

| | (1)<br>10-year | (2)<br>10-year | (3)<br>10-year | (4)<br>10-year | (5)<br>10-year | (6)<br>10-year |
|---|---|---|---|---|---|---|
| Ln GDP pc. | -0.0195 | -0.0327 | -0.0306 | -0.0264 | -0.0210 | -0.0295 |
| | (0.0050)*** | (0.0038)*** | (0.0066)*** | (0.0053)*** | (0.0044)*** | (0.0040)*** |
| Share serv. | | 0.0011 | | | | |
| | | (0.0004)*** | | | | |
| Share high-skill serv. | | | 0.0029 | 0.0032 | | 0.0034 |
| | | | (0.0012)** | (0.0012)** | | (0.0014)** |
| Great Recession | | | | -0.0046 | | -0.0032 |
| | | | | (0.0041) | | (0.0053) |
| RVA per worker | | | | | 0.1651 | 0.1724 |
| | | | | | (0.0283)*** | (0.0355)*** |
| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| N | 1131.0000 | 988.0000 | 988.0000 | 988.0000 | 696.0000 | 590.0000 |
| N country | 39 | 38 | 38 | 38 | 24 | 22 |
| $R^2$ | 0.6537 | 0.7215 | 0.7246 | 0.7296 | 0.6855 | 0.7618 |

*Notes:* This table presents the regression estimates where the dependent variable in each specification is the estimated $\beta$-convergence over a 10-year rolling window for each country in our balanced panel. “RVA per worker” is defined as the Real Value Added per number of workers in the high skill private (business) service sector. “Great Recession” is an indicator which equals one if the calculation period for the convergence measure falls between 1997 and 2012. Specifications include country fixed effects. The RVA is calculated from the GGDC database.

Table E.8: Within country convergence, structural transformation and labor productivity, ETD data

| | (1) 10-year | (2) 10-year | (3) 10-year | (4) 10-year | (5) 10-year | (6) 10-year |
|---|---|---|---|---|---|---|
| Ln GDP pc. | -0.0198 | -0.0329 | -0.0304 | -0.0264 | -0.0284 | -0.0238 |
| | (0.0047)*** | (0.0036)*** | (0.0062)*** | (0.0050)*** | (0.0044)*** | (0.0030)*** |
| Share serv. | | 0.0010 | | | | |
| | | (0.0004)*** | | | | |
| Share high-skill serv. | | | 0.0027 | 0.0029 | | -0.0005 |
| | | | (0.0011)** | (0.0012)** | | (0.0011) |
| Great Recession | | | | -0.0045 | | -0.0047 |
| | | | | (0.0040) | | (0.0049) |
| RVA per worker | | | | | 0.0361 | 0.0342 |
| | | | | | (0.0045)*** | (0.0067)*** |
| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| N | 1435.0000 | 1204.0000 | 1204.0000 | 1204.0000 | 389.0000 | 321.0000 |
| N country | 57 | 52 | 52 | 52 | 20 | 18 |
| $R^2$ | 0.6388 | 0.7218 | 0.7241 | 0.7286 | 0.7621 | 0.8360 |

*Notes:* This table presents the regression estimates where the dependent variable in each specification is the estimated $\beta$-convergence over a 10-year rolling window for each country in our unbalanced panel. "RVA per worker" is defined as the Real Value Added per number of workers in the high skill private (business) service sector. "Great Recession" is an indicator which equals one if the calculation period for the convergence measure falls between 1997 and 2012. Specifications include country fixed effects. The RVA is calculated from the ETD database.

Table E.9: Robustness to Fact 2, without Real Value Added

| | (1) 10-year | (2) 10-year | (3) 10-year | (4) 10-year | (5) 10-year | (6) 10-year |
|---|---|---|---|---|---|---|
| Ln GDP pc. | -0.0135 | 0.0007 | -0.0082 | -0.0071 | -0.0128 | -0.0092 |
| | (0.0027)$^{***}$ | (0.0071) | (0.0042)$^{*}$ | (0.0090) | (0.0045)$^{***}$ | (0.0089) |
| Share HS priv. serv. | 0.2621 | | | 0.2381 | 0.2711 | 0.2652 |
| | (0.0641)$^{***}$ | | | (0.0722)$^{***}$ | (0.0648)$^{***}$ | (0.0600)$^{***}$ |
| Share LS serv. | | -0.1282 | | -0.0902 | | |
| | | (0.1190) | | (0.1281) | | |
| Share HS pub. serv. | | | 0.0177 | | -0.0266 | |
| | | | (0.0975) | | (0.1042) | |
| Share LS and HS publ. serv. | | | | | | -0.0436 |
| | | | | | | (0.0847) |
| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | | | | | | |
| N | 1531 | 1531 | 1531 | 1531 | 1531 | 1531 |
| N country | 53 | 53 | 53 | 53 | 53 | 53 |
| $R^2$ | 0.5633 | 0.5450 | 0.5332 | 0.5690 | 0.5639 | 0.5671 |

*Notes:* This table presents the regression estimates where the dependent variable in each specification is the estimated $\beta$-convergence over a 10-year rolling window for each country in our unbalanced panel. Specifications include country fixed effects.

Table E.10: Robustness to Fact 2, with Real Value Added

| | (1) 10-year | (2) 10-year | (3) 10-year | (4) 10-year | (5) 10-year | (6) 10-year |
|---|---|---|---|---|---|---|
| Ln GDP pc. | -0.0237 | -0.0215 | -0.0272 | -0.0233 | -0.0236 | -0.0229 |
| | (0.0068)$^{***}$ | (0.0070)$^{***}$ | (0.0081)$^{***}$ | (0.0051)$^{***}$ | (0.0069)$^{***}$ | (0.0058)$^{***}$ |
| Share HS priv. serv. | 0.2689 | | | 0.2643 | 0.2718 | 0.2606 |
| | (0.0876)$^{***}$ | | | (0.1214)$^{**}$ | (0.0960)$^{***}$ | (0.0980)$^{**}$ |
| RVA per worker, HS priv. serv. | 0.0741 | 0.0723 | 0.0713 | 0.0741 | 0.0720 | 0.0726 |
| | (0.0328)$^{**}$ | (0.0319)$^{**}$ | (0.0306)$^{**}$ | (0.0330)$^{**}$ | (0.0325)$^{**}$ | (0.0335)$^{**}$ |
| Great Recession | 0.0173 | 0.0470 | 0.0501 | 0.0177 | 0.0199 | 0.0201 |
| | (0.0187) | (0.0135)$^{***}$ | (0.0220)$^{**}$ | (0.0222) | (0.0224) | (0.0253) |
| Share LS serv. | | -0.1171 | | -0.0077 | | |
| | | (0.0654)$^{*}$ | | (0.0902) | | |
| Share HS pub. serv. | | | -0.0096 | | -0.0244 | |
| | | | (0.0731) | | (0.0812) | |
| Share LS and HS pub. serv. | | | | | | -0.0173 |
| | | | | | | (0.0632) |
| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| N | 951 | 951 | 951 | 951 | 951 | 951 |
| N country | 28 | 28 | 28 | 28 | 28 | 28 |
| $R^2$ | 0.7318 | 0.7160 | 0.7100 | 0.7318 | 0.7322 | 0.7321 |

*Notes:* This table presents the regression estimates where the dependent variable in each specification is the estimated $\beta$-convergence over a 10-year rolling window for each country in our unbalanced panel. "Great Recession" is an indicator which equals one if the calculation period for the convergence measure falls between 1997 and 2012. Specifications include country and year fixed effects.

# F Robustness for Fact#3

This section reports a series of robustness tests for our finding that high-skill private services are more geographically concentrated within a country and over time. We employ different measures of concentration as the Gini index as well as the Gini index divided by the Gini indexed ratio between sectoral and population concentration. We also switch the sample from balanced to unbalanced. Finally, we report the change in Gini ratio over the full time period for countries that had the highest decrease in high-skill private service employment in terms of the Gini ratio.

Table F.11: Concentration measures in unbalanced data set

| | Gini<br>(1) | Gini Ratio<br>(2) | HHI<br>(3) |
|---|---|---|---|
| Agriculture | .37 | .97 | .15 |
| Manufacturing | .46 | 1.17 | .19 |
| LS Services | .48 | 1.21 | .2 |
| HS priv. Services | .59 | 1.54 | .3 |
| HS pub. Services | .45 | 1.15 | .2 |

*Notes*: This table displays the sectoral concentration for the period 1980-2010 the unbalanced sample. Gini-Ratio is defined as the ratio between the sectoral Gini and the Gini of overall employment. Author's calculation.

Table F.12: Countries with the highest sectoral concentration decrease

| | Country | Δ Agric. | Country | Δ Manuf. | Country | Δ LS. Serv. | Country | Δ HS. priv. Serv. | Country | Δ HS. pub. Serv. |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Portugal | -.127 | Greece | -.123 | Indonesia | -.09 | Benin | -.269 | Guatemala | -.187 |
| 2 | Colombia | -.116 | Mexico | -.114 | Mexico | -.077 | Thailand | -.203 | Thailand | -.109 |
| 3 | Indonesia | -.097 | Brazil | -.11 | Costa Rica | -.074 | Guatemala | -.183 | Mexico | -.102 |
| 4 | Greece | -.073 | Costa Rica | -.104 | Brazil | -.073 | Ireland | -.14 | Costa Rica | -.092 |
| 5 | Italy | -.065 | Benin | -.089 | Portugal | -.063 | Denmark | -.119 | Indonesia | -.09 |
| 6 | Sweden | -.05 | Togo | -.086 | Guatemala | -.058 | Mexico | -.118 | Benin | -.082 |
| 7 | France | -.042 | Uruguay | -.065 | Thailand | -.051 | Greece | -.117 | Brazil | -.073 |
| 8 | Norway | -.04 | Denmark | -.047 | Belgium | -.046 | Netherlands | -.116 | Ireland | -.058 |
| 9 | Netherlands | -.032 | Thailand | -.045 | Netherlands | -.042 | Belgium | -.115 | Bolivia | -.039 |
| 10 | Canada | -.031 | Switzerland | -.037 | Dominican Republic | -.037 | Costa Rica | -.096 | Togo | -.036 |

*Notes:* This table displays the countries with the highest sectoral concentration decrease as measured with the Gini Index. Author's calculation.

## F.1 Robustness for all Facts Excluding Small Countries

In this section, we report all the main facts from 1 to 3, excluding all the small countries. Our definition of small countries is that they have fewer than 10 million inhabitants. The concern is that our results are driven by small internal geographies that, overall, are not important in the aggregate. One hypothesis could be that many rich countries are small like Switzerland and Sweden, and since these are high-income countries, they might have converged already and, hence, are driving the lack of convergence results. Thus, besides weighing for population size, which also reduces the relevance of this group, we went an extra step and reran the analysis for the main facts, excluding them from the sample. By doing so, we are left with 27 countries for which we have regional GDP data since 1980. The list of the remaining countries is reported in table F.13. As is noticeable from the table, all the main countries are still there, including Kenya, South Africa and Tanzania, the three African countries in our GDP sample. We confirm our main findings for the three main facts, excluding the small countries.

Table F.13: Country definition

| | Country (1) | Size (2) | High Ineq. (3) | High Income (4) | OECD (5) | Early Developers (6) | Comm. Exp. (7) |
|---|---|---|---|---|---|---|---|
| 1 | Australia | Medium | 0 | 1 | 1 | 1 | 0 |
| 2 | Brazil | Large | 1 | 0 | 0 | 0 | 0 |
| 3 | Canada | Medium | 0 | 1 | 1 | 1 | 0 |
| 4 | Chile | Medium | 1 | 1 | 1 | 1 | 0 |
| 5 | China | Large | 1 | 0 | 0 | 0 | 0 |
| 6 | Colombia | Medium | 1 | 0 | 1 | 0 | 1 |
| 7 | France | Large | 0 | 1 | 1 | 1 | 0 |
| 8 | India | Large | 1 | 0 | 0 | 0 | 0 |
| 9 | Indonesia | Large | 1 | 0 | 0 | 0 | 0 |
| 10 | Italy | Large | 0 | 1 | 1 | 1 | 0 |
| 11 | Japan | Large | 0 | 1 | 1 | 1 | 0 |
| 12 | Kenya | Medium | 1 | 0 | 0 | 0 | 0 |
| 13 | Malaysia | Medium | 0 | 0 | 0 | 0 | 0 |
| 14 | Netherlands | Medium | 0 | 1 | 1 | 1 | 0 |
| 15 | Peru | Medium | 1 | 0 | 0 | 0 | 1 |
| 16 | Philippines | Medium | 1 | 0 | 0 | 0 | 0 |
| 17 | Poland | Medium | 0 | 0 | 1 | 1 | 0 |
| 18 | Republic Of Korea | Large | 1 | 0 | 1 | 1 | 0 |
| 19 | Romania | Medium | 0 | 0 | 0 | 1 | 0 |
| 20 | South Africa | Medium | 1 | 1 | 0 | 1 | 0 |
| 21 | Spain | Medium | 0 | 1 | 1 | 1 | 0 |
| 22 | Tanzania | Medium | 1 | 0 | 0 | 0 | 0 |
| 23 | Thailand | Medium | 1 | 0 | 0 | 0 | 0 |
| 24 | Turkey | Large | 1 | 1 | 1 | 0 | 0 |
| 25 | Uk | Large | 0 | 1 | 1 | 1 | 0 |
| 26 | United States | Large | 0 | 1 | 1 | 1 | 0 |
| 27 | West Germany | Large | 0 | 1 | 1 | 1 | 0 |
| | Total | | 14 | 13 | 15 | 15 | 2 |

*Notes:* This table reports the definitions for the heterogeneity analysis. These characteristics are fixed across the time periods and are collected in 1981. The definitions are as follows. High inequality: above median Gini coefficient of GDP per capita (0.175). High income: above median GDP per capita (7504). Size:

Population size. Early developers: Definition based on Henderson et al. (2017). It equals one if the country is defined as having "high education". Comm. Exp.: whether the country is a commodity exporter.

Table F.14 reports the number of countries that have GDP data and the subset of those that also have regional employment data between 1990 and 2010 as well as GGDC. Therefore, when we do joint GDP and regional-level analysis, we will be left with a smaller sample. Figure F.11 reports the central $\beta$-convergence fact and its lack thereof, showing very similar

Table F.14: Summary table for GDP and employment data

| Region | GDP | | +Emp. | | | +GGDC | | |
|---|---|---|---|---|---|---|---|---|
| | Nb. Countries | 1990-2010 | Nb. Countries | Avg. Nb. Years | 1990-2010 | Nb. Countries | Avg. Nb. Years | 1990-2010 |
| Africa | 3 | 3 | 2 | 44 | 0 | 3 | 47 | 3 |
| Asia | 9 | 9 | 9 | 48 | 8 | 9 | 48 | 9 |
| Australia and Oceania | 1 | 1 | 1 | 38 | 1 | 1 | 38 | 1 |
| East Europe | 2 | 2 | 2 | 39 | 1 | 2 | 39 | 2 |
| North America | 2 | 2 | 2 | 64 | 2 | 2 | 64 | 2 |
| South America | 4 | 4 | 4 | 49 | 3 | 4 | 49 | 4 |
| West Europe | 6 | 6 | 6 | 39 | 6 | 6 | 39 | 6 |
| Total | 27 | 27 | 26 | | 21 | 27 | | 27 |

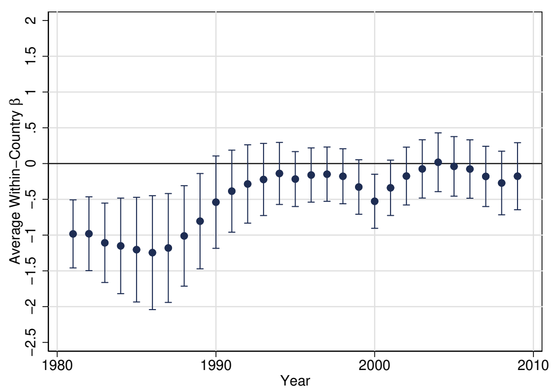
*Notes:*: This table shows the number of countries which we use mainly in our analysis and that contain GDP and employment data from our as well as GGDC/ETD on the country level and the number of average years per country. This table only displays the countries that are not "small", i.e. that have more than 10 million inhabitants in any given year. The values are split by country groups. "+ Emp." refers to countries that have employment and GDP data from our sources solely. "+ GGDC" are the sample of countries that in "+ Emp." but extended with the national employment data from GGDC and/or ETD. Author's calculation.

quantitative patterns also for the list of non-small countries, both in the case of unconditional and conditional convergence.

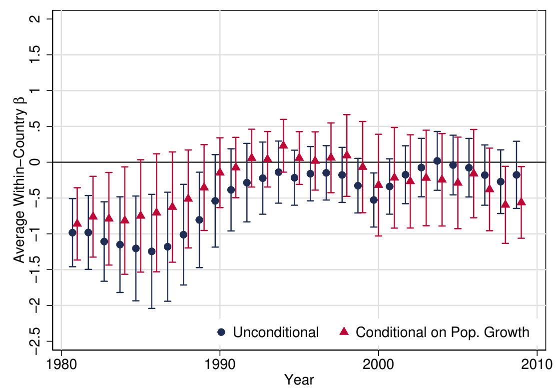
Table F.15 reports the summary statistics of countries that have ever converged in the period. According to our definition, the results do not significantly differ from those in the main text, with a slight increase in countries that had "ever converging" episodes.

We also replicate fact 2 for this smaller sample. We find the fact to be very robust to the inclusion of both country and year fixed effects, as shown both in figure F.12 and in table F.17. There is a strong, positive and statistically significant correlation between decline in $\beta$-convergence and share high-skill services, as well as with real value added per worker in high-skill services.

Figure F.11: Within-Country $\beta$ Over Time




(a) Baseline: Population-weighted regression




(b) Conditional and unconditional convergence

*Notes:* This figure reports the average within-country $\beta$ convergence for a balanced sample of 27 countries between 1981 and 2019 which possess employment data between 1991 and 2010. In panel (a) the regressions for each country are weighted by population size, panel (b) compares unconditional and conditional convergence rates, where the latter control for population growth

.

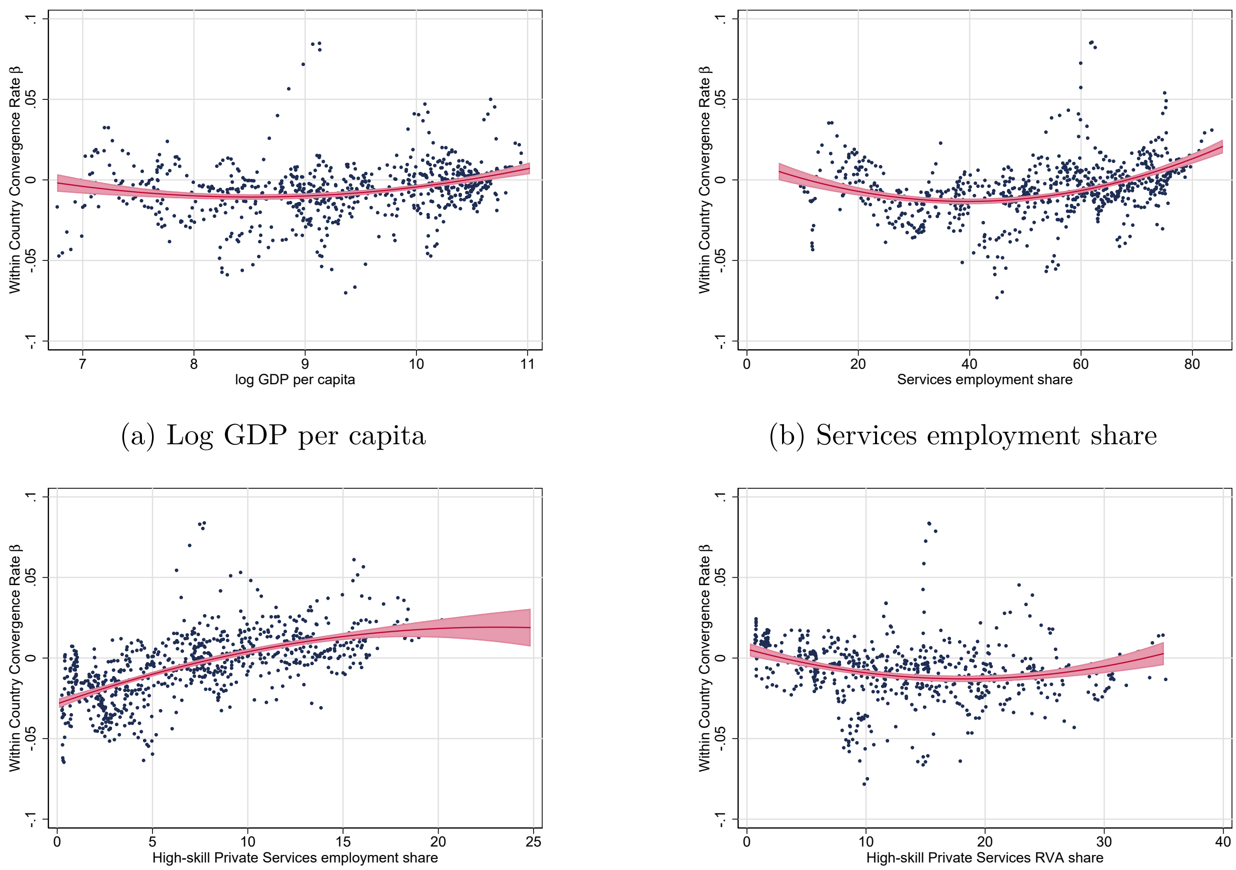
Table F.15: The Decline in Within-Country Convergence

| | Ever converged | | |
|---|---|---|---|
| | 1981-1989 | 1991-1999 | 2001-2009 |
| Share of countries | 74.1% | 59.3% | 59.3% |
| Share of GDP | 76.7% | 50.3% | 54.3% |
| Share of population | 67.9% | 60.2% | 59.1% |

*Notes:* This table reports the share of countries which has ever converged in a given decade, which is defined as having at least one $\beta$ estimate in this period that is negative and statistically significant at the 5% significance level. We further show the share of our sample's GDP and population that is represented by the respective countries. We use the 27 countries which have employment data between 1991-2010.

While our employment data is not restricted to only 27 countries, for completeness, we report fact 3 for the non-small countries in table F.18 and figure F.13. Overall, the main results are that high-skill private services are more concentrated than other services, and any other sectors still stand independently of the measure we use. This suggests that the result does not come from the fact that small countries in general have only one main hub where business services happen.

Figure F.12: Structural Transformation and Regional Convergence




(a) Log GDP per capita

(b) Services employment share

(c) High-skill private services empl. share

(d) High-skill private services RVA share

*Notes:* Population weighted beta vs. log GDP per capita (a), vs. services employment share (b), the high-skill services employment share (c) and the the high-skill services real value added share (d) for the countries which have employment from 1991-2010 (27 countries). Estimates are residualized off year and country fixed effects. The red line shows the evolution of the average country. Confidence intervals are plotted around the estimated.

## F.2 Robustness for $\sigma$-convergence

An alternative regional convergence measure is the $\sigma$-convergence and its change over time as defined in equation 5. The literature often uses this measure as the main measure of regional convergence as it allows for the study of the whole distribution. We also examine how our estimates of $\beta$-convergence relate to changes in the cross-regional dispersion of incomes ($\sigma$-convergence) in Figure F.14. We compute $\sigma$-convergence as the decade-on-decade change in the population-weighted coefficient of variation of regional GDP per capita. Across our full sample of country-periods, we find a robust negative association between $\beta$ and $\Delta\sigma$: economies exhibiting faster conditional catch-up almost invariably experience larger reductions

Table F.16: Within-Country Convergence and Structural Transformation

| | (1) 10-year | (2) 10-year | (3) 10-year | (4) $\Delta CoV$ | (5) 10-year | (6) $\Delta CoV$ |
|---|---|---|---|---|---|---|
| Ln GDP pc. | -0.0075 | -0.0094 | -0.0135 | 0.0241 | -0.0237 | 0.0060 |
| | (0.0038)* | (0.0078) | (0.0027)*** | (0.0151) | (0.0068)*** | (0.0256) |
| Share serv. | | 0.0151 | | | | |
| | | (0.0740) | | | | |
| Share HS priv. serv. | | | 0.2621 | 0.6178 | 0.2689 | 0.4949 |
| | | | (0.0641)*** | (0.3091)* | (0.0876)*** | (0.2769)* |
| Great Recession | | | | -0.0446 | 0.0173 | -0.0382 |
| | | | | (0.0230)* | (0.0187) | (0.0723) |
| RVA per worker, HS priv. serv. | | | | | 0.0741 | 0.2145 |
| | | | | | (0.0328)** | (0.4075) |
| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | | | | | ✓ | ✓ |
| N | 1531 | 1531 | 1531 | 1520 | 951 | 940 |
| N country | 53 | 53 | 53 | 53 | 28 | 28 |
| $R^2$ | 0.5330 | 0.5335 | 0.5633 | 0.4118 | 0.7318 | 0.6169 |

*Notes*: This table presents the regression estimates where the dependent variable in each specification is the estimated $\beta$-convergence over a 10-year rolling window for each country in our unbalanced panel. "RVA per worker" measures the Real Value Added per workers in the high-skill private service sector that we obtain from the GGDC database. "Great Recession" is an indicator which equals one if the time period of the 10-year-convergence regression start between the years 1997 and 2012. Specifications include country and/or year fixed effects. Robust standard errors clustered at the country level are reported in parenthesis. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table F.17: Within-Country Convergence and Structural Transformation

| | (1) 10-year | (2) 10-year | (3) 10-year | (4) 10-year | (5) 10-year | (6) 10-year | (7) 10-year | (8) 10-year |
|---|---|---|---|---|---|---|---|---|
| Ln GDP pc. | -0.0198 | -0.0244 | -0.0297 | -0.0252 | -0.0222 | -0.0277 | -0.0238 | -0.0379 |
| | (0.0050)*** | (0.0072)*** | (0.0051)*** | (0.0034)*** | (0.0036)*** | (0.0026)*** | (0.0106)** | (0.0066)*** |
| Share serv. | | 0.0358 | | | | | | |
| | | (0.0760) | | | | | | |
| Share high-skill serv. | | | 0.3986 | 0.4523 | | 0.4590 | 0.4697 | 0.3300 |
| | | | (0.1103)*** | (0.1418)*** | | (0.1498)*** | (0.2074)** | (0.1580)** |
| Great Recession | | | | -0.0056 | | -0.0050 | 0.0000 | 0.0273 |
| | | | | (0.0048) | | (0.0053) | (.) | (0.0136)* |
| RVA per worker | | | | | 0.1616 | 0.1547 | -0.0175 | 0.1255 |
| | | | | | (0.0179)*** | (0.0135)*** | (0.0326) | (0.0228)*** |
| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Year FE | | | | | | | ✓ | ✓ |
| N | 757.0000 | 757.0000 | 757.0000 | 757.0000 | 629.0000 | 629.0000 | 629.0000 | 629.0000 |
| N country | 27 | 27 | 27 | 27 | 22 | 22 | 22 | 22 |
| $R^2$ | 0.6626 | 0.6643 | 0.6978 | 0.7048 | 0.7034 | 0.7417 | 0.3850 | 0.7918 |

*Notes*: This table presents the regression estimates where the dependent variable in each specification is the estimated $\beta$-convergence over a 10-year rolling window for each country in our unbalanced panel. "RVA per worker" measures the Real Value Added per workers in the high-skill private service sector that we obtain from the GGDC database. "Great Recession" is an indicator which equals one if the time period period of the 10-year-convergence regression start between the years 1997 and 2012. Specifications include country and/or year fixed effects.

Table F.18: Regional Concentration of Sectoral Employment

| | Gini (1) | Gini Ratio (2) | HHI (3) |
|---|---|---|---|
| Agriculture | .41 | .94 | .11 |
| Manufacturing | .48 | 1.1 | .14 |
| LS Services | .48 | 1.1 | .14 |
| HS priv. Services | .55 | 1.3 | .19 |
| HS pub. Services | .44 | 1.02 | .13 |

*Notes*: This table measures the regional concentration of sectoral employment in 1991 and 2010 for 26 countries that have regional employment data between 1991 and 2010. For each country-year, we compute the Gini and Herfindahl index of sectoral employment across all regions. "Gini-Ratio" divides the Ginis of sectoral employment by the Gini of overall employment to adjust for countries' heterogeneity in the overall size distribution of regions.

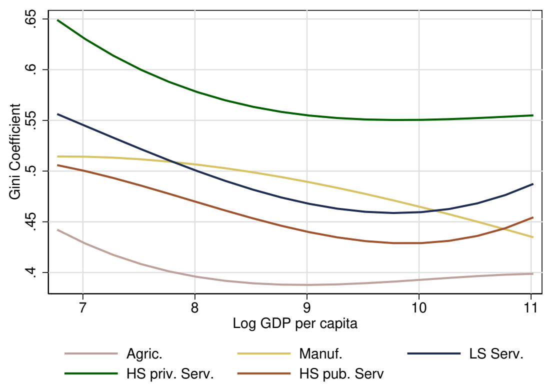
in income dispersion (more negative $\Delta\sigma$). This tight linkage—$\beta$- and $\Delta\sigma$-convergence are highly correlated—confirms that, in our context, faster convergence in growth rates does indeed translate into a narrowing of spatial inequality, and underpins our subsequent focus on $\beta$-convergence as a parsimonious summary of regional income dynamics.

We report the result for our main table of interest, in which we study the evolution of $\sigma$-convergence and its relationship with service and GDP per capita. The results reported in Table F.19 show that the two measures are aligned in a very robust manner in terms of how they relate to the service economy, particularly the high-skill service economy. Countries with a higher share of high-skill service sector are associated with less $\sigma$-convergence, and countries with higher real value added in high-skill services are associated with less $\sigma$-convergence.

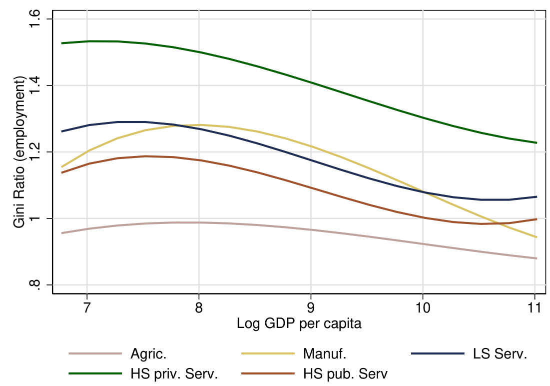
## F.3 Other Results

We report below two other findings related to $\beta$-convergence across countries to complement the central fact of the decline of $\beta$-convergence within countries. We then report an observation about the relationship between economic growth and inequality within a country and across individuals to highlight the different roles of regional and individual inequality on economic growth. Finally, we complement our fact # 2 with a "growth-style" regression in which we assess the role of alternative forces on the change in $\beta$-convergence within-country. We confirm

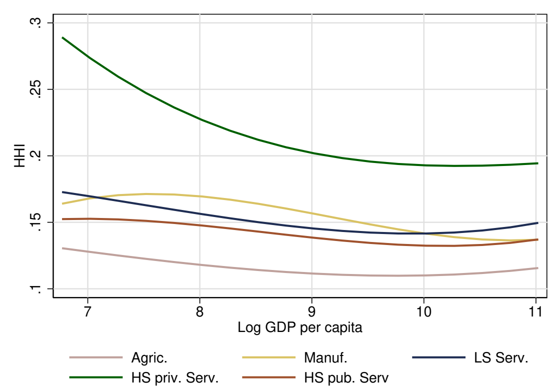
Figure F.13: Sectoral Concentration By Development




(a) Gini Coefficient




(b) Gini-Ratio (employment)

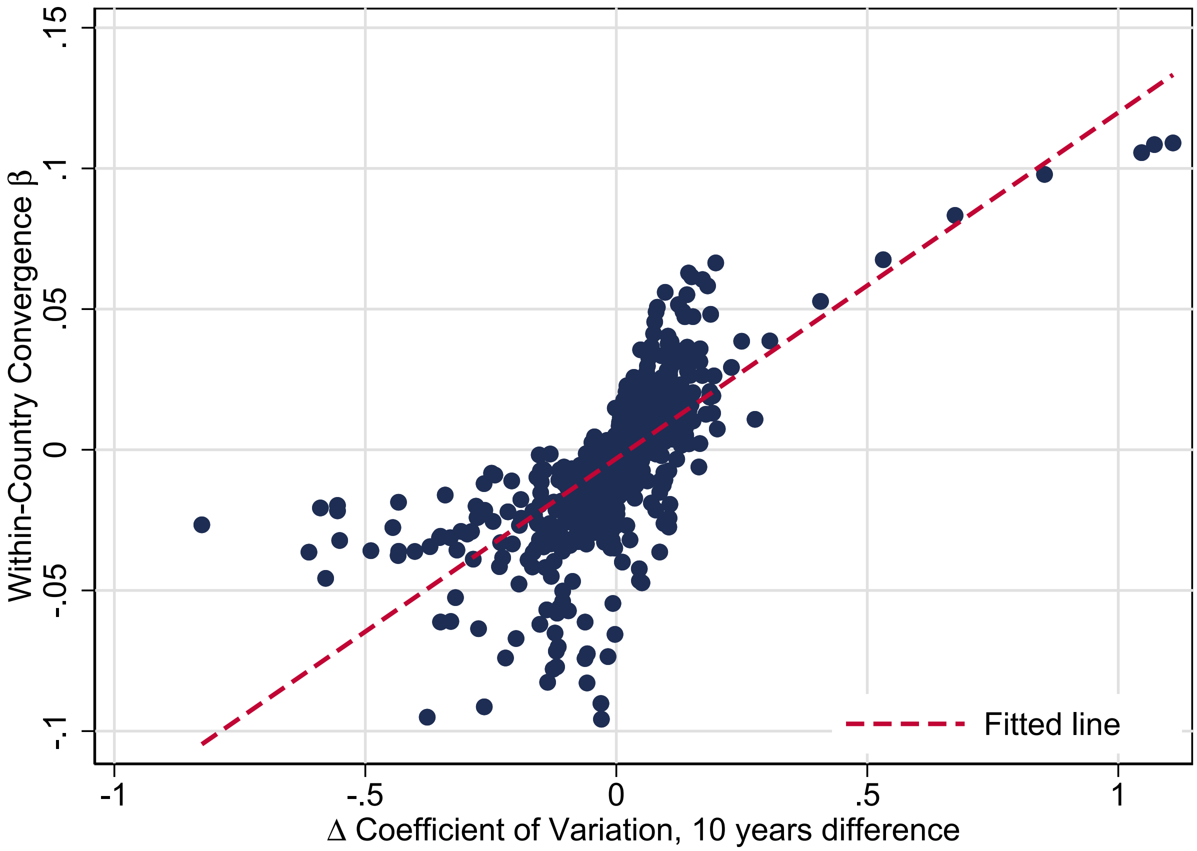



(c) HHI

*Notes:* This figure shows sectoral concentration along log GDP per capita in from 1980-2010 for the 26 countries that have employment from 1991-2010 (the 27 countries in table XX except Kenya). The solid line represents a quadratic fit residualized off country fixed effects. For each country-year, the "Gini-Ratio (employment)" is defined as the Gini Coefficient of sectoral employment divided by the Gini Coefficient of overall employment.

the hypothesis above that structural transformation has the most significant role overall.

Figure F.14: Relationship between $\Delta\sigma$ and $\beta$




*Notes:* This figure displays the change in coefficient of variation and the within-country convergence rate for a balanced sample of 39 countries between 1981 and 2019.

.

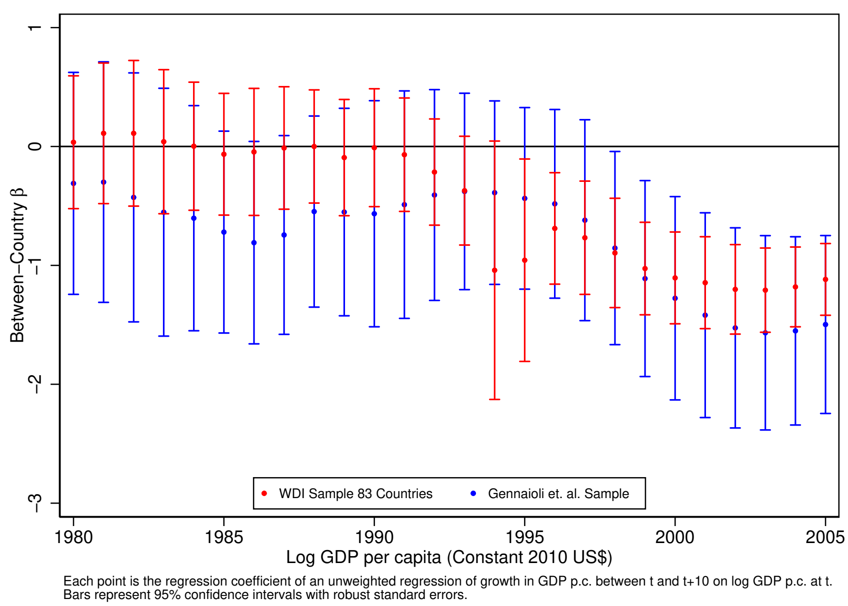
Table F.19: Within-Country Convergence and Structural Transformation

| | (1)<br>10-year | (2)<br>10-year | (3)<br>10-year | (4)<br>10-year | (5)<br>10-year | (6)<br>10-year | (7)<br>10-year |
|---|---|---|---|---|---|---|---|
| Ln GDP pc. | -0.0206 | -0.0237 | -0.0303 | -0.0259 | -0.0220 | -0.0274 | -0.0369 |
| | (0.0047)*** | (0.0077)*** | (0.0049)*** | (0.0031)*** | (0.0037)*** | (0.0027)*** | (0.0067)*** |
| Share serv. | | 0.0248 | | | | | |
| | | (0.0798) | | | | | |
| Share high-skill serv. | | | 0.4129 | 0.4636 | | 0.4532 | 0.3398 |
| | | | (0.1169)*** | (0.1498)*** | | (0.1460)*** | (0.1546)** |
| Great Recession | | | | -0.0054 | | -0.0050 | 0.0257 |
| | | | | (0.0050) | | (0.0053) | (0.0136)* |
| RVA per worker | | | | | 0.1620 | 0.1557 | 0.1275 |
| | | | | | (0.0182)*** | (0.0136)*** | (0.0228)*** |
| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year FE | | | | | | | ✓ |
| N | 716.0000 | 716.0000 | 716.0000 | 716.0000 | 716.0000 | 716.0000 | 716.0000 |
| N country | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| $R^2$ | 0.6739 | 0.6747 | 0.7073 | 0.7135 | 0.6994 | 0.7370 | 0.7874 |

*Notes*: This table presents the regression estimates where the dependent variable in each specification is the estimated $\beta$-convergence over a 10-year rolling window for each country in our unbalanced panel. "RVA per worker" measures the Real Value Added per worker in the high-skill private service sector that we obtain from the GGDC database. "Great Recession" is an indicator which equals one if the period of the 10-year-convergence regression starts between the years 1997 and 2012. Specifications include country and/or year fixed effects.

### F.3.1 $\beta$ Cross-country Convergence increased over time

Figure F.15: $\beta$ Cross-Country Convergence



### F.3.2 National economic growth is positively correlated with spatial income inequality but negatively correlated with individual income inequality

We document how economic growth correlates with inequality at individual and at regional level reporting results in table F.20. Regional inequality is captured by our $\beta$ estimates from fact 1. Individual inequality is measured with Gini coefficients and Gini growth. In column 1 we correlate GDP growth over 10 years at country level with the beta estimates. We control for year fixed effects and we cluster the standard errors at country level. We find that the coefficient is positive but it is not statistically significant. In column 2 we regress GDP growth on initial Gini coefficient. Similarly to column 1, we find a positive coefficient but no statistical significance. In column 3 we regress GDP growth on both $\beta$ estimates and Gini coefficients. The $\beta$ estimates report a coefficients very close to 0 and not statistically significant. Instead, the Gini coefficient is positively correlated and statistically significant at 90%. In column 4, to take into account both changes in individual inequality and differences in initial level of GDP, we find that the estimate on the Gini coefficient becomes negative as well as the sign on the growth of Gini coefficient. In the remaining columns we had controls for potential drivers of economic growth that might also be correlated with regional and individual inequality measures.

We start from democracy indicators to account for how institutions might drive growth. We then add controls for education years to proxy for human capital levels. Then, we complement the analysis by adding proxies for structural transformation such as agricultural share and agricultural productivity growth. To account for geography we include controls such as roads per capita and total road. We then account for trade openness of the country by adding a measure of foreign trade agreement. In each of these specifications we notice that the coefficient on $\beta$ stays positive and in the order between 0.04 and 0.12 but it is not statistically significant. Instead, the coefficient on Gini is negative, ranging between -.02 and -.09 and statistically significant in most of the cases. Finally, in the last column we add all the controls described before. This allows to control for co-founders that could drive the relationship between inequality and economic growth.

We find that the coefficient estimate on Within-country $\beta$ is equal to .22 and statistically significant at 99%. This is in stark contrast with the estimate on both the Gini coefficient the Gini coefficient growth that are respectively equal to -.08 and -32.63 and both statistically significant at 99%. Therefore, we conclude that while regional inequality (higher $\beta$) is positively correlated with economic growth, individual inequality and individual inequality growth are negatively correlated with GDP growth.

This result is important since it highlights a different role of space in affecting growth Within-country convergence is negatively related to a country's growth in agricultural productivity. This is presumably because the latter is a strong predictor of structural transformation as documented by Huneeus and Rogerson (2020). Hence, once we control for the growth in agricultural productivity, the relationship between economic growth and the change in within-country regional inequality doubles.

Table F.20: Growth and Inequality

| | $\Delta$ GDP | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Within-country $\beta$ | .023 | | -.001 | .04 | .04 | .12 | .09 | .04 | .04 | .22 |
| | .81 | | .99 | 0.74 | 0.10 | .08 | .10 | .10 | .10 | 0.02 |
| Gini | | .03 | .04 | -.02 | -.02 | -.03 | -.09 | -.03 | -.02 | -.08 |
| | | .08 | .02 | .01 | .01 | .01 | .03 | .01 | -.02 | 0.00 |
| Gini Growth | | | | -16.95 | -17.04 | 4.91 | -48.75 | -24.90 | -17.95 | -32.63 |
| | | | | 16.63 | 16.96 | 15.01 | 18.59 | 14.46 | 16.09 | 0.20 |
| ln(Initial GDP) | | | | -1.08 | -1.08 | -1.32 | -2.41 | -1.11 | -1.06 | -2.10 |
| | | | | .00 | .19 | .27 | .51 | .25 | .22 | 0.00 |
| N | 795 | 905 | 536 | 536 | 536 | 406 | 341 | 536 | 536 | 217 |
| $R^2$ | .06 | .10 | .09 | .34 | 0.34 | 0.36 | .56 | 0.35 | 0.34 | .59 |
| Controls: | | | | | | | | | | |
| Democracy | | | | | X | | | | | |
| Education | | | | | | X | | | | |
| Structural Change | | | | | | | X | | | |
| Geography | | | | | | | | X | | |
| Trade Openness | | | | | | | | | X | |
| All | | | | | | | | | | X |
| Time FE | X | X | X | X | X | X | X | X | X | X |

Notes: This table reports the estimates of running a regression of GDP growth levels on within-country $\beta$ convergence conditional on several observables in different specifications. Standard errors are clustered at country level.

### F.3.3 Understanding the Drivers of Regional Inequality

Fact #2 highlights the correlation between a shift toward service and regional convergence. To provide supportive evidence to this fact and test for alternative hypothesis, we run a horse race among several potential candidates. We find some hypotheses consistent with existing literature but we also highlight a new for role of structural transformation in shaping regional convergence in both directions. Specifically, in accordance with Caselli and Coleman (2001) and Eckert and Peters (2018), we find that structural transformation from agriculture to manufacturing pushes for regional convergence. We confirm the new result that structural transformation toward service reduces regional convergence. The literature on regional inequality has pointed out to several explanations for regional convergence.

As previously mentioned, Caselli and Coleman (2001) and Eckert and Peters (2018) highlight the role of structural transformation as a driver of regional convergence in the US. To take into account such force we include agricultural productivity growth as well as share of manufacturing in the economy and he include the role of service productivity growth to capture the transition to modern economy.

offered an explanation suggesting that open access to trade. Market access as well as free trade agreements capture aim at capturing this story in our specification. Another factor that might drive the low speed of convergence is land restrictions such as geographical factors as shown by Ganong and Shoag (2017). To capture land unavailability we include several measures such as ruggedness, % of land in desert, distance from the coast and % of fertile soil.

Differential increase and return in human capital might be one of the explanations as well as in Giannone (2017). We include average years of education as well as change in average years of education to capture human capital. Table F.22 reports the estimates of the horse race. The dependent variable in each of these specifications is the speed of convergence $\hat{\beta}$ estimated with a 10-year interval at country level for each decade between 1980 and 2020. The results of column (1) suggest a positive but non statistically significant correlation between speed of convergence and GDP per capita growth. Once we adjust for initial GDP in column (2) we find a positive correlation between initial GDP and speed of convergence suggesting that countries with richer countries experience a lower speed of convergence (or more regional inequality). To account for our main story of structural transformation we include controls for change in agricultural productivity as well changes in service productivity. The first is negatively correlated with $\beta$ convergence. We interpret this result suggesting that an increase in agricultural productivity growth will increase regional convergence. Simultaneously, an increase in service productivity growth will decrease regional convergence.

When including political scores in column (4), we find that while the coefficient is positive it is not statistically significant. In column (5), we add controls for average years of education and their respective growth over 10 years. We find these coefficients are negatively correlated with higher speed of convergence but are not statistically significant either.

In column (6), we include variables that capture internal geographical differences as well as internal mobility. We find that more roads per capita are positively correlated with higher regional convergence. We also find that higher percentage of land covered in desert is correlated with lower regional convergence. Column 7 accounts for a story of trade openness. However, while we find a positive coefficient we do find statistical significance. Column (8) accounts for the final horse race among all the potential channels and allows to control for access to trade and overall market access suggests that more foreign trade agreements are positively correlated with slower convergence speed. Once all these determinants are considered jointly, we find that faster service productivity growth, higher political score index, a higher percentage of land covered in desert and more access to trade are all explanatory variables that predict slower speed of convergence. Simultaneously, structural change and distance from the coast are correlated with faster speed of convergence. When we run a variance decomposition exercise, we find that structural transformation is the biggest contributor by a large margin that explain the variation in speed of convergence across countries and over time.

Next in table F.21, we verify that cross-country (or cross-sectional) differences in within-country convergence rates are not due to other factors like external trade agreements, the polity of countries, and their human capital endowment.

Table F.21: Determinants of Regional Convergence

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Service Share | 0.0539 | 0.0596 | 0.0696 | 0.0836 | 0.1036 |
| | (0.0563) | (0.0576) | (0.0546) | (0.0455)* | (0.0400)** |
| $\Delta$ Serv. Product. | 58.1721 | 57.7615 | 62.7862 | 56.2782 | 65.8885 |
| | (17.1730)*** | (16.7112)*** | (19.3753)*** | (13.1700)*** | (10.1276)*** |
| Roads/Cap. (km) | | -9.5731 | -3.7698 | 6.0459 | 8.6554 |
| | | (15.4056) | (14.9946) | (16.2952) | (15.1441) |
| Avg. FTAs | | | 1.1752 | 1.7897 | 2.2101 |
| | | | (1.2591) | (1.7238) | (1.6152) |
| Years of Education | | | | -0.0786 | -0.1271 |
| | | | | (0.1904) | (0.1989) |
| $\Delta$ Years of Educ. | | | | 3.1583 | -8.7005 |
| | | | | (31.4909) | (30.9531) |
| Political Score | | | | | -0.0962 |
| | | | | | (0.0654) |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.2013 | 0.2155 | 0.2191 | 0.3213 | 0.3442 |

*Notes:*This table shows the regression estimates where the dependent variable in each column is the estimate of $\beta$-convergence for 10-year rolling windows for each country in our sample. The unit of observation is country×year. Robust standard errors are reported in parenthesis.

Table F.22: Testing for Complementary Hypotheses

| | Within country $\beta$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\Delta$ GDP | 0.03 | 0.09 | 0.06 | 0.07 | 0.15 | 0.17 | 0.09 | 0.32 |
| | (0.12) | (0.12) | (0.33) | (0.12) | (0.13) | (0.12) | (0.12) | (0.19) |
| Initial GDP | | 0.59 | 0.31 | 0.37 | 0.63 | 0.76 | 0.46 | -0.77 |
| | | (0.25)** | (0.47) | (0.32) | (0.28)** | (0.44)* | (0.29) | (0.51) |
| $\Delta$ Agr. Product. | | | -20.31 | | | | | -19.62 |
| | | | (10.58)* | | | | | (12.65) |
| $\Delta$ Serv. Product. | | | 61.92 | | | | | 27.47 |
| | | | (21.96)*** | | | | | (14.03)* |
| Political Score | | | | 0.06 | | | | 0.21 |
| | | | | (0.05) | | | | (0.10)** |
| Years of Education | | | | | -0.157 | | | 0.12 |
| | | | | | (0.16) | | | (0.25) |
| $\Delta$ Years of Educ. | | | | | -35.18 | | | -1.82 |
| | | | | | (31.52) | | | (31.16) |
| Roads/Cap. (km) | | | | | | -1.67 | | -8.95 |
| | | | | | | (17.74) | | (20.55) |
| Ruggedness | | | | | | 0.04 | | 0.160 |
| | | | | | | (0.25) | | (0.14) |
| % Desert | | | | | | 0.08 | | 0.21 |
| | | | | | | (0.05)* | | (0.04)*** |
| Dist. from Coast | | | | | | -0.45 | | -1.97 |
| | | | | | | (0.60) | | (1.03)* |
| % Fertile Soil, | | | | | | 0.021 | | -0.03 |
| | | | | | | (0.02) | | (0.01)** |
| % Tropical | | | | | | 0.01 | | 0.02 |
| | | | | | | (0.01) | | (0.01)* |
| Avg. FTAs | | | | | | | 1.22 | 6.35 |
| | | | | | | | (1.72) | (1.79)*** |
| Market Access | | | | | | | | 0.00 |
| | | | | | | | | (0.00) |
| Year FE | X | X | X | X | X | X | X | X |
| N | 795 | 795 | 375 | 769 | 619 | 748 | 769 | 228 |
| R^2 | 0.0172 | 0.0746 | 0.2171 | 0.0827 | 0.0853 | 0.1141 | 0.0756 | 0.5168 |

*Notes:* This table reports the estimates of 4 conditional on several observables. Standard errors are clustered at country level. The ***, **, and * represent statistical significance at the 0.001, 0.005, and 0.01 levels respectively.

# G Data Appendix

## G.1 GDP Data

This section details the sources for GDP data, utilizing a comprehensive dataset compiled from multiple sources and covering a wide range of countries. Table G.23 provides specifics on data availability, and origin for each data point, highlighting both developed and developing economies and focusing on regional economic trends over time. Notes address data gaps in specific regions due to different factors, referencing a supplementary spreadsheet for further details on data sources and series extraction. Data cleaning methods are described, including anomaly detection and splicing techniques to address inconsistencies, particularly in Canada (2012), China (1999), Peru (2007), and Mexico (2011). Finally, a robustness check is performed by visually comparing interpolated national GDP per capita values against those from the Penn World Table 10.0, demonstrating the accuracy of the interpolation method.

### G.1.1 Data source

GDP Data for our analysis is sourced from a comprehensive dataset belonging to multiple and several sources accessible via the following table G.23. This table contains crucial sources and availability of our GDP economic indicators that underpin our study, facilitating a detailed examination of regional dynamics over time with emphasis on the developing world.

Table G.23: Data Sources and Variables

| **Country** | **Variable** | **Year** | **Year Available** | **Authors** | **Source** |
|---|---|---|---|---|---|
| Australia | GDP | 1981-1990* | | CGKK | Australian Bureau of Statistics |
| Australia | GDP per capita | 1981-1990* | | CGKK | Australian Bureau of Statistics |
| Australia | GDP | 1990-2019 | | CGKK | Australian Bureau of Statistics |
| Australia | GDP per capita | 1990-2019 | | CGKK | Australian Bureau of Statistics |
| Bolivia | GDP | 1980-1986 | | GLLS | BNIS |
| Bolivia | GDP | 1988-2019 | | DOSE | - |
| Bolivia | Population | 1950, 1976 | | GLLS | City Population |
| Bolivia | Population | 1988-2019 | | DOSE | - |

| Country | Variable | Year | Year Available | Authors | Source |
|---|---|---|---|---|---|
| Brazil | GDP | 1970, 1975, 1985, 1986-2019 | | CGKK | IPEA |
| Brazil | Population | 1970, 1980, 1991, 1996, 2000, 2007, 2010, 2022 | | CGKK | IPEA |
| Canada | GDP | 1961-2012* | 1961-2011 | GLLS | Statistics Canada |
| Canada | GDP | 2012-2019 | 1997-2019 | CGKK | Statistics Canada |
| Canada | Population | 1956, 1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006, 2011 | | GLLS | City Population |
| Canada | Population | 2019 | | CGKK | City Population |
| Chile | GDP per capita | 1960-2001 | | GLLS | Mideplan/Diaz-Vernon (2004); CBC |
| Chile | GDP | 2002 | | CGKK | Central Bank |
| Chile | GDP | 2003-2007 | | CGKK | Central Bank |
| Chile | GDP | 2008-2019 | | CGKK | Central Bank |
| Chile | Population | 1960, 1992, 2002, 2010 | | GLLS | City Population |
| Chile | Population | 2017 | | CGKK | City Population |
| China | GDP | 1952-1999* | 1952-2010 | GLLS | NBS - 1949-2008 |
| China | GDP | 1999-2019 | 1999-2020 | CGKK | OECD |
| China | Population | 1954-1956, 1970, 1985, 2007-2009, 2010 | | GLLS | City Population |
| China | Population | 2018 | | CGKK | City Population |
| Colombia | GDP | 1980-2019 | | CGKK | DANE |
| Colombia | Population | 1964, 1973, 1985, 1993, 2005, 2011 | | GLLS | City Population |
| Colombia | Population | 2018, 2020 | | CGKK | City Population |
| Estonia | GDP | 1995-2019 | | DOSE | - |
| Estonia | Population | 1995-2019 | | DOSE | - |
| India | GDP | 1980-2017 | | CGKK | - |
| India | Population | 1980-2017 | | CGKK | - |
| Indonesia | GDP | 1971 | | GLLS | Literature |
| Indonesia | GDP | 1980-2019 | | DOSE | - |

| Country | Variable | Year | Year Available | Authors | Source |
|---|---|---|---|---|---|
| Indonesia | Population | 1971, 1980, 1990, 1995, 2000, 2010 | | GLLS | City Population |
| Indonesia | Population | 2015, 2019 | | CGKK | City Population |
| Japan | GDP | 1955-1974 | | CGKK | Cabinet Office |
| Japan | GDP | 1975-1999 | | CGKK | Cabinet Office |
| Japan | GDP | 2000 | | CGKK | Cabinet Office |
| Japan | GDP | 2001-2014 | | CGKK | Cabinet Office |
| Japan | Population | 1955-1999 | | CGKK | Statistics Bureau of Japan |
| Japan | Population | 2000-2020 | | CGKK | Statistics Bureau of Japan |
| Kenya | GDP | 1970-1999 | | DOSE | - |
| Kenya | GDP | 2004, 2005, 2009 | | CGKK | UN Human Development Reports |
| Kenya | GDP | 2013-2019 | | CGKK | Kenya National Bureau of Statistics |
| Kenya | Population | 1966-1999, 2013-2017 | | DOSE | - |
| Kenya | Population | 2019 | | CGKK | Kenya National Bureau of Statistics |
| Malaysia | GDP | 1970, 1975, 1980, 1990, 1995, 2000, 2005-2010 | | GLLS | Literature |
| Malaysia | GDP | 2011-2015 | | CGKK | Department of Statistics Malaysia |
| Malaysia | GDP | 2016-2019 | | DOSE | - |
| Malaysia | Population | 1970, 1980, 1991, 2000, 2010 | | GLLS | - |
| Malaysia | Population | 2016-2019 | | DOSE | - |
| Mexico | GDP per capita | 1950-1960 | | GLLS | Literature |
| Mexico | GDP | 1970, 1975, 1980, 1993-2011* | 1970-2010 | GLLS | NSA |
| Mexico | GDP | 2011-2019 | 2003-2020 | CGKK | OECD |
| Mexico | Population | 1950, 1960, 1970, 1980, 1990, 1995, 2000, 2005, 2010 | | GLLS | City Population |
| Pakistan | GDP | 1970-1981 | | GLLS | Literature |
| Pakistan | GDP | 1981-2004 | | DOSE | - |
| Pakistan | Population | 1951, 1961, 1972, 1981 | | GLLS | City Population |

| Country | Variable | Year | Year Available | Authors | Source |
|---|---|---|---|---|---|
| Pakistan | Population | 1982-2004 | | DOSE | - |
| Panama | GDP | 1996-2019 | | DOSE | - |
| Panama | Population | 1996-2019 | | DOSE | - |
| Peru | GDP | 1970-1995 | | GLLS | Literature |
| Peru | GDP | 2001-2007* | 2001-2010 | GLLS | NSA |
| Peru | GDP | 2007-2019 | | CGKK | NSA |
| Peru | Population | 1961, 1972, 1981, 1993, 2007 | | GLLS | City Population |
| Peru | Population | 2017, 2020 | | CGKK | City Population |
| Philippines | GDP | 1975-2019 | | DOSE | - |
| Philippines | Population | 1975-2019 | | DOSE | - |
| Poland | GDP | 1995-2019 | | DOSE | - |
| Poland | Population | 1995-2019 | | DOSE | - |
| Republic of Korea | GDP per capita | 1985-2019 | | CGKK | - |
| Republic of Korea | Population | 1985-2019 | | CGKK | - |
| Romania | GDP | 1995-1996 | | GLLS | Eurostat |
| Romania | GDP | 1997-2018 | | DOSE | - |
| Romania | Population | 1977, 1992 | | GLLS | City Population |
| Romania | Population | 1997-2018 | | DOSE | - |
| Russia | GDP | 1994-2019 | | DOSE | - |
| Russia | Population | 1994-2019 | | DOSE | - |
| South Africa | GDP | 1970, 1975, 1980-1989 | | GLLS | Literature |
| South Africa | GDP | 1995-2019 | | CGKK | NSA |
| South Africa | Population | 1970, 1980, 1985, 1991, 1996, 2001, 2007 | | GLLS | City Population |
| South Africa | Population | 2011, 2019 | | CGKK | City Population |
| South Africa B | GDP | 1995-2019 | | CGKK | NSA |
| South Africa B | Population | 1970, 1980, 1985, 1991, 1996, 2001, 2007 | | GLLS | City Population |
| South Africa B | Population | 2011, 2019 | | CGKK | City Population |

| Country | Variable | Year | Year Available | Authors | Source |
|---|---|---|---|---|---|
| Switzerland | GDP share | 1960, 1970, 1980, 1990, 2000, 2010 | | Literature | |
| Switzerland | Population | 1960, 1970, 1980, 1990, 2000, 2010 | | NSA | Federal Statistical Office |
| Tanzania | GDP per capita | 1980, 1985, 1990, 1994 | | GLLS | NSA |
| Tanzania | GDP per capita | 2000-2010 | | GLLS | NSA |
| Tanzania | GDP | 2016-2019 | | CGKK | NSA |
| Tanzania | Population | 1978, 1988, 2002, 2005 | | GLLS | City Population |
| Tanzania | Population | 2012, 2019 | | CGKK | City Population |
| Thailand | GDP, GDP per capita | 1981-1995* | | CGKK | NSA |
| Thailand | GDP, GDP per capita | 1995-2019 | | CGKK | NSA |
| Turkey | GDP | 1975-1986 | | GLLS | Literature |
| Turkey | GDP | 1992-2001 | | GLLS | NSA |
| Turkey | GDP per capita | 2004-2019 | | CGKK | OECD |
| Ukraine | GDP | 1995-2003 | | DOSE | - |
| Ukraine | Population | 1995-2003 | | DOSE | - |
| Ukraine | GDP, GDP per capita | 2004-2019 | | CGKK | NSA |
| United Kingdom | GDP per capita | 1950, 1960, 1970 | | GLLS | Literature |
| United Kingdom | GDP per capita | 1995-2010 | | GLLS | Eurostat |
| United Kingdom | GDP per capita | 2011-2018 | | CGKK | NSA |
| USA | GDP per capita | 1950-2019 | | CGKK | NSA |
| USA | Population | 1950, 1960, 1970, 1980, 1990, 2000, 2010, 2019 | | CGKK | NSA |
| Uzbekistan | GDP | 2000-2019 | | DOSE | - |
| Uzbekistan | Population | 2000-2019 | | DOSE | - |
| Vietnam | GDP per capita | 1993 | | GLLS | Literature |
| Vietnam | GDP | 1995-2018 | | DOSE | - |
| Vietnam | Population | 1993 | | GLLS | Literature |
| Vietnam | Population | 1995-2018 | | DOSE | - |

For a more detailed table with details on the source and the series that has been extracted from the source, refer to this table here.

### G.1.2 Notes on Regions

In aligning with the classification standards established by Gennaioli et al. (2014), we categorize our regions for effective aggregation. In ensuring a balanced dataset, we acknowledge certain exceptions involving regions where the availability of GDP per capita data is compromised due to their unique political status during the years in question. Specifically:

- Japan: The data for Okinawa is notably absent for the years 1955-1971 due to its annexation status.
- Russia: The Chechen region shows missing GDP per capita values from 1994-2004 as a result of regional conflict and instability.
- Ukraine: For the year 2019, there are missing data points for the Autonomous Republic of Crimea and Sevastopol City which occurred during the political annexation process.

### G.1.3 Variable construction

To compute Country $i$'s region $s$ in year $t$, we apply the formula below, emphasizing the relationship between national performance and regional outputs:

$$(\text{Regional GDP per capita})_{ist} = (\text{National GDP per capita})_{it} \times \frac{(\text{Regional GDP share})_{ist}}{(\text{Regional population share})_{ist}} \tag{G.19}$$

Contrasting regional GDP and population shares, we utilize data outlined in Section G.1.1. In instances where regional population data is lacking but GDP data is present, we substitute missing figures through linear interpolation, ensuring continuity in analysis. National statistics are derived from the Penn World Table 10.0, specifically the *cgdpe* value for GDP and *pop* for the population figures. Subsequently, we calculate national GDP per capita by dividing national GDP by the population.

### G.1.4 Data cleaning

To maintain the integrity of our dataset, we implement a thorough anomaly detection process, summarized as follows:

1. We calculate the annual growth rate of Regional GDP per capita for each country-region-year, establishing a baseline for expected growth patterns.

2. Should this growth rate exceed 20% in absolute terms, we flag the country for further scrutiny, indicating a potential outlier in the data.

3. If such anomalies arise in years coinciding with data transitions, we leverage splicing techniques to harmonize any discrepancies.

The splicing process has been applied to rectify the following problematic country-years, each of which presents unique challenges requiring methodological adjustments:

- Canada (2012): In this particular year, significant changes in the data sources used for economic indicators led to noticeable discrepancies in reported figures. The variations could be attributed to a methodological shift in how regional GDP data was collected and reported. By applying splicing, we ensured consistency and reliability in the dataset, allowing for a more accurate representation of Canada's economic performance during this time.

- China (1999): The late 1990s were pivotal for China, marked by vast economic reforms and a shift towards market-oriented policies. The resultant transformations significantly impacted regional economic data, requiring us to closely review and adjust the figures for 1999. Given the rapid growth and changes in this period, splicing was essential to align the data accurately, enabling a clearer understanding of China's evolving economic landscape.

- Peru (2007): This year was notable due to ongoing developments in Peru's economy, including political changes and shifts in global commodities markets that influenced GDP reporting. The data available from different sources showed discrepancies, prompting the need for splicing. This correction not only addressed data continuity but also enhanced the reliability of the analysis regarding Peru's economic trajectory and growth within that timeframe.

- Mexico (2011): The economic data for Mexico in this year exhibited inconsistencies, likely arising from alterations in data collection methodologies linked to new statistical frameworks. The splicing technique was crucial here to bridge the gaps created by source changes and ensure that the GDP per capita estimates accurately reflected the economic conditions during this transitional period, thereby supporting meaningful comparisons with other years in our analysis.

#### G.1.5 Sample selection

Following the criteria that each country should not have yearly gaps longer than 10 years, the following country-years have been removed from our dataset to maintain the integrity and continuity of our analysis. This approach ensures that the data remains comparable across regions and over time, allowing for more robust conclusions. In the case of Kenya, we made a strategic decision to drop the more recent years to maximize coverage during the earlier observational period. The specific adjustments are as follows:

- Australia: We excluded years up to 1953, where the next available data point is 1976. This large gap of 23 years represented a significant discontinuity that could distort any longitudinal analysis. The decision to remove these early years ensures that our dataset accurately reflects periods of economic activity without large voids that could obscure trends.

- Kenya: In this instance, we decided to eliminate years from 2013 onwards. This decision was made to prioritize a complete set of data from earlier years, as the year immediately prior to 2013 is 1999, resulting in a substantial gap of 14 years. By focusing on years before 2013, we can achieve a more concentrated and reliable analysis of Kenya's economic performance over a critical period, enhancing our insights into the region's development.

- UK: The data for the UK prior to 1970 were removed, as the next data availability begins in 1995, creating a gap of 25 years. This lengthy absence of data points raises concerns over the representativeness of any analysis that might include such incomplete data. The exclusion of these years helps ensure that our study encompasses only those periods with sufficiently detailed and reliable data, thereby maintaining a high standard of accuracy and relevance in our findings.

By enforcing this criterion, we have effectively streamlined our dataset. This strategic removal of inconsistent data points minimizes the risk of misleading interpretations and strengthens the overall quality of our research outputs. The result is a dataset that is more cohesive and better suited for comprehensive economic analysis, enabling us to accurately assess trends and patterns within the regions of interest.

### G.1.6 Missing values

We interpolate missing values of regional GDP per capita for each region between its initial and final years of observation by estimating a linear regression model. This approach allows us to make informed estimates for years where data is missing, thereby ensuring that our analysis remains comprehensive and robust. The rationale behind this technique lies in its ability to leverage existing data to create plausible estimates based on observed trends.

For each region, we regress regional GDP per capita on a linear time trend, which accounts for general economic growth over time, as well as national GDP per capita obtained from the Penn World Table 10.0. This dual approach allows us to capture both the unique regional characteristics and the overarching national economic context, using the following model:

$$(\text{Regional GDP per capita})_{ist} = \beta_0^s + \beta_1^s t + \beta_2^s \text{National GDP per capita}_{it} + u_{ist} \tag{G.20}$$

For each missing year in region $s$ of country $i$, we interpolate the missing value using the predicted value based on the OLS estimates of this model.

### G.1.7 Robustness check

Figure G.16 illustrates the results of the interpolation exercise conducted on our primary GDP variable. The horizontal axis represents the national GDP per capita as reported by the Penn World Table 10.0, specifically indicated by the *cgdpe* metric. This reflects the economic output per capita measured at purchasing power parity, providing a standardized measure for comparison across countries. Meanwhile, the vertical axis shows the national GDP per capita values that we have estimated based on our interpolated regional GDP per capita figures.

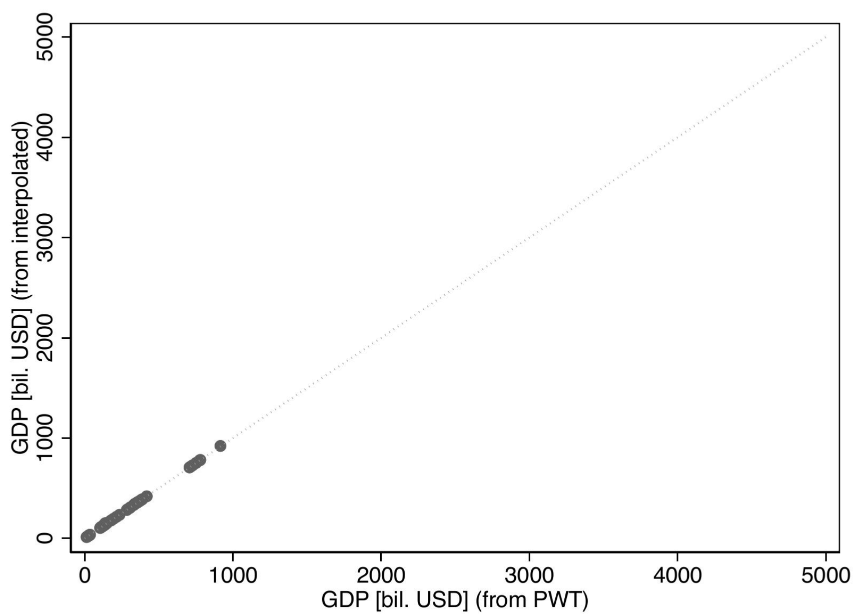
The close alignment of our estimates to the 45-degree line—a reference line where the values on the vertical axis equal those on the horizontal axis—highlights the accuracy and validity of the interpolation process. When data points cluster around this line, it suggests that the interpolated values closely mirror the established national GDP measurements, indicating that our technique has successfully preserved the underlying economic relationships during interpolation.

Moreover, the proximity of the data points to the 45-degree line implies that our interpolated estimates for regional GDP per capita effectively capture the national economic dynamics at play. This serves as strong evidence that our methodology for estimating missing values is robust and reliable, yielding outcomes that reflect the true economic conditions of each country and region during the periods analyzed.

The successful execution of this interpolation exercise not only enhances our confidence in the quality of the dataset but also enriches the subsequent analyses we can conduct. By ensuring that we have a complete and cohesive set of GDP per capita figures, we can explore regional economic trends with greater precision, draw meaningful comparisons among different regions, and contribute valuable insights to the overarching findings of our research.

In addition to demonstrating methodological rigor, Figure G.16 also serves as a visual confirmation of the underlying theoretical framework guiding our analysis. As we move forward, these well-founded interpolated values will provide a critical basis for understanding variances in regional economic performance, contributing to a holistic view of economic development patterns within and across countries.

Figure G.16: Validation of GDP data: Interpolation




## G.2 Sectoral Employment Data

We use three main data sources to construct a comprehensive data set of sectoral employment shares by region that covers a large cross-section of countries and multiple decades. First, we obtain census data from the Integrated Public Use Microdata Series (Ruggles et al., 2015, 2024), second, we obtain labor force survey data from the World Bank Global Labor

Database (GLD) and the World Bank i2d2 database.[6] The third data source is the ARDECO database from the ECJRC (Auteri et al., 2024).[7] We supplement this data with additional country-specific sources which provide information on regional employment by sector for Australia, China, Japan, South Korea and the UK.

Table G.24 documents the full coverage of our final dataset and lists which data source we use for each country and time period to obtain the sectoral employment data by regions. For multiple countries, we combine employment data from different sources (e.g., censuses and labor force surveys) to achieve the longest possible time coverage. The construction of the final data set requires careful harmonization of geographic regions over time and across different data sets as well as substantial data cleaning. We now explain the regional crosswalks, the data cleaning and the merging procedure in more detail.

Table G.24: Data sources for employment data

| | Country Name | Nb regions | Dataset 1 | Years (# obs) | Dataset 2 | Years (# obs) | GDP data yrs (# obs) |
|---|---|---|---|---|---|---|---|
| 1 | Argentina | 24 | Ipums | 1980–2001 (3) | | | |
| 2 | Australia | 8 | Other | 1984–2023 (40) | | | 1981–2019 (39) |
| 3 | Austria | 9 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 4 | Bangladesh | 6 | QLFS/ LFS | 2005–2016 (5) | | | |
| 5 | Belgium | 11 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 6 | Benin | 12 | Ipums | 1979–2013 (4) | | | |
| 7 | Bolivia | 9 | Ipums | 1976–2012 (4) | ECE | 2015–2019 (5) | 1980–2019 (40) |
| 8 | Botswana | 21 | Ipums | 1981–2011 (4) | | | |
| 9 | Brazil | 20 | Ipums | 1970–2010 (5) | PNAD/ PNADC | 2012–2020 (9) | 1970–2019 (50) |
| 10 | Bulgaria | 28 | ECJRC | 1995–2021 (18) | | | 1990–2019 (30) |
| 11 | Cameroon | 7 | Ipums | 2005–2005 (1) | | | |
| 12 | Canada | 10 | Ipums | 1971–2011 (5) | | | 1961–2019 (59) |
| 13 | Chile | 13 | Ipums | 1982–1982 (1) | CASEN | 1992–2017 (11) | 1960–2019 (60) |
| 14 | China | 27 | Ipums | 1982–1990 (2) | Stat. Yrbook | 1999–2010 (11) | 1952–2019 (68) |
| 15 | Colombia | 19 | Ipums | 1964–1993 (3) | ECH/ ENH/ GEIH | 1999–2021 (12) | 1980–2019 (40) |
| 16 | Costa Rica | 7 | Ipums | 1963–2011 (5) | | | |
| 17 | Croatia | 21 | ECJRC | 1995–2021 (24) | | | 1993–2019 (27) |
| 18 | Czech Republic | 14 | ECJRC | 1993–2021 (29) | | | 1990–2019 (30) |
| 19 | Denmark | 11 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 20 | Dominican Republic | 23 | Ipums | 1960–2010 (4) | | | |
| 21 | Ecuador | 14 | Ipums | 1982–2001 (3) | ENEMDU | 2007–2017 (2) | |
| 22 | Egypt | 24 | Ipums | 1986–1996 (2) | LFS | 2006–2019 (3) | |
| 23 | Estonia | 5 | ECJRC | 1990–2021 (28) | | | 1993–2019 (27) |
| 24 | Ethiopia | 10 | Ipums | 1994–1994 (1) | LFS | 1999–2021 (3) | |
| 25 | Fiji | 4 | Ipums | 1986–2014 (4) | | | |
| 26 | Finland | 19 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 27 | France | 27 | ECJRC | 1980–2021 (41) | | | 1980–2019 (40) |
| 28 | Germany | 16 | ECJRC | 1980–2021 (41) | | | 1991–2019 (29) |
| 29 | Ghana | 10 | Ipums | 2000–2010 (2) | | | |
| 30 | Greece | 13 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 31 | Guatemala | 22 | Ipums | 1964–2002 (5) | ENCOVI/ ENEI | 2006–2011 (2) | |
| 32 | Guinea | 33 | Ipums | 1983–2014 (2) | | | |
| 33 | Haiti | 4 | Ipums | 1982–2003 (2) | | | |
| 34 | Honduras | 18 | Ipums | 1974–2001 (2) | | | |
| 35 | Hungary | 8 | ECJRC | 1992–2021 (30) | | | 1980–2019 (40) |
| 36 | India | 27 | Ipums | 1983–2009 (6) | PLFS | 2017–2018 (2) | 1980–2019 (40) |
| 37 | Indonesia | 26 | Ipums | 1971–1990 (4) | SAKERNAS | 2007–2019 (10) | 1971–2019 (49) |
| 38 | Ireland | 6 | Ipums | 1971–1971 (1) | ECJRC | 1980–2021 (42) | 1980–2019 (40) |
| 39 | Israel | 7 | Ipums | 1995–1995 (1) | | | |
| 40 | Italy | 21 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 41 | Jamaica | 14 | Ipums | 1982–2001 (3) | | | |
| 42 | Japan | 47 | Other | 1977–2017 (9) | | | 1955–2019 (65) |
| 43 | Latvia | 6 | ECJRC | 1990–2021 (30) | | | 1992–2019 (28) |
| 44 | Liberia | 5 | Ipums | 1974–2008 (2) | | | |
| 45 | Lithuania | 10 | ECJRC | 1990–2021 (27) | | | 1992–2019 (28) |
| 46 | Malaysia | 12 | Ipums | 1970–2000 (4) | | | 1970–2019 (50) |
| 47 | Mali | 8 | Ipums | 1987–2009 (3) | | | |
| 48 | Mauritius | 10 | Ipums | 1990–2011 (3) | | | |
| 49 | Mexico | 32 | Ipums | 1960–2020 (8) | | | 1993–2019 (27) |

[6] We list the names of all labor force surveys in Table G.24 below. A data description of the GLD data and harmonization process can be found at https://github.com/worldbank/gld.

[7] The data and documentation is provided at https://knowledge4policy.ec.europa.eu/territorial/ardeco-database_en.

| | Country Name | Nb regions | Dataset 1 | Years (# obs) | Dataset 2 | Years (# obs) | GDP data yrs (# obs) |
|---|---|---|---|---|---|---|---|
| 50 | Mongolia | 21 | Ipums | 2000–2000 (1) | LFS | 2002–2022 (2) | |
| 51 | Mozambique | 11 | I2D2 | 1996–2014 (5) | | | |
| 52 | Netherlands | 12 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 53 | Nicaragua | 12 | Ipums | 1995–2005 (2) | | | |
| 54 | North Macedonia | 8 | ECJRC | 1997–2021 (10) | | | 1994–2019 (26) |
| 55 | Norway | 6 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 56 | Pakistan | 4 | LFS | 2001–2020 (11) | | | 1970–2004 (35) |
| 57 | Panama | 7 | Ipums | 1960–1980 (3) | EMO/ EH | 1989–2018 (13) | 1996–2019 (24) |
| 58 | Papua New Guinea | 20 | Ipums | 1980–2000 (2) | | | |
| 59 | Paraguay | 13 | Ipums | 1962–1992 (4) | EIH/ EPH | 1997–2017 (3) | |
| 60 | Peru | 23 | ENA | 1997–2021 (13) | | | 1970–2019 (50) |
| 61 | Philippines | 7 | Ipums | 1990–1995 (2) | LFS | 1997–2018 (18) | 1975–2019 (45) |
| 62 | Poland | 17 | ECJRC | 1991–2021 (31) | | | 1980–2019 (40) |
| 63 | Portugal | 7 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 64 | Republic Of Korea | 9 | Other | 1975–2020 (10) | | | 1980–2019 (40) |
| 65 | Romania | 39 | ECJRC | 1990–2021 (30) | | | 1980–2019 (40) |
| 66 | Senegal | 8 | Ipums | 1988–2013 (2) | | | |
| 67 | Serbia | 25 | ECJRC | 1995–2021 (24) | | | 1995–2019 (25) |
| 68 | Slovak Republic | 8 | Ipums | 1991–2011 (4) | ECJRC | 2012–2021 (10) | 1993–2019 (27) |
| 69 | Slovenia | 12 | ECJRC | 1991–2021 (31) | | | 1991–2019 (29) |
| 70 | South Africa | 4 | Ipums | 2001–2007 (2) | QLFS | 2008–2020 (13) | 1995–2019 (25) |
| 71 | Spain | 19 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 72 | Sweden | 8 | ECJRC | 1980–2021 (42) | | | 1980–2019 (40) |
| 73 | Switzerland | 25 | Ipums | 1970–1990 (3) | ECJRC | 1995–2021 (27) | 1980–2019 (40) |
| 74 | Tanzania | 18 | ILFS | 2000–2020 (4) | | | 1980–2019 (40) |
| 75 | Thailand | 68 | Ipums | 1970–1980 (2) | LFS | 1985–2021 (6) | 1981–2019 (39) |
| 76 | Togo | 3 | Ipums | 1970–2010 (2) | | | |
| 77 | Trinidad And Tobago | 4 | Ipums | 1980–2000 (3) | | | |
| 78 | Turkey | 18 | Ipums | 1985–2000 (3) | HLFS | 2009–2019 (8) | 1975–2019 (45) |
| 79 | Uganda | 36 | Ipums | 2002–2002 (1) | | | |
| 80 | UK | 10 | Other | 1981–2022 (42) | | | 1980–2019 (40) |
| 81 | United States | 51 | Ipums | 1960–2020 (8) | | | 1950–2019 (70) |
| 82 | Uruguay | 19 | Ipums | 1963–2006 (4) | ECH | 2007–2017 (4) | |
| 83 | Venezuela | 22 | Ipums | 1981–1981 (1) | EHM | 1989–2006 (5) | |
| 84 | Vietnam | 39 | Ipums | 1989–2019 (4) | | | 1993–2018 (26) |
| 85 | Zambia | 8 | Ipums | 1990–2010 (3) | | | |

Table G.25: Number of (un)balanced employment countries

| Region | Employment | | | |
|---|---|---|---|---|
| | Nb. countries | Avg. nb. years | 1980-2019 | 1990-2019 |
| Africa | 17 | 22 | 0 | 1 |
| Asia | 14 | 30 | 3 | 5 |
| Australia and Oceania | 3 | 28 | 0 | 1 |
| East Europe | 13 | 27 | 0 | 4 |
| North America | 4 | 44 | 2 | 2 |
| South America | 18 | 38 | 3 | 3 |
| West Europe | 16 | 40 | 15 | 16 |
| Total | 85 | | 23 | 32 |

### G.2.1 Harmonization of Regions over Time and across Data Sources

**Geographic Level of Aggregation.** In terms of spatial aggregation, we use the geographic level equivalent to states or provinces for all countries.

**Census Data.** The census data obtained from ipums international contains geographic identifiers, which are harmonized over time for each country ("geolev1" variable).

**Labor Force Survey Data.** For the labor force surveys, we harmonize geographic identifiers over time by creating a crosswalk for each country, which cleans potential differences in region's spelling and names and which adjusts for border changes, for example, when regions were merged or split. The harmonization over time sometimes requires us to aggregate several regions to ensure that we can compare consistent geographic units over time.

**ARDECO ECJRC data.** The ARDECO database from the ECJRC provides harmonized geographic data for different levels of the standardized NUTS regions. For each country, we select the NUTS level that corresponds most closely to the equivalent of states or provinces.

**Harmonization of Geographic Units across Data Sources.** Table G.24 shows that we combine employment data from several data sources for multiple countries to achieve the longest possible time coverage. When we use several data sets for a given country, we first create a geographic crosswalk that harmonizes region names across the relevant data sets. This harmonization across data sources can require aggregating regions to ensure that geographic units are consistently defined across all data sources. Once the employment data time series is cleaned for each country, we further merge this data with GDP data at the region-year level, which requires the creation of another final crosswalk of geographic units. The final number of regions in the harmonized dataset is listed in column 2 of Table G.24.

### G.2.2 Classification of Sectors across Datasets

We group workers into five sectors: (i) agriculture, (ii) manufacturing, (iii) low-skill services, (iv) high-skill and slow-productivity-growth services, and (v) high-skill and high-productivity-growth services. We choose the three categories within the service sector to account for the sector's large heterogeneity. Duarte and Restuccia (2019) point out that service sectors that employ high-skilled workers differ substantially in their productivity growth and income elasticities, which matters for aggregate productivity. We therefore further separate high-skill service sectors into sectors with either slow or high productivity growth based on Duarte and Restuccia (2019).

Table G.26 shows how we map the respective harmonized industry codes from ipums, the GLD labor force data, and ARDECO to these five sector categories.

For countries where we use non-harmonized employment data, we manually classify the provided (oftentimes detailed) industry codes into our five sector categories. This classification process sometimes required careful manual checking.

For example, the Korean employment data lists "Printing and Broadcasting" as one sector of employment in the 2010 and 2015 data, which combines activities that are typically reported separately with one being typically classified as manufacturing and the other as low-skill services. To choose how to best classify this hybrid category, we therefore carefully inspect the time series of sectoral employment when including "Printing and Broadcasting" either in manufacturing or in low-skill services. For the low-skill service sector, we see a large drop in

Table G.26: Sector and IPUMS Codes

| Sector | Ipums (INDGEN Code) |
|---|---|
| Agriculture | Agriculture, fishing, and forestry (10); |
| Manufacturing | Mining and extraction (20);<br>Manufacturing (30);<br>Electricity, gas, water and waste management (40);<br>Construction (50);<br>Other industry, n.e.c. (130) |
| Services low-skill | Wholesale and retail trade (60);<br>Hotels and restaurants (70);<br>Transportation, storage, and communications (80) |
| High-skill and slow-productivity-growth services | Public administration and defense (100);<br>Services, not specified (110);<br>Education (112);<br>Health and social work (113);<br>Other services (114);<br>Private household services (120); |
| High-skill and high-productivity-growth services | Financial services and insurance (90);<br>Business services and real estate (111) |

*Notes:* This table presents a shortened view of the sector classifications and corresponding IPUMS INDGEN codes.

the employment share in 2010 and 2015 if "Printing and Broadcasting" is not included, which makes us conclude that the category should be classified as low-skill services.

Another example that required additional data cleaning is China's employment data which we collect from three separate sources: First, ipums, second, the Chinese statistical yearbooks, and third, a private data base. Ipums covers data for 1980-2000, the yearbooks cover 1999-2010, and the private data set spans 1998-2021. The private data set provides a rich sectoral breakdown by region but it excludes self-employed workers, which especially reduces the agriculture employment share. The yearbooks include information on self-employed workers and provide data at the regional level but they only disaggregate employment into the three broad sectors of agriculture, manufacturing, and services without allowing for a further disaggregation of the service sector. To proceed, we use ipums data by region-and-sector for the years before 1999 and we then use the employment shares by region for agriculture, manufacturing, and services from the yearbooks for the period from 1999 to 2010. To further disaggregate employment in the service sector, we then use the corresponding employment shares for the three service sub-sectors *within* the service sector for each region-year from the private data base. We then multiply the within-service-sector shares with the overall service share from the yearbooks to create the final employment shares for all three service sectors for each region and year.

# H Data Challenges and Cleaning Methods

Creating a consistent time series of regional employment data across countries involves addressing two main types of data challenges: (1) irregularities within individual data sources over time and (2) inconsistencies that arise when merging data from different sources. Here, we explain the nature of these challenges and the cleaning steps that we take to address them, including the procedures for within-dataset corrections and cross-dataset adjustments.

## H.1 Irregularities within Datasets and Cleaning Methods

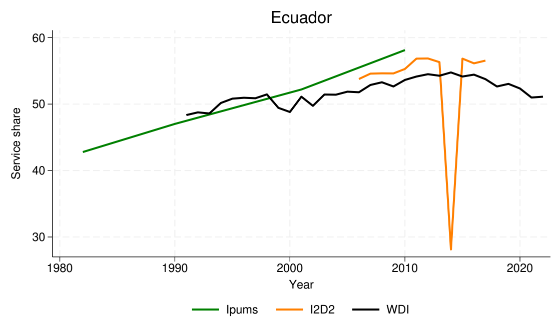
Irregularities in employment trends can occur within a given data source over time, for example, due to small sample sizes or variations in sampling procedures. We notices irregularities of two types:

**Spike Behavior.** Employment shares sometimes exhibit sharp increases (decreases) in one year which then reverse again in the following year. We label such behavior "spikes" which can distort trends and are likely data noise due to small sample size or other inconsistencies. One example of this is Ecuador, where the service employment share decreases sharply in the labor force survey data around 2013 and then rebounds again the following year (cf. Figure H.17a).

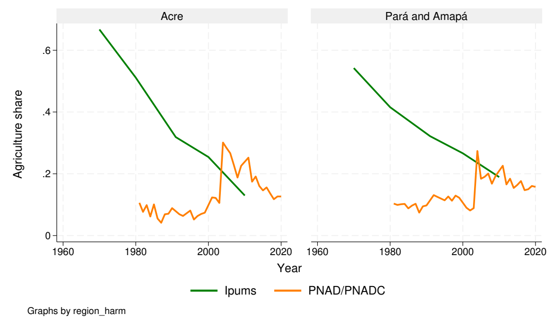
To clean the data set from such irregularities, we define an observation as a "spike" if the sectoral employment share changes in opposite directions by more than 4 (annualized) percentage points (pp) between two consecutive time periods. For example, if a sectoral employment share increases from $t-1$ to $t$ and then decreases again from $t$ to $t+1$, then we flag the time period $t$ as a "spike". If any region or sector is marked as a "spike", then we drop all observations for this country-year and we interpolate the data across this year to smooth the data series. Given that we drop the entire country-year, this cleaning step removes about 4,800 out of 25,000 observations.

**Persistent Shifts in Employment Composition.** Another challenge arises if some regions within a country exhibit sustained shifts in sectoral employment shares over time, which are very large or represent discontinuities in trends. An example is Brazil, where certain regions (e.g., Acre, Pará, and Amapá) show a persistent increase in agricultural employment share around 2003 (cf. Figure H.17b). When such changes in sectoral employment shares appear suspiciously large or at odds with national trends, we drop the given data source. In the case of Brazil, for example, we use the census data from IPUMS from 1970 to 2010 and the PNAD/PNADC from 2012 onward. We then connect the two series without a level-shift adjustment, which generates a smoother data series than using PNAD/PNADC from 1980 onward.

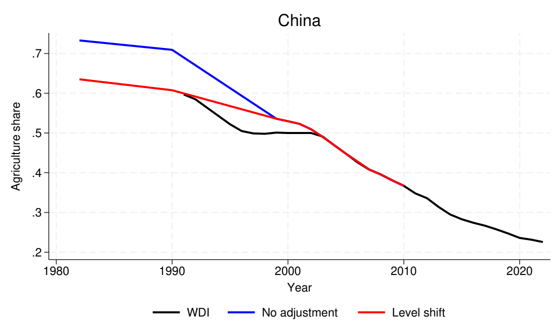
Figure H.17: Examples of Data Irregularities and Cleaning Steps




(a) Ecuador: Example of a spike




(b) Brazil: Example of shifts within sources




(c) China: Solution to shifts across sources

*Notes:* This figure shows examples of data irregularities and the cleaning method that we applied. Figure H.17a shows a spike in the service employment share for Ecuador around 2014. Figure H.17b shows a jump in the agricultural employment share in two regions in Brazil around 2003. Figure H.17c illustrates the challenges of combining data from different sources using the case of China. For China, applying a level adjustment to one data source smoothed the trends and improved the fit between the final data series and the corresponding data moments from the WDI at the national level.

**Sparse Data Coverage and Large Gaps.** Some countries have limited availability of employment data so that we have large gaps between observed time periods. Interpolating over very large time gaps can lead to inaccuracies in the measurement. For countries where gaps in data availability exceed ten years, we therefore manually assess the interpolated time series by comparing national employment trends in each sector against the annual data from the World Development Indicators (WDI). If the interpolated trends align well with the WDI at the national level, we keep the time series and we drop them otherwise. This decision is not based on a particular threshold but rather manual/visual inspection. This cleaning step dropped 4 country-year observations: Ghana in 1984, Nicaragua in 1971, Pakistan in 1973 and Romania in 1977. Excluding these episodes, there are 22 country-year observations left with gaps larger than 10 years.

## H.2 Merging Data across Datasets: Challenges and Cleaning Methods

When we combine data from multiple data sources for a given country, differences in sampling procedures or variable definitions can lead to inconsistencies in the time series of the data. To mitigate these issues, we adopt the following steps:

**Selecting a Primary Data Source.** For countries for which we combine more than one data source in our micro data, we first designate a "main" source and we then align any additional ("non-main") sources to it. To select the main source, we calculate the national employment shares of the agriculture, manufacturing, and service sector in each data source and compare their similarity to the national employment shares from the WDI by computing the Mean Squared Error (MSE). We choose the data source with the lowest average MSE as the primary dataset. When the MSE cannot be computed or is based only on very few data points, we instead manually assess the consistency between each data source and the WDI. Column 3 of Table H.27 indicates which data source we select as the "main" source for each country.

**Adjusting and Merging the Non-Main Source.** When combining multiple data sources for a given country, we either combine their raw data directly without any adjustments, or we apply a level adjustment to the non-main data source to ensure that there is no artificial jump or discontinuity in the time series trend at the point where data sources change. This is done as follows. For the "non-main" data source we compute the growth rate in employment per sector for all its years. In case the "main" and "non-main" source overlap, we compute the employment per sector forward (or backward) for the "main" source using this growth rate and then calculate the employment shares from this data. For countries where the data do not overlap between sources, we add one step and we first project the trend of sectoral employment in the main data source forward (or backward) using the last 5 years (or the closest observations to it) until we achieve an overlap between the data sources. From this point of overlap, we then proceed in the same way as for countries with overlapping data sources.

For each country, we then compare the combined time series of national employment shares to their counterparts from the WDI by computing the MSE both for the "level-adjusted" and the "raw" time series. For the final data set, we then usually choose the adjustment method which leads to the lowest MSE; however, in some cases we deviate from this rule based on a manual inspection of the time series at the national and regional level. Column 4 of Table H.27 reports for which countries we use the level-adjusted or raw data when combining data sources. Column 5 shows whether we made this decision based on the MSE or based on manual inspection of the data. For illustration, Figure H.17c plots the raw and level-adjusted data series of the national agricultural employment share in China. The figure shows that the level-adjusted data series provides a better fit to the national shares reported in the WDI.

**Flagging and Robustness Checks.** For countries where our cleaning procedure does not yield satisfactory results (e.g., due to high volatility), we flag these cases for further analysis and conduct robustness checks in subsequent empirical work.

## H.3 Interpolation of Final Dataset Across Time

We then use the clean and combined data series to linearly interpolate the sectoral employment shares for each region over missing years.

## H.4 Interpolation Results

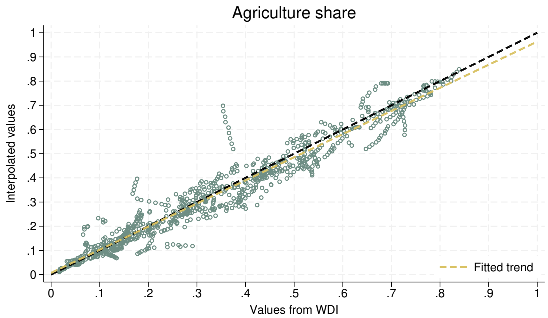
Figure H.18 compares the national employment shares in agriculture, manufacturing, and services from our final interpolated time series to their counterparts from the WDI for all country-years that are available in both data sets. The figure includes the 45 degree line (in black) and the linear fit (in yellow). Across all three sectors, the correlation between both data series is very high and the fitted trend aligns closely with the 45 degree line.

Table H.27: Multiple data sources summary

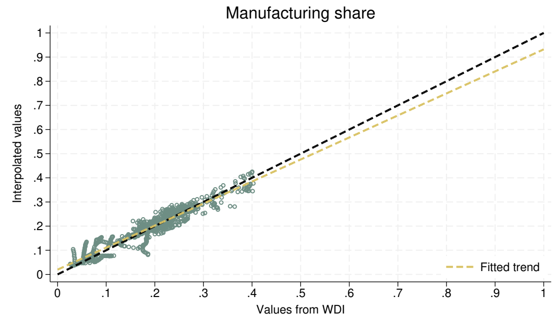
| | (1) Country | (2) Main Source | (3) Selection Rule for Main Source | (4) Adjustment of Non-main Source | (5) Selection Rule for Adjustment | (6) Overlap |
|---|---|---|---|---|---|---|
| 1 | Bolivia | Ipums | Manual | No | MSE | No |
| 2 | Brazil | PNAD/ PNADC | MSE | No | MSE | No |
| 3 | Chile | CASEN | MSE | Yes | MSE | Yes |
| 4 | China | Stat. Yrbook | MSE | Yes | MSE | Yes |
| 5 | Colombia | ECH/ ENH/ GEIH | MSE | Yes | MSE | No |
| 6 | Ecuador | ENEMDU | MSE | No | Manual | Yes |
| 7 | Egypt | LFS | MSE | Yes | Manual | Yes |
| 8 | Ethiopia | LFS | MSE | Yes | MSE | No |
| 9 | Guatemala | ENCOVI/ ENEI | MSE | Yes | MSE | No |
| 10 | India | Ipums | MSE | Yes | MSE | Yes |
| 11 | Indonesia | SAKERNAS | Manual | Yes | MSE | Yes |
| 12 | Ireland | ECJRC | MSE | Yes | MSE | Yes |
| 13 | Mongolia | LFS | MSE | Yes | Manual | No |
| 14 | Panama | EMO/ EH | MSE | Yes | MSE | Yes |
| 15 | Paraguay | EIH/ EPH | MSE | Yes | MSE | No |
| 16 | Philippines | LFS | MSE | No | MSE | Yes |
| 17 | Slovak Republic | Ipums | MSE | Yes | MSE | Yes |
| 18 | South Africa | Ipums | MSE | Yes | MSE | No |
| 19 | Switzerland | ECJRC | MSE | Yes | MSE | Yes |
| 20 | Thailand | LFS | MSE | Yes | MSE | Yes |
| 21 | Turkey | HLFS | MSE | No | Manual | No |
| 22 | Uruguay | ECH | MSE | Yes | Manual | No |
| 23 | Venezuela | EHM | MSE | Yes | MSE | Yes |

*Notes:* This table lists all countries for which we use more than one data source to construct the longest possible time series of sectoral and regional employment shares. The column "Main Source" lists which data source was chosen as the primary source. The column "Selection Rule for Main Source" specifies whether the data source was chosen as "main" based on the MSE criterion or based on a manual comparison. The column "Adjustment of Non-main Source" equals "Yes" if we apply a level adjustment to the non-main data source. The column "Selection Rule for Adjustment" specifies whether the choice of applying a level-adjustment to the non-main source (or not) was made based on the MSE criterion or based on a manual inspection. The column "Overlap" is equal to "Yes" if the two data sources overlap in at least one year.

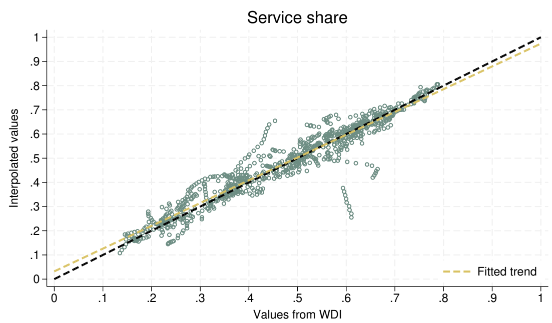
Figure H.18: Interpolated values vs. actual WDI data




(a) Agriculture share




(b) Manufacturing share

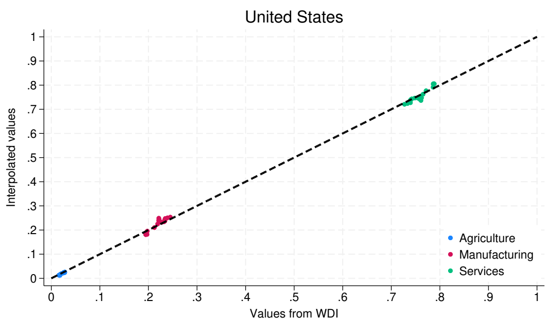



(c) Service share

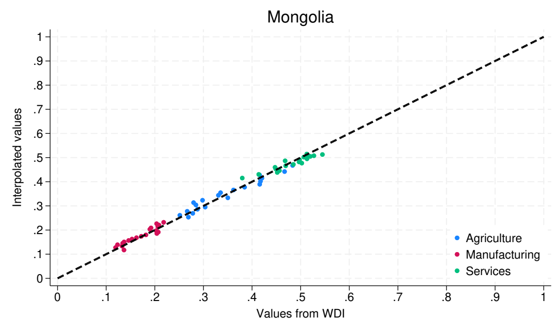
*Notes:* This figure shows at the national level how the interpolated sectoral employment shares in our data set relate to their counterparts in the WDI data. Each scatter point represents a sector-country-year. The figures plot the employment shares for (a) agriculture, (b) manufacturing, and (c) services. The black line is the 45 degree line and the yellow line shows the fitted trend of the following regression: $empshare_{s,t}^{interpol} = \alpha + \beta empshare_{s,t}^{WDI} + \varepsilon_{s,t}$ where $empshare_{s,t}^{interpol}$ is the employment share of sector $s$ at time $t$ of the interpolated data set and $empshare_{s,t}^{WDI}$ is the sectoral employment share from the WDI.

Figure H.19 shows scatter plots for selected countries. For most countries, the sectoral employment shares in our final data set match their counterpart in the WDI data very well at the national level. The close relationship holds for developed and developing countries. Yet, there are a few countries for which our final data set does not match the WDI data well (in the interpolated or raw version). An example for this is Bolivia as shown in Figure H.19c.

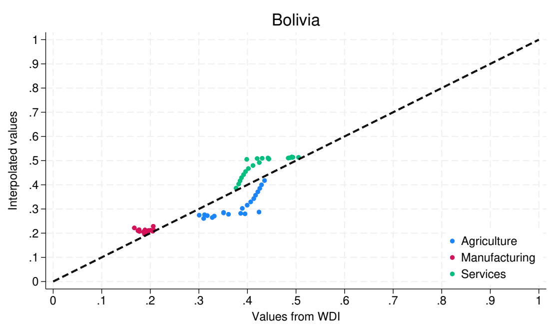
Figure H.19: Interpolated values vs. actual WDI data, country examples




(a) United States




(b) Mongolia




(c) Bolivia

*Notes:* This figure compares the sectoral employment shares of our final data set with their counterparts in the WDI data at the national level for the United States, Mongolia and Bolivia.