CLOSER TOGETHER OR FURTHER APART? WITHIN-COUNTRY REGIONAL DISPARITIES AND ADJUSTMENT IN ADVANCED ECONOMIES

Online Annex

Annex 2.1. Variables, Data Sources, and Sample Coverage

All data sources used in the chapter are listed in Annex Table 2.1.1. The chapter uses data from 43 economies which contain 655 underlying territorial level 2 (TL2) regions—see Annex Table 2.1.2 for a list of the economies in the sample and the total number of TL2 regions by economy. The country coverage for the different analytical exercises conducted is presented in Annex Table 2.1.3 (varying due to data availability). Baseline PPP-adjusted regional real gross domestic product (GDP) per capita comes from the OECD Regional Database. For coverage prior to 2000, it is spliced backwards using the data compiled by Gennaioli and others (2014), matched by region. In this chapter, advanced economies are classified as such according to the October 2019 *World Economic Outlook* definition. All other economies are classified as emerging market economies.

The OECD Regional Database contains details on ten sectors, based on the ISIC revision 4 sectoral classification. They include: agriculture (ISIC classification A); industry, including manufacturing and energy (ISIC classifications B through E); construction (ISIC classification F); trade, including distribution, transportation, accommodation, and food services (ISIC classifications G through I); information technology and communications (ISIC classification J); finance and insurance (ISIC classification K); real estate (ISIC classification L); professional services (ISIC classifications M through N); public services (ISIC classifications O through Q); and other services (ISIC classifications R through U).

The underlying data on household-equivalized income for the calculation of inequality (Figure 2.5) come from the Luxembourg Income Study (LIS), using regions defined by LIS, for the following countries and years: Australia (2014), Austria (2013), Canada (2013), Czech Republic (2013), Denmark (2013), Estonia (2013), Finland (2013), France (2010), Germany (2013), Greece (2010), Ireland (2010), Israel (2014), Italy (2014), Slovak Republic (2013), Spain (2013), Switzerland (2013), the United Kingdom (2013), and the United States (2013). See Annex 2.2 for details on the inequality decomposition.

Data from Eurostat, European Union Labor Force Survey (EU LFS) is used to tabulate migration shares within country by education level and labor market status (employed, unemployed, out of the labor force; Figure 2.13). Surveys from the following countries over 2000-16 are used: Austria; Belgium; Bulgaria; Czech Republic; Denmark; France; Greece; Hungary; Norway; Poland; Portugal; Romania; Slovak Republic; Slovenia; Spain; Sweden; and Switzerland. The responsibility for all conclusions drawn from the EU LFS data lies entirely with the authors.

Firm-level data used in the analysis of capital allocative efficiency come from the Orbis database. See Annex 2.6 for further details.

Annex Table 2.1.1. Data Sources

Source	Indicators
OECD Regional Database and Gennaioli and others (2014)	Gross Domestic Product per Capita (Constant Prices and PPP-adjusted), Unemployment rate, Dependency ratio, Life expectancy, Infant mortality, Share of labor force participants with tertiary education, Share of enrollment in tertiary education, Employment rate, Labor force participation rate, Long-term unemployment rate, Youth (aged 15-24 years-old) unemployment rate, Share of youth (aged 15-24 years-old) not in education, employment, or training (NEET), Sectoral real gross value added (constant prices), Sectoral employment (number of individuals), Working age (aged 15-64 years-old) population (number of individuals), Population (number of individuals), Inward migration (number of individuals), Outward migration, (number of individuals), Inward youth migration (number of individuals aged 15-24 years-old), Outward youth migration (number of individuals aged 15-24 years-old)
Gbohoui, Lam, and Lledo (2019)	Subnational Price Deflators
OECD Market Regulation database	Country-Level Product Market Regulation Index
OECD Employment database	Country-Level Employment Protection Legislation Index
OECD Social and Welfare database	Country-Level Unemployment Benefits Gross Replacement Ratio
IMF, World Economic Outlook database	Country-Level Trade Openness
World Bank, Doing Business Indicators	Country-Level Ease of Starting a Business
World Input-Output database	Country-Level Sectoral Imports & Exports by Destination and/or Origin
Das and Hilgenstock (2018)	Country-Level Sectoral Exposure to Routinization
History Database of Global Environment database	Grid-Level Population Count
NASA Earth Exchange Global Daily Downscaled Projections data set (NEX_GDDP)	Grid-Level Temperature and Precipitation Forecast
University of East Anglia, Climate Research Unit (CRU TS v.3.24)	Grid-Level Temperature and Precipitation Historical
Luxembourg Income Study	Individual Equivalized Household Disposable Income

Source: IMF staff compilation.

Annex Table 2.1.2. Territorial Grid of Country Sample

ISO 3166-1 alpha-3 Code	Country Name	Territorial Level 2 Regions (TL2)
AUS	Australia	States/Territories (8)
AUT	Austria	Bundesländers (9)
BEL	Belgium	Regions (3)
BGR	Bulgaria	Oblasti (6)
BRA	Brazil	Estados (27)
CAN	Canada	Provinces/Territories (13)
CHE	Switzerland	Grandes Régions (7)
CHL	Chile	Regiones (15)
CHN	China	Shěng/Zìzhìqū/Zhíxiáshì (31)
COL	Colombia	Departamentos (34)
CZE	Czech Republic	Oblasti (8)
DEU	Germany	Länder (16)
DNK	Denmark	Regioner (5)
ESP	Spain	Comunidades/Ciudades Autónomas (19)
FIN	Finland	Suuralueet (5)
FRA	France	Régions (22)
GBR	United Kingdom	Regions (12)
GRC	Greece	Regions (13)
HRV	Croatia	Regions (2)
HUN	Hungary	Planning Statistical Regions (7)
IDN	Indonesia	Provinsi (33)
IND	India	States/Territories (33)
IRL	Ireland	Regions (2)
ISL	Iceland	Regions (2)
ISR	Israel	Districts (6)
ITA	Italy	Regioni (21)
JPN	Japan	Groups of Prefectures (10)
KOR	Korea	Regions (7)
MEX	Mexico	Estados (32)
NLD	Netherlands	Provinces (12)
NOR	Norway	Landsdeler (7)
NZL	New Zealand	Regional Councils (14)
PER	Peru	Regiones (25)
POL	Poland	Voivodeships (16)
PRT	Portugal	Comissões de Coordenação e Desenvolvimento Regional e Regiões Autónomas (7)
ROU	Romania	Development Region (8)
RUS	Russia	Oblasts/Republics/Okrugs/Krais (83
SVK	Slovak Republic	Zoskupenia krajov (4)
SVN	Slovenia	Kohezijske regije (2)
SWE	Sweden	Riksomraden (8)
TUR	Turkey	Regions (26)
USA	United States	States (51)
ZAF	South Africa	Provinces (9)

Source: OECD Regional Database and IMF staff compilation.
Note: Number of subnational regions by country shown in parentheses next to the country's TL2 designation.

Annex Table 2.1.3. Sample of Economies Included in Analytical Exercises

Exercise	Economies[1]
Ratio of real GDP per capita in subnational regions at the 90th percentile to the 10th percentile (Figure 2.1, panels 1 and 3)	Panel 1: Australia; Austria; Canada; Czech Republic; Denmark; France; Germany; Greece; Italy; Japan; Korea; Netherlands; New Zealand; Norway; Portugal; Spain; Sweden; Switzerland; United Kingdom; and United States. Panel 3: Brazil*; Bulgaria*; Chile*; China*; Colombia*; Hungary*; India*; Indonesia*; Mexico*; Peru*; Poland*; Russia*; South Africa*; and Turkey*.
Average speed of regional convergence (Figure 2.1, panels 2 and 4)[2]	Panel 2: France; Germany; Italy; United Kingdom; United States. Panel 4: Brazil*; Chile*; China*; Colombia*; India*; Indonesia*; Mexico*; and Peru*.
Subnational regional disparities in AEs (Figure 2.2)	Australia; Austria; Canada; Czech Republic; Denmark; Finland; France; Germany; Greece; Italy; Japan; Korea; Netherlands; New Zealand; Norway; Portugal; Slovak Republic; Spain; Sweden; Switzerland; United Kingdom; and United States.
Subnational regional unemployment and activity (Figure 2.3)	Australia; Austria; Belgium; Canada; Czech Republic; Denmark; Finland; France; Germany; Greece; Ireland; Italy; Netherlands; New Zealand; Norway; Portugal; Slovak Republic; Slovenia; Spain; Sweden; Switzerland; United Kingdom; and United States.
Lagging regions versus others (Figures 2.4 and 2.8)	Australia; Austria; Belgium; Canada; Czech Republic; Denmark; Finland; France; Germany; Greece; Iceland; Ireland; Israel; Italy; Japan; Korea; Netherlands; New Zealand; Norway; Portugal; Slovak Republic; Slovenia; Spain; Sweden; Switzerland; United Kingdom; and United States.
Labor market productivity counterfactual exercise (Figure 2.9)	Australia; Canada; Czech Republic; Denmark France; Germany; Greece; Italy; Netherlands; New Zealand; Spain; Sweden; United Kingdom; and United States.
Import competition shock (Figures 2.10 and 2.12)	Australia; Austria; Canada; Czech Republic; Denmark; Finland; France; Germany; Greece; Ireland; Italy; Netherlands; Portugal; Slovak Republic; Slovenia; Spain; Sweden; United Kingdom; and United States.
Automation shock (Figures 2.11 and 2.12)	Australia; Austria; Belgium; Canada; Czech Republic; Denmark; Finland; France; Germany; Greece; Ireland; Italy; Korea; Netherlands; New Zealand; Norway; Portugal; Spain; Slovak Republic; Slovenia; Sweden; Switzerland; United Kingdom; and United States.
Subnational regional migration and labor mobility (Figure 2.13)[3]	Panel 1: Austria; Belgium; Bulgaria*; Czech Republic; Denmark; France; Greece; Hungary*; Norway; Poland*; Portugal; Romania*; Slovak Republic; Slovenia; Spain; Sweden; Switzerland. Panels 2 and 3: Austria; Belgium; Bulgaria*; Czech Republic; Denmark; France; Greece; Hungary; Norway; Poland*; Portugal; Romania*; Slovak Republic; Slovenia; Spain; Sweden; and Switzerland.
Orbis/firm-level data and allocative efficiency analysis (Figure 2.14)	Australia; Austria; Belgium; Bulgaria*; Denmark; Finland; France; Germany; Hungary*; Italy; Norway; Poland*; Portugal; Romania*; Slovak Republic; Spain; Sweden; Switzerland; United Kingdom; and United States.

Source: IMF staff compilation.
[1]Asterisk (*) denotes emerging market and developing economies as classified by October 2019, *World Economic Outlook*.
[2]Panel 2 is a balanced AE sample beginning in 1970. Panel 4 is a balanced EMDE sample beginning in 2000.
[3]Panels 2 and 3 use European Union Labour Force Survey data 2000–16.

Subnational regions are classified as lagging or other (nonlagging) according to two criteria, based on: (1) their level of real GDP per capita compared to the within-country regional median in 2000 (above/below); and (2) their average annual real GDP per capita growth compared to their overall country's average real GDP per capita growth over the period 2000-16 (above/below). If a subnational region is classified as below on both these criteria, it is considered a lagging region in our analysis—initially relatively poor and not catching up to the country average during 2000-16. About 22 percent of regions in advanced economies are classified as lagging, with some countries having about one-third of regions lagging while others have none.

Annex 2.2 Calculation and Decomposition of Income Inequality

The generalized entropy (GE) index is a measure of inequality in the distribution of a strictly positive variable (individual or household income) that can be calculated for a given parameter a, whose value indicates the weight given to different parts of the income distribution. The index is more sensitive to lower incomes for lower values of a and to the existence of higher incomes for higher values. Unlike some other inequality indices like the well-known Gini coefficient, GE indices have the advantage that they are additively decomposable—the overall value of the index can be partitioned according to the properties of individual units, such as region, gender, age, and so on (Shorrocks 1980). GE(a=0) is also known as the mean log deviation (MLD). The MLD is zero when all units have identical incomes and takes larger positive values as incomes become more unequal. It is calculated as:

$GE(0)$ for the entire population

$$GE(0) = -\frac{1}{n}\sum_{i} \ln(\frac{y_i}{\overline{y}})$$

where n is the number of households, y_i is income of household i, and $\overline{y}$ is the mean of y_i. For a country-level index, a decomposition into components due to average differences across regions and within region differences is possible.

$GE(0)$ decomposition across regions (indexed by k)

$$GE(0) = \underbrace{\sum_{k} v_k GE(0)_k}_{within} + \underbrace{\sum_{k} v_k \ln(\frac{1}{\lambda_k})}_{between}$$

where $v_k = \frac{n_k}{n}$ is the population share of region k, $GE(0)_k$ is $GE(0)$ calculated for region k, and $\lambda_k = \frac{\overline{y}_k}{\overline{y}}$ is the relative mean income of region k. The within term captures the inequality due to the variability of income within each region, while the between term captures the inequality due to the variability of income across different regions.

Annex 2.3 Shift-Share Analysis of Regional Labor Productivity

This section decomposes the labor productivity (defined as real gross value-added per worker) difference between individual regions and the national average into three components: 1) a sectoral employment mix component; 2) a pure productivity differential component; and 3) an allocative efficiency component. The allocative efficiency component in this exercise is different from the sensitivity of capital investment to returns described and estimated in Annex 2.6. In the labor productivity decomposition here, allocative efficiency is captured by the product of the deviations of the regional employment share and labor productivity by sector from the national employment share and labor productivity by sector. It is clear that lagging regions, defined in terms of the convergence criterion, have lower labor productivity than other (non-lagging) regions (Figure 2.8, panel 1). The goal here is to establish whether lagging regions have lower regional productivity simply because they specialize in sectors that are low-productivity at the county level, or if they have lower labor productivity even after fixing the sectoral employment mix at the national level. The data allow for a breakdown of employment and labor productivity by region across ten sectors (see Annex 2.1 for the sector listing).

The difference between regional and national labor productivity, $(y_r - y)$, is decomposed into the three components listed above following Esteban (2000), as in the below equation. The following notation is used: y_r denotes labor productivity in region r, y is national labor productivity, $y_{r,s}$ denotes labor productivity in region r and sector s, and y_s for national labor productivity in sector s. The national employment share in sector s (l_s), and employment share in region r and sector s ($l_{r,s}$), are defined analogously, so that:

$$(y_r - y) = \sum_s l_{r,s} y_{r,s} - \sum_s l_s y_s = \underbrace{\sum_s (l_{r,s} - l_s)\, y_s}_{\mu_r,\ \text{employment mix}} + \underbrace{\sum_s l_s (y_{r,s} - y_s)}_{\pi_r,\ \text{pure productivity differential}} + \underbrace{\sum_s (l_{r,s} - l_s)\,(y_{r,s} - y_s)}_{\alpha_r,\ \text{allocative efficiency}}$$

In short, the equation above can be rewritten as:

$$(y_r - y) = \mu_r + \pi_r + \alpha_r$$

where μ_r represents the sectoral employment mix component, π_r represents the pure productivity differential component, and α_r represents the allocative efficiency component, all for region r. Note that all the labor productivities above are expressed in international dollars per worker (constant PPP).

To ease comparison across countries, the relative weight of the variance of each of the three components (μ_r, π_r, α_r) in the overall observed variance in $(y_r - y)$ is computed, as well as a term summing up their covariances. Specifically, the variance decomposition is provided in the equation below. Figure 2.7 presents the sample analog of this decomposition, calculated at the region level based on variation over time, and finally averaged across regions within a country in order to show a summary of results at the country level.

$$Var(y_r - y) = Var(\mu) + Var(\pi) + Var(\alpha) + 2[Cov(\mu, \pi) + Cov(\mu, \alpha) + Cov(\pi, \alpha)]$$

This shift-share analysis is used to further our understanding of the underlying forces that explain the difference between lagging and other regions. For example, regression analysis shown in Annex Table 2.3.1 indicates that lagging regions have both an unfavorable sectoral employment mix, and lower (pure) productivity vis-à-vis the national average, as expected given findings from Figure 2.8 in the main text. Moreover, the very large magnitude of the coefficient when looking at the pure productivity differential, indicated that this component plays a major role in explaining the labor productivity differences between the lagging and other regions.

Annex Table 2.3.1. Lagging Regions and Shift-Share Decompositions

	Industry mix	Pure productivity differential	Allocative efficiency
Lagging region	-1682.4***	-7355.7***	-586.8
	(-3.39)	(-5.04)	(-1.91)
Constant	-1139.5***	123.5	140.8
	(-5.06)	(0.19)	(1.01)
Number of observations	251	251	251
	0.174	0.178	0.230

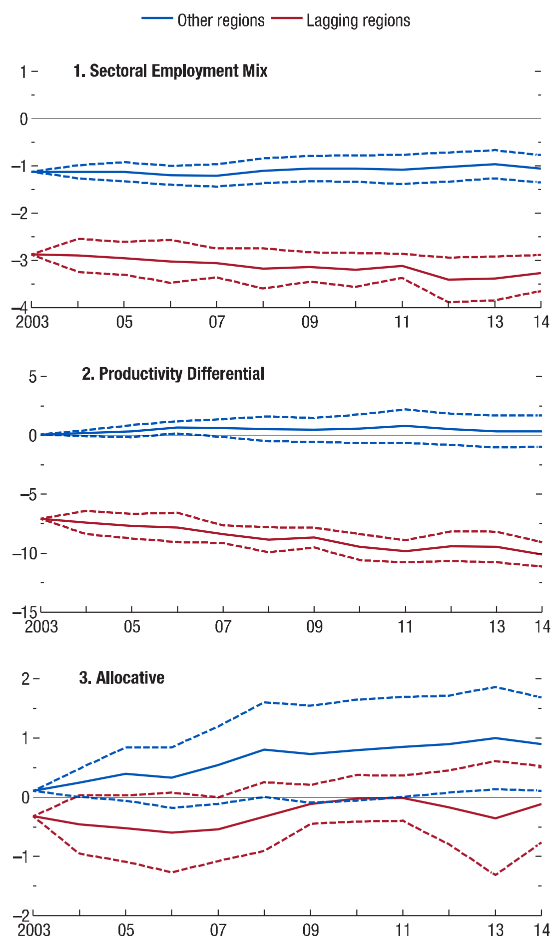
Source: IMF staff calculations.
Note: The three shift-share components are all expressed in USD (constant PPP). In this regression each component is averaged over the period 2003-2016, so that the number of observations equals the number of regions in the sample. Lagging region is a dummy variable equal to 1 if the region has below median initial income and the average growth is below the average country level growth over the period 2000-2016. All regressions control for country fixed effects. T-statistics shown in parentheses.
* p<0.05,** p<0.01,*** p<0.001

Finally, Figure 2.3.1 compares the evolution of each of the shift-share components over time, for lagging and other regions respectively. It reinforces the two conclusions based on regression analysis in Annex Table 2.3.1: 1) both the pure productivity differential and the sectoral employment mix contribute to lagging regions' poor labor productivity, but also 2) differences in within-sector labor productivity, also referred to as pure productivity differential, are the dominant force that explains why regions are lagging.

Over time, the gap between the lagging and other regions has widened for the pure productivity differential, a worrisome trend. The level differences between the lagging and other regions for the sectoral employment mix component and the productivity differential component are large and continue to be large. Finally, allocative efficiency does not seem to be different across the lagging and other regions.

Annex Figure 2.3.1. Changes in Shift-Share Components Over Time

(Thousand US dollars per worker, constant PPP)



Sources: IMF staff calculations.
Note: The lines plot year fixed effects from a regression of each shift-share component to show changes from the level of each shift-share component in 2003. Dashed lines indicate 90% confidence intervals.

Annex 2.4 Counterfactual Exercise for Labor Productivity Changes in Lagging Versus Other Regions

This section explains the methodology underlying the exercise to reveal how the effectiveness of regional adjustment and reallocation affects the labor productivity (defined as real gross value-added per worker) difference between lagging versus other regions, as shown in Figure 2.9 in the main text.

The exercise is conducted country-by-country and holds regional sectoral employment shares fixed at their values in an initial year with wide data availability (2002) and explores how regional labor productivity would have evolved if only sectoral labor productivities at the regional level changed as they did. The counterfactual labor productivity for country c for region group $k \in \{\text{lagging, other}\}$ is denoted as $\tilde{y}_{c,t}^{k}$ and constructed as:

$$\tilde{y}_{c,t}^{k} = \frac{\sum_{r \in \Lambda_{c,k}} \left[y_{r,c,2002} \left(\sum_s e_{r,c,s,2002} \right) \left(\sum_s \left(\frac{e_{r,c,s,2002}}{\left(\sum_m e_{r,c,m,2002} \right)} \frac{y_{r,c,s,t}}{y_{r,c,s,2002}} \right) \right) \right]}{\sum_{r \in \Lambda_{c,k}} \left(\sum_s e_{r,c,s,2002} \right)},$$

Here $\Lambda_{c,k}$ denotes the set of regions in country c with property k (lagging or other) and $e_{r,c,s,t}$ and $y_{r,c,s,t}$ denote the employment (in number of individuals) and the labor productivity of sector s in region r of country c in year t, respectively.

The actual labor productivity for country c for region group k is denoted by $y_{c,t}^{k}$ and constructed as:

$$y_{c,t}^{k} = \frac{\sum_{r \in \Lambda_{c,k}} y_{r,c,t} \left(\sum_s e_{r,c,s,t} \right)}{\sum_{r \in \Lambda_{c,k}} \left(\sum_s e_{r,c,s,t} \right)}.$$

The chart shows the average of the ratios of labor productivity in lagging versus other regions (either actual or counterfactual, denoted below with tilde) over the sample of advanced economies (with at least one lagging region):

$$\frac{y_{c,2002}^{\text{Lagging}}}{y_{c,2002}^{\text{Other}}}, \frac{\tilde{y}_{c,2014}^{\text{Lagging}}}{\tilde{y}_{c,2014}^{\text{Other}}}, \text{ and } \frac{y_{c,2014}^{\text{Lagging}}}{y_{c,2014}^{\text{Other}}}.$$

Annex 2.5 Regional Labor Market Effects of Local Labor Demand Shocks: Trade and Technology Shocks

The chapter estimates the regional labor market responses to identified trade and technology shocks, differentiating between their effects on the average exposed region versus a lagging region. Differences in the initial sectoral employment industry mix across regions imply different exposures to shocks, variation across which has been exploited extensively in the prior literature (Bartik 1991; Blanchard and Katz 1992; Topalova 2010; Autor and Dorn 2013; Autor, Dorn, and Hanson 2013a; Dauth, Findeisen, and Suedekum 2014; Jakubik and Stolzenburg 2018). This section provides details on the construction of the shocks and presents the econometric specification behind results shown in Figures 2.10-12. It also presents further details on the incidence of shocks across countries and a discussion of the robustness of the findings to excluding advanced economies with large commodity exports or subnational regions identified as resource-rich with large shares of oil and minerals production. Employment in a region is exhaustively allocated across ten sectors (see Annex 2.1 for the sectoral listing).

Trade: Import Competition in External Markets from China

The trade shock is based on import competition from China in external markets, per local worker. Definition of the import competition shock mirrors that of Autor, Dorn and Hanson (2013a).

$$\Delta IPW_{r,c,t} = \sum_s \frac{L_{r,c,s,2000}}{L_{r,c,2000}} \frac{\Delta M_{o,s,t}}{L_{c,s,2000}}$$

where $\Delta M_{o,s,t}$ is the log difference of imports from China. To capture rising competitiveness of Chinese exporters, rather than domestic import demand, the chapter uses imports from China to a set of advanced economies, excluding home.[1] Thus, imports are indexed with o (for other countries), rather than c (for home country). The log difference of Chinese imports in sector s in other advanced countries is divided by the number of workers in the same sector in the home country. Finally, this is weighted by the subnational regional industry mix $(\frac{L_{r,c,s,2000}}{L_{r,c,2000}})$ for region r in the year 2000, which is the earliest available year in the sample. This mitigates possible simultaneity bias, since contemporaneous employment by region could be affected by anticipated China trade.

The impact of the trade shock on regional performance is estimated using the local projection method (Jordà 2005), modelled as follows:

$$y_{r,c,t+h} - y_{r,c,t-1} = \beta_h z_{r,c,t} + \boldsymbol{\gamma}_h{}' \boldsymbol{X}_{r,c,t} + \alpha_{r,c,h} + \alpha_{t,h} + \epsilon_{r,c,h,t}. \quad (2.5.1)$$

For horizons $h = \{0,1,\dots,4\}$, the change in outcome variable y in country c region r from year $t-1$ to year $t+h$ is a function of the trade shocks $z_{r,c,t}$, controlling for lagged subnational log real GDP per capita, lagged national real GDP per capita, and lagged log population density (all

[1] The set of advanced economies are the same as in Autor, Dorn and Hanson (2013) and include: Australia, Denmark, Finland, Germany, Japan, Spain, Switzerland. In addition to those, the United States is also included. For regions in one of these eight countries, imports to the country itself are excluded.

subsumed in $\boldsymbol{X}_{r,c,t}$), as well as subnational region and year fixed effects, $\alpha_{r,c,h}$ and $\alpha_{t,h}$, respectively. The outcome variables considered are the regional unemployment rate, the regional labor force participation rate, and log regional in- and out-migration (in log number of individuals). The coefficient β_h captures the impact of the import competition shock at horizon h. In Figures 2.10 and 2.12, this effect is scaled by the standard deviation of the shocks (a representative adverse import competition shock).

To show how the import competition shock affects lagging regions, equation (2.5.1) is modified to add an interaction term between the trade shock and a lagging region indicator, and control for the lagging indicator. As elsewhere in the chapter, lagging regions are defined as those with both initial real GDP per capita that is below the country's regional median in 2000, and average growth of regional real GDP per capita below the country's growth rate over 2000-16.

Finally, to explore the role of policies, the equation in (2.5.1) is modified to add a policy interaction term:

$$y_{r,c,t+h} - y_{r,c,t-1} = \beta_{1,h} w_{c,t} \cdot z_{r,c,t} + \beta_{2,h} w_{c,t} + \beta_{3,h} z_{r,c,t} + \gamma_h{}' X_{r,c,t} + \alpha_{r,c,h} + \alpha_{t,h} + \epsilon_{r,c,h,t}, \tag{2.5.2}$$

where $w_{c,t}$ indicates the national-level policy indicator of interest. In the analysis here, the stringency of employment protection regulation and the generosity of the unemployment insurance scheme are considered, one at a time.

To assess how the effect of the import competition may vary depending on the existing policy environment in the country, Figure 2.12 plots $\beta_{1,h} w_{\text{low}} + \beta_{3,h}$ and $\beta_{1,h} w_{\text{high}} + \beta_{3,h}$ at different horizons h, where w_{low} and w_{high} denote the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles of the distribution of the selected structural policy indicator w in the advanced economies sample respectively (each scaled by the standard deviation of the import competition shock).

Technology: Machinery and Equipment Capital Prices and Vulnerability to Automation

Earlier studies have found that falling costs of automation can generate labor market polarization (Autor and Dorn 2013). This implies that regions that are more exposed to automation, defined as those with a larger share of the labor force being involved in routine jobs (that is, performing tasks that are repetitive and codifiable and hence more likely to be implemented by a robot), face a larger labor market adjustment need when the cost of automation falls.

The impact of automation is captured through an interaction between the exposure to routinization and a fall in the relative price of investment goods. The subnational regional exposure to routinization $RTI_{r,c,t}$ for region r in country c in year t is constructed as the average sectoral routinization score weighted by sectoral employment ($L_{r,c,s,t}$) shares within the region:

$$RTI_{r,c,t} = \sum_s \left(\frac{L_{r,c,s,t}}{\left(\sum_{s=1}^{S_{i,t}} L_{r,c,s,t}\right)} RTI_s \right),$$

where the routinization score for sector s denoted by RTI_s is time invariant and constructed from task-specific routinization scores and the task share within United States' sectors (Das and Hilgenstock 2018).

As with the trade shock, the impact of automation on local labor market performance is estimated using local projection methods (Jordà 2005):

$$y_{r,c,t+h} - y_{r,c,t-1} = \beta_{1,h}\Delta p_{r,c,t} + \beta_{2,h}RTI_{r,c,t-1} + \beta_{3,h}\Delta p_{c,t} \cdot RTI_{r,c,t-1} + \boldsymbol{\gamma}_h{}'\boldsymbol{X}_{r,c,t} + \alpha_{r,c,h} + \alpha_{t,h} + \epsilon_{r,c,h,t}, \quad (2.5.3)$$

where the controls are identical to that of equation 2.5.1 examining the impact of import competition shocks. For horizons $h = \{0,1, \ldots ,4\}$, the change in outcome variable y in country c region r from year $t-1$ to year $t+h$ is a function of the change in relative price of investment goods $\Delta p_{c,t}$, routinization score $RTI_{r,c,t-1}$, and the interaction $\Delta p_{c,t} \cdot RTI_{r,c,t-1}$, controlling for lagged subnational log real GDP per capita, lagged national real GDP per capita, and lagged log population density (all subsumed in $\boldsymbol{X}_{r,c,t}$), as well as subnational region and year fixed effects, $\alpha_{r,c,h}$ and $\alpha_{t,h}$, respectively. The outcome variables considered are the regional unemployment rate, the regional labor force participation rate, and log regional in- and out-migration (in log number of individuals). The coefficient $\beta_{3,h}$ captures the effect of the automation shock, arising from a fall in the relative price of investment goods (machinery and equipment prices) with an increase in regional vulnerability. In Figures 2.11 and 2.12, this effect is scaled by the product of the standard deviations of $RTI_{r,c,t-1}$ and $\Delta p_{c,t}$ times minus one (as a representative adverse automation shock).

To show how automation shocks affect lagging regions, equation (2.5.3) is modified to add interaction terms between a lagging region indicator and automation-related variables, while also controlling for the lagging region indicator. As elsewhere in the chapter, lagging regions are defined as those with both initial real GDP per capita that is below the country's regional median in 2000, and average growth of regional real GDP per capita below the country's growth rate over 2000-16.

Finally, to study the role of national-level policies on the regional impact of automation, the equation in (2.5.3) is modified as below:

$$\begin{aligned} y_{r,c,t+h} - y_{r,c,t-1} = {} & \beta_{1,h}\Delta p_{c,t} + \beta_{2,h}RTI_{r,c,t-1} + \beta_{3,h}\Delta p_{c,t} \cdot RTI_{r,c,t-1} \\ & + \delta_{1,h}w_{c,t} \cdot \Delta p_{c,t} + \delta_{2,h}w_{c,t} \cdot RTI_{r,c,t-1} + \delta_{3,h}w_{c,t} \cdot \Delta p_{c,t} \cdot RTI_{r,c,t-1} \\ & + \theta_h w_{c,t} + \boldsymbol{\gamma}_h{}'\boldsymbol{X}_{r,c,t} + \alpha_{r,c,h} + \alpha_{t,h} + \epsilon_{r,c,h,t}, \end{aligned} \quad (2.5.4)$$

where the automation shock-related variables are interacted with a national-level policy variable $w_{c,t}$, similar to equation 2.5.2.

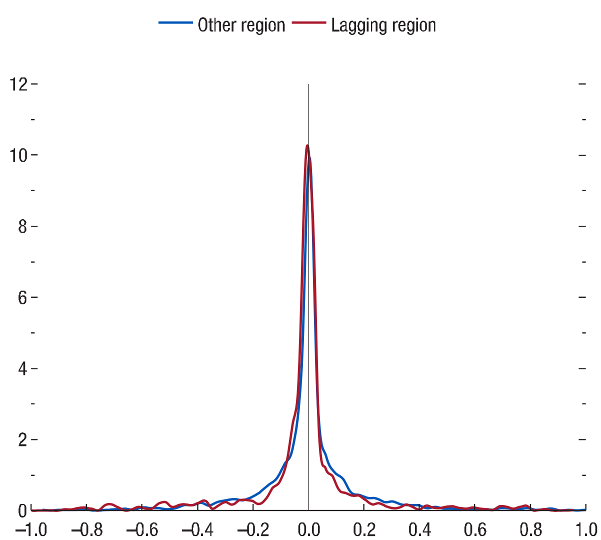
The regional effect of the automation shock then depends on the level of the selected national structural policy indicator w . Evaluated at 25th and 75th percentiles of the distribution of the indicator w in the sample of advanced economies, Figure 2.12 shows for each horizon h,

$\delta_{3,h} w_{\text{low}} + \beta_{3,h}$ and $\delta_{3,h} w_{\text{high}} + \beta_{3,h}$ (scaled by the representative adverse automation shock as above).

Trade and Technology Shock Incidence across Lagging and Other Regions

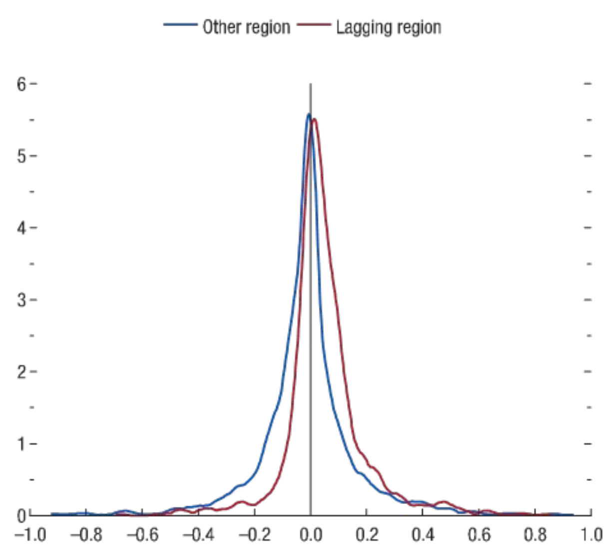
See Annex Figures 2.5.1 and 2.5.2 for the distributions of the shocks over the samples of lagging and other regions, after controlling for country-year fixed effects. Kolmogorov-Smirnov tests of the differences in shock distributions between lagging and other regions are statistically significant and suggest that lagging regions are actually somewhat less likely to receive adverse trade and technology shocks (as defined here).

Annex Figure 2.5.1. Distribution of Import Competition Shocks
(Density)



Sources: OECD Regional Database; and IMF staff calculations.
Note: This figure plots the kernel densities of the residuals from a regression of import competition shocks on country-year fixed effects, according to whether or not the shocks were to a lagging or other region. The shocks are constructed as described in Annex Section 2.5 for the sample of advanced economies over 2000–16. Lagging regions in a country are defined as those with real GDP per capita below the country regional median in 2000 and with average growth below the country's growth over 2000–16. A Kolmogorov-Smirnov test of the difference in the distributions of shocks for lagging versus others is statistically significant and indicates milder adverse import competition shocks for lagging regions.

Annex Figure 2.5.2. Distribution of Automation Shocks
(Density)



Sources: OECD Regional Database; and IMF staff calculations.
Note: This figure plots the kernel densities of the residuals from a regression of automation shocks on country-year fixed effects, according to whether or not the shocks were to a lagging or other region. The shocks are constructed as described in Annex Section 2.5 for the sample of advanced economies over 2000–16. Lagging regions in a country are defined as those with real GDP per capita below the country regional median in 2000 and with average growth below the country's growth over 2000–16. A Kolmogorov-Smirnov test of the difference in the distributions of shocks for lagging versus others is statistically significant and indicates milder adverse automation shocks (that is, less negative) for lagging regions.

Robustness to Exclusion of Commodity Exporters or Resource-Rich Regions

Resource-rich regions within a country with high shares of oil and mining production in their economic activity may pull away from the rest due to commodity booms. Other regions in such countries could be classified as lagging by comparison as a consequence, even if their living standards are not necessarily low. To ensure that the findings are robust to excluding such regions, reestimation of the stylized facts and regression analyses were undertaken using two alternative samples: (1) excluding advanced economy commodity exporters as identified in chapter 2 of the September 2015 *World Economic Outlook* (Australia, Canada, and Norway) plus the United Kingdom; and (2) excluding oil and mineral producing regions identified through a mix of data (share of real GDP from oil and mining; oil production per capita; and so on) and judgment (drawing upon internal expertise on commodity production worldwide), reflecting the sometimes poor data availability on geographically disaggregated commodity production. Commodity producing regions so-identified account for about 5 percent of the sample, appearing in 6 advanced economies (Australia, Canada, Netherlands, Norway, United Kingdom, and United States). The stylized facts regarding lagging versus other regions and regression results under these two alternative samples showed no material difference to the baseline results using the full sample (results available upon request).

Annex 2.6 Firm-Level Analysis of Capital Allocative Efficiency

This annex describes the approach used to estimate the distribution of firm's capital allocative efficiency across regions by sector within a country. This section uses Orbis data, provided by Bureau van Dijk, a Moody's Analytics company. Orbis contains information on millions of companies across the globe although the coverage varies by country. The main strength of the dataset lies in the availability of harmonized cross-country financial information for both privately held and publicly listed firms since the mid-90s. The data were obtained through the "Orbis Historical" product that provides the best time series coverage.

The analysis focuses on the sample of countries for which the firms included in Orbis represent at least 40% of the total output reported in official sources. The United States is included in the sample, even when having a somewhat lower coverage in some years, given its relevance in the global economy. For most countries the data span 2000 to 2015. However, some countries have a slightly shorter time series: Austria, Germany, Korea, and Slovak Republic.

The "raw" data requires intensive cleaning prior to estimation. The cleaning procedure follows closely Kalemli-Özcan and others (2015), Gopinath and others (2017) and Gal (2013). The cleaning steps first involve dealing with basic reporting mistakes (that is, negative sales, total assets, employment, cost of employees, tangible fixed assets or liabilities; missing or zero values for the cost of materials, operating revenue, total assets and missing NACE industry code). Second, the cleaning procedure conducts further quality checks that verify the age of the firm, the ratio of short-term to long-term liabilities, the ratio of employees to capital, tangible fixed assets to total assets, capital to shareholder funds, and total assets to shareholder funds. The procedure also applies filters on the annual growth rates of sales, operating revenues and number of employees by company size category. Finally, the main balance sheet variables are deflated and PPP-adjusted (that is, adjusted by purchasing power parity) to allow cross-comparability. The details of the data construction are provided in Díez, Fan, and Villegas-Sánchez (2019). The industry classification from Orbis is mapped to the OECD industry classification. The postal code of the firm is used to map it to the region in which the firm is based in.[1]

Assuming a Cobb-Douglas function, the marginal product of capital and labor is equal to the average product of capital and labor, respectively. The marginal of product of labor and capital can therefore be defined as value added divided by labor or capital. The allocative efficiency of firms is defined as their sensitivities of the growth in capital and labor to their respective marginal returns.

The following regression equation is estimated:

$$\Delta Y_{i,t} = \beta_{r,c,s,t} * MP_{i,t-1} + \alpha_{r,c,s,t} + \alpha_i + \epsilon_{i,t}$$

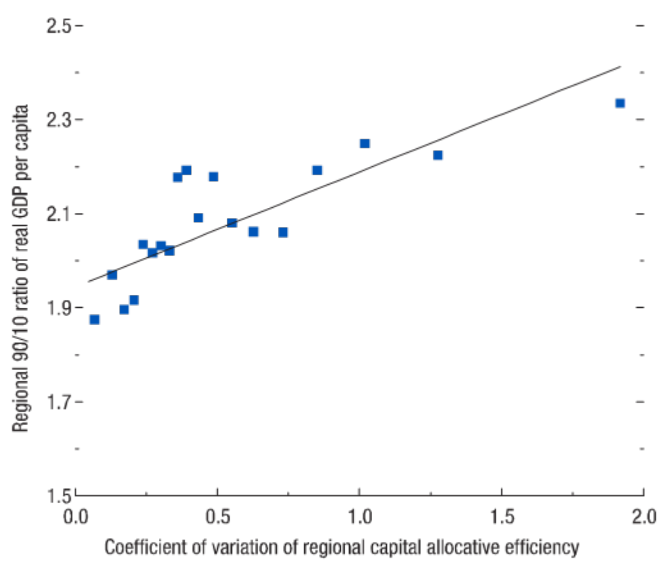
[1] Firms which report more than one postal code, the postal code that occurs most frequently is used. The postal code is merged through the GeoNames Postal Code Database to the OECD Regional Database. If the postal code can be mapped to more than one region, the firm is assigned to all the regions in which it is based.

where $\Delta Y_{i,t}$ is the log difference of the capital stock (approximately capital growth) for firm i in year t. $\alpha_{r,c,s,t}$ is a region-country-sector-year fixed effect, MP is the marginal revenue product of capital, defined as the log of value added divided by capital, α_i is a firm-level fixed effect. $\beta_{r,c,s,t}$ reflects the average sensitivity of capital to its marginal revenue product across all firms within each region-country-sector-year. These coefficients can be interpreted as measures of the capital allocative efficiency of each region-sector-year; higher values indicate a greater responsiveness of the firm's capital investment decision to the marginal return.

The regression is based on an unbalanced panel of 2,580,600 firms from 24 countries from 1985 to 2016 in 264 regions in 10 sectors for the relationship between the marginal product of capital and investment (capital growth). A high coefficient indicates that firms increase capital more in response to an increase in the marginal product of capital. In the next step, the distribution of subnational $\beta_{r,c,s,t}$ by country are examined. The focus is on the coefficient of variation of $\beta_{r,c,s,t}$ across regions within each country-sector-year, denoted by $\sigma^Y_{c,s,t}$ and defined as the ratio of the standard deviation of the coefficients across regions divided by the mean value (all coefficients are strictly positive). The coefficients of variation of capital allocative efficiencies within each country indicate the extent of regional disparities in allocative efficiencies by sector.

$\sigma^Y_{c,s,t}$ is strongly correlated with the dispersion of firm-level revenue total factor productivity (TFPR) across regions (an alternative measure of the extent of regional factor misallocation under some assumptions), even after controlling for sector-year and country-sector fixed effects. The dispersion of firm-level TFPR across regions is calculated by first taking the average firm-level TFPR measured estimated using the procedure from Gandhi, Navarro, and Rivers (2018) within each region-country-sector-year and then taking the standard deviation across country-sector-years. Moreover, as seen in Annex Figure 2.6.1 and mentioned in the main text, there is a correlation between regional disparities in economic activity and the coefficient of variation of regional capital allocative efficiency, suggesting that some of the regional disparities signifies regional misallocation.

Annex Figure 2.6.1. Dispersion in Allocative Efficiency is Correlated with Greater Regional Disparities



Sources: OECD Regional Database; Orbis; and IMF staff calculations.
Note: The figure illustrates the regression slope for the relationship between regional 90/10 ratios of real GDP per capita and the coefficient of variation of regional capital allocative efficiency after controlling for country-sector-year fixed effects. Dots show the binned underlying data from the regression, based on the method from Chetty, Friedman, and Rockoff (2013). See Annex 2.1 for the country sample.

Figure 2.14 in the main text is drawn from the results of the following regression model:

$$\sigma^Y_{c,s,t} = \beta_1 w_{c,t} + \alpha_{s,t} + \alpha_{c,t} + \epsilon_{c,s,t}$$

where $\sigma^{Y}_{c,s,t}$ is the coefficient of variation by country-sector (as described above), $\alpha_{s,t}$ is a sector-year fixed effect, $\alpha_{c,t}$ is a country-year fixed effect. National level structural policy indicators are considered one-by-one and denoted by w. The set includes: (i) an index that reflects the stringency of national-level product market regulation; (ii) higher trade openness as measured by the sum of national exports and imports divided by country GDP; or (iii) an indicator for the ease of starting a business in the country (from the World Bank's Doing Business Database). Standard errors are clustered on the country-year level. The policy variables are normalized in the figure by their standard deviation, such that β_1 can be interpreted as the effect in units of coefficient of variation for a one standard deviation change in the policy indicator.