Econometrica Supplementary Material

SUPPLEMENT TO "GROWING LIKE INDIA—THE UNEQUAL EFFECTS OF SERVICE-LED GROWTH"

(*Econometrica*, Vol. 91, No. 4, July 2023, 1457–1494)

TIANYU FAN
Department of Economics, Yale University

MICHAEL PETERS
Department of Economics, Yale University and NBER

FABRIZIO ZILIBOTTI
Department of Economics, Yale University and NBER

APPENDIX A: THEORY

IN THIS SECTION, we discuss the technical material referred to in the text.

A-1. *Proof of Proposition 1*

To derive expression (5), note that the definition of p_{rnt} implies that

$$\int_n \beta_n \ln p_{rnt}\,dn = \ln P_{rFt}\int_n \beta_n\lambda_{nF}\,dn + \ln P_{rGt}\int_n \beta_n\lambda_{nG}\,dn + \ln w_{rt}\int_n \beta_n\lambda_{n\mathrm{CS}}\,dn - \int_n \beta_n\lambda_{n\mathrm{CS}}\ln \mathcal{A}_{rnt}\,dn.$$

Using the definitions of ω_s and $A_{r\mathrm{CS}t}$, we obtain $\int_n \beta_n \ln p_{rnt}\,dn = \omega_F \ln P_{rFt} + \omega_G \ln P_{rGt} + \omega_{\mathrm{CS}} \ln(A_{r\mathrm{CS}t}^{-1} w_{rt})$. Similarly, $\int_n \kappa_n \ln p_{rnt}\,dn = \nu_F \ln P_{rFt} + \nu_G \ln P_{rGt} + \nu_{\mathrm{CS}} \ln(A_{r\mathrm{CS}t}^{-1} w_{rt})$, where ν_s is defined in (6). Substituting these expression in (2) and recalling that $P_{rGt} = (A_{r\mathrm{CS}t}^{-1} w_{rt})$ yields the expression for $\mathcal{V}(e, \mathbf{P}_{rt})$ in (5).

To derive expression (7), note that sector s receives a share λ_{ns} of total revenue of good n. Hence,

$$\vartheta(e,\mathbf{P}_{rt}) = \frac{\int_n \lambda_{ns} e \vartheta_n^{\mathrm{FE}}(e,\mathbf{P}_{rt})\,dn}{e} = \omega_s + \nu_s\left(\frac{e}{P_{rFt}^{\omega_F} P_{rGt}^{\omega_G}\left(A_{r\mathrm{CS}t}^{-1} w_{rt}\right)^{\omega_{\mathrm{CS}}}}\right)^{-\varepsilon},$$

which is the expression in (7). In Section WA-1.1 in the Web Appendix, we extend this analysis to the case of a CES production function for final goods.

A-2. *Estimation of Parameters and Productivity (Sections 5.1 and 5.2)*

In this section, we describe in more detail how we estimate the productivity fundamentals $\{A_{rst}\}$ and the structural parameters ω_{CS} and ν_F. Given regional data on educational

Tianyu Fan: tianyu.fan@yale.edu
Michael Peters: m.peters@yale.edu
Fabrizio Zilibotti: fabrizio.zilibotti@yale.edu

 https://doi.org/10.3982/ECTA20964

attainment and sector-region data on earnings, we calculate $\{w_{rt}, H_{rFt}, H_{rGt}, H_{r\text{CS}t}\}_r$ in a model-consistent way. Human capital in location r is given by $H_{rt} = L_{rt}\sum_s \exp(\rho \times s)\ell_{rt}(s)$, where ρ is the return to education, and $\ell_{rt}(s)$ denotes the share of people in region r with s years of education at time t. Sectoral labor supply is then given by

$$H_{rst} = \frac{\sum_i 1[i \in s]w_{it}}{\sum_i w_{it}} \times H_{rt},$$

where w_{it} is the wage of individual i in region r at t and $1[i \in s]$ is an indicator function if individual i works in sector s. The average regional skill price w_r can be calculated as $w_{rt} = (\sum_{i\in r} w_{it})/H_{rt}$.

Step 1: Estimation of ω_{CS} and $\bar{\nu}_F$. The two structural parameters are jointly identified from aggregate market clearing conditions. The local market clearing equations (11)–(12), imply the two aggregate resources constraints for tradable goods $s = F, G$:

$$\sum_{r=1}^{R} w_{rt}H_{rst} = \sum_{r=1}^{R}\sum_{j=1}^{R} \pi_{rsjt}\left(\omega_s + \bar{\nu}_s\left(\frac{A_{j\text{CS}t}^{\omega_{\text{CS}}}\mathbb{E}_{jt}[q]w_{jt}^{1-\omega_{\text{CS}}}}{P_{jFt}^{\omega_F}P_{jGt}^{\omega_G}}\right)^{-\varepsilon}\right)w_{jt}H_{jt}. \tag{A-1}$$

One of the constraints is redundant due to Walras's law. We can substitute the local market clearing condition for CS (11) into the aggregate resources constraint for agriculture to obtain

$$\sum_{r=1}^{R} w_{rt}H_{rFt} = \omega_F\sum_{r=1}^{R} w_{rt}H_{rt} - \frac{\bar{\nu}_F}{\bar{\nu}_{\text{CS}}}\sum_{r=1}^{R}\left(\omega_{CS} - \frac{H_{rCSt}}{H_{rt}}\right)w_{rt}H_{rt}. \tag{A-2}$$

Given data on $\{w_{rt}, H_{rst}\}$, (A-2) yields a single equation in three unknowns: ω_F, $\frac{\bar{\nu}_F}{\bar{\nu}_{\text{CS}}}$, and ω_{CS}. We externally calibrate ω_F. Also, it is clear from the set of CS market clearing conditions in (11) that $\bar{\nu}_{\text{CS}}$ is not separately identified from the average CS productivity *level* $A_{*\text{CS}t}$. As such, the level is not interesting for us; it is legitimate to normalize $\bar{\nu}_{\text{CS}} = -1$. Conditional on a choice for ω_F, we can then use (A-2) in 1987 and 2011 to uniquely pin down ω_{CS} and $\bar{\nu}_F$.

Step 2: Estimation of the Local Price Vector $\{p_{rFt}, p_{rGt}, p_{r\text{CS}t}\}_r$. Let p_{rst} denote the local price of sector s goods in region r. The consumer price index in r is then given by $P_{rst}^{1-\sigma} = \sum_j(\tau_{rj}p_{jst})^{1-\sigma}$. Given the structural parameters, there is a unique local price vector that rationalizes all market clearing conditions (11)–(12).

Using the trade shares $\pi_{rsjt} = \tau_{rj}^{1-\sigma}p_{rst}^{1-\sigma}/P_{jst}^{1-\sigma}$, we can write the market clearing condition for tradable goods (12) as

$$w_{rt}H_{rst} = p_{rst}^{1-\sigma}\left(\sum_{j=1}^{R}\tau_{rj}^{1-\sigma}P_{jst}^{\sigma-1}\bar{\vartheta}_{jst}w_{jt}H_{jt}\right), \quad \text{for } s \in \{F, G\}.$$

Rearranging terms yields

$$p_{rst} = w_{rt}^{\frac{1}{1-\sigma}} H_{rst}^{\frac{1}{1-\sigma}} \left(\sum_{j=1}^{R} \tau_{rj}^{1-\sigma} P_{jst}^{\sigma-1} \bar{\vartheta}_{jst} w_{jt} H_{jt} \right)^{\frac{1}{\sigma-1}}, \quad \text{for } s \in \{F, G\}.$$

Because these equations are homogenous of degree of zero in p_{rst}, we achieve identification by (i) setting the average level of the price of goods as the numeraire $((\sum_r (p_{rGt})^{1-\sigma})^{\frac{1}{1-\sigma}} = 1)$, (ii) normalizing the level of food prices to unity in 1987 $((\sum_r (p_{rF1987})^{1-\sigma})^{\frac{1}{1-\sigma}} = 1)$, and (iii) pinning down the change in aggregate food prices relative to goods prices between 1987–2011 by targeting the published data analogue P_{FGt}^{Data}:

$$\sum_{r=1} \frac{w_{rt} H_{rt}}{\sum_{j=1} w_{jt} H_{jt}} \times \frac{P_{rFt}}{P_{rGt}} = P_{FGt}^{\text{Data}}.$$

We compute the equilibrium price vector as the fixed point of these conditions.

Step 3: Determining the Scale of the Nominal Wage. We proxy income by expenditure. The NSS data on expenditure is reported in rupees. Given the price vector computed in Step 2, we thus scale the observed expenditure in 1987 and 2011 to match a given growth of real GDP per capita. Since we use final goods as the numeraire, we take real GDP per capita to be denominated in goods.

Step 4: Estimation of $\{A_{rst}\}_r$. Given the nominal wage and the local price vector, sectoral productivity is simply given by $A_{rst} = w_{rt}/p_{rst}$. Using the expression for p_{rst} above, we arrive at

$$A_{rst} = w_{rt}^{\frac{\sigma}{\sigma-1}} H_{rst}^{\frac{1}{\sigma-1}} \left(\sum_{j=1}^{R} \tau_{rj}^{1-\sigma} P_{jst}^{\sigma-1} \bar{\vartheta}_{jst} w_{jt} H_{jt} \right)^{\frac{1}{1-\sigma}}, \quad \text{for } s \in \{F, G\},$$

which is equation (18) in the main text

A-3. *The Elasticity of Substitution (Section 5.3)*

In this section, we derive the elasticity of substitution. For simplicity, we suppress the region and time subscripts and denote sectoral prices by P_s. The Allen–Uzawa elasticity of substitution between sectoral output s and k is given by $\text{EOS}_{sk} \equiv \frac{\frac{\partial^2 e(P,V)}{\partial P_s \partial P_k} e(P,V)}{\frac{\partial e(P,V)}{\partial P_s} \frac{\partial e(P,V)}{\partial P_k}}$. The expenditure function is given by

$$e(P, V) = \left(V + \sum_s \nu_s \ln P_s \right)^{1/\varepsilon} \varepsilon^{1/\varepsilon} \prod_{s \in \{F,G,\text{CS}\}} P_s^{\omega_s}.$$

In Section WA-1.2 in the Web Appendix, we prove that

$$\text{EOS}_{sk} = 1 - \varepsilon \frac{(\vartheta_s - \omega_s)(\vartheta_k - \omega_k)}{\vartheta_s \vartheta_k}.$$

A-4. *The Equivalent Variation (Section 6)*

To measure welfare changes, we calculate equivalent variations (EV) relative to the 2011 status quo. Consider the indirect utility of an individual in r with human capital q:

$$\mathcal{V}(qw_r, \mathbf{P_r}) = \frac{1}{\varepsilon}\left(\frac{qw_r}{\prod_s P_{rs}^{\omega_s}}\right)^{\varepsilon} - \sum_s \nu_s \ln P_{rs}. \tag{A-3}$$

We implicitly define the EV for an individual with skills q, $\varpi^q(\hat{x}_r|x_r)$ by

$$\mathcal{V}\big(qw_r\big(1+\varpi^q(\hat{x}_r|x_r)\big), \mathbf{P_r}\big) \equiv \mathcal{V}(q\hat{w}_r, \hat{\mathbf{P}}_\mathbf{r}), \tag{A-4}$$

where $x_r \equiv (w_r, \mathbf{P_r})$ and $\hat{x}_r \equiv (\hat{w}_r, \hat{\mathbf{P}}_\mathbf{r})$ denote the vector of prices in the status quo and the counterfactual respectively. Hence, ϖ_r^q is the percentage change in income that an individual with human capital q living in district r in 2011 would require to attain the same level of utility as in the counterfactual allocation.

Using equations (A-3) and (A-4), we can solve for $\varpi^q(\hat{x}_r|x_r)$ as

$$1+\varpi^q(\hat{x}_r|x_r) = \prod_s \left(\frac{\hat{w}_r/\hat{P}_{rs}}{w_r/P_{rs}}\right)^{\omega_s} \times \left(1 - \left(\sum_s \nu_s \ln\left(\frac{\hat{P}_{rs}}{P_{rs}}\right)\right)\varepsilon\left(\frac{q\hat{w}_r}{\prod_s \hat{P}_{rs}^{\omega_s}}\right)^{-\varepsilon}\right)^{1/\varepsilon}. \tag{A-5}$$

The EV comprises two parts. The first part, $\prod_s((\hat{w}_r/\hat{P}_{rs})/(w_r/P_{rs}))^{\omega_s}$, is akin to the usual change in real wages. This would be the entire EV if preferences were homothetic, that is, if $\nu_s = 0$. The second part captures the unequal effects of productivity growth under nonhomothetic preferences.

In a similar vein, we can calculate the utilitarian welfare effects at the district level. Exploiting the aggregation properties of PIGL, we can determine the change of *regional* spending power $\overline{\varpi}_r(\hat{x}_r|x_r)$ that the representative agent in district r facing prices $\mathbf{P_r}$ would require to attain indifference. As before, $\overline{\varpi}_r(\hat{x}_r|x_r)$ is implicitly defined by

$$\mathcal{U}\big(\mathbb{E}_r[q]w_r\big(1+\overline{\varpi}_r(\hat{x}_r|x_r)\big), \mathbf{P}_r\big) = \mathcal{U}\big(\mathbb{E}_r[q]\hat{w}_r, \hat{\mathbf{P}}_\mathbf{r}\big), \tag{A-6}$$

where $\mathcal{U}$ is defined in (10). One can show that $\overline{\varpi}_r(\hat{x}_r|x_r)$ satisfies an expression similar to the one given in (A-5). As a measure of aggregate welfare, we report the average EV using district population as weights:

$$\overline{\varpi} = \sum_r \overline{\varpi}_r \frac{L_{r2011}}{\sum_r L_{r2011}}.$$

This is a purely statistical measure that does not rest on an aggregation result.

A-5. *Generalizations of Theory (Section 7.4)*

In this section, we describe the extensions discussed in Section 7.4 in more detail. Further technical analyses are available in Section WA-3 in the Web Appendix.

A-5.1. *Open Economy*

In this section, we describe the environment and calibration strategy of the open-economy extension. We defer the technical analysis to Section WA-3 in the Web Appendix.

We assume households, both in India and in the rest of the world, consume industrial goods sourced from many countries. Different national varieties, which are, in turn, CES aggregates of regional varieties enter into a CES utility function as imperfect substitutes. To capture that India might have a specific comparative advantage in ICT services, we assume India exports both domestic goods and ICT services. For simplicity, we assume ICT services are not sold in the Indian domestic market. In our estimation, we assume balanced trade, but we allow India to run a trade deficit in goods and a surplus in ICT services, which is in line with the data.

To calibrate this model, we need information on the revenue of ICT services, the exports and imports of goods, and an estimate of the trade elasticity. We measure ICT revenue from the income share of ICT workers. We classify as ICT service workers all those employed in the following service industries: (i) telecommunications, (ii) computer programming, (iii) consultancy and related software publishing activities, and (iv) information service activities. In our NSS data, these activities constituted 0.72% of total employment in 2011 (in 1987, it was less than 0.1%). ICT workers earn, on average, higher wages than other workers. When one considers the earning share, they account for 1.56% of total earnings in 2011 (in 1987, it was 0.11%). In terms of exports, according to the World Bank, the export of goods and merchandise increased from 11.3 billion (4.1% of GDP) in 1987 to 302.9 billion (16.6% of GDP) in current USD. The manufacturing sector accounted for 66% of such merchandise exports in 1987 and for 62% in 2011. According to the OECD, the domestic value-added in gross exports amounts to 83.9% of exports for India, and we assume this percentage to be constant over time. In accordance with these data, we assume the value-added export of trade increased from 13.9% in 1987 to 53.6% in 2011 as a share of the GDP in the manufacturing sector. Finally, we set the trade elasticity to 5 (Simonovska and Waugh, 2014).

A-5.2. *Imperfect Substitution and Skill Bias in Technology*

In this section, we describe the environment and calibration strategy of the *Imperfect Substitution and Skill Bias in Technology* extension. We defer the technical analysis to Section WA-3 in the Web Appendix.

In this extension, workers with different educational attainments are imperfect substitutes in production. Table WA-III in the Web Appendix shows that agricultural workers have, on average, lower educational attainment than those employed in service industries. Thus, an increase in the skill endowment could be responsible for the reallocation of workers from agriculture to CS (see, e.g., Porzio, Rossi, and Santangelo (2022) or Hendricks and Schoellman (2023)). By ignoring such skill-based specialization, our Ricardian model could potentially exaggerate the importance of technology for the development of the service sector.

We work with two skill groups and define workers to be skilled if they have completed secondary school. We assume the production functions to be of the usual CES form:

$$Y_{rst} = A_{rst}\left(\left(H^{-}_{rst}\right)^{\frac{\rho-1}{\rho}} + \left(Z_{rst}H^{+}_{rst}\right)^{\frac{\rho-1}{\rho}}\right)^{\frac{\rho}{\rho-1}} \quad \text{for } s = F, \text{CS}, G,$$

where H^{+} and H^{-} denote high- and low-skilled workers, respectively. Note that the technology admits differences in both TFP (A_{rst}) and skill bias (Z_{rst}) across sector-districts and time. We assume the elasticity of substitution ρ to be constant across sector-districts and externally calibrate $\rho = 1.8$ (see, e.g., Ciccone and Peri (2005) and Gancia, Müller, and Zilibotti (2013)). Our conclusions do not hinge on the particular calibration of ρ.

We continue to allow for heterogeneous productivity across workers of the same educational group. A worker's wage is a draw from a skill-specific Pareto distribution with the same tail parameter as in our baseline analysis.[1] As in our baseline analysis, this model is exactly identified, and for given structural parameters we can rationalize the data of sectoral earnings shares by skill group and average earnings by skill group for each region in India by choice of A_{rst} and Z_{rst}. Because sectoral productivity is now determined by two parameters, we set both A_{rs2011} and Z_{rs2011} to the respective 1987 level when running counterfactuals.

A-5.3. *Spatial Mobility*

In this section, we describe the environment and calibration strategy of the *Spatial Mobility* extension. The model is in the vein of economic geography models à la Redding and Rossi-Hansberg (2017), in which individuals' migration decisions are modeled as a discrete choice problem, with individuals receiving idiosyncratic preference shocks and locations differing in a scalar amenity. Specifically, we assume that individuals make their location choices prior to knowing their particular skill realization q and draw q from region-specific skill distribution $F_{rt}(q)$. Letting $v_{rt}(q)$ denote the utility of an individual with skills q in region r at time t, the value of settling in location r is given by

$$V_{rt}^i = \mathcal{B}_{rt}\left(\int v_{rt}(q)\,dF_{rt}(q)\right)u_{rt}^i, \tag{A-7}$$

where $\mathcal{B}_{rt}$ is a location amenity and u_{rt}^i is an idiosyncratic preference shock for location r, which we assume to be Frechet-distributed; $P(u_{rt}^i \leq u) = e^{-u^{-\eta}}$. The share of people located in region r at time t is thus given by

$$L_{rt} = \frac{\left(\mathcal{B}_{rt}\int v_{rt}(q)\,dF_{rt}(q)\right)^{\eta}}{\sum_j\left(\mathcal{B}_{jt}\int v_{jt}(q)\,dF_{jt}(q)\right)^{\eta}}L. \tag{A-8}$$

In Section WA-3.3 in the Web Appendix, we formally lay out the model and characterize its equilibrium. In particular, we discuss how we cardinalize consumers' expected consumption utility $\int v_{rt}(q)\,dF_{rt}(q)$ using the equivalent variation ϖ_{rt} to measure location amenities $\mathcal{B}_{rt}$ and idiosyncratic preferences u_{rt}^i in monetary terms. We also show that all our estimates of both structural parameters and sectoral productivities are exactly the same as in the model with immobile labor, because we can use (A-8) to rationalize the observed population distribution through an appropriate choice of amenities $\mathcal{B}_{rt}$.

To perform counterfactuals, we need an estimate of the spatial labor supply elasticity η, which in our context captures a long-run migration elasticity. In the absence of exogenous variation in local wages, this elasticity is hard to estimate directly. We therefore discipline

[1]It is impossible to separately identify the lower bound of the Pareto distribution of human capital draws from the level of the technology. Therefore, we normalize the lower bound to unity for both skill groups. Because we are only interested in changes over time in TFP, this normalization is immaterial.

this elasticity by ensuring that in a counterfactual where we set productivity to its 1987 level in all sectors, the amount of spatial reallocation is as high as what occurred in India between 1987 and 2011. While we think of this choice as an upper bound on the elasticity of spatial supply, we also tested the robustness of our results to higher-elasticity scenarios.

With our calibrated model at hand, we then compute the welfare impact of service-led growth in the presence of spatial mobility in the following way. Combining the equilibrium conditions laid out in Proposition 2 with the spatial labor supply equation (A-8), we can compute equilibrium wages and prices for any change in local productivity. Given these wages and prices, we then simulate the optimal migration behavior of 1 million individuals, given their initial realization of idiosyncratic preference shocks, u^i_{rt}. The counterfactual welfare change for an individual i that was located in region r in 2011 but moved to location j after the counterfactual productivity change is then given by $V^i_{jCF}/V^i_{r2011} - 1$, where V^i_{rt} is given in (A-7) and V^i_{jCF} is the corresponding utility value in the counterfactual allocation. In Table IX in the main text, we report the population-weighted average either at the national level or by urbanization quantile. Note that in the absence of mobility, the utility V^i_{rCF} exactly coincides with our baseline results, given that we cardinalized the location value v_{rt} in monetary terms.

APPENDIX B: Data and Measurement

In this section, we extend the discussion of our empirical analysis in Sections 2 and 4.

B-1. *International Evidence*

In Figure 1, we showed that most service employment in India is concentrated in sectors that serve consumers. Figure B-1, which displays the share of service employment catering to consumers against GDP per capita, shows that this pattern is not a prerogative of India. India is in line with the international pattern, conditional on its GDP per capita.

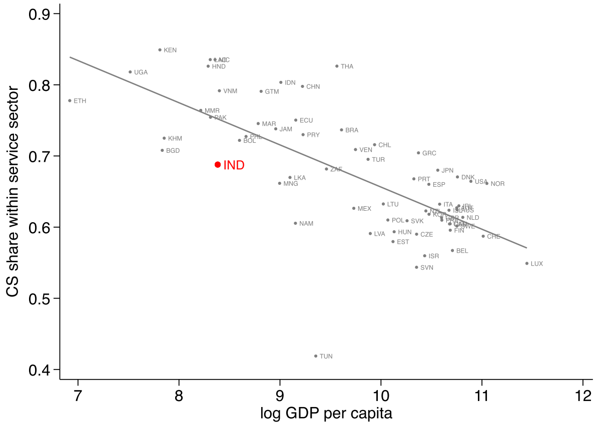

FIGURE B-1.—The Composition of Services and Economic Development. The figure shows a cross-country scatter plot. On the vertical axis, it plots the share of “Retail and Leisure and Health” (the first group of service industries in panel b of Figure 1) in total service employment excluding Education and PA in 2010. On the horizontal axis, it plots the GDP per capita. The data are from the International Labor Organization, which uses the ISIC classification.

B-2. *Data Sources*

In this section, we describe the five data sets we use in more detail.

B-2.1. *National Sample Survey*

The National Sample Survey (NSS) is a representative survey that has been conducted by the government of India to collect socioeconomic data at the household level since 1950. Each round of the survey consists of several schedules that cover different topics like consumer expenditure, employment and unemployment, participation in education, etc. We focus on the "consumer expenditure" module and the "employment and unemployment" module and use data from rounds 43, 55, 60, 64, 66, and 68 of NSS, which span the years 1987 to 2011. The survey covers all of India except for a few regions due to unfavorable field conditions.[2] For 1987 (2011), our data comprises about 126,000 (101,000) households and 650,000 (455,000) individuals.

We use the "employment and unemployment" module to measure sectoral employment shares and total earnings. An individual is defined as being employed if his/her usual principal activity is one of the following: (i) worked in household enterprises (self-employed), (ii) worked as a helper in household enterprises, (iii) worked as a regular salaried/wage employee, (iv) worked as casual wage labor in public works, and (v) worked as casual wage labor in other types of work. We describe the details of our sectoral employment classification in Section B-4 below.

We proxy income by total expenditure. More specifically, we measure total household expenditure and divide it by the number of household members older than 15 and under 65. We then attribute this average household expenditure to each household member as their labor earnings. We winsorize the expenditure data at 98th percentiles to reduce measurement error.

As we describe in more detail in Section B-2.5, the NSS provides two measures of expenditure. The so-called uniform reference period (URP) measure simply measures total expenditure as expenditure within the last 30 days. The mixed reference period (MRP) measure asks respondents for their total expenditure within the last year for a subset of durable goods to account for the lumpiness of purchases. As a measure of total spending, we thus prefer the MRP classification. For the year 2011, the MRP measure is directly contained in the employment module. For the year 1987, the employment module only contains the URP measure. To have a consistent measure in both years, we merge the 1987 expenditure module and the 1987 employment module at the household level and compute the MRP measure directly from the data on detailed spending categories. In practice, this choice is inconsequential because the URP measure and the MRP measure are highly correlated across space.

We estimate human capital using the information on educational attainment and Mincerian returns; see Section 4. In Table B-I, we report the resulting distribution of human capital across time, space, and sectors of production. In Table WA-III in the Web Appendix, we report the same composition when we classify PS and CS workers according to the NIC classification.

B-2.2. *Economic Census*

The India Economic Census (EC) is a complete count of all establishments, that is, production units engaged in the production or distribution of goods and services, not for the

[2]For example, the Ladakh and Kargil districts of Jammu and Kashmir, some interior villages of Nagaland, and villages in Andaman and Nicobar Islands are not covered in some rounds of the survey.

TABLE B-I

EDUCATIONAL ATTAINMENT.

	Less Than Primary	Primary, Upper Primary, and Middle	Secondary	More Than Secondary
		Aggregate Economy (1987–2011)		
1987	66.81%	22.01%	7.99%	3.19%
2011	40.33%	30.10%	18.79%	10.79%
		By Sector (2011)		
Agriculture	53.72%	29.23%	14.45%	2.60%
Manufacturing	32.63%	35.31%	20.68%	11.39%
CS	22.87%	30.44%	27.33%	19.36%
PS	20.75%	28.57%	28.08%	22.61%
		By Urbanization (2011)		
Rural	46.97%	29.89%	16.30%	6.84%
Urban	33.69%	30.30%	21.27%	14.73%

Note: The table shows the distribution of educational attainment over time (first panel), by sector of employment (second panel) and across space (third panel). The breakdown of rural and urban districts is chosen so that approximately half of the population live in rural districts and half live in urban districts.

purpose of sole consumption, located within the country. The censuses were conducted in the years 1977, 1980, 1990, 1998, 2005, 2013, and 2019. The micro-level data in 1990, 1998, 2005, and 2013 are publicly available.

The EC collects information such as firms' location, industry, ownership, employment, source of financing, and the owner's social group. It covers all economic sectors, excluding crop production and plantation. The EC in 2005 and 2013 exclude some public sectors like public administration, defense, and social security. In terms of geography, the EC covers all states and union territories of the country except for the year 1990, which covers all states except Jammu and Kashmir.

In Table B-II, we report some summary statistics of the EC in various years. In the most recent year, 2013, the EC has information on almost 60 million firms. The majority of them are very small: they employ, on average, around two employees, and 55% of them have a single employee. The share of firms with more than 100 employees is 0.06%.

TABLE B-II

THE ECONOMIC CENSUS: SUMMARY STATISTICS.

Year	Number of Firms	Total Employment	Employment Distribution			
			Avg.	1	Empl. < 5	> 100
1990	24,216,788	74,570,278	3.08	53.77%	91.24%	0.12%
1998	30,348,887	83,308,611	2.75	51.18%	91.71%	0.10%
2005	41,826,989	100,904,121	2.41	55.76%	93.17%	0.11%
2013	58,495,359	131,293,868	2.24	55.47%	93.44%	0.06%

Note: The table reports the number of firms, total employment, average employment, and the share of firms with one, less than five, and more than 100 employees.

B-2.3. *Service Sector in India: 2006–2007*

The Service Sector in India (2006–2007) data set is part of an integrated survey by the NSSO (National Sample Survey Organization) in its 63rd round. In the 57th round (2001–2002), the data set was called "Unorganized Service Sector." With the inclusion of the financial sector and large firms, the data set was renamed "Service Sector in India" and is designed to be representative of India's service sector. In Table B-III, we compare this Service Survey with the Economic Census for a variety of subsectors within the service sector. Table B-III shows that the service survey is consistent with the EC, that is, average firm size and the share of firms with less than five employees are quite comparable in most subsectors.

The Service Survey covers a broad range of service sectors, including hotels and restaurants (Section H of NIC 04); transport, storage and communication (I); financial intermediation (J); real estate, renting and business activities (K); education (M); health and social work (N); and other community, social and personal service activities (O). Excluded are the following subsectors: railways transportation; air transport; pipeline transport; monetary intermediation (central banks, commercial banks, etc.); trade unions; government and public sector enterprises; and firms that appeared in the Annual Survey of Industries frame (ASI 2004–2005). In terms of geography, the survey covers the whole of the Indian Union except for four districts and some remote villages.[3] The survey was conducted in a total number of 5573 villages and 7698 urban blocks. A total of 190,282 enterprises were ultimately surveyed.

For our analysis, we use two pieces of information: the number of employees and whether the main customer is another firm or a household.

B-2.4. *INAES 1999–2000*

The Informal Nonagricultural Enterprises Survey (INAES) is part of the 55th survey round of the NSSO. It covers all informal enterprises in the nonagricultural sector of the economy, excluding those engaged in mining, quarrying and electricity, gas and water supply.[4] The survey provides information on operational characteristics, expenses, value-added, fixed assets, loans, and factor income. For our analysis, we use two pieces of information: the number of employees and whether the main customer is another firm or a household. We use this data set to allocate employment in the construction sector to either consumer or producer services.

B-2.5. *Household Expenditure Survey*

The regressions in Table III are based on individual expenditure data from the National Sample Survey, Round 68, Schedule 1.0. The data set contains detailed information on a large set of spending categories. In Table B-IV, we report the categories we use in this paper.

[3]The survey covered the whole of India except: (i) Leh (Ladakh), Kargil, Punch, and the Rajauri districts of Jammu and Kashmir, (ii) interior villages situated beyond 5 km of a bus route in Nagaland, and (iii) villages of the Andaman and Nicobar Islands that remain inaccessible throughout the year.

[4]The organized sector comprises all factories registered under Sections 2(m)(i) and 2(m)(ii) of the Factories Act of 1948; 2(m)(i) includes manufacturing factories that employ 10 or more workers with electric power, and 2(m)(ii) includes manufacturing factories which 20 or more workers without electric power. The unorganized sector comprises all factories not covered in the organized sector. The informal sector is a subset of the unorganized sector. The unorganized sector includes four types of enterprises: (i) unincorporated proprietary enterprises, (ii) partnership enterprises, (iii) enterprises run by cooperative societies, trusts, private entities, and (iv) public limited companies. The informal sector only includes firms in categories (i) and (ii).

TABLE B-III

ECONOMIC CENSUS AND SERVICE SURVEY.

NIC2004	Sector	Number of Firms		Average Employment		Less Than 5 Employees	
		EC	Service Survey	EC	Service Survey	EC	Service Survey
55	Hotels and restaurants	1,491,809	30,744	2.53	2.49	90%	91%
60	Land transport; transport via pipelines	1,309,459	41,065	1.68	1.24	97%	99%
61	Water transport	7772	174	4.43	1.92	89%	98%
63	Transport activities; travel agencies	186,867	2101	3.43	3.33	86%	85%
64	Post and telecommunications	697,390	22,885	2.14	1.41	96%	99%
65–67	Financial intermediation	292,154	16,331	5.63	3.81	69%	82%
70	Real estate activities	69,538	3648	2.20	1.64	93%	96%
71	Renting of machinery and household goods	361,633	5387	2.02	1.77	94%	97%
72	Computer and related activities	66,122	1060	6.04	13.45	83%	86%
73	Research and development	2088	5	16.73	4.58	66%	89%
74	Other business activities	515,669	10,610	2.83	1.92	90%	95%
85	Health and social work	780,731	11,930	3.41	1.99	88%	95%
91	Activities of membership organizations	984,328	2837	1.86	1.32	94%	98%
92	Recreational, cultural, and sporting activities	219,823	2698	2.98	2.91	85%	82%
93	Other service activities	1,413,359	26,132	1.75	1.54	97%	99%

Note: The table reports statistics about the number of firms and their employment from the Economic Census 2005 and Service Survey 2006.

TABLE B-IV

BROAD CLASSIFICATION OF NSS EXPENDITURE SURVEY.

No.	Description	No.	Description	No.	Description
1	Cereals	13	Served processed food	25	Conveyance
2	Cereal substitute	14	Packaged processed food	26	Rent
3	Pulses and products	15	Pan	27	Consumer taxes
4	Milk and milk products	16	Tobacco	**28**	**Subtotal (1–27)**
5	Salt and sugar	17	Intoxicants	29	Clothing
6	Edible oil	18	Fuel and light	30	Bedding
7	Egg, fish, and meat	19	Medical (noninstitutional)	31	Footwear
8	Vegetables	20	Entertainment	32	Education
9	Fruits (fresh)	21	Minor durable-type goods	33	Medical (institutional)
10	Fruits (dry)	22	Toilet articles	34	Durable goods
11	Spices	23	Other household consumables	**35**	**Subtotal (29–34)**
12	Beverages	24	Consumer services excl. conveyance		

Note: The table reports the classification of broad expenditure items in the Expenditure Survey.

We classify categories 1–17 as food. We also use the spending categories 20 and 24 on services in the pooled regressions of columns 9 and 10 in Table III. In Section WA-5.2 in the Web Appendix, we report a more detailed breakdown of consumer services across subcategories.

Spending on category c is measured as spending within a particular reference period. For all categories, individuals report total spending during the last 30 days. For durable goods as well as medical and educational spending (i.e., categories 29–34), the subjects additionally report total spending in the last year. This second concept of expenditure aims to account for the lumpiness of purchases. Therefore, for this group, we take $1/12$ of annual spending as our measure of monthly expenditure. We measure total spending as the sum of all spending across all categories to calculate the spending share on food and consumer services. In Section WA-5.2 in the Web Appendix, we report a set of descriptive statistics on the cross-sectional distribution of spending, food shares, and CS shares.

In the regressions of Table III, we control for additional household-level covariates. These include the total size of the household and the number of members aged 15–65. We also control for additional household demographics such as

- the type of the household, which for rural areas is one of (i) self-employed in agriculture, (ii) self-employed in nonagriculture, (iii) regular wage/salary earner, (iv) casual worker in agriculture, and (v) casual worker in nonagriculture, (vi) other and in urban areas one of (i) self-employed (ii) regular wage/salary earner, (iii) casual worker, (iv) other;
- the household's religion—Hinduism, Islam, Christianity, Sikhism, Jainism, Buddhism, Zoroastrianism, or other;
- the household's social group—scheduled tribe, scheduled case, backward class, and other;
- whether the household is eligible to receive a rationing card.

B-3. *Geography: Harmonizing Regional Borders*

In this section, we describe the procedure we use to harmonize the geographical boundaries to construct a consistent panel of districts. The borders of numerous Indian districts have changed between 1987 and 2011. The left panel of Figure B-2 plots the districts'

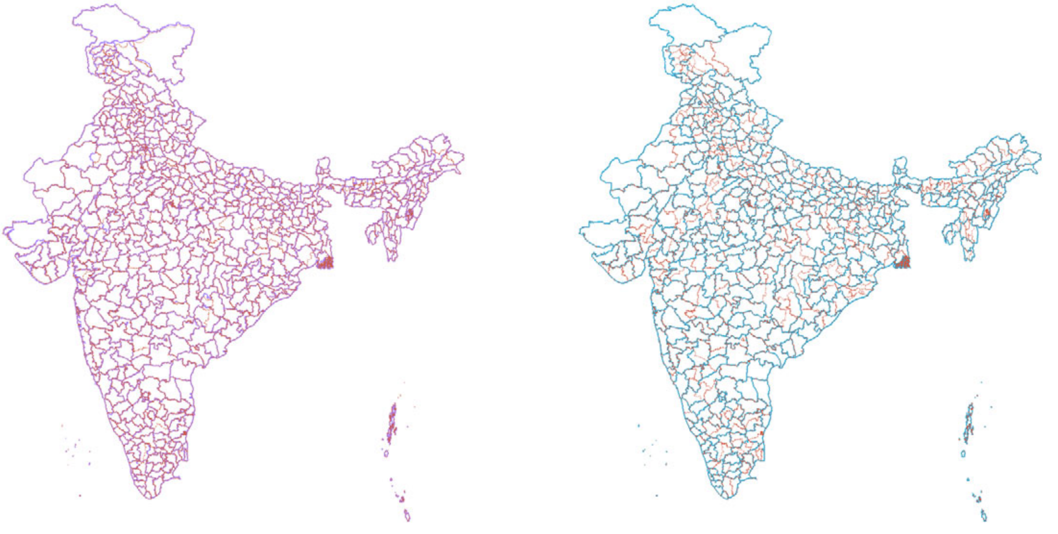

FIGURE B-2.—District Borders in India 1987–2011. The left figure plots the districts' boundaries in 2001 and 2011. The purple line represents the boundaries in 2001 and the dashed red line represents the boundaries in 2011. The right figure shows the official Indian districts in 2011 (dashed red lines) and the time-invariant geographical units we construct (solid blue lines) upon which our analysis is based.

boundaries in 2001 and 2011. The purple line represents the boundaries in 2001, and the red line represents the boundaries in 2011.

The most common type of redistricting is a *partition* in which one district has been separated into several districts in subsequent years. The second type is a *border move* in which the shared border between two districts has been changed. The third is a *merge* in which two districts were merged into a single district.

To attain a consistent geography, we take a region to be the smallest area that covers a single district or a set of districts with consistent borders over time. In the case of a partition, the region is constructed as the district in the pre-partition year. In the case of a border move, we construct the union of two districts. The right panel of Figure B-2 shows the official Indian districts in 2011 (dashed red lines) and our modified districts (solid blue lines). We exclude from the analysis two small districts that existed in 2011 but not in 1987. We also exclude districts with less than 50 observations because the small sample would yield imprecise estimates of the sectoral employment shares.

B-4. *Classification of Industries*

We distinguish four sectors: agriculture, manufacturing, consumer services, and producer services. To map these categories to the data, we first construct in Section B-4.1 six broad industries. Then, in Section B-4.2, we assign employment in services and construction to CS and PS, respectively.

B-4.1. *Broad Industry Classification*

We classify economic activities into six industries: (i) Agriculture, (ii) Manufacturing, (iii) Construction and Utilities, (iv) Services, (v) Information and Communications Technology (ICT), and (vi) Public Administration and Education. The classification relies on the official National Industrial Classification (NIC). Because the NIC system changes over time, we construct a concordance table between 2-digit industries of different versions of the NIC based on official documents and detailed sector descriptions. This concordance system allows us to compare sectoral employment patterns over time. We report the classification in Tables WA-VIII and WA-X in the Web Appendix.

B-4.2. *Attributing Employment to CS and PS*

We separate CS and PS using the Service Survey (see Section B-2.3), which reports the identity of the main *buyer* of a given firm. We refer to firms that mainly sell to other firms as PS firms and firms that mainly sell to consumers as CS firms.

Ideally, we would calculate the employment share of PS firms in each subsector of the service sectors and in each region. Unfortunately, the sample size of the Service Survey is not sufficiently large to estimate these averages precisely. Therefore, we generate the regional variation in employment shares by using regional variation in the firm-size distribution and differences in the employment share of PS firms by firm size. Empirically, within each subsector, large firms are much more likely to sell to firms, rather than consumers. In Figure WA-5 in the Web Appendix, we plot the employment share of PS firms as a function of firm size in the data. We show in Table WA-XI in the Web Appendix that the same pattern is present within 2- and 3-digit industries. We operationalize our procedure as follows:

1. We first aggregate the different 2-digit subsectors within services into seven broader categories, that we also refer to as *industries*: (i) retail and wholesale trade, (ii) hospitality, (iii) transport and storage, (iv) finance, (v) business services (including ICT), (vi) health, and (vii) community services. The mapping between the official NIC classification and these seven industries is reported in Table WA-IX in the Web Appendix.
2. For each industry k within the service sector and size bin b, we calculate the employment share of PS firms as

$$\omega_{kb}^{\mathrm{PS}} = \frac{\displaystyle\sum_{f\in(k,b)} 1\{f \in \mathrm{PS}\} l_f}{\displaystyle\sum_{f\in(k,b)} l_f}.$$

   Here, f denotes a firm, $1\{f \in \mathrm{PS}\}$ is an indicator that takes the value 1 if firm f is a PS firm, and l_f denotes firm employment. In practice, we take three size bins, namely "1 or 2 employees," "3–20 employees," and "more than 20" employees. We weigh observations with the sampling weights provided in the Service Survey.[5]
3. We then use the Economic Census (see Section B-2.2) and calculate the share of employment of firms in size bin b in industry k in region r as $\ell_{kbr} = \frac{\sum_{f\in(k,b,r)} l_f}{\sum_{f\in(k,r)} l_f}$.
4. We then combine these two objects to calculate the share of employment of PS firms in region r in industry k as $s_{rk}^{\mathrm{PS}} = \sum_b \ell_{kbr}\omega_{kb}^{\mathrm{PS}}$.

[5] In some industries, there are not enough firms with more than 20 employees to estimate $\omega_{kb}^{\mathrm{PS}}$ precisely. If there are fewer than five firms and $\omega_{kb}^{\mathrm{PS}}$ is smaller than $\omega_{kb}^{\mathrm{PS}}$ in the preceding size bin (i.e., $\omega_{k3}^{\mathrm{PS}} < \omega_{k2}^{\mathrm{PS}}$), we set $\omega_{k3}^{\mathrm{PS}} = \omega_{k2}^{\mathrm{PS}}$. Hence, for cells with few firms, we impose the share of PS firms is monotonic in firm size.

5. Finally, we use s_{rk}^{PS} to calculate the share of employment in PS and CS in region r as

$$\varpi_r^{\mathrm{PS}} = \frac{\sum_k s_{rk}^{\mathrm{PS}} l_{rk}^{\mathrm{NSS}}}{\sum_k l_{rk}^{\mathrm{NSS}}} \quad \text{and} \quad \varpi_r^{\mathrm{CS}} = \frac{\sum_k (1 - s_{rk}^{\mathrm{PS}}) l_{rk}^{\mathrm{NSS}}}{\sum_k l_{rk}^{\mathrm{NSS}}},$$

where l_{rk}^{NSS} denotes total employment in industry k in region r as measured from the NSS.

Five industries are not covered by the Service Survey. For firms in publishing and air transport, we assign all employment to PS; for firms in retail trade (except motor vehicle and the repair of personal goods), we assign all employment to CS; and for firms in wholesale trade and firms engaged in the sale and repair of motor vehicles, we use the average PS share from the subsectors for which we have the required information. We use the information on $\omega_{kb}^{\mathrm{PS}}$ from Service Survey 2005–2006, and apply it to EC 1990 and EC 2013 to get the region-sector PS shares in 1990 and 2013, respectively. Finally, we apply region-sector PS shares in 1990 and 2013 to NSS 1987 and 2011, respectively.[6]

B-4.3. *Construction and Utilities*

We merge employment in construction and utilities with services. To separate CS from PS, we follow a similar strategy as for the service industries. We use the INAES 1999–2000 discussed in Section B-2.4.

From the description of the NIC, some subsectors are clearly for public purposes. We, therefore, classify 5-digit level industries within the construction sector into public and private and drop all subsectors that we classify as public. These account for roughly 9.1% of total construction employment. See Table WA-XII in Section WA-5.2 in the Web Appendix for a detailed classification.

For all subsectors attributed to the private sector, we estimate the CS and PS share based on the information in the INAES. The survey has information on firms in the construction sector and reports the identity of the main buyer of the firm. In particular, we observe in the data whether the firm sells to: (i) the government, (ii) a cooperative or marketing society, (iii) a private enterprise, (iv) a contractor or intermediary, (v) a private individual, or (vi) others. We associate all firms that answer (ii), (iii), or (iv) with PS firms and all firms that answer (v) with CS firms. We then calculate the PS share of a given private subsector as total PS employment relative to total CS and PS employment in the respective subsector, that is, for subsector k we calculate the PS share as $\omega_k^{\mathrm{PS}} = \frac{\sum_{f \in k} 1\{f \in \mathrm{PS}\} l_f}{\sum_{f \in k} 1\{f \in \mathrm{PS},\mathrm{CS}\} l_f}$, where l_f denotes firm employment, and $1\{f \in \mathrm{PS}\}$ is an indicator for whether firm f is a PS firm.

In Table B-V, we report the relative employment shares of public employment (as classified in Table WA-XII in the Web Appendix), CS, and PS in the construction sector as a whole. The share of public employment is around 10%. Among the private subsectors, 12.9% of employment is associated with the provision of producer services. To calculate total employment in PS and CS industries within the private sectors of the construction sector for each year, we apply the 5-digit PS shares ω_k^{PS} to the NSS employment data and

[6]For 14 missing regional PS shares in 1987, we use the corresponding regional PS shares in 1999.

TABLE B-V

COMPOSITION OF THE CONSTRUCTION SECTOR.

	1999	2004	2007	2009
Public employment	0.073	0.102	0.073	0.136
CS employment	0.806	0.781	0.809	0.755
PS employment	0.121	0.116	0.118	0.109
PS/(PS + CS)	0.131	0.130	0.127	0.126

Note: The table shows the relative employment shares of PS, CS, and public employment in the construction sector in different years. We associate public employment to sectors classified as "public" in Table WA-XII in the Web Appendix. The main text explains the classification of employment in the private subsectors to CS and PS. The last row reports the relative employment share of PS within the private subsectors.

calculate shares within private sectors as

$$\varpi_t^{\mathrm{PS}} = \frac{\sum_k \omega_k^{\mathrm{PS}} l_{tk}^{\mathrm{NSS}}}{\sum_k l_{tk}^{\mathrm{NSS}}} \quad \text{and} \quad \varpi_t^{\mathrm{CS}} = \frac{\sum_k (1-\omega_k^{\mathrm{PS}}) l_{tk}^{\mathrm{NSS}}}{\sum_k l_{tk}^{\mathrm{NSS}}}$$

In summary, we attribute 9.1% of employment in construction and utilities to the public sector. For the rest of the construction and utilities, we allocate 12.9% of workers to PS.

B-5. *Trade Costs*

To calibrate the matrix of trade costs, τ_{rj}, we leverage the findings of Alder (2023), who estimates bilateral transport times between all Indian districts using the Dijkstra algorithm. He computes the fastest route between the centroids of each pair of Indian districts exploiting the existing transportation network together with estimates of travel times by different transport modes. Then he maps travel times to iceberg costs. In particular, he assumes that the iceberg trade costs between districts r and j is determined by the following equation:

$$\tau_{rj} = 1 + \alpha T_{rj}^{0.8}, \tag{B-1}$$

where T_{rj} denote the estimated travel time between r and j, and α is a scaling parameter. This specification captures the idea that trade costs increase less than proportionally with travel times, reflecting economies of scale in transportation. We calibrate $\alpha = 0.04$ to match the average trade costs across Indian states estimated by Van Leemput (2021).[7]

B-6. *Urbanization and Spatial Structural Change*

In Figure B-3, we show the structural transformation in India across time and space. We focus on urbanization as our measure of spatial heterogeneity.[8] This is a mere de-

[7]We compute the average state-level trade cost by aggregating (B-1) using the district population as weights. Alder (2023) calibrates α to match a median trade cost of 1.25, based on earlier studies. The results we obtain from either calibration are indistinguishable for our purposes; see Section WA-4 in the Web Appendix for details.

[8]The urbanization rate is the share of the population living in urban areas according to the definition of the NSS. The NSS defines an urban location in the following way: (i) all locations with a municipality, corporation,

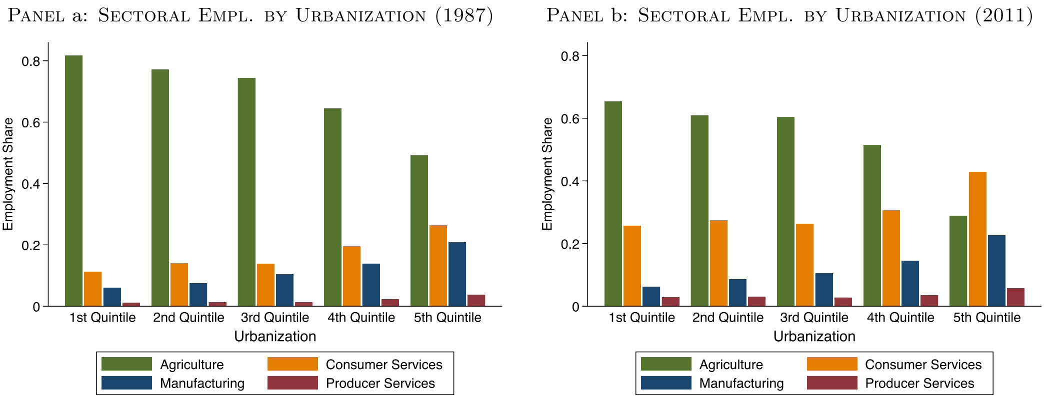

FIGURE B-3.—Sectoral Employment over Time and Space. The figure plots the sectoral employment shares by urbanization quintile in 1987 and 2011.

scriptive device because there is a strong positive correlation between urbanization and expenditure per capita in the NSS data in 2011. Figure B-3 displays sectoral employment shares by urbanization quintiles. The average urbanization rates of the five quintiles are respectively 6.4%, 12.1%, 19.5%, 29.2%, and 56.4%. Richer urban districts have lower employment shares in agriculture and specialize in the production of services and industrial goods. Over time, the share of agriculture declines. Between 1987 and 2011, the structural transformation was especially fast in more urbanized districts. In 1987, agriculture was the main sector of activity, even in the top quintile of urbanization. By contrast, in 2011, more than half of the working population was employed in CS and PS. This difference is larger when one considers earnings instead of employment because earnings are higher in service industries and in cities.

APPENDIX C: ESTIMATION

In this section, we discuss the details of the estimation.

C-1. *Estimating the Engel Elasticity ε*

C-1.1. *Nonlinear Estimation*

In Section 5.1, we estimate the Engel elasticity ε under the assumption that the asymptotic expenditure on food is small. This allowed us to estimate ε from log-linear regression of food shares and total expenditure. In this section, we estimate the ε without this assumption and focus directly on the nonlinear expression for food expenditure shares given in equation (13).

Equation (13) implies that the log food share satisfies the equation

$$\ln\big(\vartheta^{\mathrm{FE}}_{\mathcal{F}}(e,\mathbf{p}_\mathbf{r})-\beta_{\mathcal{F}}\big)=\ln\left(\kappa_{\mathcal{F}}\exp\left(\int_n \beta_n \ln p_{rn}\,dn\right)^{-\varepsilon}\right)-\varepsilon\ln e.$$

---

or cantonment and locations defined as a town area, (ii) all other locations that satisfy the following criteria: (a) a minimum population of 5000, (b) at least 75% of the male population is employed outside of agriculture, and (c) a density of population of at least 1000 per square mile.

TABLE C-I

INCOME ELASTICITY FOR FOOD: NONLINEAR ESTIMATION.

	Dependent Variable: ln(food expenditure share $-\beta_{\mathcal{F}}$)						
$\beta_{\mathcal{F}}$	0	0.01	0.02	0.03	0.04	0.05	0.06
	Panel A: OLS estimates						
$\ln e$	−0.319	−0.327	−0.336	−0.345	−0.355	−0.366	−0.378
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
N	91,474	91,474	91,474	91,474	91,474	91,474	91,474
R^2	0.4283	0.4278	0.4273	0.4266	0.4258	0.4247	0.4233
	Panel B: IV estimates						
$\ln e$	−0.395	−0.405	−0.416	−0.427	−0.439	−0.452	−0.466
	(0.013)	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)
N	85,916	85,916	85,916	85,916	85,916	85,916	85,916
R^2	0.3099	0.3097	0.3095	0.3093	0.3089	0.3084	0.3076

Note: The table shows the estimated coefficient ε of the regression (C-1) for different choices of $\beta_{\mathcal{F}}$. All variables are defined as in Table III. For all regressions, we trim the top and bottom 5% of the income distribution, and we control for region fixed effects, a (within-district) urban/rural dummy, a set of fixed effects for household size, and the number of workers within the household. In panel A, we report the OLS estimates. In panel B, we report the IV estimates. Standard errors are clustered at the district level. In all specifications, we consider a balanced sample excluding individuals whose food expenditure is below 6%. The results in the unbalanced sample including all individuals are almost identical.

We can thus consider the empirical regression

$$\ln\big(\vartheta_{\mathcal{F}}^{h} - \beta_{\mathcal{F}}\big) = \delta_r + \varepsilon \times \ln e_h + x_h'\psi + u_{rh}, \tag{C-1}$$

where $\vartheta_{\mathcal{F}}^{h}$ denotes the food share of household h living in region r, e_h denotes total household spending, δ_r is a region fixed effect, and x_h is a set of household characteristics. We now use (C-1) to estimate both $\beta_{\mathcal{F}}$ and ε without restricting $\beta_{\mathcal{F}} = 0$. We stress that we do not use the estimate of $\beta_{\mathcal{F}}$ in our analysis. $\beta_{\mathcal{F}}$ is the *final good* expenditure share on food, which is part of the final consumption vector, while our structural estimation relies on preference parameters of the value-added demand system. Hence, the value of $\beta_{\mathcal{F}}$ only matters insofar as it affects the estimate of ε. Also, focusing on the transformed dependent variable $\ln(\vartheta_{\mathcal{F}}^{h} - \beta_{\mathcal{F}})$ is computationally convenient because we can estimate (C-1) as a linear regression. This makes it easy to control for the regional fixed effects δ_r.

In Table C-I, we report the results. We focus on the specification with household controls of column 2 (for the OLS) and column 6 (for the IV) of Table III in the main text. The table shows the estimates of ε and the associated R^2 for different choices of $\beta_{\mathcal{F}}$. In panel A, we report the OLS estimates; in panel B, we report the IV estimates. The first column is the case of $\beta_{\mathcal{F}} = 0$, which is our baseline estimate.

Two results emerge. First, the estimate of ε is not sensitive to $\beta_{\mathcal{F}}$ in a range where the asymptotic expenditure on food items does not exceed 6% (the expenditure share on food items in the US is 5%). Second, a comparison of the R^2 shows that the specification with $\beta_{\mathcal{F}} = 0$ delivers the best fit to the data, even though the difference across columns is small.

C-1.2. *Consumer Service Expenditure Regression*

In columns 9 and 10 of Table III, we pool data on food shares and data on service expenditure shares. To measure service expenditures, we follow the official classification of the

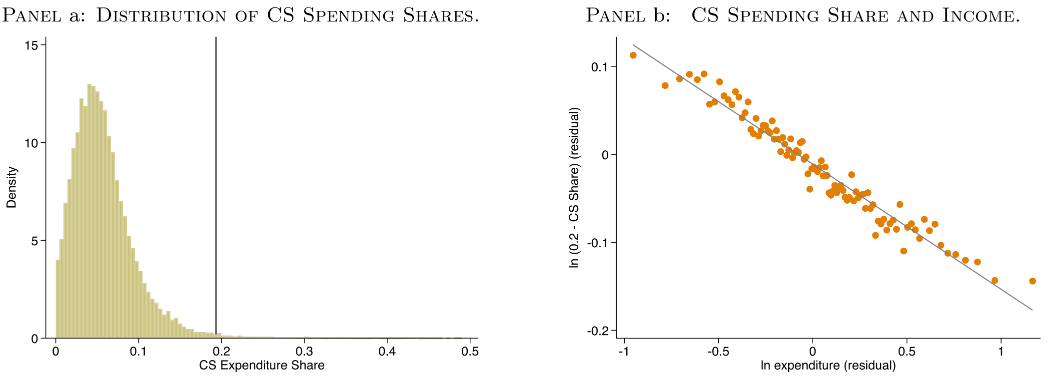

FIGURE C-1.—Consumer Service Spending. In the left panel, we display the cross-sectional distribution of spending share on services. In the right panel, we display a binscatter plot of the relationship between (the log of) total expenditure and (the log of) the differences between the actual expenditure share on consumer services and the asymptotic expenditure share 0.2, that is, $\ln(0.2 - \vartheta^h_{rCSt})$.

NSS expenditure module. As seen in Tables WA-IV and WA-V in the Web Appendix), these expenditures include, for example, domestic servants, barber shops, or tailor services. We also add entertainment expenses such as movie theaters or club fees.

In the left panel of Figure C-1, we plot the cross-sectional distribution of service expenditure shares in our data. The figure shows that the variation is sizable, and most consumers in India spend between 0 and 15% of their income on consumer services. The 99% quantile of the distribution, shown as the solid line, is 0.2.

It is useful to recall that, since CS spending is a luxury, our theory implies that $\kappa_S < 0$ and that the asymptotic expenditure share β_S exceeds the observed spending share ϑ^h_{Srt} for all households. Equation (13) thus implies that

$$\ln\big(\beta_S - \vartheta^{\mathrm{FE}}_S(e, \mathbf{p}_r)\big) = \ln \kappa_S + \varepsilon \ln\left(\exp\left(\int_n \beta_n \ln p_{rn}\, dn\right)\right) - \varepsilon \ln e. \tag{C-2}$$

Hence, the relationship between $\vartheta^{\mathrm{FE}}_S(e, \mathbf{p}_r)$ and total expenditure e is positive; the relationship between $\ln(\beta_S - \vartheta^{\mathrm{FE}}_S(e, \mathbf{p}_r))$ and $\ln e$ is negative, and in fact log-linear with a slope coefficient of ε.

To identify ε from a regression based on (C-2), we need to estimate β_S. Because β_S is the asymptotic expenditure share, we take it to be the 99% quantile of the expenditure share distribution in India, which turns out to be 0.2. This value is shown as the solid line in the left panel of Figure C-1. Given this value for β_S, we estimate ε from the same regression as in our baseline analysis contained in the main text, that is,

$$\ln\big(\beta_S - \vartheta^h_S\big) = \delta_r + \varepsilon \times \ln e_h + x'_h \psi + u_{rh}, \tag{C-3}$$

where the region fixed effect δ_r absorbs the constant κ_S and the vector of regional prices.

Table C-II reports the results. The first two columns contain different specifications of estimating (C-3) via OLS. The implied elasticity is negative but smaller than what we estimate for the specification based on food expenditure. In the last two columns, we report the IV specification, where—as in the baseline—we instrument total expenditure e with full set occupation fixed effects. Doing so increases the elasticity substantially, and we now estimate a value of around 0.3, which is still slightly lower but in the same ballpark as the IV estimate based on food expenditure.

TABLE C-II
INCOME ELASTICITY FOR CONSUMER SERVICES

	Dep. Variable: ln(0.2 − CS Exp Share)			
	(1)	(2)	(3)	(4)
$\ln e$	−0.115	−0.097	−0.263	−0.328
	(0.010)	(0.010)	(0.023)	(0.039)
Trim (top and bottom 5%)	✓	✓	✓	✓
Addtl. Controls		✓		✓
IV			✓	✓
N	90,672	90,625	85,312	85,269
R^2	0.132	0.138	0.027	0.003

Note: Standard errors, clustered at the district level, in parentheses. All variables are defined as in Table III. For all regressions, we trim the top and bottom 5% of the income distribution, and we control for region fixed effects. In columns (2) and (4), we also control for a (within-district) urban/rural dummy, a set of fixed effects for household size, and the number of workers within the household. In columns (3) and (4), we instrument household expenditure with occupational dummies as in Table III.

Finally, in the right panel of Figure C-1, we graphically display the relationship between (the log of) household expenditure and the adjusted expenditure share. While the relationship shows more noise relative to the specification based on the food expenditure shown in Figure 3, it is again approximately linear.

C-2. *Estimating the Shape of the Human Capital Distribution (ζ)*

We estimate the tail parameter of the distribution of efficiency units, ζ, from the distribution of income. Our model implies that total income and expenditure of individual h is given by $e^h_{rt} = q^h w_{rt}$, where q follows a Pareto distribution $f_{rt}(q) = \zeta \underline{q}^{\zeta}_{rt} q^{-(\zeta+1)}$. This implies that

$$\ln\bigl(f_{rt}(q)\bigr) = \ln\bigl(\zeta \underline{q}^{\zeta}_{rt}\bigr) - (\zeta + 1)\ln(q). \tag{C-4}$$

We estimate ζ from a regression of the (log of the) upper tail density on log efficiency units that we calculate as $q^h_{rt} = \frac{e^h_{rt}}{w_{rt}}$. In Table C-III, we report the estimated ζ based on (C-4). We report both the estimate based on the full sample (column 1) and the estimates by urbanization quintile (columns 2–6). We also report our estimates based on two measures of earnings: total expenditures per capita (as in our main analysis) and total income, which is also reported in the NSS data.

The estimated tail parameter for the aggregate economy is slightly below three, is stable across years, and does not depend on the exact measure of earnings. Moreover, it is declining in urbanization rate, indicating that urban locations have higher inequality. Our estimates also indicate that inequality was lower in 2011 than in 1987. For our quantitative model, we set ζ to an average value of three. In Section 7, we show that our results are robust to a variety of choices for ζ. For simplicity, we abstract from the heterogeneity in ζ across urbanization quantiles.

TABLE C-III
IDENTIFICATION OF ζ.

	Variable	Full Sample	Quartiles of Urbanization 1st	2nd	3rd	4th	5th
1987	Income	2.82	3.11	3.06	3.25	2.93	2.92
	Expenditure	2.84	3.64	3.57	3.21	3.03	2.79
2011	Income	2.85	4.04	3.47	3.13	2.90	2.71
	Expenditure	2.90	3.80	3.57	3.16	2.96	2.63

Note: The table reports the estimate of ζ based on (C-4). In the first columns, we report the estimates for the years 1987 and 2011. In the remaining columns, we perform our estimation separately for different quantiles of the urbanization distribution.

C-3. *The Relative Price of Agricultural Goods*

Our estimation uses the relative price of agricultural goods (relative to manufacturing goods) to identify the relative productivity in the agricultural sector (relative to manufacturing). The Ministry of Planning and Program Implementation (MOSPI) of the Government of India reports value-added by 2-digit sectors at current prices and constant prices from 1950–2013.[9] We then construct the sectoral price index as the ratio between sectoral value-added in current prices relative to constant prices. We normalize both price indexes in the year 2005 to unity. We then calculate the relative price of agricultural products as $p_t^{\mathrm{AM}} = p_t^A / p_t^M$. To check the validity of our results, we also use two additional data sources to calculate this relative price. The first is the GGDC 10-Sector Database,[10] which provides long-run data on sectoral productivity performance in Africa, Asia, and Latin America. This data set reports the annual series of value-added at current national prices and value-added at constant 2005 national prices. We follow the same procedures to calculate the relative price.

The second is the Wholesale Price Index (WPI) from the Office of the Economic Advisor.[11] The WPI tracks ex-factory prices for manufactured products and market prices for agricultural commodities.[12] Again, we use the same method to calculate the relative prices, and normalize the relative price in the year 2005 to 1.

In Figure C-2, we plot the relative price of agricultural goods to manufacturing goods. Since the pattern from the different data sources is very similar and 2005 is the reference year in the data, we combine ETD (2005–2011) and GGDC (1987–2005) to get a relative value-added price change of 1.52.

C-4. *Estimates of CS Productivity Growth*

In Section 5.2, we showed: (i) CS productivity is systematically higher in urbanized locations (see Figure 4), and (ii) productivity growth is spatially dispersed (see Table V). In

[9]Data are available at http://www.mospi.gov.in/data. See "Summary of macroeconomic aggregates at current prices, 1950–51 to 2013–14" and "Summary of macro economic aggregates at constant (2004–05) prices, 1950–51 to 2013–14."

[10]The data are available at https://www.rug.nl/ggdc/productivity/10-sector.

[11]The data are available at https://eaindustry.nic.in/.

[12]One issue with this is that the base year (and the basket of goods) changes during different time periods. Two series are relevant to our research. The first one is the series with the base year 1993, which is available from 1994 through 2009. The second one is the series with the base year 2004, which is available from 2005 through 2016.

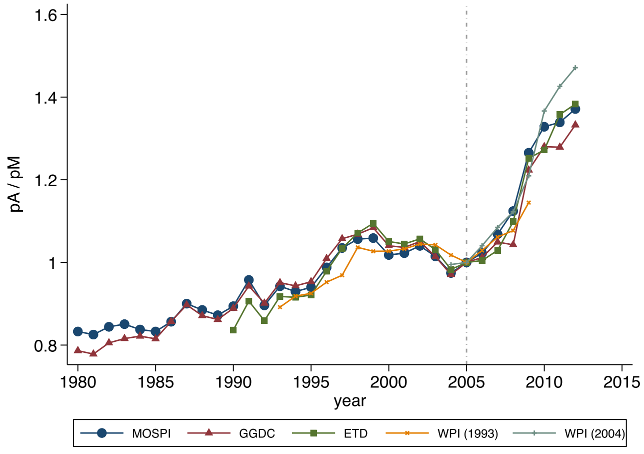

FIGURE C-2.—Relative price of agricultural to manufacturing goods. The figure shows the relative prices of agricultural products from the different sources mentioned in the main text. "MOSPI" refers to the data from the Indian Government that is used in our analysis. "GGDC" stems from the GGDC 10-Sector Database. "ETD" is the new revised version of the GGDC database. "WPI (1993)" and "WPI (2004)" are based on the Wholesale Price Index with a 1993 base year and a 2004 base year, respectively.

this section, we provide more details on the correlates of our estimates of CS productivity growth and how they depend on the demand system we use.

Consider first Table C-IV, where we regress sectoral productivity growth in region r, that is, $\ln A_{rs2011} - \ln A_{rs1987}$, on the 1987 urbanization rate in region r. Urban locations experienced higher productivity growth, especially in CS and the Industrial Sector (which, recall, includes some business services).[13]

In Figure C-3, we show the extent to which our productivity estimates depend on our estimated demand system. Specifically, we depict the distribution of CS productivity *growth*, $\frac{\ln A_{r\mathrm{CS}2011} - \ln A_{r\mathrm{CS}1987}}{2011-1987}$, as a function of the Engel elasticity ε. We consider five values of this elasticity that span the range of estimates based on our results in Table III: our baseline estimate (0.395, column 6), the estimate for high-income households (0.415, column 7), the estimate for urban locations (0.358, column 8), the OLS estimate (0.321, column 2), and the estimate based on food and service expenditure (0.23, column 9), which is the smallest estimate in our analysis. Figure C-3 shows that the estimated distribution of growth rates is quite stable. For the smallest ε of 0.23, the dispersion is slightly larger, reflecting the fact that local employment shares depend on $A_{r\mathrm{CS}t}^{\omega_{\mathrm{CS}}\varepsilon}$ (see (17)). Because the importance of service-led growth is decreasing in ε, we focus our robustness analysis on the range where $\varepsilon > 0.2$.

C-5. *Nontargeted Moments: Additional Results*

As we mention in the main text, we can use the data from the expenditure survey to validate our estimates of agricultural productivity, and hence food prices. The expenditure

[13]We also ran the regressions in Table C-IV based on the 2011 urbanization rate. The positive correlation between productivity growth and urbanization is, if anything, stronger.

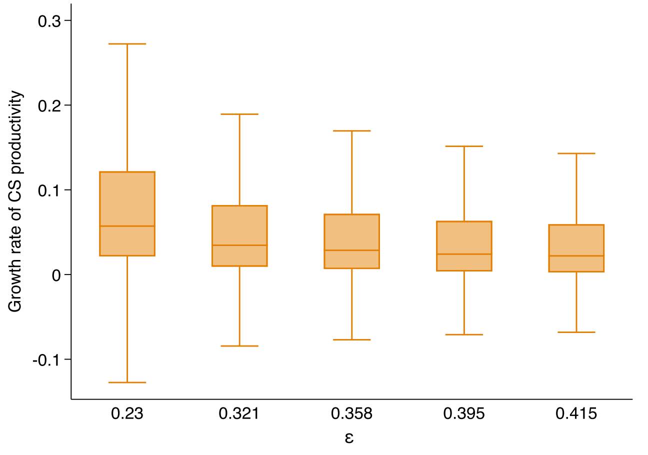

FIGURE C-3.—CS Productivity Growth and the Engel elasticity ε. The figure shows the cross-sectional distribution CS productivity growth rate, $\frac{\ln A_{rCS2011}-\ln A_{rCS1987}}{2011-1987}$, as a function of ε. We always display a boxplot that indicates the median, the interquartile range, and the upper and lower adjacent values.

survey reports both total expenditure and the total quantity bought for a variety of food items. We thus compute the price of product n in region r, p_{nr}, as the ratio between total expenditure and total quantity and then run the regression

$$\ln p_{nr} = \delta_r + \delta_n + u_{nr}, \tag{C-5}$$

where δ_r and δ_n are region and product fixed effects. The estimated fixed effect $\hat{\delta}_r$ thus describes the average food price in region r.

In Figure C-4, we show the correlation between the estimated $\hat{\delta}_r$ and the regional price of agricultural goods in the model, that is, $\ln p_{rFt}$. The two measures are positively correlated, even though we do not use the data on local food prices as targets of our estimation. In the model, the variation in local food prices reflects local agricultural productivity, local wages, and food prices of close-by locations (which have low transport costs).

TABLE C-IV

PRODUCTIVITY GROWTH AND URBANIZATION.

	Productivity Growth		
	Agriculture	Industry	Cons. Serv.
1987 urbanization	0.277	0.423	2.365
	(0.080)	(0.087)	(0.398)
Weight (1987 Pop)	✓	✓	✓
N	360	360	360
R^2	0.033	0.062	0.090

Note: The table reports the results of univariate regressions of sectoral productivity growth, $\ln(A_{rs2011}/A_{rs1987})$, on the urbanization rate in 1987. We weigh all regressions by the population size in 1987.

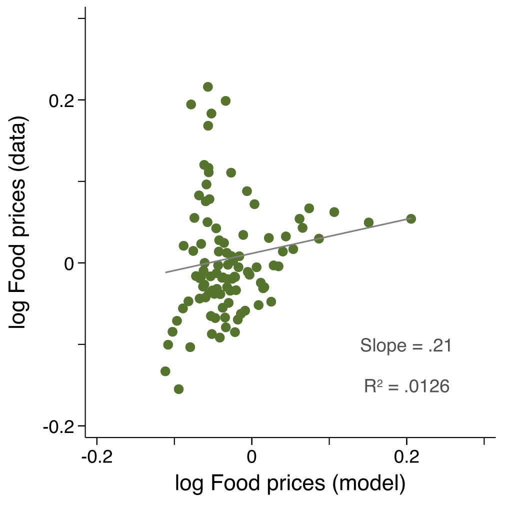

FIGURE C-4.—Food Prices: Model vs. Data. The figure shows a binscatter plot of regional log food prices in the data ($\hat{\delta}_r$ from (C-5)) and the model ($\ln p_{rF}$).

C-6. *Outliers in Quantitative Analysis*

In the quantitative analysis of Section 6, we winsorize a small number of outliers. For a small number of regions, we estimate very large changes in CS productivity. Because CS employment in our model is bounded by ω_{CS}, our theory can only rationalize employment shares close to ω_{CS} with an exceedingly high level of CS productivity.

In Table C-V, we report different quantiles of the regional distribution of welfare changes for the different counterfactuals. Consider, for example, the agricultural sector. If agricultural productivity had not grown since 1987, the most adversely affected region would have seen its welfare decline by 56% in terms of an equivalent variation. Conversely, some regions would have seen their welfare increase. The last row of Table C-V shows that some regions would have seen very large gains if CS productivity had not grown. These are regions where CS productivity *declined* between 1978 and 2011. As explained above, this pattern is entirely driven by a few districts being close to the theoretical threshold of ω_{CS}. For comparison, in the last row, we report the estimated distribution of the welfare effects in our baseline analysis, where we truncate the productivity growth

TABLE C-V

DISTRIBUTION OF WELFARE LOSSES.

	Regional Welfare Changes (%)									
	Min	1%	2%	3%	5%	95%	97%	98%	99%	Max
Agriculture	−56.0	−45.1	−43.3	−42.1	−39.6	3.8	7.7	13.7	17.8	48.0
Industry	−33.7	−28.7	−26.7	−25.8	−24.3	−5.8	−3.4	−2.3	−1.2	28.4
Cons. Serv.	−99.3	−97.1	−91.6	−87.3	−78.0	19.4	46.3	171.4	360.2	1814.2
Cons. Serv. (Baseline)	−94.4	−93.6	−88.8	−86.7	−77.7	19.3	37.5	42.2	73.5	95.5

Note: The table reports the lower and upper percentiles of the regional distributions of sectoral welfare losses.

TABLE C-VI

WELFARE LOSSES WITH DIFFERENT TRIMMING CUTOFFS.

	Trimming Cutoff					
	No Trimming	1%	2%	3%	4%	5%
Wefare Loss	−17.6%	−19.2%	−19.9%	−20.5%	−20.8%	−20.9%
Employment Share	0	0.5%	1.9%	3.2%	5.4%	8.0%

Note: The table reports the aggregate welfare effects of productivity growth in the CS sector for different trimming rules. A trimming cutoff $x\%$ means that we set the $x\%$ highest and lowest productivity growth rates to $1 - x\%$ and $x\%$, respectively.

distribution at the bottom and top 3%. This has large effects on the welfare effects in the right tail of the distribution.

These extreme values at the bottom of the regional productivity growth distribution have aggregate effects. For our baseline analysis, we trim the top and bottom 3% of the productivity growth distribution and set regional productivity growth in such regions to the 3% and 97% quantile, respectively. In Table C-VI, we report the change in aggregate welfare losses in the absence of CS productivity growth as a function of this trimming cutoff. Without any trimming, the aggregate effect is −17.6%. Once such outliers are truncated, we recover our baseline results of a welfare loss of about −20.5%. In the last row of Table C-VI, we report the aggregate employment share of the affected districts. The changes in the aggregate effects of CS growth are not driven by a few large districts but by a small number of small districts with very large changes in CS productivity.

C-7. *Details of Robustness Analysis (Section 7)*

In Figure C-5, we report the results of our analysis discussed in Section 7, where we allow for heterogeneity in the Engel elasticity ε. In the left panel of Figure C-5, we assume our baseline estimate of $\varepsilon = 0.395$ in Bangalore and $\varepsilon = 0.29$ in rural Bankura as suggested by column 8 of Table III. Doing so yields a mild reduction in spatial inequality, but the quantitative effect is small.

In the right panel, we allow for heterogeneous ε across the income ladder. In particular, we estimate productivity growth in CS based on the benchmark Engel elasticity of 0.395. Then we consider (a zero measure of) households with income above and below the median with elasticities of 0.415 and 0.218, respectively, corresponding to the estimates of column 7 in Table III. The right panel of Figure C-5 highlights that this *amplifies* the differential welfare impact of service-led growth between rich and poor households. The reason is intuitive: rich agents consume more and care more about the provision of CS. This suggests that a model with increasing Engel elasticities by income is likely to deliver even more unequal welfare effects of service-led growth.

In the main text, we focused on the robustness of our results with respect to the Engel elasticity. Here, we report our results for ω_F and ζ. We always recalibrate the entire model, when changing one of the parameters.

We summarize our results in Figure C-6, where we plot the implied impact of sectoral productivity growth as a function of the respective parameters. In the left panel, we report for completeness the effect of ε. As discussed in the main text, for the impact of service-led growth to become small, one would need to believe in an estimate of the Engel elasticity, which is much larger than suggested by both the micro data on Engel curves and the macro data on productivity growth.

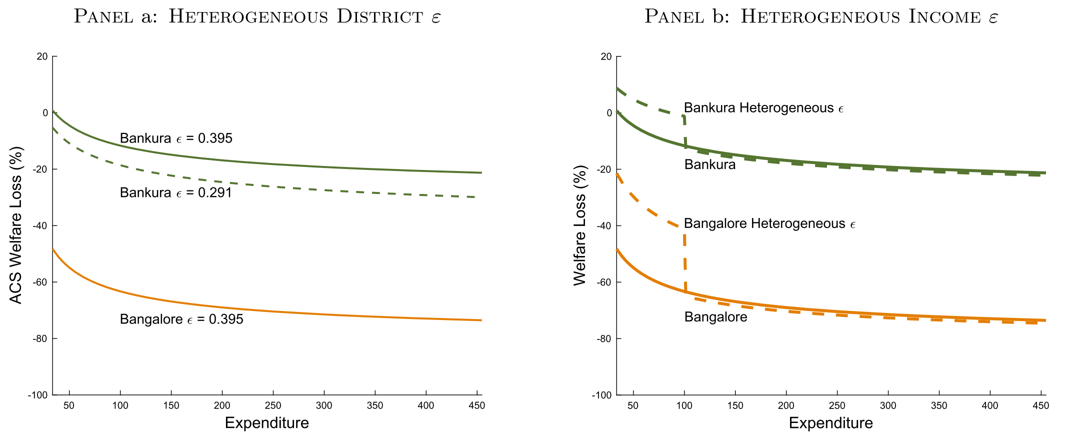

FIGURE C-5.—Heterogeneous Engel Elasticities. In the left panel, we allow for heterogeneous ε across locations. We assume that ε of individuals in Bangalore (Bankura) is 0.395 (0.291), which is in line with the results reported in Table III. In the right panel, we allow for different ε across individuals. In line with Table III, we assume that individuals above (below) the median income have ε of 0.415 (0.218).

In the middle panel, we focus on ω_F, which we calibrate to 1% so as to match the value-added share of the US farming sector in 2017. However, the value-added share of agriculture is larger than 1% in many industrial countries (e.g., 2% in Italy and France, 3% in Spain.) Therefore, we consider a range of larger ω_F. Panel (b) of Figure C-6 shows that the implied welfare impact of productivity growth in the CS sector is, if anything, slightly larger the higher ω_F. Our choice of $\omega_F = 0.01$ is therefore conservative.

Finally, in panel (c) of Figure C-6, we show the effect of the tail of the skill distribution, ζ. Note that this only changes the mapping from the "aggregate" demand parameter $\bar{\nu}_s$ to the micro parameter ν_s. All our productivity estimates are independent of ζ. Figure C-6 shows that the higher ζ, the higher the importance of CS growth relative to agricultural productivity. This reflects the importance of nonhomothetic demand. The smaller ζ, the higher income inequality. And because higher inequality increases aggregate demand for CS for a given average wage, less productivity growth is "required" to explain the increase in CS employment if ζ were small. Figure C-6 shows this intuition is borne out but that the effects are quantitatively moderate.

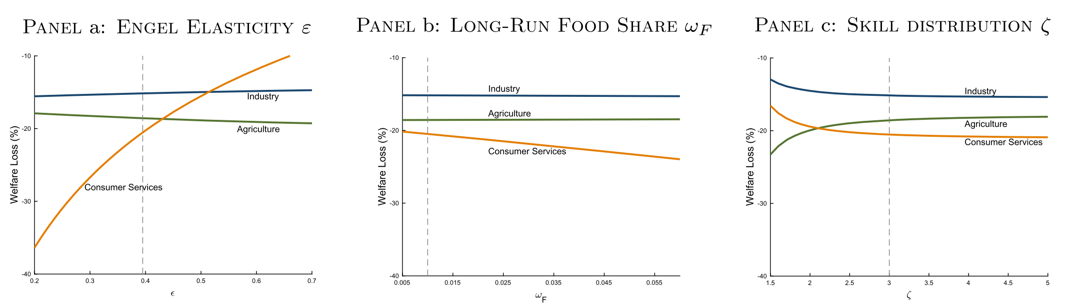

FIGURE C-6.—Robustness Analysis. Panels (a), (b), and (c) show the aggregate welfare effects as a function of the preference parameters ε, ω_F, and the tail parameter of the skill distribution ζ. The vertical dashed line corresponds to the parameter value in our benchmark analysis.

TABLE C-VII

THE IMPORTANCE OF SERVICE-LED GROWTH—ROBUSTNESS.

	Agriculture					Industry				
	Aggregate Effects	Urbanization Quintiles		Income Quantiles		Aggregate Effects	by Urbanization Quintiles		by Income Quantiles	
		1st	5th	10th	90th		1st	5th	10th	90th
Baseline	−18.6	−19.5	−15.1	−21.7	−14.9	−15.2	−11.6	−20.7	−12.3	−20.6
Alternative calibrations of ε (Section 7.2)										
$\varepsilon = 0.415$ (High Income Households)	−18.6	−19.6	−15.1	−21.9	−14.9	−15.1	−11.5	−20.7	−12.2	−20.6
$\varepsilon = 0.321$ (OLS estimator)	−18.3	−19.3	−14.9	−21.1	−15.0	−15.3	−11.8	−20.8	−12.6	−20.6
Alternative measurement choices (Section 7.2)										
Allocate PS share based on WIOD	−18.4	−19.3	−15.4	−21.3	−15.3	−16.9	−12.6	−23.6	−13.4	−23.5
Allocate ICT and Business to PS	−18.7	−19.7	−15.8	−21.5	−15.7	−16.2	−12.0	−22.9	−12.5	−22.8
Allocate Construction to Industry	−18.3	−20.8	−12.4	−22.5	−13.5	−19.1	−11.7	−30.4	−13.2	−29.5
Alternative modeling assumptions (Section 7.4)										
Open economy	−18.7	−19.5	−15.4	−21.7	−15.5	−17.7	−14.4	−22.8	−15.0	−22.5
Imperfect skill substitution	−22.8	−25.0	−17.5	−24.6	−18.9	−14.3	−10.3	−20.3	−9.8	−21.9
Spatial labor mobility	−18.1	−18.8	−15.0			−15.1	−11.8	−20.2		

We also analyzed the effect of the skill return ρ. Our estimate of 5.6% is on the lower end of typical Mincerian regressions. For this reason, we consider alternative calibrations in which the return to education is higher, up to an annual 10% that is an upper bound to the range of the typical estimates. Our results are essentially insensitive to this parameter. Similarly, our results are virtually unchanged for different values of the elasticity of substitution σ.

In Table C-VII, we report the analogue to Table IX, that is, the welfare effects of agricultural and industrial productivity growth. Table C-VII shows that our baseline results are not significantly affected by either the alternative modeling assumptions or the alternative measurement choices.

REFERENCES

ALDER, SIMON (2023): "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," SSRN Working Paper 2856050. [16]

CICCONE, ANTONIO, AND GIOVANNI PERI (2005): "Long-Run Substitutability Between More and Less Educated Workers: Evidence From U.S. States, 1950–1990," *Review of Economics and Statistics*, 87 (4), 652–663. [6]

GANCIA, GINO, ANDREAS MÜLLER, AND FABRIZIO ZILIBOTTI (2013): "Structural Development Accounting," in *Advances in Economics and Econometrics: Theory and Applications (10th World Congress of the Econometric Society)*, ed. by Daron Acemoglu, Manuel Arellano, and Eddie Dekel. Cambridge Univ. Press, 373–418. [6]

HENDRICKS, LUTZ, AND TODD SCHOELLMAN (2023): "Skilled Labor Productivity and Cross-Country Income Differences," *American Economic Journal: Macroeconomics*, 15 (1), 240–268. [5]

PORZIO, TOMMASO, FEDERICO ROSSI, AND GABRIELLA SANTANGELO (2022): "The Human Side of Structural Transformation," *American Economic Review*, 112 (8), 2774–2814. [5]

REDDING, STEPHEN J., AND ESTEBAN ROSSI-HANSBERG (2017): "Quantitative Spatial Economics," *Annual Review of Econ.*, 9 (1), 21–58. [6]

SIMONOVSKA, INA, AND MICHAEL E. WAUGH (2014): "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 92 (1), 34–50. [5]

VAN LEEMPUT, EVA (2021): "A Passage to India: Quantifying Internal and External Barriers to Trade," *Journal of International Economics*, 131, 103473. [16]

Co-editor Charles I. Jones handled this manuscript.

Manuscript received 8 July, 2022; final version accepted 28 February, 2023; available online 16 March, 2023.