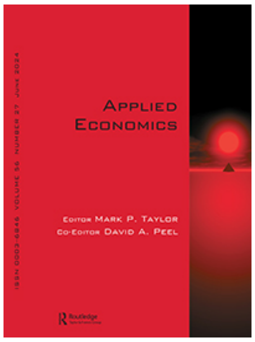

Applied Economics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/raec20

What can we learn from industry-level (aggregate) production functions?

Ben Filewod

APPLIED ECONOMICS
https://doi.org/10.1080/00036846.2024.2337780

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

What can we learn from industry-level (aggregate) production functions?

Ben Filewod

Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, London, UK

ABSTRACT
Recent work has revived two intertwined challenges to aggregate production functions (the 'identity' and 'aggregation' problems). This paper examines both problems in the context of aggregate industry-by-country analysis, first demonstrating the relevance of the identity problem for industry-level analysis and tracing its origin in the System of National Accounts. Using a case study of materials quality in global forestry and logging, the paper then compares estimates from fully physical versus conventional (monetary) production functions to isolate the aggregation problem and show that credible inference depends on appropriately modelling heterogeneity in production processes. Materials quality is measured via finite mixture modelling applied to global satellite data. Attempting to estimate the parameters of a common production technology yields poor results, because of differences in production processes between countries. The paper offers a practical approach for dealing with heterogeneity via Data Envelopment Analysis and heterogeneous coefficient panel estimators, and concludes with guidance to help applied industry-level analysis recognize and avoid both the identity and aggregation problems.

KEYWORDS
Aggregation; value-added identity; sector; input quality; production functions; data envelopment analysis

JEL CLASSIFICATION
C43; E23; L73; Q23

I. Introduction

Production functions represent relationships between inputs and output using a transformation function $Y = f(K, L, etc\ldots)$. Real production processes are heterogeneous and typically involve physical quantities, but aggregate production functions[1] estimated with value data are widely used for both research and policy formulation. Applied economists working with such functions must pay careful attention to problems surrounding data quality, model specification, and estimation strategy. Recently, two further problems (the 'aggregation' and 'identity' problems) have been revived and extensively developed (Felipe and Fisher 2003; Felipe and McCombie 2013, 2014).

The salience (if not the consequence) of these problems has been acknowledged (Heun et al. 2017; Temple 2006), but debate has focused solely on their implications for total-economy production functions (Felipe and McCombie 2010; Temple 2006, 2010). Yet the issues explicitly apply at a range of scales, and independent investigation of some of the claims made is lacking. This paper therefore investigates both problems empirically in an industry-level setting (i.e. production functions for which the unit of analysis is industry by country). Industry-level analysis is a mainstay of both academic and applied inquiry (e.g. Eberhardt, Helmers, and Strauss 2013; Harrigan 1999), but falls into a theoretical grey area between the literatures on microeconomic (firm-level) and macroeconomic (total-economy) production functions; applied work at the industry level tends to emphasize microeconometric concerns (e.g. Vandenberghe 2017). The main contribution of this paper is to pursue two fundamental problems from the macroeconomic literature in this policy-relevant setting, demonstrate their relevance, and suggest strategies for accommodating them.

The 'aggregation' problem arises because it is impossible to mathematically aggregate neoclassical production functions in a realistic way (Fisher 1969). This is a practical as well as a theoretical

CONTACT Ben Filewod b.filewod@lse.ac.uk Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, London, UK

[1]The term 'aggregate production function' generally refers to total-economy production functions, but may also used to indicate any production function estimated using value data. To avoid confusion, this paper refers to production functions as 'engineering' (i.e. physical process-level), 'firm-level', 'industry-level', or 'total-economy'.

problem: analysing the behavioural implications of an aggregate production function (e.g. substitution elasticities) may be nonsensical if there is no precise real-world counterpart, and glossing over heterogeneity in production processes may alter inference. Aggregation problems were widely demonstrated in the twentieth century (Felipe and Fisher 2003), but aggregate production functions continue to be seen as a useful approximation of reality (as per Solow's 'parable' argument (Solow 1957)) and are widely used for forecasting, measurement, and hypothesis testing. In the latter case, whether aggregation poses problems in a hypothesis-testing context should depend on whether an aggregate representation of the data-generating process can identify the phenomenon of interest.

While production functions may be estimated at a range of scales, ranging from the individual physical process to firms, industries, or nations, only the first of these avoids some degree of aggregation. Textbook presentations of production theory emphasize the trivial case in which all firms are profit-maximizing price-takers who operate efficiently in competitive markets with homogeneous inputs (e.g. Mas-Collel et al. 1995:5.E). An aggregate production function is then a scaled-up version of the individual production technology, which is the same for all firms. Imposing more realistic conditions (notably heterogeneous capital) prevents mathematical aggregation (Felipe and Fisher 2003). Fisher (1969) provided a conclusive analysis, showing that (assuming maximization of output given inputs, constant returns to scale (CRS), labour homogeneous and mobile, and capital heterogeneous and fixed) an aggregate production function can exist if and only if the underlying 'micro' production functions differ only by capital-augmenting technologies, and that no aggregation is possible without CRS. These conditions are implausibly strict: differences in average product per worker between firms in an industry, for example, would preclude the existence of an aggregate production function at the industry level (Felipe and Fisher 2003).

The 'identity' problem arises because using monetary values to aggregate over heterogeneous output and inputs may involve an accounting identity. The monetary value of output must be distributed somewhere: in the textbook two-factor case, value added ≡ payments to labour plus payments to capital. Applied work often seeks to maintain this identity for theoretical coherence (Heun et al. 2017), for example by adding the values of additional production inputs to the output measure, and it is a central assumption used to produce industry-level data within the System of National Accounts (SNA). However, estimating a production function from value data may be tautological because the usual neoclassical forms (Cobb-Douglas (Shaikh 1974) and Translog and CES (Felipe and McCombie 2013)) can be derived as transformations of the identity used to generate the data.

The clearest demonstration of the 'identity' problem in a panel context is due to Shaikh (1974). In response to Solow's (Solow 1957) proposal for measuring technical change, Shaikh demonstrated the equivalence:

$$VA_t \equiv W_t + \Pi_t \Leftrightarrow Q_t = B_t(cK_t^{\beta} L_t^{1-\beta}) \qquad (1)$$

Appendix A reproduces Shaikh's proof. VA_t represents output at time t, and W_t and Π_t give the cost shares of labour and capital, respectively. The left-hand side of eq 1 is an accounting identity, motivated by the assumption of competitive markets and CRS, which says output is fully dispersed to production factors (and W_t and Π_t are equivalently the value of labour and capital services). The right-hand side is indistinguishable from a Cobb-Douglas production function with a time-dependent technology term (B_t), but in the equivalence in eq 1 it is derived as the factor-share-weighted average of the growth rates of the wage and profit rates. Translog and CES production functions can be derived similarly by using different assumptions about the growth path of wages and profit (Felipe and McCombie 2013), meaning that the 'identity' problem implies that fitting a production function to value data is simply a matter of choosing a functional form that best approximates the growth path of wages and profits in the data to hand.

The implications of the equivalence in eq 1 have been throroughly developed elsewhere (e.g. Felipe and McCombie 2010, 2013; Shaikh 1974). Broadly, production functions fit to value data will not reveal new information about production processes. It should also always be possible to find a production function that gives results matching *a priori* expectations about what is correct, i.e. fitting the data extremely well, exhibiting CRS, and having marginal products that closely match factor cost shares (see Appendix A and following section). The 'aggregation' and 'identity' problems are therefore intertwined, since the latter ensures that estimates of aggregate production functions appear reasonable (which justifies dismissing concerns about aggregation). These arguments have led to very strong claims, including that total-economy production functions estimated with value data are spurious in general (Felipe and Fisher 2003; Felipe and McCombie 2006) and that measures of total factor productivity *à la* Solow are simply data artefacts (Shaikh 1974).

Although both the 'aggregation' and 'identity' problems have been thoroughly elaborated, they are rarely considered in applied work today. This paper takes an empirical approach to study both problems in a policy-relevant applied setting (industry-level production functions). Section II analyses the identity problem by revisiting a recent industry-level study ((Vandenberghe 2017), published in this journal), demonstrating the claims above and locating the source of the problem in the System of National Accounts. Sections III-V analyse the aggregation problem via a case study of materials quality in a natural resource industry (forestry and logging; ISIC Rev.4 A02). First, a non-parametric approach (Section IV) is used to evaluate heterogeneity and the probable importance of aggregation in this industry. Second, comparisons between production functions estimated with economic and physical data (Section V) are used to isolate the aggregation problem and show that heterogeneous coefficient estimators are a partial solution (contrasts are also made between estimates using only efficient sectors, and two alternate specifications for materials quality). Section VI concludes with guidance for industry-level applied work.

The main finding of the paper is that both 'aggregation' and 'identity' problems can be debilitating for industry-level analysis: in Section II the identity renders production function estimates spurious, while in Section V the treatment of aggregation alters inference. In response the paper suggests some simple practical strategies, namely careful attention to data, incorporation of non-parametric analysis, and flexible panel estimators. A secondary contribution is new information about the role of materials input quality in global forestry and logging (measured using the approach of Filewod and Kant, 2021), and roughly analogous to management regime, i.e. plantation, frontier, or semi-natural forest). New data on average products and efficiency are presented, and the production function estimates (when heterogeneous technology is modelled) shed some light about whether materials quality matters in this industry. This is a timely question, since policy debates in the forestry sector frequently turn on the trade-offs implicit in harvesting forests of different quality classes (e.g. Himes et al. 2022).

II. The identity problem at the industry level

To assess whether the identity problem matters for industry-level aggregate production functions, this section revisits a recent study (Vandenberghe 2017) of the productivity implications of changing labour and capital mixes in Europe (another, very similar, example is Ilmakunnas and Miyakoshi 2013). The data, obtained from the 2007 release of the EU-KLEMS database (O'Mahony and Timmer 2009; Timmer, O'Mahony, and van Ark 2007), are annual observations for 34 industries in 16 countries of real gross value added (GVA), capital services (CAP), hours worked (H_EMP), shares of total hours worked by educational level and age bracket, and the share of ICT capital in total capital compensation. The aim is to estimate an aggregate production function in which variation in input quality matters, following the method of Hellerstein and Neumark (Hellerstein and Neumark 2007). The aggregation problem is of interest, since estimating a single production function in effect assumes that different quality inputs affect output in the same way across diverse industries, as is the identity problem, since in EU-KLEMS capital services are defined[2] to exhaust value added (i.e. CAP ≡ GVA − LAB).

[2]Specifically (O'Mahony and Timmer 2009: F380): "The nominal rate of return is determined *ex post* as it is assumed that the total value of capital services for each industry equals capital compensation. . .derived as gross value added minus labour compensation".

If the identity problem holds, regressing GVA on CAP and LAB in logs should yield coefficient estimates (β_{CAP}, β_{LAB}) that are either identical or very close to factor shares (depending on whether factor shares of output are constant over time (Felipe and Holz 2001)), while regressing the raw variables should return perfectly collinear results. Quality adjusting inputs should not be possible, since there will be no residual variation to explain. Table 1 collects results from a series of regressions that assess these claims, using the same data as the original study.[3] The average (country by sector by year) factor shares are 0.319 for CAP and 0.681 for LAB.

Pooling observations, a regression of the raw data (Model 1) reveals the identity problem (i.e. perfect collinearity). Log-transforming variables (Model 2) allows estimation, with coefficient estimates extremely close to factor shares and a high r^2. Critically, these results match the claims of the 'identity' problem and are evidence of it, and are *not* confirmation that the model is reasonable. In Model 3, variables are deflated to constant dollars (local currency) using EU-KLEMS price indices and one-way (country by sector) fixed effects are included. Coefficient estimates and (within) r^2 seem reasonable, but all that has been added is 'noise' from imperfect deflation. In Model 4, labour services is replaced with total hours worked (H_EMP): this is a conventional fixed-effects production function with labour measured in physical units, but recall that CAP is still derived as a residual using an accounting identity. Replacing LAB with H_EMP doesn't fix the 'identity' problem because the variation on which the coffecient on labour is estimated remains the same (as shown by the very similar results from Models 3 and 4).

Model 5 reproduces the specification in Vandenberghe (2017) (eq 26), less the terms describing age shares of the labour mix (H_MS and H_HS are shares of medium- and high-skilled workers in total hours worked, CAPIT is the share of ICT capital in total capital compensation, and numeraire quality categories are omitted according to the Hellerstein and Neumark approach). Because a residual measure of capital input is still used, H_EMP and the quality terms for labour and capital are identified primarily from variation in labour's share of output (labour services, LAB) and the noise introduced by imperfect deflation. In the shift from Model 1 to Model 2, the identity problem allows apparently reasonable estimates to be obtained from collinear data, matching *a priori* expectations but in fact containing no new information (high fit, coefficient estimates close to factor cost shares, and close-to-CRS are guaranteed). In Models 2–5, progressive refinements obscure but do not alter the problem, and there is no basis for interpreting the result as a hypothesis test about input quality.

It is worth emphasizing that analysts working at the industry level already face a dizzying array of methodological issues, chiefly from the firm-level literature, and are unlikely to be intimately involved in generating their dataset. This gives the 'identity' problem urgency, since the obvious source of industry-level data (the System of National Accounts, or SNA) conforms to an accounting identity by design. SNA data on capital inputs are derived as a residual, according to the accounting identity in the LHS of eq 1 and assuming the existence of an aggregate production function (with competitive markets and CRS) as justification. This is because capital inputs are extremely difficult to measure: capital stocks are generally owned rather than rented, and it is not obvious how to price them in a given year. In SNA 2008, the value of capital services (the theoretically

Table 1. Demonstrative regressions using EU-KLEMS data.

Model	Specification	Model	β_{CAP}	β_{LAB}	r^2	Within r^2
1	VA ~ CAP + LAB	pooling	1	1	1	-
2	ln(VA) ~ ln(CAP) + ln(LAB)	pooling	0.371	0.705	0.992	-
3	ln(VA95) ~ ln(CAP95) + ln(LAB95)	individual FE	0.468	0.375	0.999	0.562
4	ln(VA95) ~ ln(CAP95) + ln(H_EMP)	individual FE	0.504	0.322	0.999	0.551
5	ln(VA95) ~ ln(CAP95) + ln(H_EMP) + ln(H_MS) + ln(H_HS) + ln(CAPIT)	two-way FE	0.241	0.470	0.999	0.312

[3]Downloaded from http://www.euklems.net/euk08i.shtml#top. Data on 31 (not 34) industries are used, and Canada and the United States are omitted (see Supplemental Information); temporal coverage is the same

preferred measure of capital inputs) is calculated as the user cost of capital multiplied by the productive capital stock so as to equal the (residual) value of gross operating surplus within the primary distribution of income current account (Schreyer 2001; Schreyer and Organisation for Economic Co-operation and Development, and SourceOECD (Online service) 2009). The stock of capital is observable (via investment series) but the user cost of capital is not. It can be derived via the approach of Jorgensen (see Schreyer 2001: 5.4), but this requires knowing the internal rate of return, an issue on which "theory provides no guidance" (Schreyer 2001: 5.4.2). National accountants therefore choose either to set the value of capital services equal to gross operating surplus and solve for the internal rate of return, or select an arbitrary value (for example, the rate of return on government bonds). The former 'endogenous' calculation ensures that the value of capital services is indistinguishable from the value of gross operating services, so that the accounting identity in eq 1 is satisfied (the latter 'exogenous' approach is primarily viewed as a check on data quality).

III. Description of 'aggregation' case study

To assess whether the aggregation problem matters for industry-level aggregate production functions, the remainder of the paper analyzes a case study in which production functions are used for hypothesis testing. To clearly separate the aggregation and identity problems, conventional estimates (using monetary data) are compared against estimates using *entirely* physical data. The point of the physical estimates is to see how aggregation alters inference when the identity problem is ruled out, not to analyse the implied production technology. The case study research question is: 'does materials quality matter in the forestry and logging sector?'. Apart from the inherent interest noted in Section I, the forestry and logging sector provides an ideal setting to examine the aggregation problem alone because entirely physical measurement is reasonable (output is a commodity, and a measure of materials quality may be constructed from satellite data as described below).

Estimated production functions are of the form $Y = f(K, L, M)$. In the conventional versions, Y is measured as (industry) value-added, K as the value of capital stock, L as number of persons employed, and M as the area of harvested forest (the materials input, which is transformed by the forestry and logging industry into timber). In the physical versions, Y is measured as industrial roundwood production (m^3) and K is measured as final energy consumption (terajoules). The baseline model estimates a Cobb-Douglas production function via fixed-effects. There are clearly serious limitations to this minimal set-up, including the suitability of a Cobb-Douglas specification (Heun et al. 2017), the failure of fixed effects estimation to eliminate simultaneity (Griliches and Mairesse 1995), and (for the physical production functions) variation in the capital input not captured by final energy use.[4] These (and other) shortcomings are intended to focus attention on the aggregation problem alone.

For both conventional and physical production functions, quality variation in M is measured by unsupervised classification (finite mixture modelling) using global 30 m satellite observations of forest cover loss (Hansen et al. 2013), post-processed to identify harvested areas and classified using the approach of (Filewod and Kant 2021) (see Appendix B). This measure captures high-level (i.e. relevant for industry-level analysis) variation in the quality of harvested forest land, identified using pixel-level variation along two fundamental micro-economic quality dimensions from the forest economics literature (travel costs, proxied by least-cost travel time to cities, and site productivity, proxied by canopy height). Three quality classes are retrieved, corresponding to three bivariate skew-normal distributions ('components') fit to a global stratified sample of harvested forest pixels via expectation-maximization. Class/component 2 (59.4% of sampled pixels) captures harvest occurring close to markets in forests that are likely to be intensively managed (e.g. tree farms, plantations, semi-natural planted forests), while class/component 1 (33.1%) captures harvest occurring in more

[4]Time-invariant differences in capital stock (e.g. some of the difference in energy efficiency of machinery between countries) should be absorbed by the individual fixed effects.

remote areas (e.g. semi-natural planted forests, managed natural forests), and class/component 3 (7.5%) captures harvest occurring in very remote areas with higher standing volume (e.g. frontier forests). Using these data, the production function estimates essentially ask whether the type of forest harvested matters for productivity in the forestry and logging sector.

Economic variables were obtained from the 2016 release of the World Input-Output Database (WIOD) (Timmer et al. 2015) Socio-Economic Accounts (SEA) (Gouma et al. 2018) and converted to 2010 $USD. Data on emissions-relevant (i.e. final) energy use were taken from WIOD's energy accounts (Corsatea et al. 2019), and data on industrial roundwood production from FAOSTAT. Each variable is observed once per country per year, with 28 countries retained for analysis following quality filtering of geospatial results (Figure 1). Means and standard deviations for all variables are given in Appendix C.

IV. Exploring producer heterogeneity

An aggregate (industry-level) production function assumes that all countries produce on the efficient frontier of a common production technology. If this is not the case, then non-parametric approaches to productivity analysis such as index numbers or Data Envelopment Analysis (DEA) may be preferable (van Biesebroeck 2007, 2008). In this section, DEA is used to assess potential heterogeneous technology in the case study dataset, before proceeding to production function estimation in Section V.

There are many flavours of DEA. Here, a basic input-oriented DEA model is used to estimate the minimal production possibility set consistent with the data. Each observation (country) is compared to the efficient frontier, which is a linear envelope defined by the most efficient observations. With three inputs and N observations, the basic input-oriented DEA model under variable returns (Bogetoft and Otto 2011: eq 5.1) is:

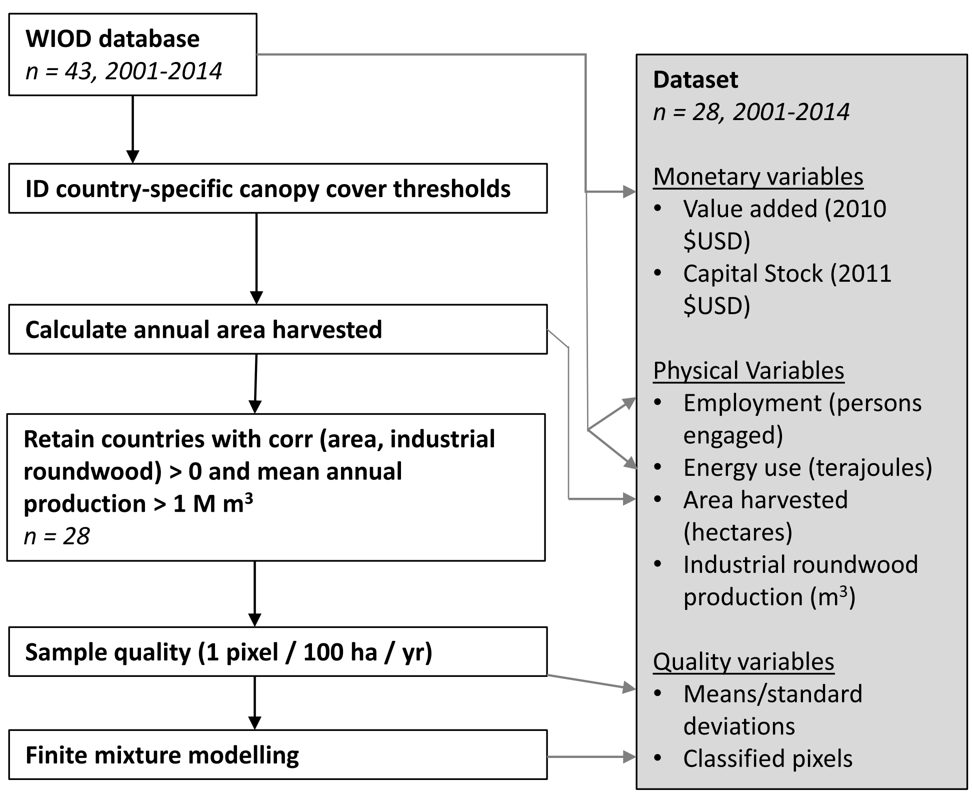

Figure 1. Case study dataset construction.

$$\begin{aligned}
&\min_{\theta^*,\lambda_1,\ldots,\lambda_N} && \theta^* - \delta\left(\sum_i z_i^- + z_y^+\right) \\
&\text{subject to} && \theta^* x_i^* + z_i^- = \sum_{n=1}^{N} \lambda^n x_i^n \qquad i \in (K, L, M) \\
& && y^* - z_y^+ = \sum_{n=1}^{N} \lambda^n y^n \\
& && \sum_{n=1}^{N} \lambda^n = 1 \quad z_i^-, z_y^+ \geq 0 \quad \theta^* \leq 1
\end{aligned} \tag{2}$$

The input efficiency of the industry being evaluated (θ^*) is the scaling factor by which all inputs could be reduced without reducing output (Farrell efficiency), as demonstrated by the production plans of other industries. Weights λ^n compare the industry to a linear combination of most similar industries. The z variables are 'complementary slacks', which capture possible gains in efficiency beyond what can be achieved by a proportional rescaling of inputs; δ is the corresponding penalty for slack. The DEA analysis employs the weakest possible assumptions (in this context, free disposability, convexity, and variable returns to scale (Bogetoft and Otto 2011)).[5] Because eq (2) is non-stochastic, 2001–2014 values are averaged to reduce the impact of interannual variation (Ruggiero 2007) and two outliers are excluded (Austria, due to apparent measurement error in the energy input, and Russia, due to missing data).

The last four columns of Table 2 collect DEA efficiency scores (**E**; equal to 1 if the country is at the efficient frontier) and complementary slacks (**S**; summed across inputs to compare relative efficiency between countries). The analysis is performed for the physical data only, first for the basic dataset (**E** and **S**; output is industrial roundwood and inputs are final energy, person-hours, and harvested area) and second ($\mathbf{E}_{aug}$ and $\mathbf{S}_{aug}$) for a modified dataset where harvested area has been allocated into the three quality classes described in Section III. The remainder of Table 2 provides average products (for K, L, and M) and summary information about materials input quality. Compared to DEA, average products give limited information (and implicitly assume constant returns to scale (Bogetoft and Otto 2011)), but do provide a straightforward first assessment of the plausibility of assuming a single production technology.

The results of this benchmarking exercise clearly demonstrate heterogeneous technology in the forestry and logging sector. Only 11 of 26 countries operate at the efficient frontier ($\mathbf{E} = 1$) in the unaugmented model (for $\mathbf{E}_{aug}$ this rises to 13). Average products exhibit wide variation across countries, some of which is likely due to measurement error.[6] The key implication is that the aggregation problem should matter in this dataset. One response (implemented in Section V) is to retain only (theoretically consistent) efficient producers when estimating production functions. Another is to abandon parametric analysis altogether, instead using detailed DEA or index number approaches in combination with narrative economic history to understand the importance of materials quality in global forestry and logging. However, the suitability of non-parametric methods depends on the importance of measurement error (van Biesebroeck 2007), and this would forego formal hypothesis testing.

V. Production function estimates

In this section, a variety of Cobb-Douglas production function estimates are compared to empirically assess the importance of the aggregation problem for industry-level analysis, using the hypothesis-testing case study described in Section III. Estimating fully physical production functions (which are clearly not subject to an accounting identity) isolates the aggregation problem; a comparison of results obtained using fixed effects versus heterogeneous panel estimators (the Swamy and Mean Groups estimators) then shows how inference changes when an aggregate production technology is not assumed. An alternative set of estimates uses only the DEA-efficient firms identified in Section IV, for which a single production technology is more reasonable. A final set of contrasts considers two alternate approaches to

[5]Free disposability is innocuous if the data do not include undesirable outputs. Convexity implies that unobserved (convex) combinations of inputs are feasible production plans, which is increasingly plausible at higher levels of aggregation. Variable returns is weaker than the alternate assumption that production can be scaled.

[6]For example, roundwood output per area harvested ranges from 0.068 (Indonesia) to 10.414 (Slovenia) because satellite observations of harvested area at 30 m resolution include some forest cover loss not due to harvest (Indonesia) and exclude harvest from non-clearcut silviculture (Slovenia). Since measurement error across countries is extremely difficult to eliminate, data exploration is a vital step in sector-level analysis.

Table 2. Physical input intensities and quality characteristics (2001–2014 means) for A02 panel.

	Flag	Mm^3: 1000 hectares	Mm^3: 1000 persons	Mm^3: Tj	n	Access (hours)	height (metres)	C1/C2/C3	E	S	E_{aug}	S_{aug}
AUS	2	0.227	2.341	0.00183	16980	4.4	18.4	38/57/5	0.89	5716.9	0.98	8687.59
AUT	1	1.978	0.519	0.44221	1583	0.9	26.7	0/100/0	NA	NA	NA	NA
BEL	1	2.496	1.733	0.00348	1400	0.5	21.5	0/100/0	1	0	1	0
CAN	2	0.181	3.203	0.00325	127286	6.8	17	70/22/8	1	0	1	0
CHE	1	8.948	1.007	0.02009	1400	0.6	24.6	0/100/0	1	0	1	0
CHN	3	0.438	0.01	0.00048	46419	3	23.6	17/75/7	0.95	275886.12	1	291217.08
CZE	1	4.286	0.491	0.00254	1404	0.5	21.5	0/100/0	0.94	1150.03	0.94	1150.03
DEU	1	4.775	1.192	0.00732	2681	0.6	21.8	0/100/0	1	0	1	0
DNK	1	0.919	0.338	0.00231	1399	1	15.9	3/97/0	1	0	1	0
FRA	1	0.939	0.85	0.00185	5695	0.7	18.8	0/100/0	0.64	5733.05	0.7	6271.01
GBR	1	0.974	0.528	0.00105	2449	1	16.1	3/97/0	0.35	953.05	0.36	1008.8
HRV	1	1.98	0.294	0.00421	1400	0.7	22.1	0/100/0	0.73	0	0.74	0.5
IDN	3	0.068	0.02	0.00133	131441	3.2	24	17/72/11	0.13	340.9	0.15	451.25
IND	3	1.496	0.002	7e-04	4740	1.9	22.7	6/89/6	0.48	35231.69	0.54	40508.54
IRL	1	0.691	0.224	0.00142	1400	0.7	14.1	0/100/0	0.42	0	0.43	1.2
JPN	3	0.889	0.189	0.00041	2989	0.5	25.3	0/99/0	0.27	7504.9	0.27	7762.08
KOR	3	0.438	0.055	0.00051	1422	0.5	23.6	0/100/0	0.15	0	0.15	0.39
LTU	1	0.454	0.415	0.00386	1613	0.9	18.5	0/100/0	0.76	0	0.91	8.46
NOR	1	0.35	1.576	0.02756	3441	1.4	18.9	4/96/0	1	0	1	0
POL	1	1.9	0.516	0.00307	2479	0.7	18.7	0/100/0	0.6	9.56	0.6	9.56
PRT	1	0.464	0.808	0.00538	3328	0.9	16.8	0/100/0	0.81	0	1	15.84
ROU	1	1.172	0.261	0.01038	1675	1.4	26.5	0/100/0	1	0	1	0
RUS	3	0.154	NA	NA	163247	7.9	19.7	60/25/14	NA	NA	NA	NA
SVK	1	2.939	0.293	0.00846	1400	0.8	25.2	0/100/0	1	0	1	0
SVN	1	10.414	0.368	0.00255	1400	0.6	24.9	0/100/0	1	0	1	0
SWE	1	0.363	2.043	0.01325	27146	2.2	17.9	29/71/0	1	0	1	0
TUR	3	0.753	0.081	0.00251	2601	1	18.4	0/100/0	0.39	35.83	0.42	40.14
USA	2	0.293	0.905	0.00448	179747	1.9	21.3	7/92/1	1	0	1	0

Columns 3–5 give output:input ratios (average products), expressed as million m^3 of industrial roundwood output per unit of materials, labour, or capital (energy) input. Columns 6–9 give the the number of harvested pixels sampled to characterize materials input quality (n), the mean scores for each underlying quality dimension used for unsupervised classification (see Appendix B for details), and the shares of harvested area falling into each quality class (C1-C3).

including a quality term in the production function. All specifications are compared against estimates using conventional (monetary) data, giving 24 estimates in total. In the physical production function industries use capital (energy), person-hours, and forests to produce a commodity, while in the economic version industries use capital (stocks), person-hours, and forests to produce profits.

Broadly speaking, there are two approaches to test hypotheses about input quality in a production function framework: either variables describing the quality dimensions of inputs are simply added to the estimating equation, without a formal statement of the upstream production technology, or the production function may be modified to accommodate heterogeneous inputs. The former approach (hereafter termed 'characterization'), has a long pedigree (e.g. Garnero, Kampelmann, and Rycx 2014; Griliches 1967) but lacks theoretical coherence.[7] The second approach ('augmentation') is consistent with an upstream aggregate production function. Early efforts (e.g. Griliches 1967) included inputs of different quality as separate factors of production. More recently, Hellerstein and Neumark (HN henceforth) (Hellerstein and Neumark 1995; Hellerstein, Neumark, and Troske 1999) showed that input quality classes may be modelled as substitutes whose productivity is estimated relative to a numeraire.

Implementing the HN approach with three quality classes, taking logs, and (approximately) linearalizing by applying $ln(1-x) \approx x$ (an approach taken by HN and others (Griliches 1967; Hellerstein and Neumark 2007)) yields the estimating equation:

$$y_{it} = a + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \beta_{c2} c2_{it} + \beta_{c3} c3_{it} + e_{it} \tag{3}$$

Lower case variables denote logs, and $c2_{it}$ and $c3_{it}$ are the fraction of materials input falling into quality class 2 and 3 for producer i in year t. The analagous 'characterization' approach is simply:

[7]Added variables are frequently seen as statistical controls, irrespective of theoretical coherence: HN (Hellerstein and Neumark 1995), for example, also employ terms for ownership structure, location, and firm age.

$$y_{it} = a + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \beta_{\bar{H}} \bar{H}_{it} + \beta_{\bar{A}} \bar{A}_{it} + e_{it} \tag{4}$$

This is a log-linearized Cobb-Douglas production function with three factors, to which two variables describing quality have been added ($\bar{H}$ and $\bar{A}$, the country mean scores of harvested pixels on the two underlying quality dimensions used for finite mixture modelling, i.e. canopy height and travel time to cities)

Baseline estimates of eq 3–4 are obtained using one-way (country) fixed effects (Table A1); following Lagrange multiplier tests and Chamberlain's test as recommended by (Baltagi 2013). A parallel set of estimates use heterogeneous coefficient models, which relax the assumption that a single coefficient vector describes the data generating process for all units. These models are intended to accommodate the producer heterogeneity at the root of the aggregation problem[8]; they are also empirically motivated, since Chow tests for poolability in the fixed effects models firmly rejects for both the physical and economic datasets ($p \leq 0.001$ in both cases) and estimating country-specific models by OLS reveals wide variation in coefficients. The first heterogenous panel estimator is that of Swamy (1970), who proposed a random coefficient estimator in which the common data generating process is identified as the average of the individual coefficients weighted using the producer-level variances. Collecting coefficients in β and letting X represent the data, the model may be written (Croissant and Millo 2018):

$$y_{it} = \beta X_{it} + \delta_i X_{it} + e_{it}, \quad \delta_i = \beta_x - \beta \sim N(0, \Lambda) \tag{5}$$

If the deviations (δ_is) from the average effects are not correlated with the idiosyncratic error e_{it}, eq 5 can be estimated by generalized least squares. Alternatively, the Mean Groups estimator (Pesaran and Smith 1995) identifies each mean effect β as the simple average of the individual effects from i individual regressions ($\bar{\beta} = \frac{1}{N}\sum_{i=1}^{N} \beta^i$). Whereas Swamy imposed parametric assumptions to derive an expression for the variance of $\bar{\beta}$, the Mean Groups approach assumes only the exogeneity of the regressors and independent errors to express var($\bar{\beta}$) as the empirical covariance of the β^i (Croissant and Millo 2018).

Fixed effects estimates using physical data (models 1–6) are poor: using the full dataset (odd numbers in Table 3) coefficients on capital and labour are not significant, and overall model fits (reported as adjusted 'within' r^2) are low (significant coefficients on M are implied by the quality filtering shown in Appendix B, Figure A3). Fixed-effects results using economic data (models 7–12) are somewhat better, with significant and plausible coefficients on all inputs. However, point estimates (and implied elasticities of scale) are very low, ranging from 0.083 to 0.284 for capital, 0.078–0.313 for labour, and 0.064–0.137 for materials. An obvious explanation for this is that endogenous input selection is known to bias coefficient estimates downwards in fixed effects models, "with the capital coefficient falling faster than the labor coefficient and often actually reaching zero" (Griliches and Mairesse 1995): 11. For the physical models, omitted variables (notably the use of energy to represent the capital input, despite prior support (Santos et al. 2016) for this approach) and measurement error may also play a role: the coefficient estimates on capital and labour in the fixed effects models are almost never significant, suggesting that these models are not viable platforms for testing hypotheses about materials quality. Using only DEA-efficient countries (even numbers in Table 3) improves goodness-of-fit for physical models only but decreases it for the conventional models, because DEA was performed using the physical dataset only. While results are plausible (taking endogeneity bias into account) for the conventional models, this should be surprising given the marked heterogeneity shown in Section IV. A weak version of the identity problem may be at play due to the use of monetary measures of output and capital.

The picture that emerges from the heterogeneous panel models (Table 4) is rather different. The results for the physical models are greatly improved, with all inputs showing significant

[8]The idea of using these methods to tackle the heterogeneous technology highlighted by the theoretical aggregation literature appears to have been suggested as early as Zellner, 1966 (cited in Swamy 1970).

Table 3. Results of fixed effects regressions.

	Basic model		'characterization'		'augmentation'	
Physical production function: RWD = TJ + EMP + AREA						
	(1)	(2)	(3)	(4)	(5)	(6)
K	0.013	-0.009	0.010	-0.002	0.011	0.001
	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)	(0.0132)
L	0.013	0.150	0.034	0.121	0.035	0.002
	(0.037)	(0.073) *	(0.037)	(0.074)	(0.038)	(0.061)
M	0.122	0.126	0.121	0.138	0.125	0.133
	(0.014) ***	(0.019) ***	(0.0140 ***	(0.020) ***	(0.014) ***	(0.019) ***
$\bar{H}$			0.008	0.025		
			(0.007)	(0.012) *		
$\bar{A}$			-0.045	-0.033		
			(0.022) *	(0.047)		
c2					0.462	1.168
					(0.221) *	(0.412) **
c3					0.594	-2.368
					(0.533)	(1.155) *
adj r2	0.125	0.270	0.139	0.220	0.131	0.252
Economic production function: VA = K + EMP + AREA						
	(7)	(8)	(9)	(10)	(11)	(12)
K	0.188	0.220	0.120	0.083	0.193	0.284
	(0.039) ***	(0.084) **	(0.039) **	(0.082)	(0.038) ***	(0.074) ***
L	0.219	0.078	0.313	0.286	0.307	0.171
	(0.065) **	(0.146)	(0.063) ***	(0.140) *	(0.065) ***	(0.108)
M	0.120	0.137	0.095	0.064	0.128	0.123
	(0.024) ***	(0.036) ***	(0.023) ***	(0.036).	(0.024) ***	(0.031) ***
$\bar{H}$			0.069	0.114		
			(0.012) ***	(0.022) ***		
$\bar{A}$			-0.054	0.041		
			(0.036)	(0.085)		
c2					1.913	0.940
					(0.370) ***	(0.712)
c3					0.818	7.534
					(0.897)	(2.030) ***
adj r2	0.096	0.055	0.194	0.202	0.157	0.126

Even-numbered models are estimated using the full dataset (n=28); odd-numbered models use only DEA-efficient countries (n=11 for the three-input models (2,8) and n=13 for the quality-augmented models (4,6,10,12).
*** [0, 0.001], ** (0.001, 0.01], * (0.01, 0.05]. Where "[" indicates that the bound of the interval is included and "(" does not.

effects on output and plausible magnitudes. The output elasticity of labour is about twice that of capital (measured as final energy consumption). This dramatic change from the (physical) fixed effects results in Table 3 to the heterogeneous coefficient results in Table 4 is a clear demonstration of the importance of the aggregation problem at the sector level, in a setting where the intertwined identity problem is unambiguously ruled out. For the conventional (monetary) models in Table 4, the relative importance of inputs changes (i.e. is roughly equal), and the labour input is never significant. Note that the coefficients on capital and labour in the baseline conventional estimates (models 7 and 10) sum to nearly 1, and that this is consistent with the (monetary) capital stock measure having been constructed so as to sum with labour's share of output to equal value added.

To summarize: comparing fixed effects and heterogenous panel estimates for the physical production functions excludes the identity problem and highlights the aggregation problem. Because the production processes of different countries are heterogeneous, attempting to estimate the parameters of a common production technology yields poor results. When the estimation strategy permits variable coefficients, the effect of inputs on outputs is strongly significant. Handling heterogeneity appropriately is clearly necessary, as strongly suggested by the benchmarking results in Section IV, but this constrains the types of questions that can be asked of the data. Specifically, heterogeneous technology precludes a behavioural interpretation of estimated parameters: mean coefficients cannot be interpreted as output elasticities because the modelling strategy starts with the assumption that no common production technology exists. On the other hand, appropriately modelling coefficient heterogeneity does provide a valid framework for hypothesis testing.

Table 4. Results of heterogeneous panel models.

	Swamy			Mean Groups		
Physical production function: RWD = TJ + EMP + AREA						
	(1)	(2)	(3)	(4)	(5)	(6)
K	0.125	0.119	0.164	0.116	0.122	0.185
	(0.027) ***	(0.045) **	(0.061) **	(0.037) **	(0.041) **	(0.057) **
L	0.394	0.308	0.281	0.414	0.297	0.263
	(0.069) ***	(0.010) **	(0.102) **	(0.088) 8***	(0.091) **	(0.090) **
M	0.120	0.104	0.103	0.132	0.103	0.100
	(0.027) ***	(0.031) ***	(0.025) ***	(0.031) ***	(0.027) ***	(0.021) ***
$\bar{H}$		0.011			0.011	
		(0.014)			(0.012)	
$\bar{A}$		0.043			0.109	
		(0.130)			(0.110)	
c2			-0.054			-0.185
			(0.515)			(0.478)
c3			-0.351			NA
			(0.923)			
multiple_r2	0.978	0.998	0.988	0.995	0.997	0.997
Economic production function: VA = K + EMP + AREA						
	(7)	(8)	(9)	(10)	(11)	(12)
K	0.520	0.303	0.114	0.547	0.282	0.097
	(0.108) ***	(0.105) **	(0.135)	(110) ***	(0.098) **	(0.129)
L	0.473	0.334	-0.025	0.451	0.325	-0.209
	(0.288)	(0.285)	(0.285)	(0.294)	(0.274)	(0.274)
M	0.148	0.083	0.129	0.165	0.096	0.140
	(0.027) ***	(0.039) *	(0.048) **	(0.042) ***	(0.032) **	(0.043) **
$\bar{H}$		0.126			0.140	
		(0.033) ***			(0.030) ***	
$\bar{A}$		0.043			0.151	
		(0.249)			(0.223)	
c2			5.490			6.036
			(1.585) ***			(1.536) ***
c3			1.027			NA
			(0.999)			
multiple_r2	0.943	0.995	0.963	0.986	0.993	0.994

For each combination of estimator and dataset, the first (e.g. (1)) is the baseline specification, the second (2) uses the 'characterization' approach to include materials input quality, and the third (3) uses the HN 'augmentation' approach.
*** [0, 0.001], ** (0.001, 0.01], * (0.01, 0.05]. Where "[" indicates that the bound of the interval is included and "(" does not.

Answering the case study question – does materials quality matter in global forestry and logging? – is complicated by a number of methodological issues not considered here (e.g. choice of functional form). Disregarding the fixed effects models due to the inadmissibility of assuming a single (aggregate) production technology, the heterogeneous coefficient models in Table 4 give contradictory results. Results from the physical models show that materials quality does not matter for industrial roundwood output in either a 'characterization' or an 'augmentation' (HN) approach. Conversely, results from the conventional (monetary) models show that quality does matter: countries producing from forests classed in quality bin 2 tend to see higher value added (models 9 and 12), and the height of harvested forests is significantly and positively associated with increased value added (models 8 and 11). These results can be reconciled by considering the different outcome variables used, recalling that the measure of materials quality used here is essentially a description of the forest management regime. The physical results show that more than one regime can be associated with high industrial roundwood output (e.g. both short-rotation plantations and frontier forests can produce high harvested volumes). The conventional 'characterization' results (models 8, 11) show that countries whose harvested forests are more mature tend to see higher profits since mature forests yield more wood (and typically higher value) per unit harvesting effort. The monetary 'augmentation' results (9, 12) are probably driven by two outliers, Canada and Russia, which are the only countries primarily producing quality categories *other* than component two and are widely known as producers of unprocessed (lower-value-added) products.

VI. Concluding remarks

Industry-level production function analysis is likely to remain an important area of applied work, especially when rich firm-level datasets are scarce. However, the theoretical basis for estimating industry-level production functions is less well studied than either the firm-level or total-economy case, potentially leading applied workers to overlook key concerns arising from disparate literatures.

This paper considers two such concerns: the identity and the aggregation problems. It finds (Section II) that the former is potentially a serious problem for industry-level analysis, given the tendency to rely on SNA data (which explicitly conforms to an accounting identity) and the current preference for using capital services to measure capital inputs. In theory, the identity problem can be avoided by paying careful attention to data. This requires increased attention to the origins of industry-level data, and in particular to the construction of observations on capital and their relationship with the output measure. Regressions of value-added on capital services should be ruled out. Similar regressions on capital stocks are suspect, but may be admissible if the capital stock series are built up independently of the value-added measure.

The benchmarking results in Section IV and the production function results in Section V argue for taking the aggregation problem into account in industry-level work. Industries (and firms) employ a wide variety of production processes, and using aggregate production functions for hypothesis testing (or other purposes) requires appropriately modelling heterogeneity in production technology. Fortunately, techniques for assessing heterogeneity in production processes are readily available (e.g. average products, DEA), and, in a panel data context, heterogeneous coefficient estimators provide a straightforward modelling solution.

A minimal interpretation of these results is that both the identity problem and the aggregation problem *can* matter for industry-level analysis. The wide array of methodological issues associated with production function estimation can easily create a form of 'attention bias', leading to significant issues being neglected. Nevertheless, applied economists working with industry-level data should carefully consider the importance of heterogeneous technology in their data and the construction of any monetary variables used; given the degree to which subjective choices about estimation strategy can influence results (Heun et al. 2017), it will also be important to search for complementary sources of information (e.g. expert interviews, narrative histories) when interpreting results.

Based on these observations, the following guiding questions can be considered for production function estimation with industry-level data:

- **Do the data conform to an upstream accounting identity?** If yes, production function estimation is tautological, and different data must be sought. In general, data originating in the System of National Accounts conform to an accounting identity by design. Any measure of capital services should be carefully scrutinized, since capital services are not directly observed and are usually derived as an accounting residual. If stock measures of capital are used, a thorough exploration of their origin (e.g. in tax returns) is recommended.
- **Do the data involve an element of aggregation?** At the industry level, the answer is likely always yes – and heterogeneity in production technology should be explored empirically and modelled if necessary. Note that essentially all data, including firm-level data, involve aggregation. At issue is whether this aggregation matters for the research question(s) of interest. At minimum, average products should be calculated as a quick and simple means of assessing production heterogeneity.
- **Have a broad range of methodological issues been considered?** If not, potentially serious problems may have been neglected and inference may be contingent on an arbitrary methodological decision. This is particularly pressing in areas of less active methodological research, because the burden on analysts to synthesize multiple areas of inquiry is correspondingly higher (e.g. prior work may not have established preferred model specifications or estimation strategies; methods from other areas may be applied out of context without regard to underlying assumptions, etc.).

Looking ahead, criticisms of aggregate production functions have gone well beyond industry-level work (e.g. Felipe and McCombie 2013), while new methods for total-economy aggregation have emerged (Baqaee and Farhi 2018). This paper has attempted to take the core issues seriously in a specific applied context, providing some new results and guidance for future work; since both the 'aggregation' and 'identity' problems appear to be both significant and neglected, a metascientific assessment of their prevalence in published production function studies could be a valuable next step.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Social Science and Humanities Research Council of Canada, and the ForestLAB grant at the Grantham Research Institute of Climate Change and the Envronment at the London School of Economics and Political Science.

ORCID

Ben Filewod http://orcid.org/0000-0002-1572-3151

References

Badi, H. B. 2013. *Econometric Analysis of Panel Data*. 5th ed. John Wiley & Sons, Ltd.

Baqaee, D., and E. Farhi. 2018. *The Microeconomic Foundations of Aggregate Production Functions*. Cambridge, MA: National Bureau of Economic Research. November w25293. Tech. rep. w25293. https://doi.org/10.3386/w25293.

Bogetoft, P., and L. Otto. 2011. *Benchmarking with DEA, SFA, and R*. Vol. 157 New York, NY: Springer. https://doi.org/10.1007/978-1-4419-7961-2.

Ceccherini, G., G. Duveiller, G. Grassi, G. Lemoine, V. Avitabile, R. Pilli, A. Cescatti, et al. 2020. "Abrupt Increase in Harvested Forest Area Over Europe After 2015." *Nature* 583 (7814): 72–77. July. https://doi.org/10.1038/s41586-020-2438-y.

Corsatea, T. D. 2019. *World Input-Output Database Environmental Accounts: Update 2000 2016. Tech. Rep. JRC116234*. European Comission Joint Research Centre.

Croissant, Y., and G. Millo. 2018. *Panel Data Econometrics with R*. First ed. Hoboken, NJ: John Wiley & Sons.

Eberhardt, M., C. Helmers, and H. Strauss. 2013. "Do Spillovers Matter When Estimating Private Returns to R&D?" *Review of Economics and Statistics* 95 (2): 436–448. May. https://doi.org/10.1162/REST_a_00272.

Feenstra, R. C., R. Inklaar, and M. P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–3182. October. 20130954. https://doi.org/10.1257/aer.20130954.

Felipe, J., and F. M. Fisher. 2003. "Aggregation in Production Functions: What Applied Economists Should Know." *Metroeconomica* 54 (2–3): 208–262. May. https://doi.org/10.1111/1467-999X.00166.

Felipe, J., and C. Holz. 2001. "Why Do Aggregate Production Functions Work? Fisher's Simulations, Shaikh's Identity and Some New Results." *International Review of Applied Economics* 15 (3): 261–285. https://doi.org/10.1080/02692170110052338.

Felipe, J., and J. McCombie. 2006. "The Tyranny of the Identity: Growth Accounting Revisited." *International Review of Applied Economics* 20 (3): 283–299. https://doi.org/10.1080/02692170600735963.

Felipe, J., and J. S. McCombie. 2013. *The Aggregate Production Function and the Measurement of Technical Change*. Edward Elgar.

Felipe, J, and J. S. L. McCombie. 2010. "What Is Wrong with Aggregate Production Functions. On Temple's 'Aggregate Production Functions and Growth economics'." *International Review of Applied Economics* 24 (6): 665–684. November. https://doi.org/10.1080/02692171.2010.512146.

Felipe, J., and J. S. L. McCombie. 2014. "The Aggregate Production Function: 'Not Even Wrong." *Review of Political Economy* 26 (1): 60–84. January. https://doi.org/10.1080/09538259.2013.874192.

Filewod, B., and S. Kant. 2021. "Identifying Economically Relevant Forest Types from Global Satellite Data." *Forest Policy and Economics* 127:127. https://doi.org/10.1016/j.forpol.2021.102452.

Fisher, F. M. 1969. "The Existence of Aggregate Production Functions." *Econometrica* 37 (4): 553. October. https://doi.org/10.2307/1910434.

Garnero, A., S. Kampelmann, and F. Rycx. 2014. "The Heterogeneous Effects of Workforce Diversity on Productivity, Wages, and Profits." *Industrial Relations: A Journal of Economy & Society* 53 (3): 430–477. July. https://doi.org/10.1111/irel.12064.

Geoffrey, J. M., and S. Rathnayake. 2014. "On the Number of Components in a Gaussian Mixture Model: Number of Components in a Gaussian Mixture Mode." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 4 (5): 341–355. September. https://doi.org/10.1002/widm.1135.

Gouma, R. 2018. "WIOD Socio-Economic Accounts 2016: Sources and Methods." *Technical Report*: 27.

Griliches, Z. 1967. "Production Functions in Manufacturing: Some Preliminary Results." *The Theory and Empirical Analysis of Production NBER Book Series Studies in Income and Wealth* 275–340.

Griliches, Z., and J. Mairesse. 1995. Production Functions: The Search for Identification. *Tech. Rep. W5067*. Cambridge, MA: National Bureau of Economic Research. March w5067. https://doi.org/10.3386/w5067.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, et al. 2013. "High-Resolution Global Maps of 21st-Century Forest

Cover Change." *Science* 342 (6160): 850–853. November. https://doi.org/10.1126/science.1244693.

Harrigan, J. 1999. "Estimation of Cross-Country Differences in Industry Production Functions." *Journal of International Economics* 47 (2): 267–293. https://doi.org/10.1016/S0022-1996(98)00011-7.

Hashem Pesaran, M., and R. Smith. 1995. "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels." *Journal of Econometrics* 68 (1): 79–113. https://doi.org/10.1016/0304-4076(94)01644-F.

Hellerstein, J. K., and D. Neumark. 1995. "Are Earnings Profiles Steeper Than Productivity Profiles? Evidence from Israeli Firm-Level Data." *The Journal of Human Resources* 30 (1): 89. https://doi.org/10.2307/146192.

Hellerstein, J. K., and D. Neumark. 2007. "Production Function and Wage Equation Estimation with Heterogeneous Labor." In *Hard-To-Measure Goods and Services: Essays in Honor of Zvi Griiliches*, 42. University of Chicago Press.

Hellerstein, J. K., D. Neumark, and K. R. Troske. 1999. "Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations." *Journal of Labor Economics* 17 (3): 409–446. https://doi.org/10.1086/209926.

Heun, M., J. Santos, P. Brockway, R. Pruim, T. Domingos, and M. Sakai. 2017. "From Theory to Econometrics to Energy Policy: Cautionary Tales for Policymaking Using Aggregate Production Functions." *Energies* 10 (2): 203. February. https://doi.org/10.3390/en10020203.

Himes, A., M. Betts, C. Messier, and R. Seymour. 2022. ""Perspectives: Thirty Years of Triad Forestry, a Critical Clarification of Theory and Recommendations for Implementation and Testing." *Forest Ecology & Management* 510 (4): 120103. https://doi.org/10.1016/j.foreco.2022.120103.

Ilmakunnas, P., and T. Miyakoshi. 2013. "What Are the Drivers of TFP in the Aging Economy? Aging Labor and ICT Capital." *Journal of Comparative Economics* 41 (1): 201–211. February. https://doi.org/10.1016/j.jce.2012.04.003.

Oliveira Prates, M., C. Rˆomulo Barbosa Cabral, and V. H. Hugo Lachos. 2013. "**Mixsmsn**: Fitting Finite Mixture of Scale Mixture of Skew-Normal Distributions." *Journal of Statistical Software* 54 (12). https://doi.org/10.18637/jss.v054.i12.

O'Mahony, M., and M. P. Timmer. 2009. "Output, Input and Productivity Measures at the Industry Level: The EU KLEMS Database." *The Economic Journal* 119 (538): F374–F403. June. https://doi.org/10.1111/j.1468-0297.2009.02280.x .

Romulo Barbosa Cabral, C., V. H. Hugo Lachos, and M. O. Prates. 2012. "Multivariate Mixture Modeling Using Skew-Normal Independent Distributions." *Computational Statistics & Data Analysis* 56 (1) : 126–142. January. https://doi.org/10.1016/j.csda.2011.06.026.

Ruggiero, J. 2007. "A Comparison of DEA and the Stochastic Frontier Model Using Panel Data." *International Transactions in Operational Research* 14 (3): 259–266. May. https://doi.org/10.1111/j.1475-3995.2007.00585.x.

Santos, J. 2016. "Does a Small Cost Share Reflect a Negligible Role for Energy in Economic Production?." *MPRA Paper 70850*.

Schreyer, P. 2001. *Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth*. Paris: OECD. Statistics. OCLC: 248096819.

Schreyer, P., and Organisation for Economic Co-operation and Development, and SourceOECD (Online service), eds. 2009. *Measuring Capital: OECD Manual 2009*. 2. ed. Paris: OECD. en OCLC: ocn441208732.

Shaikh, A. 1974. "Laws of Production and Laws of Algebra: The Humbug Production Function." *The Review of Economics and Statistics* 56 (1): 115. February. https://doi.org/10.2307/1927538.

Simard, M., N. Pinto, J. B. Fisher, and A. Baccini. 2011. "Mapping Forest Canopy Height Globally with Spaceborne Lidar." *Journal of Geophysical Research* 116 (G4). November. https://doi.org/10.1029/2011JG001708.

Solow, R. M. 1957. "Technical Change and the Aggregate Production Function." *The Review of Economics and Statistics* 39 (3): 312. August. https://doi.org/10.2307/1926047.

Swamy, P. A. V. B. 1970. "Efficient Inference in a Random Coefficient Regression Model." *Econometrica* 38 (2): 311–323. https://doi.org/10.2307/1913012.

Temple, J. 2006. "Aggregate Production Functions and Growth Economics." *International Re- View of Applied Economics* 20 (3): 301–317. July. https://doi.org/10.1080/02692170600736052.

Temple, J. 2010. "Aggregate Production Functions, Growth Economics, and the Part-Time Tyranny of the Identity: A Reply to Felipe and McCombie." *International Review of Applied Economics* 24 (6): 685–692. November. https://doi.org/10.1080/02692171.2010.511454.

Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer, and G. J. de Vries. 2015. "An Illustrated User Guide to the World Input–Output Database: The Case of Global Automotive Production." *Review of International Economics* 23 (3): 575–605. August. https://doi.org/10.1111/roie.12178.

Timmer, M. P., M. O'Mahony, and B. van Ark. 2007. "EU KLEMS Growth and Productivity Ac- Counts: Overview November 2007 Release." *In* 200 (11): 64–78. https://doi.org/10.1177/0027950107080390.

van Biesebroeck, J. 2007. "Robustness of Productivity Estimates." *The Journal of Industrial Economics* 55 (3): 529–569. https://doi.org/10.1111/j.1467-6451.2007.00322.x.

van Biesebroeck, J. 2008. "The Sensitivity of Productivity Estimates: Revisiting Three Important Debates." *Journal of Business & Economic Statistics* 26 (3): 311–328. https://doi.org/10.1198/073500107000000089.

Vandenberghe, V. 2017. ""The Productivity Challenge. What to Expect from Better-Quality Labour and Capital inputs?" En." *Applied Economics* 49 (40): 4013–4025. August. https://doi.org/10.1080/00036846.2016.1273504.

Weiss, D. J., A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, et al. 2018. "A Global Map of Travel Time to Cities to Assess Inequalities in Accessibility in 2015." *en In: Nature* 553 (7688): 333–336. January. https://doi.org/10.1038/nature25181.

Appendices

Appendix A: Shaikh's (1974) critique

This appendix reproduces the derivation of the identity problem (in a panel data context) given in Anwar Shaikh's 1974 paper (Shaikh 1974). For convenience, the notation has been slightly altered and all algebraic steps are made explicit.

Consider time series data on the value of output VA (value-added) and total payments to factors of production W and R (wages and profits, respectively), all in monetary terms, as well as any index numbers L and K for labour and capital inputs. By definition, we have

$$VA_t \equiv W_t + R_t \tag{A1}$$

for each period t. This is an accounting identity, typically motivated by the argument that payments to factors of production exhaust profits in competitive markets. Write this equation in per-unit-labor terms by dividing through by L_t and multiplying the last term by one:

$$\frac{VA_t}{L_t} = \frac{W_t}{L_t} + \frac{R_t}{L_t}\left(\frac{K_t}{K_t}\right)$$

Letting q_t and k_t represent the ratios of output and capital to labour, and $w_t = \frac{W_t}{L_t}$ and $k_t = \frac{R_t}{K_t}$ (the wage and profit rates):

$$q_t = w_t + r_t k_t \tag{A2}$$

Then differentiate with respect to time (using $\frac{\partial x}{\partial t} = \dot{x}$ and dropping the time subscript for simplicity):

$$\dot{q} = \dot{w} + \dot{r}k + r\dot{k}$$

multiply by 1 and regroup terms to express the right hand side in rates:

$$\dot{q} = \dot{w}\frac{w}{w} + \dot{r}k\frac{r}{r} + r\dot{k}\frac{k}{k} \;=\; \left(\frac{\dot{w}}{w}\right)w + \left(\frac{\dot{r}}{r}\right)rk + \left(\frac{\dot{k}}{k}\right)kr$$

Divide through by q:

$$\frac{\dot{q}}{q} = \left(\frac{\dot{w}}{w}\right)\frac{w}{q} + \left(\frac{\dot{r}}{r}\right)\frac{rk}{q} + \left(\frac{\dot{k}}{k}\right)\frac{rk}{q}$$

Note that we can write the share of profits in output as $s = \frac{R_t}{VA_t}$, and (by identity 6) the share of wages in output as $1 - s$. Futhermore, $\frac{R_t}{VA_t} = \frac{r_t k_t}{q_t}$. Making these substitutions gives:

$$\frac{\dot{q}}{q} = (1-s)\frac{\dot{w}}{w} + s\frac{\dot{r}}{r} + s\frac{\dot{k}}{k} \tag{A3}$$

Let $\frac{\dot{B}}{B} = (1-s)\frac{\dot{w}}{w} + s\frac{\dot{r}}{r}$, i.e. the factor-weighted share of the growth rate of wages and profits, and note that equation 8 then appears identical to equation 2a in Solow's famous paper on the measurement of technical change (Robert 1957). If factor shares of output happen to be stable over time (**assumption 1**) then s is a constant. Letting $s = \beta$ to emphasize this, 8 may be integrated over time as follows:

$$\int \frac{\dot{q}}{q}dt = \int \frac{\dot{B}}{B}dt + \int \beta\frac{\dot{k}}{k}dt$$

$$\ln q = \int \frac{\dot{B}}{B}dt + \beta \ln k + \ln c$$

where c is the constant of integration. Taking exponents:

$$q = exp\left(\int \frac{\dot{B}}{B}dt\right)ck^{\beta} \tag{A4}$$

If the growth rates of wages w and profits r are constant as well (**assumption 2**) then $\frac{\dot{B}}{B}$ is essentially a function of time. Making another change of notation to emphasize this point gives:

$$q = [B(t)]ck^{\beta} \tag{A5}$$

This appears to be a Cobb-Douglas production function in per-unit-labor terms with shift parameter B. Substituting $q = \frac{VA}{L}$ and $k = \frac{K}{L}$ and rearranging:

$$VA = B_t c K^{\beta} L^{1-\beta} \tag{A6}$$

Shaikh's point is that 11 is a Cobb-Douglas production function, but has been derived from an accounting identity under weak (and empirically reasonable) assumptions about the constancy of factor shares in output. A more detailed discussion may be found in Chapter 2 of Felipe and McCombie (2013) including earlier formulations of the argument in a cross-sectional context (in which the relevant assumptions are the constancy of factor shares and growth rates across space).

The equivalence in 6 relies on two assumptions to show the equivalence between an accounting identity and a specific function form. The assumptions are, first, that factor shares of output are constant, and second, that the growth of the wage and profit growth rates is constant. These are empirically plausible, particularly in short panels, and simulation work shows that the argument is fairly robust to violating the former (Felipe and Holz 2001). In fact, removing them does not appear to affect the substance of the argument: these assumptions are simply required to derive a Cobb-Douglas production function from the value-added identity. Different assumptions about the time paths of factor shares and growth rates allow different functional forms to be derived, for example the translog and the CES (Felipe and McCombie 2013: Appendix 2A1). Viewed in this light, model selection in production function estimation (when data obey the accounting identity) may be seen as choosing a representation for these time paths, rather than identifying an aggregate production technology (Felipe and Holz 2001).

Appendix B: Dataset construction

This appendix provides an overview of the case study dataset construction. For a full description of the approach to assessing materials quality, see Filewod and Kant (2021).

Output, capital, and labour

Economic variables were obtained from the World Input-Output Database (WIOD) project (Timmer et al. 2015), which collects and harmonizes data from national statistical institutes and regional aggregates (e.g. Eurostat) for 43 major global economies. Series for output (value added) were deflated to 2010 values using country-specific price indices and then converted to US dollars using exchange rates taken from the Penn World Tables (Feenstra, Inklaar, and Timmer 2015). The WIOD Socio-Economic Accounts (SEA) do not provide price indices for capital stock, so instead a total-economy capital stock deflator from the Penn World Tables (base year 2011) was applied. Data on emissions-relevant (i.e. final) energy use were obtained from WIOD's energy accounts (Corsatea et al. 2019), which reconcile the IEA energy balances with the SNA accounting framework. Compared to the energy balances, the energy accounts allocate energy use to the country where it physically occurs (versus by the country of residence of reporting units) and attribute emissions from the transport sector to associate industries (versus reporting emissions from the transport sector).

While the WIOD offers an attractive combination of sector detail and country coverage, as compared to other multi-regional input-output tables, two limitations should be noted. First, achieving a common sector classification for the years 2001–2014 requires the WIOD SEA to disaggregate or impute input data for several non-European countries. A variety of procedures are used to take full advantage of the available data per country, centred around the use of industry shares in value-added to disaggregate data and linear models of ratios of value-added to other variables (e.g. employment) to impute missing years. Both approaches effectively assume aspects of the production technology (i.e. that ratios of inputs to output follow a constant trend, or that input intensities in disaggregated industries are identical to input intensities in industry aggregates). The ability of panel data estimators to identify mean effects while controlling for country heterogeneity, including mis-measurement, may partially control for violations of these assumptions. To facilitate robustness tests, a data quality indicator was added to indicate the degree of transformation imposed by the WIOD SEA (column 'flag' in Table 2: 1 = EU countries, for which full sector and time coverage is typically available, 2 = AUS, CAN, U.S.A. for which no desegregation is required, 3 = other countries, requiring disaggregation or substantial imputation). Second, the treatment of autoproduction in the WIOD energy accounts does not use a mass balance approach; instead, energy use from wood fuel is divided amongst forestry sectors using monetary shares of forestry output consumed per sector.

Series of industrial roundwood production (summed production of sawlogs or veneer logs, pulpwood, and other industrial roundwood) were obtained from the Food and Agriculture Organization of the United Nations' statistical database (FAOSTAT). This FAO reporting category is preferred to the 'roundwood' reporting category because the inclusion of fuelwood in the latter introduces bias in countries with significant fuelwood production by households.

Materials

The materials input for the forestry and logging sector is economically exploited forest land (i.e. harvested area), which varies in quality on two fundamental dimensions (Filewod and Kant 2021): location and (ecological) site productivity. Both annual area harvested (per country) and its quality can be derived from global geospatial datasets ((Hansen et al. 2013, Simard et al. 2011, Weiss et al. 2018)) based on earth observation satellite data (for location, measured here as accessibility to cities (Weiss et al. 2018)), earth observation data play a secondary role to human-generated data, e.g. Open Street Map). While the satellite-derived loss areas and quality metrics represent a significant advance over previously available data, the identification of harvested area from area lost is noisy. This is of most concern for harvest systems (e.g. single-tree selection, thinning) that do not produce cover loss at 30 m scale. Data were processed and analysed in R (v3.6) and the parallelized, cloud-based Google Earth Engine.

Countries typically do not report annual areas harvested, and identifying the causes of cover loss is a significant challenge. Here, a number of heuristics are applied to the Global Forest Change dataset (Hansen et al. 2013) to convert annual global maps of forest cover loss to annual maps of harvested area. Specifically, loss due to fire is excluded using annual composites of area burned generated from the MODIS 250 m product MCD54A1 v.6. Loss due to canopy dynamics is avoided by imposing a minimum patch size threshold of 11 8-connected pixels ($\sim$ 1 ha). Loss occurring in protected areas is excluded using annual masks of protected areas that prohibit resource extraction (IUCN categories Ia-IV). Finally, loss occurring in densely settled areas is excluded. In addition, two data cleaning steps are implemented using scripts published by Ceccherini et al. (Ceccherini et al. 2020) to calibrate forest areas (and resultant loss detection) against FAO data. First, country-specific forest cover thresholds are applied that minimize differences between year 2015 forest cover and data from the FAO's 2015 Forest Resources Assessment (mean % error: 5.04, median: 0.49). Second, a minimum mapping unit is applied at all processing steps to remove isolated tree patches that fall below common UN size thresholds used to define forests ($\geq$0.5 ha, implemented here by retaining only 30 m forested pixels for which there are 5 or more 30 m pixels within a 100 m square kernel). These data cleaning steps were found to minimally affect estimates of annual area lost.

Of course, this approach will include an unknown amount of cover loss not due to harvest but not excluded by the masking procedure, as well as excluding harvest that does not result in stand-replacing disturbance at the $\sim$30 m scale. Changes in sensor sensitivity and processing algorithms also complicate time-series analysis with the underlying dataset

(Hansen et al. 2013). As a quality control, annual estimates of harvested area were compared against FAO data on industrial roundwood[9] production (Figure A3). Only countries for which Pearson's $r \geq 0$ were retained for analysis; of these, two (CYP, LUX) were excluded because their small forest sectors (mean 2000–2014 production volumes $\leq$ 1 million m^3) do not offer meaningful comparisons against major industrial producers. This quality filtering reduced the number of countries in the dataset from 43 to 28.

To assess the quality of harvested forests, I rely on the von Thünen land use model to motivate the importance of travel costs and site productivity in determining economic rents. These variables can be proxied by wall-to-wall maps of canopy height (Simard et al. 2011) and travel time (Weiss et al. 2018), which are randomly sampled (at the location of forest harvest observations) with a density of one 30 m pixel per 100 ha of harvested area for each country and year (with a minimum sample of 100 pixels per country per year). Random sampling for small countries is distributed over the entire country footprint (FAO GAUL0 administrative boundaries). For large countries (AUS, BRA, CAN, CH, IND, IDN, RUS, U.S.A.), resource limits in Earth Engine preclude both random sampling and annual area loss calculations. For these countries, both calculations are mapped over FAO GAUL2 administrative areas (e.g. counties). This introduces an (essentially negligible) bias due to rounding, since GAUL2 units with <51 harvested pixels ($\sim$4.6 ha at the equator) are not sampled. The joint and marginal sampling distribution for all countries is given in Figure A1.

The resulting data allow summary metrics (means, standard deviations) to be calculated for each quality dimension. Alternatively, harvest observations can be classified by the (economic) class of forest in which they occur. I implement both approaches. Classification is via finite mixture modelling, which treats observations as arising from an underlying data generating process that may be modelled using multiple parametric distributions. An expectation-maximization algorithm is used to select the number of components (model order) as well as their parameters. Gaussian components are most widely used but, as discussed in (Filewod and Kant 2021), theory suggests that skew components should be used when the underlying data generating process is the harvesting decision of firms and observations score on measures of location and productivity (*ceteris paribus*, firms preferentially harvest accessible forests with higher stocking volume). Here, observations are classified using a finite mixture of skew normal distributions (Romulo Barbosa Cabral, Hugo Lachos, and Prates 2012, Oliveira Prates, Rˆomulo Barbosa Cabral, and Hugo Lachos 2013); the choice to use asymmetric components responds to the well-known problem of overestimating model order when using symmetric components to reproduce an irregular density surface. Because components overlapped heavily, model order was selected using the popular (Geoffrey and Rathnayake 2014) Bayesian Information Criterion (BIC) (a widely used alternate measure, Integrated Complete Likelihood, favours well-separated components).

In essence, this approach uses variation (in accessibility-height feature space) in harvested forests to identify distinct data-generating processes (the skew-normal components), which are interpreted as distinct management regimes (corresponding to quality tiers). Figure A1 shows the samplign distribution of the data used for finite mixture modelling, and 3 gives the resulting classification. Components are strongly overlapping, and the maximum-probability assignment of pixels to components is clearly ambiguous for many individual pixels. However, the retrieved components (classes) are readily interpretable. Component 2 (59.4% of sampled pixels) captures harvest occurring close to markets in forests that are likely to be intensively managed (e.g. tree farms, plantations, semi-natural planted forests). Component 1 (33.1%) captures harvest occurring in more remote areas (e.g. semi-natural planted forests, managed natural forests). Component 3 (7.5%) captures harvest occurring in more remote forests with higher standing volume (e.g. frontier forests).

[9]Roundwood Production comprises all wood obtained from removals on public or private land whether round, split, squared or in other forms, including wood recovered following natural loss and wood removed for fuel. Volumes are reported under bark. An alternative measure is Industrial Roundwood, which excludes removals for fuel (i.e. retaining only sawlogs and veneer logs, pulpwood round or split, and other industrial roundwood).

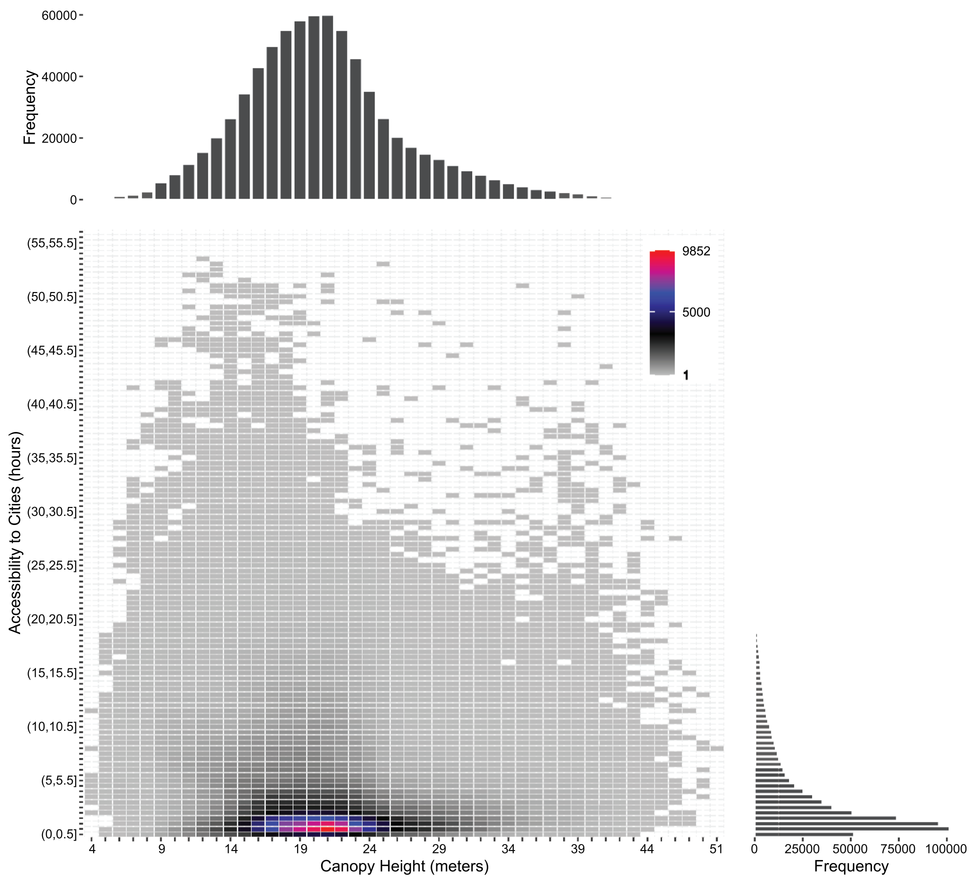

Figure A1. Sampling distribution for height and access data (data for 2001–2014 inclusive, $n = 740{,}165$).

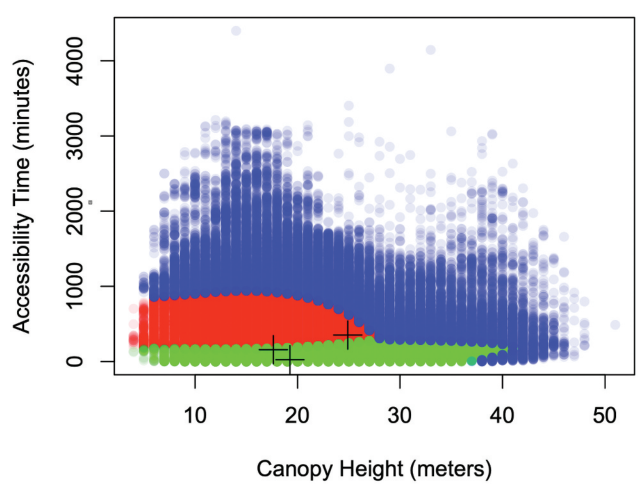

Figure A2. Sample classification using a finite mixture of skew-normal distributions. Colours indicate maximum probability assignment of sampled pixels ($n = 740{,}165$) to components; crosses indicated component means (C1 (μ_{height}, $\mu_{accessibility}$): 17.6, 167.3, C2: 19.3, 24.5, C3: 24.9, 352.4).

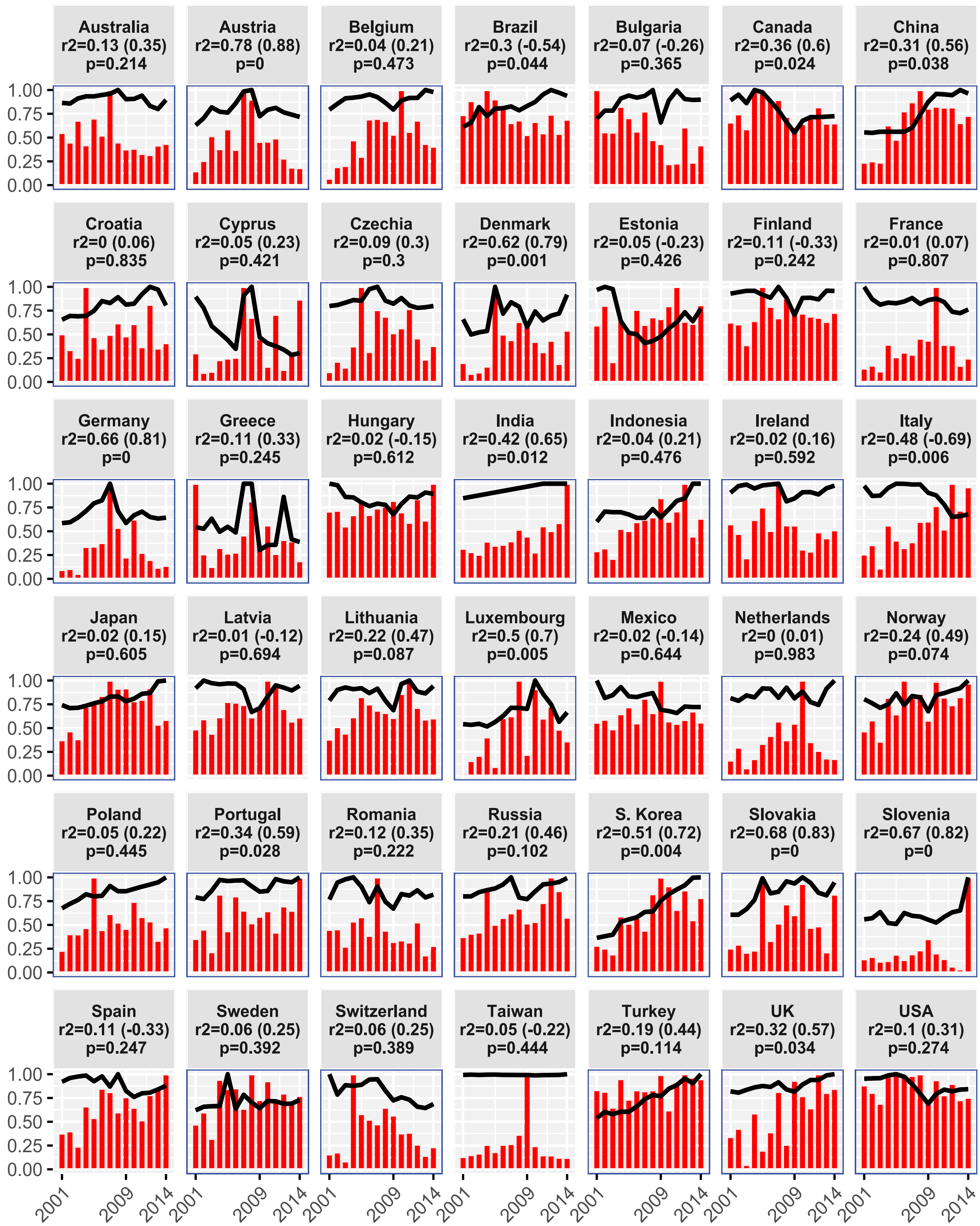

Figure A3. Post-processed global-forest watch forest cover loss (red bars) compared against FAOSTAT roundwood production statistics black lines). Pearson's correlation coefficient r is in brackets (blue boxes highlight countries for which $r \geq 0$) and p is obtained via OLS. Values have been normalized using maximums per series. Figure style follows(Ceccherini et al. 2020), and results presented there (for European countries only) are similar for common years.

Appendix C: Means and standard deviations for case study data

Table A1. Summary statistics for 2001–2014 A02 industry.

	VA (millions)	K (billions)	EMP (1000s)	TJ (TJ)	RWD (m^3)	AREA (1000 ha)	n=	access (hours)	height (m)	C1 (%)	C2 (%)	C3 (%)
AUS	885.7	3.8	11.5	14875.6	25.6	122.9	16980	4.4	18.4	0.4	0.6	0
	(337.4)	*(0.7)*	*(2.8)*	*(3646)*	*(1.5)*	*(45)*		*(8.7)*	*(8.7)*	*(0.1)*	*(0.1)*	*(0.1)*
AUT	1363.6	2.8	25.6	62	13.3	8.6	1583	0.9	26.7	0	1	0
	(221.8)	*(0.6)*	*(1.5)*	*(38.1)*	*(1.7)*	*(4.9)*		*(4.5)*	*(4.5)*	*(0)*	*(0)*	*(0)*
BEL	128.6	0.3	2.4	1817.6	4.2	2.6	1400	0.5	21.5	0	1	0
	(19.2)	*(0.1)*	*(0.1)*	*(1123.5)*	*(0.3)*	*(1.3)*		*(3.9)*	*(3.9)*	*(0)*	*(0)*	*(0)*
CAN	3837.5	2.7	50.8	57119.6	162.9	910.7	127286	6.8	17	0.7	0.2	0.1
	(957.1)	*(0.8)*	*(2.7)*	*(17330.6)*	*(27.6)*	*(164.7)*		*(4.7)*	*(4.7)*	*(0)*	*(0)*	*(0)*
CHE	312.9	0.9	3.6	2913.7	3.7	0.6	1400	0.6	24.6	0	1	0
	(37.7)	*(0.2)*	*(0.4)*	*(1955.4)*	*(0.5)*	*(0.4)*		*(4.5)*	*(4.5)*	*(0)*	*(0)*	*(0)*
CHN	30756.4	614.3	16286.2	303005.1	125.2	332.3	46419	3	23.6	0.2	0.8	0.1
	(5362.6)	*(97.4)*	*(7721.5)*	*(160518.8)*	*(32.7)*	*(129.2)*		*(5.8)*	*(5.8)*	*(0.1)*	*(0.1)*	*(0)*
CZE	971.3	63.6	29.3	5628.6	14.2	4.8	1404	0.5	21.5	0	1	0
	(275.9)	*(8.9)*	*(4.2)*	*(715)*	*(1.1)*	*(2.7)*		*(4.3)*	*(4.3)*	*(0)*	*(0)*	*(0)*
DEU	3024.9	7.1	41.6	8367.9	49.2	17.6	2681	0.6	21.8	0	1	0
	(691.4)	*(0.8)*	*(3.2)*	*(3682.1)*	*(8)*	*(14.5)*		*(4.7)*	*(4.7)*	*(0)*	*(0)*	*(0)*
DNK	268.2	3.2	4.9	714.1	1.6	2.5	1399	1	15.9	0	1	0
	(39.5)	*(0.8)*	*(0.9)*	*(146.3)*	*(0.3)*	*(1.6)*		*(3.5)*	*(3.5)*	*(0)*	*(0)*	*(0)*
FRA	3035.9	7.6	33.4	16536.6	28.2	40.7	5695	0.7	18.8	0	1	0
	(207.2)	*(1.6)*	*(3.7)*	*(4075.6)*	*(2.3)*	*(26.4)*		*(3.5)*	*(3.5)*	*(0)*	*(0)*	*(0)*
GBR	593.8	1.2	16.2	9342.4	8.3	16.4	2449	1	16.1	0	1	0
	(134.2)	*(0.4)*	*(3.1)*	*(3519.4)*	*(0.6)*	*(8.5)*		*(4.6)*	*(4.6)*	*(0)*	*(0)*	*(0)*
HRV	234.5	5.3	11.6	1347.1	3.4	1.9	1400	0.7	22.1	0	1	0
	(52.1)	*(2.1)*	*(1.4)*	*(606.4)*	*(0.5)*	*(0.8)*		*(4.1)*	*(4.1)*	*(0)*	*(0)*	*(0)*
IDN	6324.2	144898.2	2862.5	44204.5	55.2	941.6	131441	3.2	24	0.2	0.7	0.1
	(760.5)	*(88597.3)*	*(334.2)*	*(8705.4)*	*(9.4)*	*(352.6)*		*(7.7)*	*(7.7)*	*(0)*	*(0)*	*(0)*
IND	23579.1	6840.9	23643.3	67213.8	46.5	34.8	4740	1.9	22.7	0.1	0.9	0.1
	(3102.3)	*(3335.7)*	*(971.9)*	*(6823.5)*	*(2.7)*	*(15)*		*(7)*	*(7)*	*(0)*	*(0)*	*(0)*
IRL	312.2	1.4	13.2	2212.3	2.5	4.2	1400	0.7	14.1	0	1	0
	(123.4)	*(0.9)*	*(4.7)*	*(986)*	*(0.2)*	*(1.6)*		*(4.2)*	*(4.2)*	*(0)*	*(0)*	*(0)*
JPN	4487.3	7743.4	95.1	44976.3	17.3	21.4	2989	0.5	25.3	0	1	0
	(975)	*(816.6)*	*(24.1)*	*(11410.2)*	*(1.9)*	*(6.3)*		*(5.2)*	*(5.2)*	*(0)*	*(0)*	*(0)*
KOR	1257	4635	51.5	5931.2	2.8	7.1	1422	0.5	23.6	0	1	0
	(302.6)	*(677.4)*	*(9.5)*	*(2328.8)*	*(0.9)*	*(3)*		*(5.6)*	*(5.6)*	*(0)*	*(0)*	*(0)*
LTU	158.9	0.8	12.6	1259.8	4.7	11	1613	0.9	18.5	0	1	0
	(49.1)	*(0.2)*	*(4.2)*	*(268.5)*	*(0.4)*	*(2.7)*		*(3)*	*(3)*	*(0)*	*(0)*	*(0)*
NOR	837.8	11.3	5.2	1167.8	8.1	24.6	3441	1.4	18.9	0	1	0
	(95.7)	*(1.1)*	*(0.8)*	*(723.9)*	*(0.9)*	*(6.5)*		*(2.8)*	*(2.8)*	*(0)*	*(0)*	*(0)*
POL	1483.4	15.1	59.2	9931.1	30.2	17.5	2479	0.7	18.7	0	1	0
	(221.9)	*(2.7)*	*(8.6)*	*(1457.5)*	*(3.2)*	*(6.4)*		*(3.4)*	*(3.4)*	*(0)*	*(0)*	*(0)*
PRT	903.2	3.3	12	1863.6	9.6	23.7	3328	0.9	16.8	0	1	0
	(108.4)	*(0.8)*	*(0.3)*	*(364.1)*	*(0.8)*	*(8.4)*		*(4.7)*	*(4.7)*	*(0)*	*(0)*	*(0)*
ROU	599.7	15.4	44.4	1323.3	10.7	10.6	1675	1.4	26.5	0	1	0
	(186.1)	*(7.9)*	*(10.6)*	*(757.2)*	*(1.2)*	*(4.9)*		*(5.2)*	*(5.2)*	*(0)*	*(0)*	*(0)*
RUS	0	0	0	0	167.1	1166.1	163247	7.9	19.7	0.6	0.3	0.1
	(0)	*(0)*	*(0)*	*(0)*	*(14.4)*	*(355.6)*		*(3.5)*	*(3.5)*	*(0.1)*	*(0.1)*	*(0.1)*
SVK	505.1	2.9	26.3	1216.6	7.6	3.2	1400	0.8	25.2	0	1	0
	(220.6)	*(0.3)*	*(2.3)*	*(680.8)*	*(1.2)*	*(1.7)*		*(4.9)*	*(4.9)*	*(0)*	*(0)*	*(0)*
SVN	203.2	1.3	5.8	849.7	2.1	0.4	1400	0.6	24.9	0	1	0
	(54.6)	*(0.4)*	*(0.6)*	*(134.9)*	*(0.4)*	*(0.4)*		*(4.6)*	*(4.6)*	*(0)*	*(0)*	*(0)*
SWE	4446.2	75.7	33.1	5342.1	65.4	194	27146	2.2	17.9	0.3	0.7	0
	(1165.6)	*(14.4)*	*(6.8)*	*(1454.3)*	*(8.7)*	*(49.2)*		*(2.4)*	*(2.4)*	*(0)*	*(0)*	*(0)*
TUR	1594	6.9	176.9	5475.7	13.9	18.6	2601	1	18.4	0	1	0
	(149)	*(3.9)*	*(29)*	*(650)*	*(2.8)*	*(2.6)*		*(4.9)*	*(4.9)*	*(0)*	*(0)*	*(0)*
USA	20990.6	32	413.6	91881.1	373.2	1291.2	179747	1.9	21.3	0.1	0.9	0
	(1694.2)	*(3.6)*	*(30.5)*	*(33873.9)*	*(39.2)*	*(173.3)*		*(4.5)*	*(4.5)*	*(0)*	*(0)*	*(0)*