Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No. 2216

September 2026

# Measuring the prevalence and direction of innovation: a pilot approach using LLMs

Prashant Gart
Peter Lambert
Ralf Martin
Denis Medvedev

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Programme on Innovation and Diffusion

Economic and Social Research Council

**Abstract**
This paper pilots a method for measuring firms' product and process innovation at scale. Our approach uses LLMs to complete a detailed innovation survey for a sample of 600,000 companies operating across 40 countries. The preliminary results contain 6 million individual instances of innovative activity along with rich textual descriptions. These responses carry signal, for example a one-standard-deviation increase in a firm's product innovation index is associated with a 5% higher measured TFP residual. The data reveal that innovation is widespread: two-fifths of companies engage in product innovation and one-quarter in process innovation, far exceeding the share of patenting firms. Firms are more likely to have labor-augmenting than labor-saving introductions, while about 14% have green product innovation and 12% have green process innovation. Greenness and net labor augmentation are positively associated across countries. Beyond our specific application, this paper illustrates the potential for using LLMs to generate survey-level insights at population scale.

Keywords: innovation direction; green innovation; automation; LLMs; AI measurement; patents
JEL codes: O30; O31; O33; O34; Q55; J24

This paper was produced as part of the Growth Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We thank seminar participants of LSE POID seminar for helpful comments. This work was supported by the ESRC via the PRINZ project. The paper proposes a new of gathering data with AI tools, AI (OpenAI API) was used for data gathering. We also used AI also in coding the analysis and drafting the paper.
Prashant Garg, Bocconi University. Peter John Lambert, University of Warwick and London School of Economics. Ralf Martin, Imperial College London, CEPR and Centre for Economic Performance at the London School of Economics. Denis Medvedev, World Bank.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

# 1 Introduction

The rate and direction of innovative activities shape the long-run arc of human prosperity (Acemoglu, 2002; Romer, 1990), but the measurement of innovation is itself highly contested. These measurement challenges matter for policy-makers seeking to foster more inclusive, pro-worker, and pro-environmental economic prosperity. In this paper, we offer an approach to the measurement of firm-level innovative activities that can help inform both the prevalence and the direction of innovation. At its highest aspiration, the approach we propose can deliver the precision of a survey instrument at population scale.

Much of the prior research measures innovative activity using patent data for scale and rich textual features (Griliches, 1990), or innovation surveys for representativeness and to cover a broader set of activities (Mairesse & Mohnen, 2010). The former is selective in innovations and companies that seek intellectual property protection, while the latter suffers from standard cost burdens of collecting detailed survey responses. Our approach instead leverages AI tools to construct the Global AI Innovation Survey (GAIIS). By combining information on real firms with the available large-scale pretraining and reasoning capabilities of frontier large language models (LLMs), we were able to ask the models to respond to an innovation survey on behalf of approximately 600,000 companies across 40 countries. For each company in our sample, we repeatedly prompt an LLM with a consistent set of questions about recently introduced products and the adoption of any new processes in-house. This pilot exercise results in a cross-section of nearly 6 million reported product introductions or process changes, along with rich textual detail about each of these innovative activities.

We further classify these activities according to three features of interest to policy-makers: green innovation (Aghion et al., 2016; Popp, 2002), and labor-saving and labor-augmenting innovation (Acemoglu & Restrepo, 2019). Green refers to innovation related to climate mitigation or adaptation. Labor-saving refers to innovation that automates tasks or reduces labor requirements, while labor-augmenting refers to innovation that complements workers or raises their productivity. These labels are assigned to each extracted innovation using a common LLM prompt, with details in Appendix A. These are just a few dimensions we extract.[1]

To assess whether the extracted measures of innovation contain any firm-level economic signal, we merge the LLM-derived measures with balance-sheet and income statement data and run a simple Solow-style production-function regression. This exercise follows an old tradition of relating firm-level innovation measures to measures of a firm's productivity (Crépon et al., 1998). We find that these LLM-based innovation indicators carry information about firm productivity

[1]As part of this study, we will release a database of firm-level innovations that can be used to study a much wider range of innovation phenomena (www.prinzproject.io/innobase).

beyond what is captured by labor, capital, and sector and country controls. Our empirical results show that a one-standard-deviation increase in an LLM-based product innovation indicator is associated with about 5% higher total factor productivity, while a comparable process innovation indicator is associated with about 2.5% higher TFP.

Our measures yield four main findings. First, LLM-based measures suggest dramatically more innovation than patent-based measures: we find evidence of product innovation in over 40% of firms and evidence of process innovation in over 25%, compared with a patenting rate below 4%.

Second, firms are predominantly engaged in labor-augmenting innovation. On the product side this amounts to 24% of firms and on the process side to about 15%. The labor-saving share is roughly one-third of the augmenting share for products and one-half for processes.

Third, like innovation overall, green innovation is under-represented by patents: fewer than 1% of firms have green (Y02) patents. Using our LLM-based innovation indicators, we find evidence of green product innovation in 14% of firms and green process innovation in 12%. LLM-based green firm-shares are positively correlated with green patent shares at the country, sector, and country-sector levels.

Fourth, there is substantial variation in the augmenting bias and the greenness of innovation across countries and sectors. On greenness, South Korea and Japan rank first and second for green product innovation, followed by China. Saudi Arabia and South Korea lead green process innovation. Across sectors, Software & IT comes last on net-augmenting process innovation; even so, slightly more firms there engage in labor-augmenting than in labor-saving innovation.

The ability to identify granular instances of innovation across our sample allows us to measure its prevalence and direction, and for the most part our findings indicate that many of these activities have a pro-worker and pro-environmental leaning. A number of limitations temper the optimism this might engender. First, while we measure millions of instances of innovation, their consequences are not directly measured. Much of the discourse on the future direction of technology is concentrated on large-scale general-purpose technologies such as AI and robotics. Our measures are perhaps better read as the broad patterns that individual firms are deploying at a more incremental and local scale: a revealed measure of everyday innovation, rather than innovation in frontier scientific laboratories. Second, our approach is heavily biased toward firms with a greater online footprint. Countries and sectors where firms maintain less publicly accessible material may look less innovative under this lens, and cross-country comparisons should be read accordingly.

Limitations notwithstanding, this paper offers a snapshot of the many and varied ways firms are improving their efficiency through process innovation or creating greater value for their customers through product innovation. While innovation is our main application, the approach can be used across domains where the information needed to respond to a survey is plausibly found inside the training data of a large frontier language model.

This paper complements a long tradition of measuring innovation using patents and surveys. Griliches (1990) reviews the strengths and limitations of patents as economic indicators. Trajtenberg (1990) uses citation-weighted patent counts to measure innovation value, while Hall et al. (2001) develop the data infrastructure for large-scale patent-citation analysis. Patent data offer scale and rich technological detail, but exclude innovation outside the patent system.

Innovation surveys, guided by the *Oslo Manual* (OECD & Eurostat, 2018), cover product and process innovation beyond patents and formal R&D. Crépon et al. (1998) link survey measures of innovation to R&D and productivity, while Mairesse and Mohnen (2010) review their use in econometric analysis. The World Bank Enterprise Surveys analyzed by Cirera and Maloney (2017) show that firms reporting product or process innovation substantially outnumber those undertaking formal R&D: among innovating firms, fewer than half, and in most countries fewer than 30%, undertake formal R&D. Surveys such as CIS and the World Bank Enterprise Surveys can close important coverage gaps, but are costly, infrequent, and bounded in the detail they collect. Private surveys have also been used to study technological adoption, for example Baksy et al. (2025) use a marketing analytics company's private register of robot adoption. Our approach seeks to combine survey-style measurement with broad coverage and innovation-level textual detail.

Our classifications build on empirical work measuring the environmental and labor direction of innovation. Popp (2002) uses patent data to study energy-saving innovation, while Aghion et al. (2016) distinguish clean from dirty auto patents in a model of directed technical change. Our work also connects to a larger clean-technology patenting literature which measures things like climate-mitigation and renewable-energy related innovation across countries, industries, and firms (Dechezleprêtre et al., 2011, 2013; Noailly, 2012; Noailly & Smeets, 2015). The task-based framework of Acemoglu and Restrepo (2018, 2019) distinguishes technologies that substitute labor from technologies that create or reinstate labor tasks. Empirical work operationalizes this distinction through patent-task similarity (Webb, 2019), classifiers identifying automation patents (Hémous et al., 2025; Mann & Püttmann, 2023), and links between patents, job tasks, and new occupational titles (Autor et al., 2024; Kogan et al., 2023).

Other work infers technology direction from job postings, occupational task data, and administrative data on technology adoption.[2] Cirera et al. (2026) directly measure the technologies

[2] Some examples of such work include the use of job-posting or employee/worker-profiles (Acemoglu, Autor, et al., 2022; Babina et al., 2024); task- or skill-based exposure measures (Brynjolfsson et al., 2018; Eloundou et al., 2024; Felten et al., 2021; Garg et al., 2026) (and the omitted variable issues they can induce, for example Lambert and Schindler (2026)); and administrative or official survey data on technology adoption (Acemoglu, Anderson, et al., 2022; Aghion et al., 2025; Bessen et al., 2025).

establishments use across business functions through establishment surveys, although their focus is technology adoption rather than innovation. These sources are valuable, but usually observe labor demand, occupational exposure, or selected technologies rather than broad firm-level evidence on innovation. Our contribution is to apply common environmental and labor classifications to a broad inventory of product and process innovations, including those outside the patent system.

Finally, our approach relates to the use of text and LLMs for economic measurement. The text-as-data literature develops methods for turning unstructured language into economic variables (Gentzkow et al., 2019). Within the LLM literature, one strand uses models to annotate supplied text (Gilardi et al., 2023; Hansen et al., 2023); another uses them as synthetic respondents, generating answers conditional on demographic or other characteristics (Argyle et al., 2023), and others build relational databases using LLM parsing of text (Fetzer et al., 2024; Garg & Fetzer, 2025). Related work applies this simulation approach to innovation surveys by conditioning on firm characteristics (Park & Yang, 2026). Our aim differs: we use information associated with specific firms, together with the knowledge encoded in an LLM, to construct inventories of their reported product and process innovations.

These uses raise distinct measurement questions. Bisbee et al. (2024) show that matching survey averages need not recover response variation or relationships between variables. Ludwig et al. (2026) distinguish prediction from estimation and emphasize the need for validation data when LLM outputs measure economic concepts. Our contribution is to apply a common survey instrument to firm-specific information at scale, retaining innovation-level descriptions that can be classified and linked to firm outcomes.

The rest of the paper is organized as follows. Section 2 discusses our data and measurement infrastructure. Section 3 presents the production-function exercise. Section 4 explains how we tag innovations as labor-augmenting, labor-saving, or green, and examines their prevalence, variation across countries and sectors, and relationships to one another. Section 5 concludes.

## 2 Data and Measurement

Our register of companies comes from Moody's ORBIS database. We start from a draw of up to 100,000 firms per country for 40 countries, restrict to firms with active status, and keep those with the employment and fixed assets populated. This yields a core sample of about 300,000 firms. To widen coverage beyond firms with complete financials, we add an equally sized random draw of other active Orbis firms in the same countries. The combined set covers 605,830 firms, which we call the full snapshot. The filtering approach leads to an unbalanced sample across countries, and we do not claim that this full snapshot is a stratified sample.

When we aggregate firm-level results to a country or sector, we report unweighted results. For cross-country and cross-sector visualizations, we restrict to countries with 25 or more firms in the regression sample, and to industries with at least 100 firms in the sample. We also use the same country/sector set across the regression-sample and full-snapshot views so that the two visualizations may be directly compared on a common subsample.

To measure innovation, we take information from each firm and prompt a frontier proprietary LLM repeatedly with the same set of innovation-survey-style questions, supplying each firm's name and country and asking it to list the products and processes the firm has recently introduced. The model answers from what it has learned about the firm from public information during training. Each individual introduction returned by the LLM is then classified along three dimensions: whether it is *green* (related to climate-change mitigation or adaptation), whether the labor effect is *augmenting* or *labor-saving* (or unclear), and whether the introduction is on the product or process side. We aggregate up from each individual introduction to firm-level binary tags: a firm is tagged as "green-product" if it has at least one green product introduction, "augmenting-process" if it has at least one labor-augmenting process introduction, and so on. This gives us six firm-level direction tags (green, augmenting, labor-saving, each on the product and process side) plus the union "any green" across the two sides. Appendix A summarizes the extraction and classification protocol.[3]

This approach relies on the foundation LLM having access to company information inside its training corpus. To assess whether this imposes a strong limitation, we run a number of benchmarking exercises. For comparison with the patent record, we use Orbis-attributed patent counts with PATSTAT 2021 Spring as the underlying source and the firm–patent mapping coming from OrbisIP. "Clean" patents are those in CPC class Y02 (climate-mitigation technologies). Patent counts are stocks through approximately 2020. The LLM indicators are constructed using the extraction protocol in Appendix A, which asks for product and process introductions since 2015. The comparison is therefore between patent stocks and model-extracted innovation activity, not matched-period innovation flows.[4]

[3]The prompts are in English, while the underlying training material about firms is multilingual. We do not find systematic differences in extracted innovation incidence between English-language and non-English-language jurisdictions of comparable Orbis coverage.

[4]We restore the green and total patent counts from a pre-winsorization snapshot of the Orbis-attributed counts. Our headline analysis parquet winsorizes patent counts at the regression-sample 99th percentile, which incidentally sets the green-count column to zero for every firm because fewer than 1% of firms have any green patent. The pre-winsorization counts give a usable green-share measure for cross-country comparisons.

# 3 Regression Analysis of Productivity Residuals

We demonstrate that the data we have collected contain an economically meaningful signal by conducting Solow-style total-factor-productivity regressions. That is, we show that indicators constructed from our LLM-collected innovation data are strongly related to residual variation in firms' revenue after controlling for inputs and granular fixed effects. We consider this a parsimonious validation test of our approach.

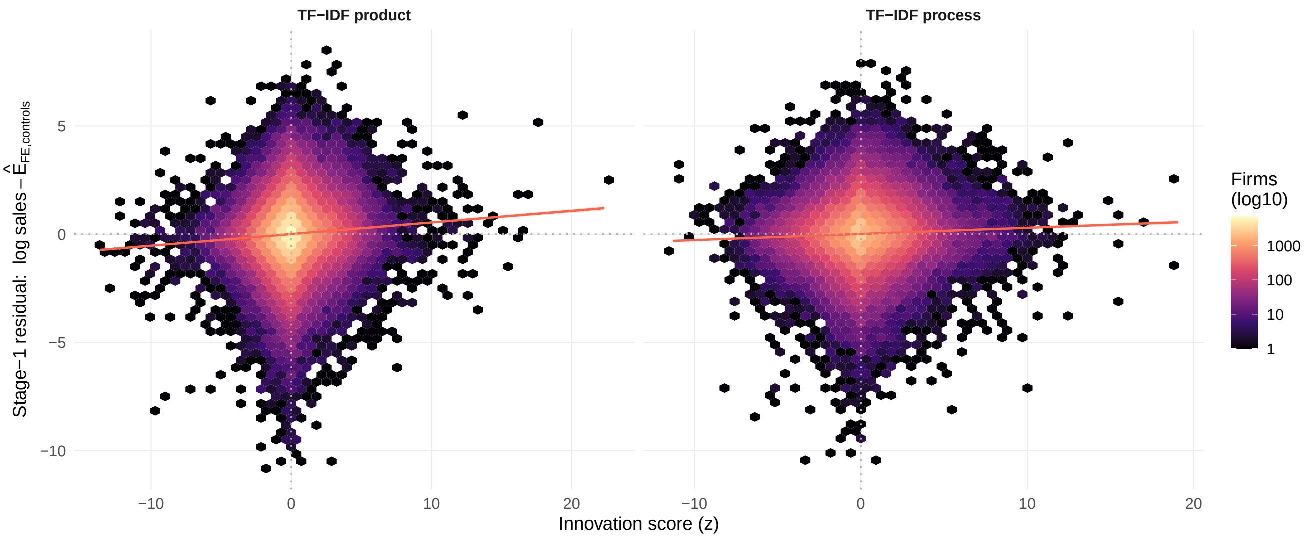
To set up this test, we construct two firm-level innovation scores, one for the product side and one for the process side. Each score answers a single question: how much of a firm's otherwise unexplained revenue can be predicted from the words used to describe its innovations? The outcome we try to predict is a productivity residual. We regress log sales on log employment, log capital, a capital-missing indicator, $\log(1 + \text{official patent count})$, and country$\times$NACE2 fixed effects, and keep what is left over. This residual is the portion of a firm's revenue that its measured inputs, its patent stock, and its country–sector cell do not account for. The predictor is the firm's own innovation text. We represent that text using TF-IDF (term frequency–inverse document frequency) weighting, which we then relate to the residual via a lasso regression.

Fitting the lasso on all firms at once would undermine the test. Each firm's own residual would have helped choose the coefficients later used to score that same firm, so the score would fit the outcome it is asked to explain partly by construction, and a positive correlation between the two would be mechanical rather than informative. We therefore use $K = 3$ cross-fitting: firms are split into three random folds, and each firm is scored by a lasso trained only on the two folds that do not contain it. Every score is thus an out-of-sample prediction from a model that never saw that firm's sales, so any relationship we go on to document has to come from vocabulary–productivity patterns that generalize across firms. Finally, firms for which the LLM returned no text in a given category have a score of exactly zero by construction. In the regressions we control for the potential selectivity of this group by including an indicator for "has any extracted text" as an additional control variable.

In Figure 1 we plot the residualized log sales against each of our cross-fitted innovation scores. Both panels show a positive and significant relationship. This is meaningful per se and does not just follow by construction because the indicators were constructed on a held-out sample, so the reported correlations relate log sales to predictions that never saw them. Appendix Figure F1 reports the fold-averaged lasso coefficients in the form of a word cloud. The terms that show up loaded most heavily are recognizable innovation vocabulary.

Table 1 explores if each indicator holds differential information by including both simultaneously along with other indicators in production-function regressions. Column (1) is the non-text baseline: log employment, log capital, the capital-missing indicator, log(1+official patent count),

**Figure 1:** Stage-1 residualized log sales vs cross-fitted TF-IDF innovation score.




**Notes:** Product (left) and process (right). Hex-bin density of all text-bearing firms; red line is an OLS fit with a 95% confidence band. Stage-1 residualization regresses log sales on log employment, log capital, the capital-missing flag, $\log(1 + \text{official patent count})$, and country$\times$NACE2 fixed effects.

and country$\times$NACE2 fixed effects. Columns (2) and (3) add the cross-fitted product and process innovation scores separately, each with its has-text indicator and the corresponding count regressor (log(1+number of product introductions) in column (2); log(1+number of process introductions) in column (3)). Column (4) includes both scores and both count regressors together. The patent and size controls enter symmetrically in every column. Standard errors are two-way clustered at country and NACE2.

Three patterns stand out from these results. First, we see that both the product and also the process innovation scores hold sizable explanatory power for log sales after conditioning on patents, firm size, and country$\times$NACE2 fixed effects. The LLM-extracted text content is informative about a firm's productivity beyond standard production-function inputs. Second, the process-score coefficient is around half the product-score coefficient in column (4), but remains highly statistically significant. Third, the text-based scores add predictive content beyond the official metrics of patenting intensity. Product and process scores remain positive and precisely estimated, even after controlling for patent counts, while the patent coefficient is positive but imprecisely estimated. In quantitative terms the results imply that a one-standard-deviation higher measure of a firm's product innovation score is associated with about a 5% higher level of TFP.

An alternative check on whether the extracted product/process introductions correspond to real activity can be found in Appendix C, which reports the results of a small audit where, for a stratified sample of 1,200 extracted product/process innovations we attempt to corroborate our LLM innovation data via automated live web access. That is, an agent tries to locate a public

**Table 1:** Production function regressions.

| | (1) Baseline | (2) + Prod score | (3) + Proc score | (4) + Both |
|---|---|---|---|---|
| Product innovation score (z) | | 0.056*** | | 0.053*** |
| | | (0.008) | | (0.008) |
| Process innovation score (z) | | | 0.029*** | 0.024*** |
| | | | (0.004) | (0.004) |
| Has any product text | | 0.023 | | 0.038 |
| | | (0.032) | | (0.039) |
| Has any process text | | | −0.004 | −0.010 |
| | | | (0.028) | (0.036) |
| log(1 + official patent count) | 0.032 | 0.032 | 0.033 | 0.033 |
| | (0.023) | (0.022) | (0.023) | (0.022) |
| log(1 + product introductions) | | −0.015 | | −0.043 |
| | | (0.046) | | (0.056) |
| log(1 + process introductions) | | | 0.031 | 0.039 |
| | | | (0.036) | (0.049) |
| Log employment | 1.015*** | 1.014*** | 1.015*** | 1.014*** |
| | (0.041) | (0.041) | (0.041) | (0.041) |
| Log capital | 0.153*** | 0.152*** | 0.153*** | 0.152*** |
| | (0.032) | (0.032) | (0.032) | (0.032) |
| Capital missing | 1.578*** | 1.573*** | 1.576*** | 1.571*** |
| | (0.348) | (0.346) | (0.347) | (0.346) |
| Num.Obs. | 309 222 | 309 222 | 309 222 | 309 222 |
| R2 | 0.861 | 0.861 | 0.861 | 0.861 |
| R2 Adj. | 0.860 | 0.860 | 0.860 | 0.860 |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is log revenue (log of firm sales). The sample consists of regression-sample firms with positive sales and employment. Standard errors two-way clustered at country and NACE2 are reported in parentheses. The product and process innovation scores are K=3 cross-fitted TF-IDF lasso predictions of log-revenue residuals (after partialling out the controls and fixed effects), constructed as described in Section 3. The scores are zero by construction for firms with no extracted text in the relevant category; the has-text indicators absorb the corresponding level shift between text-bearing and no-text firms.

source of information supporting the innovation claim in our data. A source documenting the specific introduction is found for around one-in-ten items. Process and product innovations can be corroborated from publicly available web searches at roughly the same rate. The probability of a verified online source grows with firm size, from about 7 percent among the smallest firms to 30 percent among the larger entities. The overall corroboration rate as well as the higher rate for larger firms might be considered low. However, we repeat the TFP residual validation exercise for separate size groups (Appendix D) and find little difference between the size groups in how well LLM innovation indicators relate to firm-level productivity. Hence, we conclude that the low corroboration rate is not necessarily a sign of low quality of the extracted data. It could in part reflect that the leading large language models had access to proprietary data in training.[5]

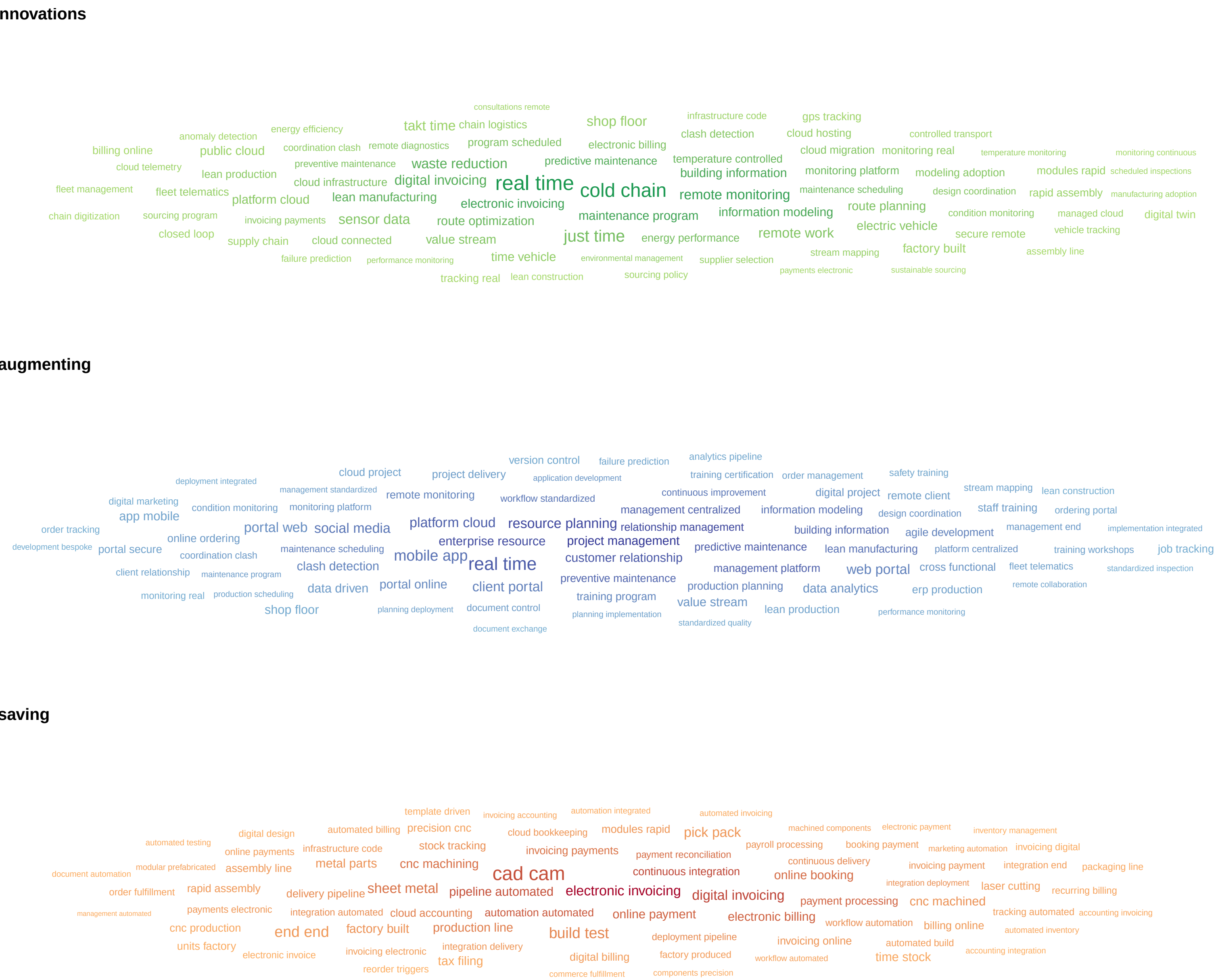
# 4 Direction of Innovation

We now turn to the substantive content of the LLM extraction. We have classified each individual introduction (innovation description) along three direction tags: green, labor-augmenting, and labor-saving. Do those tags pick up genuine direction-of-innovation patterns or simply reflect classifier idiosyncrasies? A useful first sanity check is to ask which words actually populate each tag. Figure 2 reports a contrastive bigram analysis on the $\sim$5.9 million extracted introductions: for each tag, we score bigrams by their log-odds-with-Dirichlet-prior z-score (Monroe et al., 2008), which measures how distinctively a phrase appears in that tag relative to the rest of the corpus. The top-ranked bigrams are then rendered as a word cloud with size and color proportional to the z-score.[6] The three tags read very differently. The green tag is dominated by phrases such as "cold chain", "waste reduction", "electronic invoicing", and "predictive maintenance" — a mix of resource-efficiency and digital-process themes that aligns naturally with what one would expect a green-classified introduction to be about. The labor-augmenting tag features project-management, resource-planning, and training vocabulary; the labor-saving tag features electronic invoicing, CNC machining, and workflow automation. The classifier appears to be picking up real semantic distinctions, not noise.

[5]For instance, OpenAI has licensing agreements with News Corp covering current and archived content from publications including the *Wall Street Journal*, *Barron's*, and *MarketWatch* (OpenAI, 2024b); with the *Financial Times* (OpenAI, 2024a); and with Axel Springer covering otherwise paid content from publications including *Business Insider* and *Politico* (OpenAI, 2023). Archived business reporting of this kind is likely to be particularly informative about individual firms while remaining gated to a public web search.

[6]We make an anonymized version of the full database of all extracted product and process introductions available at www.prinzproject.io/innobase. This is still an early-stage beta, and we plan to continue to refine and improve the product in due course.

**Figure 2:** Contrastive bigrams characterizing the three innovation-themes.

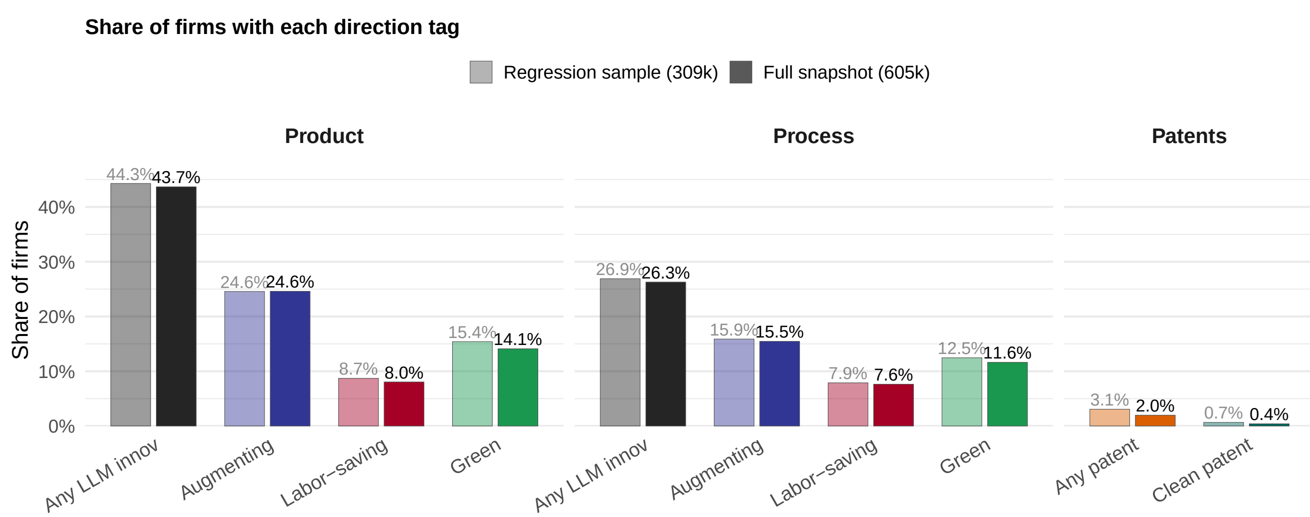



**Notes:** For each slice (green, labor-augmenting, labor-saving) bigrams are scored by their log-odds-with-Dirichlet-prior z-score (Monroe et al., 2008), comparing bigram frequency in the slice to bigram frequency in the rest of the 5.9-million-event corpus. The top-80 bigrams by z-score are shown per panel; word size and color intensity are both proportional to the z-score.

With the tag content sanity-checked, Figure 3 reports the headline composition of firm-level direction tags. The figure has three panels: (i) product, (ii) process, and (iii) patents, which reports the share of firms linked to patents. Within each panel, a faded bar shows the common regression-sample share, and the solid bar reports the shares from the full-snapshot of companies in our pilot study. This shows a rather minor gap of less than half a percentage point in most cases. That said, the regression sample displays systematically a larger (or exactly equal) share across categories. This is not surprising as the regression sample will be skewed to larger companies. However, we tentatively conclude from this comparison that our LLM approach might be able to capture reliable information about companies that are less well represented in conventional datasets.[7]

By contrast, the gap between the LLM-based and patent-based view of innovation is sizable, with roughly 40% of firms having at least one product introduction and over 25% having at least one extracted process introduction, against the share of patenting firms below 4%. Within the LLM-based tags, augmenting introductions are the most common direction tag on both sides. In the full snapshot, the labor-saving firm-share is roughly one-third of the augmenting share for products and one-half for processes. Green tags cover about 14% of firms on the product side and 12% on the process side.

**Figure 3:** Share of firms carrying each direction tag.

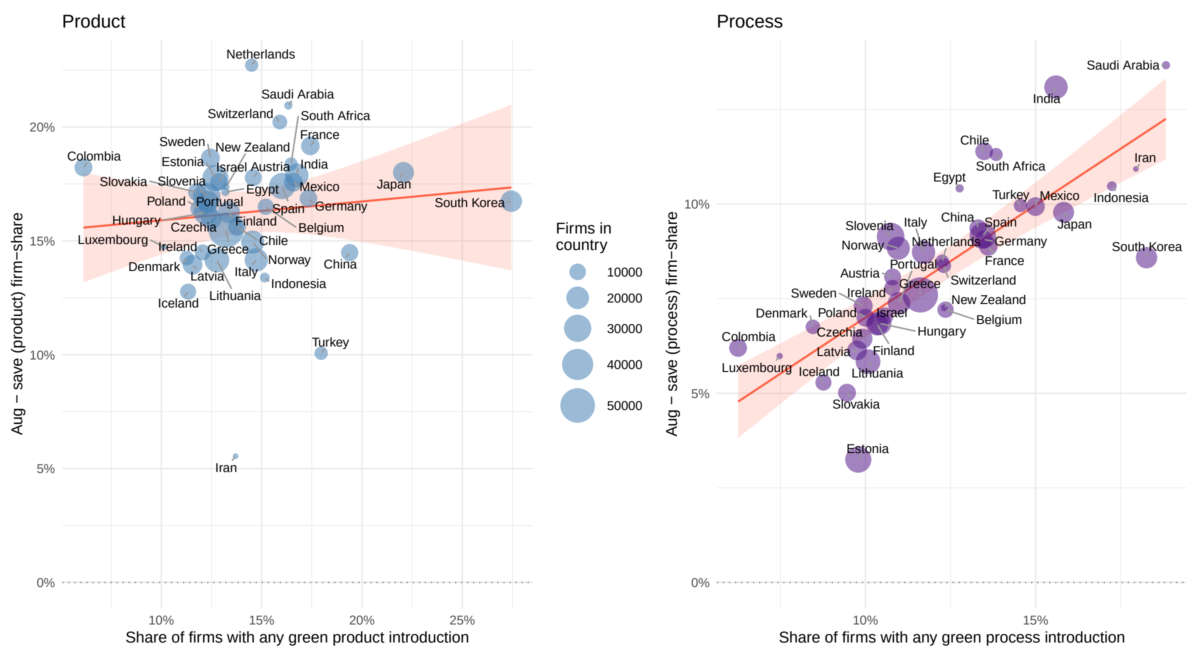



**Notes:** The figure has three panels: "Product" (left) reports the share of firms with any extracted product introduction ("Any LLM innov") alongside the share with a labor-augmenting, labor-saving, or green product introduction; "Process" (middle) reports the corresponding shares on the process side; "Patents" (right, narrower) reports the firm-level share with any official Orbis-attributed patent ("Any patent") and the share with at least one "clean" patent (CPC class Y02, climate-mitigation technologies). Within each tag, the faded bar is the regression sample (309k firms with required financials) and the solid bar is the full 605k snapshot. Both samples include firms with no extracted text. Differences between the bars reflect differences in sample composition.

[7] A potentially interesting exercise we have left for future research is to try to recover basic information about these firms such as size classifications — e.g. an SME indicator — or ownership types.

Does the direction of innovation vary systematically across countries? Figure 4 plots, for each country, the share of firms with at least one green introduction (x-axis) against the net-labor direction — the difference between the augmenting and the labor-saving firm-shares (y-axis) — separately for product introductions (left) and process introductions (right). The sample is the full 605k snapshot, and dot size is total firm count in the country.

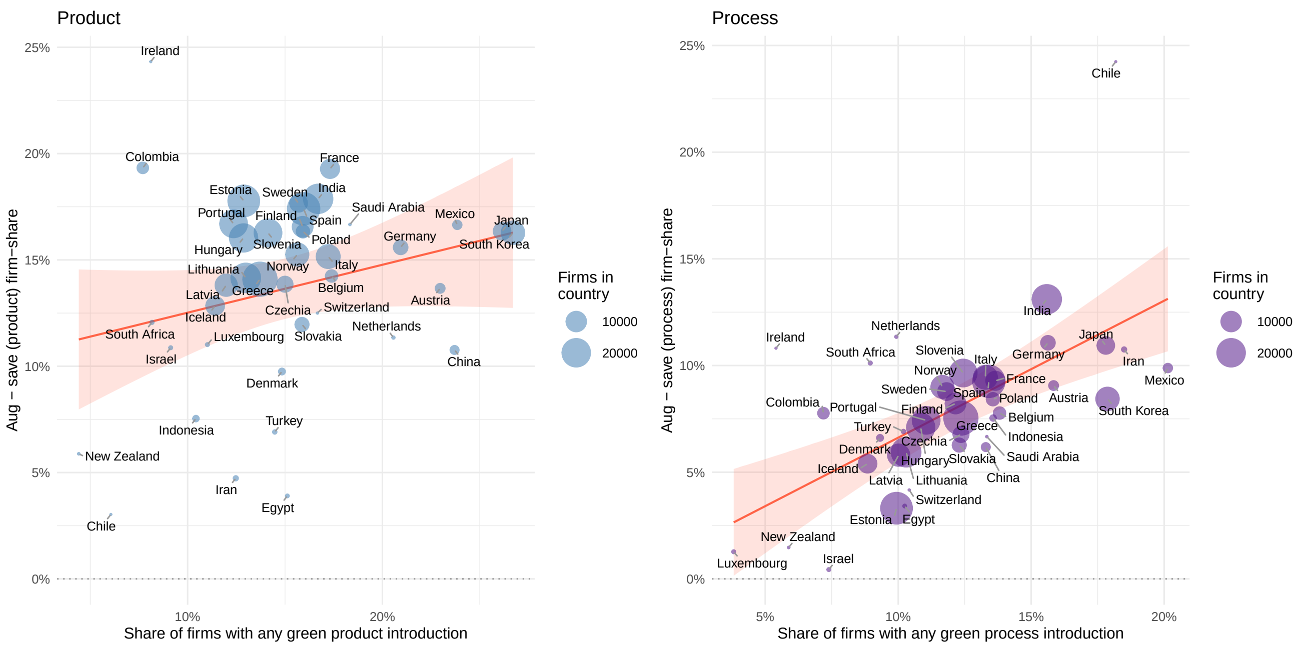
**Figure 4:** Direction of innovation: green firm-share vs net-labor direction (augmenting − saving), product (left) and process (right).




**Notes:** Full 605k snapshot. Country labels use full names; red line is an OLS fit with shaded 95% confidence band. Each dot is a country. The x-axis is the share of firms with at least one extracted green product (left) or process (right) introduction. The y-axis is the difference between the augmenting-firm share and the labor-saving-firm share, where "augmenting" and "labor-saving" are LLM-classifier tags applied to each extracted introduction and aggregated to firm level (the firm carries a tag if it has at least one such introduction). Sample restricted to countries with at least 25 firms in the regression sample. Firms with no extracted text are included in the denominator and contribute zeros to the numerator.

Three patterns emerge. First, there is substantial cross-country variation in green firm-shares. Second, the cross-country relationship between greenness and net-labor direction is positive in both panels — countries with greener product or process introductions also tilt their innovation toward labor-augmenting activities on net. Third, the relationship is tighter on the process side than on the product side. The fact that greenness and labor-augmentation co-move at the country level is, to us, a non-obvious finding; it suggests that the same kinds of firms (or country-level innovation environments) that lean toward climate-relevant innovation also lean toward complementing rather than displacing labor. For comparability, Figure 5 reports the same scatter on the 309k regression-sample firms; the positive relationship is also visible in the regression sample.

**Figure 5:** Direction of innovation: green firm-share vs net-labor direction (augmenting − saving), regression sample.




**Notes:** Product introductions are shown on the left and process introductions on the right. Each dot is a country. The x-axis is the share of regression-sample firms with at least one extracted green product or process introduction. The y-axis is the difference between the share of firms with a labor-augmenting introduction and the share with a labor-saving introduction. The introduction-level tags are aggregated to the firm level, with a firm carrying a tag if it has at least one introduction assigned that tag. The sample comprises approximately 309k firms with the required financial fields and is restricted to countries with at least 25 such firms. Firms with no extracted text remain in the denominator and contribute zeros to the relevant numerators. Country labels use full names; the red line is an OLS fit with a shaded 95% confidence band.

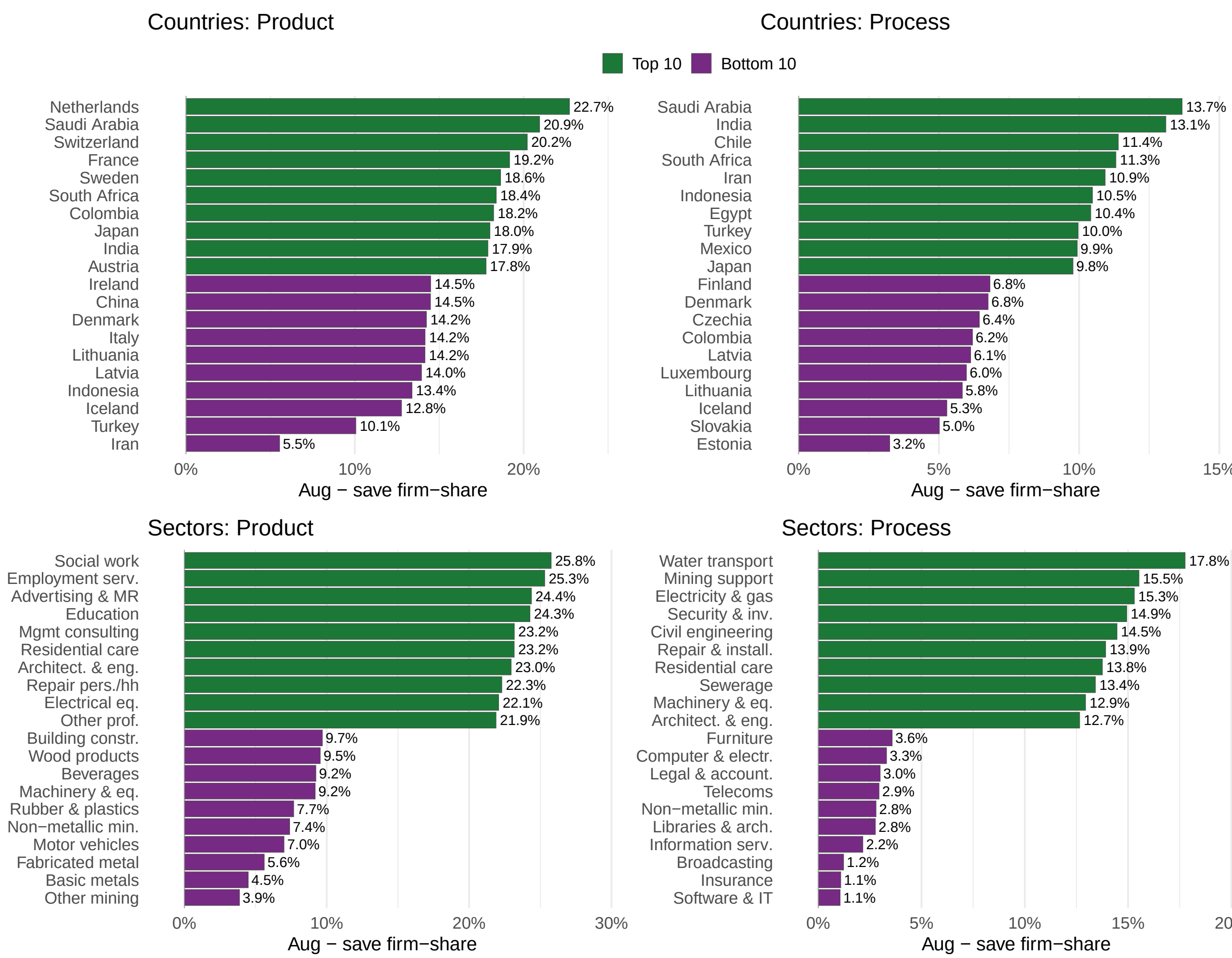
The scatter plot is informative about the cross-country relationship but masks who actually sits at each end of the distribution. Figures 6 and 7 report top-10 / bottom-10 horizontal bar rankings of countries and NACE2 sectors on the net-augmenting firm-share and on the green firm-share respectively. Each figure has four subpanels: countries on the top row, NACE2 sectors on the bottom row; product introductions on the left, process introductions on the right.[8] South Korea and Japan lead green product innovation in the full snapshot, followed by China. Saudi Arabia and South Korea lead green process innovation. On the labor-direction side, Software & IT predictably comes last among sectors on net-augmenting process innovation — perhaps unsurprising given how much of recent process innovation in software is precisely about automating workflows — but, importantly, even there the firm-share engaged in labor-augmenting innovation slightly exceeds the firm-share engaged in labor-saving innovation. The country and sector sets are the same ones used in the body-text scatter plots: 40 countries with at least 25 firms in the regression sample and 86 NACE2 sectors with at least 100 firms in the regression sample.

How much of the cross-country variation in Figures 6 and 7 reflects sectoral composition rather than within-sector country differences? Sector fixed effects explain more cross-firm variation than country fixed effects for every direction tag, with the gap largest on green-product (3.5 versus 0.8 percent) and smallest on augmenting-process (0.6 versus 0.5 percent), matching the wider dispersion of the sector panels (Appendix Table J1). The country rankings themselves are nonetheless stable under sector adjustment.[9] The largest adjustments hit industrial economies whose green and labor-saving tags partly reflect their sectoral mix, with Korea, Japan, Iran, Saudi Arabia, and Turkey appearing repeatedly among the countries most affected by sectoral composition.

[8] All rankings are computed on the full 605k snapshot.

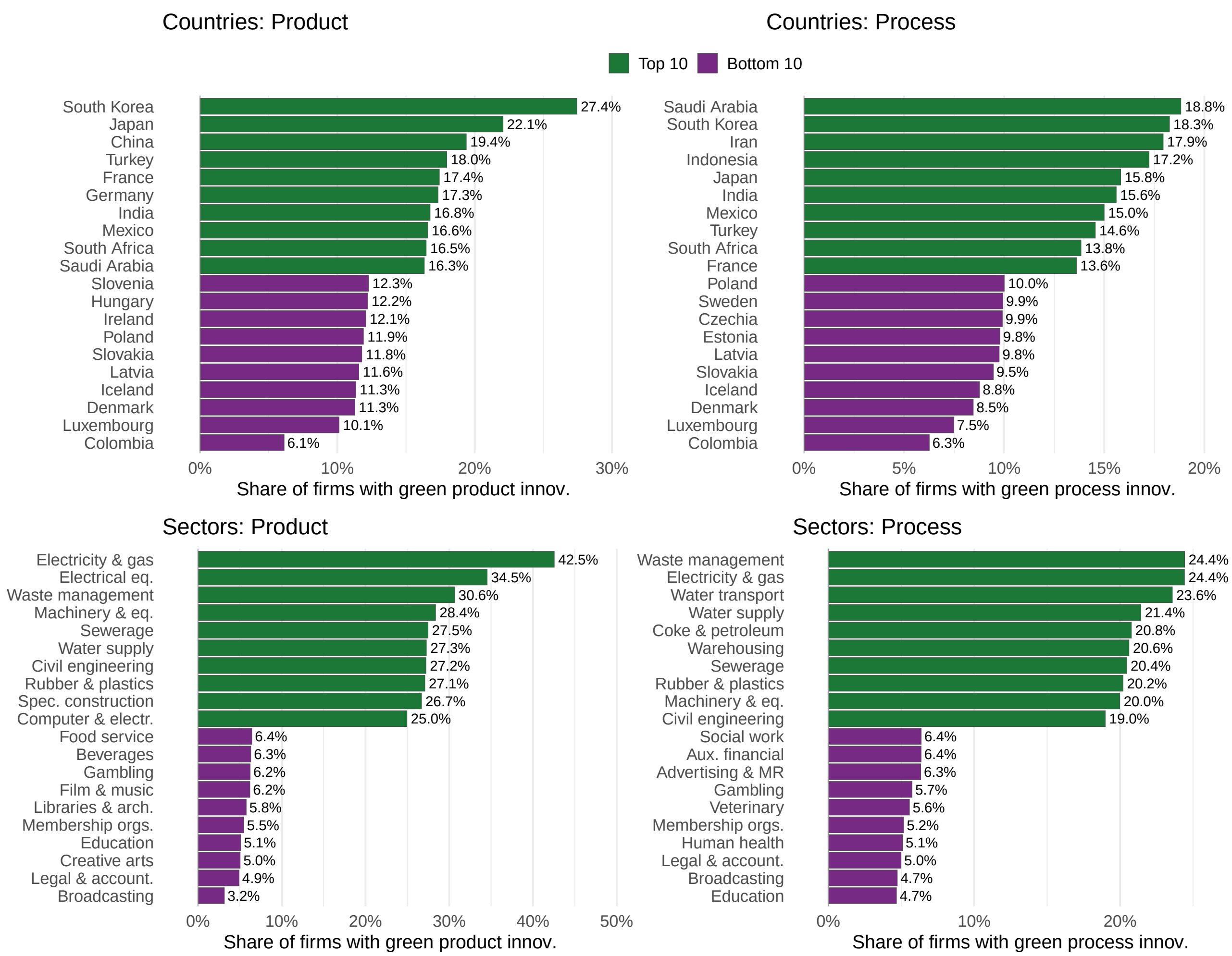
[9] Spearman rank correlations between raw and sector-adjusted country shares run from 0.89 on green-product to 0.995 on augmenting-process; see Appendix Figure J1.

**Figure 6:** Top 10 and bottom 10 countries (top) and NACE2 sectors (bottom) by the net labor augmenting–saving firm-share difference.




**Notes:** Countries appear in the top row and NACE2 sectors in the bottom row. Within each row, product (left) and process (right) panels rank independently. Top 10 in green, bottom 10 in purple. Computed on the full 605k snapshot. The ranking universe contains 40 countries and 86 NACE2 sectors.

**Figure 7:** Top 10 and bottom 10 countries (top) and NACE2 sectors (bottom) by the green innovation firm-share




**Notes:** Countries appear in the top row and NACE2 sectors in the bottom row. Product innovation appears on the left and process innovation on the right. Top 10 in green, bottom 10 in purple. Computed on the full 605k snapshot. The ranking universe contains 40 countries and 86 NACE2 sectors.

## 4.1 Augmenting vs Labor-Saving

The net-labor measure used so far compresses two underlying shares into a single difference, which is convenient for visualizing directional patterns but can hide what is happening to each component on its own. Figure 8 unpacks the difference by plotting the augmenting firm-share against the labor-saving firm-share directly, with a 45-degree reference line indicating equality between the two. Two patterns are visible. First, the two shares co-vary strongly across countries on the product side and even more strongly on the process side: countries that innovate a lot tend to do so on both sides of the labor-direction divide, not exclusively in one. Second, however, the OLS fit lies almost everywhere below the 45-degree line, and the gap is widest at high innovation intensities. Countries with the highest augmenting-share have a substantially smaller labor-saving-share. This is the cross-country counterpart of the headline result in Figure 3: labor-augmenting innovation dominates labor-saving innovation, and the dominance is more pronounced precisely where overall innovation intensity is highest.

**Figure 8:** Augmenting vs labor-saving firm-shares by country, product (left) and process (right).


**Notes:** Full 605k snapshot. The dotted line is 45-degree equality; the red line is an OLS fit with a 95% confidence band. Each dot is a country. Augmenting and labor-saving firm-shares come from the LLM classifier's labor-effect label applied to each extracted introduction, aggregated to firm-level tags. Sample restricted to countries with at least 25 firms in the regression sample. Firms with no extracted text are included in the denominator and contribute zero to both numerators.

A useful external check comes from the Global Automation Atlas (Garg et al., 2026), which uses task-level assessments to construct country–industry measures of the balance between technology that augments labor and technology that substitutes for it. The Atlas therefore measures the same economic direction as our labor labels, but from a different unit of observation:

tasks rather than model-reported product and process introductions. After aligning country coverage and industry classifications, its augmentation-minus-substitution index has a Spearman rank correlation of 0.58 with our net labor-direction measure across 78 two-digit industries ($p < 0.001$); the relationship remains positive for product and process introductions separately and at finer industry definitions (Appendix B). This confirms that the industry patterns in our measure capture labor direction rather than terminology specific to firm announcements; it does not imply equal levels across the two measures or validate individual labels.

## 4.2 Greenness in the LLM Extraction vs Greenness in Patents

A central motivation of this paper is that patents capture only a small fraction of innovation activity. However, does this necessarily imply that patent-derived indicators are misleading for instance when assessing the direction of innovation at the level of countries? We explore this in Figure 9 plotting, for each country, the share of firms with at least one green product or process introduction (x-axis) against the country's green patent share, defined as CPC Y02 patents over total patents (y-axis). The two panels report the same scatter on the regression sample (left) and the full snapshot (right); the country set is identical across the two panels.

We find a positive and statistically significant relationship. Repeating the exercise at the sector and country-by-sector levels (Appendix Figures H5 and I5) leads to the same outcome suggesting the conclusion that patent analysis can provide an indication of innovation direction that is relevant beyond the sample of patenting firms.

**Figure 9:** Green firm-share vs green patent share by country.

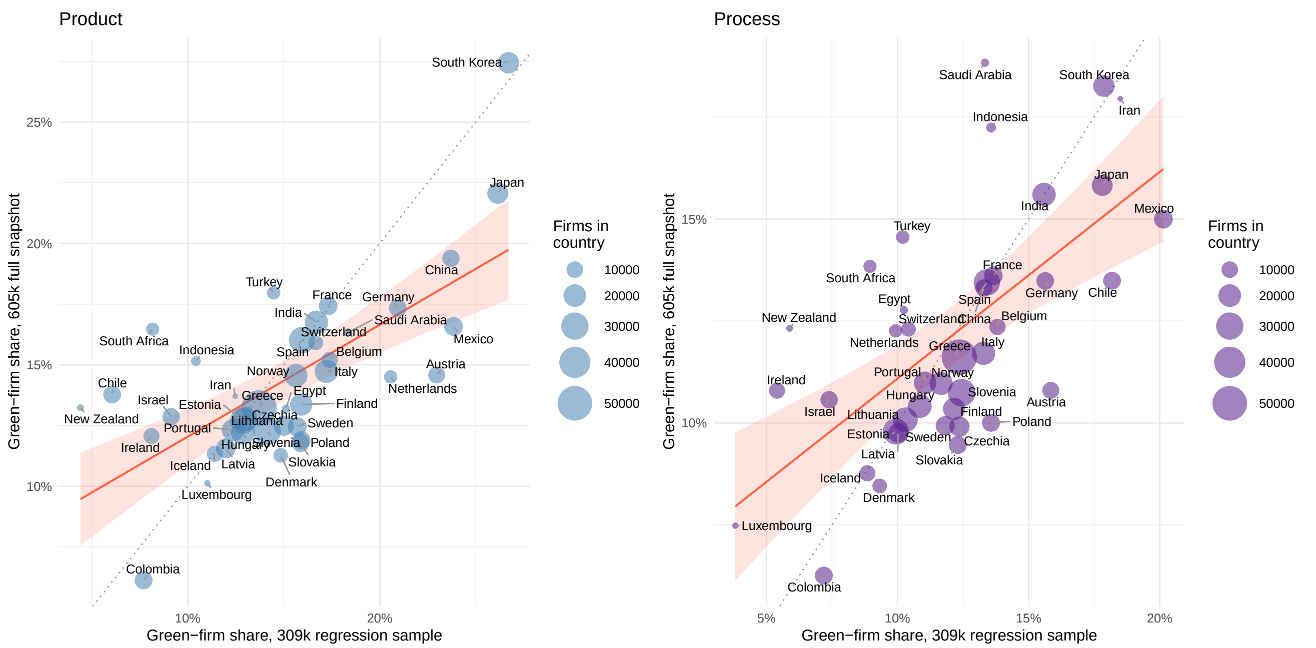

**Notes:** Green firm-share covers any green product or process introduction. The regression sample appears on the left and the full 605k snapshot on the right. Patent counts are the un-winsorized stocks from Orbis (firm-patent mapping via OrbisIP, PATSTAT 2021 Spring). The dotted line marks numerical equality between shares with different denominators; it is not an equal-incidence benchmark. The red line is an OLS fit with a 95% confidence band. Each dot is a country. The x-axis is the share of firms in the country with at least one extracted green product or process introduction. The y-axis is $\sum_f \texttt{off\_patent\_clean\_count}_f / \sum_f \texttt{off\_patent\_count}_f$, summed over firms in the country, with patent counts restored from the pre-winsorization snapshot (the headline analysis parquet zeros out `off_patent_clean_count` via aggressive count winsorization). Restricted to countries with at least 25 patents and at least 25 firms in the regression sample.

## 4.3 Cross-Sample Robustness

We have shown several patterns by reporting the regression sample and the full snapshot side by side, finding consistent patterns. A remaining question is if the coherence between the samples extends to lower levels of aggregation such as individual countries.

We explore this by plotting country-level indicators from both samples against each other. Figure 10 reports the share of firms with green innovation and Figure 11 the difference between firms with augmenting and labor-saving innovations. In both cases we find a strong positive correlation with some heterogeneity in either direction.

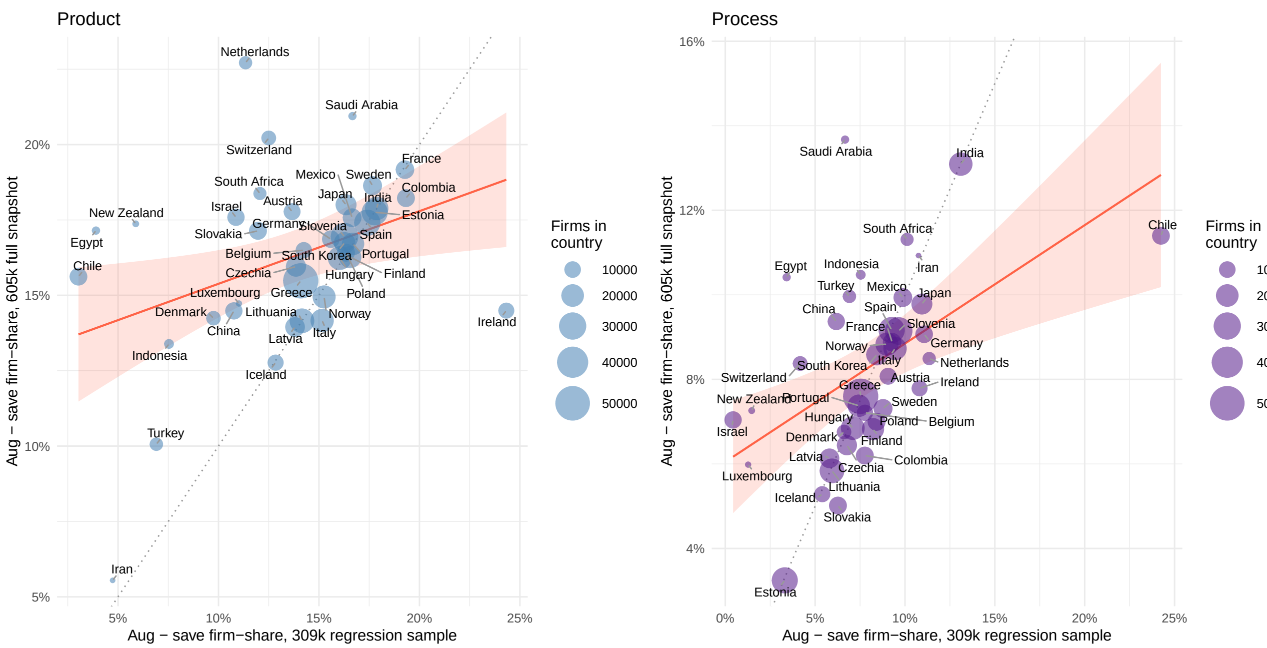
**Figure 10:** Cross-sample comparison of green firm-shares.




**Notes:** The x-axis reports the green firm-share in the 309k regression sample and the y-axis reports the same measure in the 605k full snapshot. Product innovation appears on the left and process innovation on the right. The dotted line is 45-degree equality; the red line is an OLS fit.

A related concern is that the uneven country counts, rather than genuine cross-country variation, drive the direction results, since a few countries contribute far more firms than others. To check this we cap the sample at 20,000 firms per country, drawing at random within each country above the cap and keeping smaller countries in full. The cap binds for nine countries and retains about 538,000 of the 605,000 firms, so it removes the most extreme over-weighting while staying close to the original sample. The results are essentially unchanged. Headline direction shares move by less than half a percentage point; the production-function score coefficients stay positive and significant and shrink only modestly, from 0.053 to 0.052 on the product side and from 0.024 to 0.022 on the process side; the country-level green-versus-green-patent relationship remains positive and significant in both the regression sample and the full snapshot; and country

**Figure 11:** Cross-sample comparison of augmenting–saving firm-shares.




**Notes:** The x-axis reports the augmenting–saving firm-share in the 309k regression sample and the y-axis reports the same measure in the 605k full snapshot. Product innovation appears on the left and process innovation on the right. The dotted line is 45-degree equality; the red line is an OLS fit.

rankings are almost perfectly preserved, with rank correlations of at least 0.999. A more stringent cap at 10,000 firms per country leaves the signs and the main patterns intact but shrinks the validation coefficients by roughly a fifth and renders the full-snapshot country green-patent correlation insignificant, so we treat the 10,000 cap as a stress test and the 20,000 cap as the more representative robustness check. Appendix E reports the full baseline-versus-cap comparison.

# 5 Conclusion

In this pilot study, we set out to measure the prevalence and direction of innovation at the firm level for a much broader sample of firms and countries than patent or survey data alone permit. We do so by prompting an LLM with questions similar to what would be queried with employees of a firm in an innovation survey. If the LLM's training data includes information about the firm, this prompting provides us with a channel to access and structure the training data information. The concern could be that the responses are generic, based on the industry of the firm rather than the firm itself, or that answers could entirely be hallucinated.

To address this we show that the LLM-derived data includes an information signal at the firm level by linking the LLM-derived innovation indicators to independently obtained performance data of these companies and showing a statistically significant and economically meaningful association even after stripping out sector and country effects. We then draw three substantive findings on the innovation behavior of the firms in the sample.

First, LLM-based measures detect substantially more innovation than patent-based indicators. We find evidence of product innovation in more than 40% of firms in our sample and process innovation in more than 25%, compared with patent incidence below 4%. The same gap appears for green innovation: about 14% of firms have a green product introduction, compared with fewer than 1% holding a green patent. Second, firms are predominantly engaged in labor-augmenting rather than labor-saving innovation. This is true on both the product and process sides and, with limited exceptions, across countries and sectors. Third, there is substantial cross-country and cross-sector heterogeneity. Some of this variation accords with prior expectations—Korea and Japan are among the greenest economies, while Software & IT ranks lowest on net-augmenting process innovation—while other patterns are less obvious, most notably the systematic co-movement of greenness and labor augmentation across countries.

The predominance of labor-augmenting innovation is relevant to debates about technological displacement, but it does not necessarily establish the distributional consequences of innovation. Those consequences depend on the economic importance of particular innovations, the firms that develop and adopt them, and how the resulting productivity gains are shared. Our contribution is to make the direction of innovation visible across a much broader set of firms and activities than patent data alone can capture. The exercise shows that LLMs can recover economically meaningful firm-level information at scale and can potentially complement conventional innovation surveys as AI models and underlying information environments evolve.

Several caveats are important. First, our sample is not representative of the underlying population of firms, particularly informal and micro-enterprises. Second, the quality of the LLM responses is constrained both by the information contained in the model's training data and by the model's ability to recover relevant firm-specific facts from its internal representations. Future work will aim to augment our sample with human evaluation. Third, our analysis gives equal weight to firms and to individual innovations. The resulting measures therefore speak primarily to the prevalence and direction of innovative activity across firms, rather than to the economic importance of the technologies themselves. A widely diffused incremental improvement and a transformative general-purpose technology each enter our count-based measures as instances of innovation, even though their aggregate consequences may differ enormously. Our results should therefore be read as a broad picture of the direction of everyday firm-level innovation rather than as a measure of the technological frontier.

The last caveat is that our pilot study produces a single cross-section of innovation activities. We plan to extend the approach over time, allowing us to build a time series of firm-level innovation indicators. This would also support analysis of policy and economic shocks. To date, our pilot suggests that LLM-based measurement can provide a useful complement to existing data on the rate and direction of innovation. We believe such an approach will prove valuable across many contexts where detailed measurement akin to surveys is harder to collect, but where large foundation models have plausibly been exposed to rich information on the respondents of interest.

# References


Acemoglu, D. (2002). Directed Technical Change. *Review of Economic Studies*, *69*(4), 781–809. https://doi.org/10.1111/1467-937X.00226

Acemoglu, D., Anderson, G., Beede, D., Buffington, C., Childress, E., Dinlersoz, E., Foster, L., Goldschlag, N., Haltiwanger, J., Kroff, Z., Restrepo, P., & Zolas, N. (2022, November). *Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey* (NBER Working Paper No. w30659). Cambridge, MA. https://doi.org/10.3386/w30659

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial Intelligence and Jobs: Evidence from Online Vacancies. *Journal of Labor Economics*, *40*(S1), S293–S340. https://doi.org/10.1086/718327

Acemoglu, D., & Restrepo, P. (2018). The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, *108*(6), 1488–1542. https://doi.org/10.1257/aer.20160696

Acemoglu, D., & Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, *33*(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Aghion, P., Bunel, S., Jaravel, X., Mikaelsen, T., Roulet, A., & Søgaard, J. (2025). How Different Uses of AI Shape Labor Demand: Evidence from France. *AEA Papers and Proceedings*, *115*, 62–67. https://doi.org/10.1257/pandp.20251047

Aghion, P., Dechezleprêtre, A., Hémous, D., Martin, R., & Van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, *124*(1), 1–51. https://doi.org/10.1086/684581

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, *31*(3), 337–351. https://doi.org/10.1017/pan.2023.2

Autor, D., Chin, C., Salomons, A., & Seegmiller, B. (2024). New Frontiers: The Origins and Content of New Work, 1940–2018. *The Quarterly Journal of Economics*, *139*(3), 1399–1465. https://doi.org/10.1093/qje/qjae008

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*, 103745. https://doi.org/10.1016/j.jfineco.2023.103745

Baksy, A., Chandler, D., & Lambert, P. J. (2025). *Anatomy of automation: CNC machines and industrial robots in UK manufacturing, 2005-2023* (CEP Discussion Paper No. 2131). London.

Bessen, J., Goos, M., Salomons, A., & Van Den Berge, W. (2025). What Happens to Workers at Firms that Automate? *Review of Economics and Statistics*, *107*(1), 125–141. https://doi.org/10.1162/rest_a_01284

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., & Larson, J. M. (2024). Synthetic replacements for human survey data? The perils of large language models. *Political Analysis*, *32*(4), 401–416. https://doi.org/10.1017/pan.2024.5

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What Can Machines Learn and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, *108*, 43–47. https://doi.org/10.1257/pandp.20181019

Cirera, X., Comin, D., & Cruz, M. (2026). Technology Sophistication Across Establishments. *The Quarterly Journal of Economics*, *141*(3), 2025–2085. https://doi.org/10.1093/qje/qjag018

Cirera, X., & Maloney, W. F. (2017). *The innovation paradox: Developing-country capabilities and the unrealized promise of technological catch-up*. World Bank Publications.

Crépon, B., Duguet, E., & Mairesse, J. (1998). Research, innovation and productivity: An econometric analysis at the firm level. *Economics of Innovation and New Technology*, *7*(2), 115–158. https://doi.org/10.1080/10438599800000031

Dechezleprêtre, A., Glachant, M., Haščič, I., Johnstone, N., & Ménière, Y. (2011). Invention and transfer of climate change–mitigation technologies: A global analysis. *Review of Environmental Economics and Policy*, *5*(1), 109–130.

Dechezleprêtre, A., Glachant, M., & Ménière, Y. (2013). What drives the international transfer of climate change mitigation technologies? Empirical evidence from patent data. *Environmental and Resource Economics*, *54*(2), 161–178.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, *384*(6702), 1306–1308. https://doi.org/10.1126/science.adj0998

Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, *42*(12), 2195–2217. https://doi.org/10.1002/smj.3286

Fetzer, T., Lambert, P. J., Feld, B., & Garg, P. (2024). *AI-generated production networks: Measurement and applications to global trade* (CEPR Discussion Paper No. No. 19708).

Garg, P., Crosta, T., & Baier, J. (2026, July 22). *Global Automation Atlas* (2). https://doi.org/10.48550/ARXIV.2605.17086

Garg, P., & Fetzer, T. (2025, January 12). *Causal Claims in Economics* (arXiv No. 2501.06873). https://doi.org/10.48550/arXiv.2501.06873

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, *120*(30), e2305016120. https://doi.org/10.1073/pnas.2305016120

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, *28*(4), 1661–1707. https://www.jstor.org/stable/2727442

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The NBER patent citation data file: Lessons, insights and methodological tools* (Working Paper No. 8498). National Bureau of Economic Research. https://doi.org/10.3386/w8498

Hansen, S., Lambert, P. J., Bloom, N., Davis, S., Sadun, R., & Taska, B. (2023, March). *Remote Work across Jobs, Companies, and Space* (NBER Working Paper No. w31007). Cambridge, MA. https://doi.org/10.3386/w31007

Hémous, D., Olsen, M., Zanella, C., & Dechezleprêtre, A. (2025). Induced Automation Innovation: Evidence from Firm-Level Patent Data. *Journal of Political Economy*, *133*(6), 1975–2028. https://doi.org/10.1086/734778

Kogan, L., Papanikolaou, D., Schmidt, L. D., & Seegmiller, B. (2023, November). *Technology and Labor Displacement: Evidence from Linking Patents with Worker-Level Data* (NBER Working Paper No. w31846). Cambridge, MA. https://doi.org/10.3386/w31846

Lambert, P. J., & Schindler, Y. (2026). *The Broken Ladder: AI, Remote Work, and Early-Career Hiring*. https://doi.org/10.2139/ssrn.6787638

Ludwig, J., Mullainathan, S., & Rambachan, A. (2026). Large language models: An applied econometric framework. *Annual Review of Economics*, *18*, 283–316. https://doi.org/10.1146/annurev-economics-120925-105620

Mairesse, J., & Mohnen, P. (2010). Using innovation surveys for econometric analysis. In *Handbook of the Economics of Innovation* (pp. 1129–1155, Vol. 2). Elsevier. https://doi.org/10.1016/s0169-7218(10)02010-1

Mann, K., & Püttmann, L. (2023). Benign Effects of Automation: New Evidence from Patent Texts. *Review of Economics and Statistics*, *105*(3), 562–579. https://doi.org/10.1162/rest_a_01083

Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2008). Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis*, *16*(4), 372–403. https://doi.org/10.1093/pan/mpn018

Noailly, J. (2012). Improving the energy efficiency of buildings: The impact of environmental policy on technological innovation. *Energy Economics*, *34*(3), 795–806.

Noailly, J., & Smeets, R. (2015). Directing technical change from fossil-fuel to renewable energy innovation: An application using firm-level patent data. *Journal of Environmental Economics and Management*, *72*, 15–37.

OECD & Eurostat. (2018). *Oslo Manual 2018: Guidelines for collecting, reporting and using data on innovation* (4th ed.). https://doi.org/10.1787/9789264304604-en

OpenAI. (2023, December 13). *Partnership with axel springer to deepen beneficial use of AI in journalism*. Retrieved July 14, 2026, from https://openai.com/index/axel-springer-partnership/

OpenAI. (2024a, April 29). *We're bringing the financial times' world-class journalism to ChatGPT*. Retrieved July 14, 2026, from https://openai.com/index/content-partnership-with-financial-times/

OpenAI. (2024b, May 22). *A landmark multi-year global partnership with news corp*. Retrieved July 14, 2026, from https://openai.com/index/news-corp-and-openai-sign-landmark-multi-year-global-partnership/

Park, J. H., & Yang, J.-S. (2026). LLMs as complementary tools for innovation surveys research: Pattern replication and contextual relevance. *Scientometrics*. https://doi.org/10.1007/s11192-026-05571-8

Popp, D. (2002). Induced innovation and energy prices. *American Economic Review*, *92*(1), 160–180.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, *98*(5), S71–S102. https://doi.org/10.1086/261725

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *RAND Journal of Economics*, *21*(1), 172–187. https://doi.org/10.2307/2555502

Webb, M. (2019). *The Impact of Artificial Intelligence on the Labor Market* (SSRN Working Paper No. 3482150). https://doi.org/10.2139/ssrn.3482150

# Online Appendix

# A LLM extraction and direction classification protocol

This appendix summarizes how we turn public firm information into the product, process, green, labor-augmenting, and labor-saving measures used in the paper. The same procedure is applied to every firm.

For each firm, we prompt `gpt-5-mini` with the firm's name and country and ask it to identify products and processes the firm has introduced since 2015. These two fields are the model's only input: we supply no retrieved web text, no website content, and no Orbis attributes. The model's response therefore reflects what it has learned about the firm from public information during training, conditional on the name and country provided.

The extraction prompt asks the model to identify concrete product and process introductions and to avoid generic inference from the firm's industry. The core instruction is:

> Read the information about the firm. Identify products and processes that the firm has recently introduced, launched, adopted, or substantially upgraded. Return only concrete introductions supported by the supplied information. Do not infer innovations from the firm's industry alone. Separate products from processes. For each item, provide a short description, and the introduction year if available.

The model returns a structured record with separate product and process entries. The main fields are:

| Field | Description |
|---|---|
| `name` | Official product or service name (proper noun). |
| `general_description` | Concise, info-dense description of the introduction; brand-agnostic. |
| `release_year` | Introduction or release year (`YYYY`, 2015 onward; `NA` if unknown). |
| *Firm-level fields:* | |
| `sector` | Short sector label (2–5 words), e.g. "consumer electronics". |
| `employees_last_known` | Bucket: `1-5`, `6-20`, `21-100`, `101+`, or `NA`. |
| `hq_city` | Primary city where the firm is headquartered. |

We classify an item as a product introduction when it is a new or substantially upgraded good or service offered by the firm. We classify an item as a process introduction when it is a new or substantially upgraded production, logistics, delivery, digital, or internal operating process used by the firm. Malformed records and duplicate records are removed in post-processing.

Each extracted introduction is then classified along the direction dimensions used in the paper. The core classification instruction is:

> Classify the introduction along the following dimensions. First, is it green, meaning related to climate-change mitigation or adaptation, energy efficiency, emissions reduction, renewable energy, recycling, or related environmental objectives? Second, what is its likely labor direction: labor-saving, labor-augmenting, or unclear? Use labor-saving when the introduction substitutes for labor, automates tasks, reduces labor requirements, or replaces worker activities. Use labor-augmenting when the introduction complements workers, raises worker productivity, improves decision-making, creates new tasks, or expands the scope or quality of work performed by labor. Use unclear when the description does not provide enough evidence.

The classification output is:

| Field | Allowed values / description |
|---|---|
| `green` | Yes, no, or unclear. |
| `labor_direction` | Labor-augmenting, labor-saving, or unclear. |
| `classification_reason` | Short explanation for the assigned labels. |

The main analysis aggregates these introduction-level records to firm-level indicators. A firm is coded as having green product innovation if at least one extracted product introduction is classified as green. A firm is coded as having labor-augmenting process innovation if at least one extracted process introduction is classified as labor-augmenting. We construct the same indicators for green, labor-augmenting, and labor-saving introductions separately for products and processes. Country, sector, and country-by-sector measures are firm-weighted shares of firms carrying each tag.

Firms with no extracted introduction are treated as zeros in the firm-share measures. These zeros mean that the model returned no relevant introduction under this protocol. They should not be read as definitive evidence that the firm undertook no innovation of any kind.

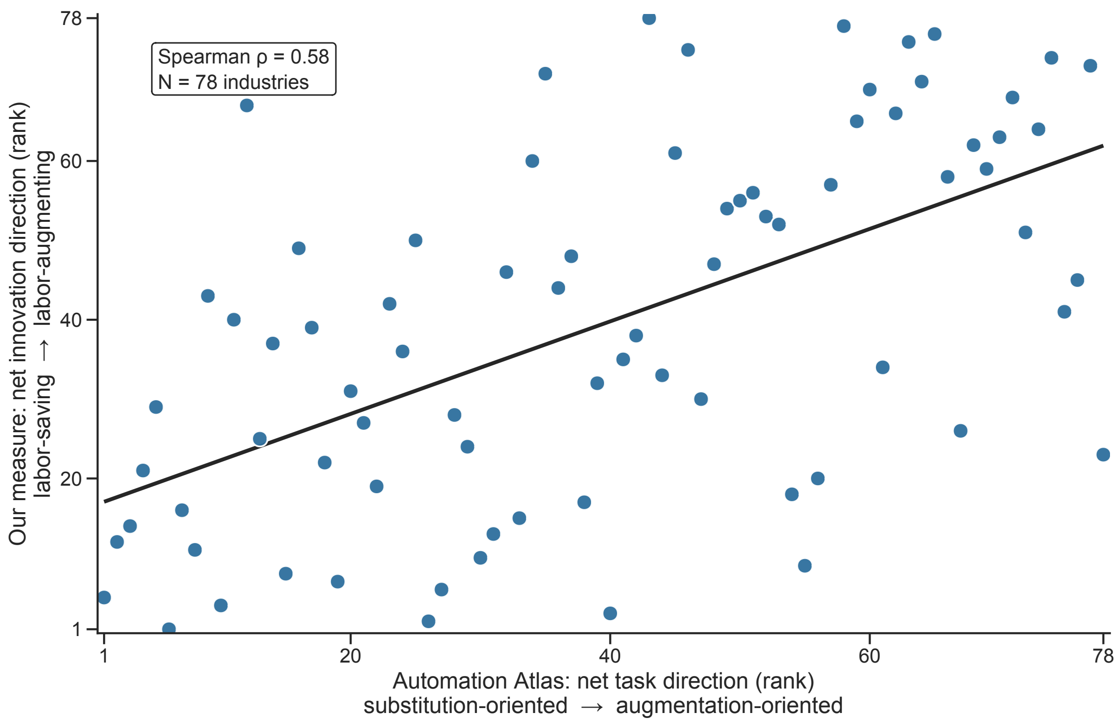
# B External benchmark for labor direction

This appendix compares the labor direction implied by the extracted firm-level innovation records with the task-based augmentation–substitution measure in the Global Automation Atlas (Garg et al., 2026). The two measures are constructed from different underlying information and have different numerical scales: our measure records whether firms have at least one product or process introduction classified as labor-augmenting or labor-saving, whereas the Atlas aggregates task-level assessments. The exercise therefore compares industry rankings, not levels.

We map the firms' NACE4 industries to ISIC4 and retain country–industry cells observed in both sources. For each firm, the primary direction measure is an indicator for at least one labor-augmenting product or process minus an indicator for at least one labor-saving product or process. The Atlas counterpart is its weighted augmentation share minus weighted substitution share. Both measures are averaged over the same matched country–ISIC4 firm observations and then collapsed to the stated industry level; every matched firm receives equal weight. We retain industries represented by at least 200 firms.

Figure B1 displays the two-digit comparison. Industries that rank as more augmentation-oriented in the Atlas also rank as more labor-augmenting in the firm-level innovation data. Table B1 shows that this association is positive for product and process introductions separately and remains positive at finer industry definitions. This is convergent evidence across a firm-announcement measure and a task-based benchmark; it neither compares their levels nor validates individual labels.

**Figure B1:** Industry rankings of net labor direction.




**Notes:** Each dot is a two-digit industry. The horizontal axis ranks industries by the Automation Atlas task-based augmentation-minus-substitution measure; the vertical axis ranks them by the corresponding firm-level product-or-process measure. The black line is an unweighted linear fit shown only as a visual guide. The Spearman statistic compares ranks because the two measures have different scales.

**Table B1:** Convergent comparison of labor direction with Automation Atlas.

| Industry detail | Industries | Product | Process | Any product or process |
|---|---|---|---|---|
| 2-digit | 78 | 0.52 | 0.36 | 0.58 |
| 3-digit | 184 | 0.47 | 0.30 | 0.52 |
| 4-digit | 270 | 0.40 | 0.29 | 0.45 |

**Notes:** Entries are Spearman rank correlations between the firm-level net labor-direction measure and Automation Atlas's task-based augmentation-minus-substitution measure. The product and process columns use the corresponding firm tags. The final column assigns each firm its product-or-process union. Both measures are aggregated over the same matched country–ISIC4 firm observations before collapse to the stated ISIC industry detail; cells with fewer than 200 firms are excluded. All reported correlations are positive and statistically significant at the 1% level.

# C External verification of extracted introductions against public web sources

This appendix examines the extent to which the extracted introductions can be corroborated in publicly available web sources, and where such corroborating evidence is located. We audit a sample of 1,200 extracted introductions (product and process), stratified by innovation direction and weighted toward small firms in line with the underlying sample. For each introduction, a language model with web-search access attempts to locate a public web page that documents the specific introduction. The audit is conducted twice over the same items, under two mutually exclusive source restrictions: in the first arm the model may cite only the firm's own website; in the second it may cite only third-party sources, with the firm's own site excluded. An automated compliance check, based on the firm's Orbis-recorded web domain together with a firm-name heuristic, indicates that the restrictions were respected in 96% of website-arm and 97% of third-party-arm verifications, so the two channels can be compared symmetrically.

We classify an introduction as verified when the located page describes the named product or process, whether with its introduction year, without a year, or in reworded form. A page that confirms only that the firm exists and operates in the stated area does not meet this standard; we report this weaker existence-level corroboration separately.

Table C1 reports the results. Three findings emerge. First, 9.7% of the audited introductions can be matched to a public page documenting the specific product or process. Second, the firm's own website and third-party sources corroborate at nearly identical rates (6.2% and 6.4%) but exhibit limited overlap: of the 116 verified introductions, 39 are documented only on the firm's own site, 41 only in third-party sources, and 36 in both. The two channels are therefore complementary rather than substitutable — restricting the search to either channel alone would omit roughly one third of the publicly verifiable introductions, which implies in particular that a verification strategy confined to company websites would capture only part of the available evidence. Third, corroborating the firm's existence and field of activity is substantially more common than corroborating the specific introduction (32.5% of items via third-party directories

**Table C1:** Verification of extracted introductions against public web sources, by channel.

| | Items | Share of sample |
|---|---|---|
| *Specific introduction documented (N = 1,200):* | | |
| Firm's own website | 75 | 6.2% |
| Third-party sources | 77 | 6.4% |
| Either channel | 116 | 9.7% |
| *Decomposition of the 116 verified items:* | | |
| Firm website only | 39 | 3.2% |
| Both channels | 36 | 3.0% |
| Third-party only | 41 | 3.4% |
| *Memo: firm existence confirmed (weaker standard):* | | |
| Firm's own website | | 10.4% |
| Third-party sources | | 32.5% |

**Notes:** "Specific introduction documented" pools exact (item and year), item-without-year, and reworded matches to the named product or process. "Firm existence confirmed" means the source confirms the firm operates in the stated area without documenting the specific introduction; it is dominated by business directories and company registries. Each restricted arm is run over the same 1,200 items; compliance with the source restriction is 96–97%.

and registries, against 9.7% for the introduction itself). The two constitute different evidentiary standards, and only the latter bears on the content of the extracted measure.

**Table C2:** Specific-introduction verification by firm size.

| Employee band | Items | Verified (either channel) |
|---|---|---|
| 1–5 employees | 685 | 6.6% |
| 6–20 employees | 236 | 10.6% |
| 21–100 employees | 130 | 16.9% |
| 101–1,000 employees | 54 | 29.6% |

**Notes:** Verification to the specific introduction, either channel, by Orbis employee band; 95 items without an employee bucket are omitted. Under the weaker any-source standard (including existence-only confirmation), the corresponding rates are 28.3%, 40.7%, 42.3%, and 61.1%.

Verification rates increase strongly with firm size (Table C2), from 6.6% among firms with at most five employees to 29.6% in the 101–1,000 employee band, consistent with larger firms maintaining a more extensive public web presence. Rates are similarly higher in countries with denser firm-level web coverage: under the weaker any-source standard, third-party corroboration falls from 38.6% in high-coverage countries to 20.7% in low-coverage ones, and own-website corroboration from 13.6% to 4.5%. Non-verification is more common among small firms, a pattern consistent with limited public visibility; the audit does not establish why any particular introduction could not be corroborated.

Table C3 reports the distribution of source types among verified introductions. On the firm's own website, corroboration is drawn from general, product, service, and press pages; among third-party sources, trade and news coverage accounts for the largest share, followed

**Table C3:** Source types among verified introductions, by channel.

| Firm's own website (75) | | Third-party sources (77) | |
|---|---|---|---|
| General firm website page | 63 | Trade / news article | 40 |
| Service / solution page | 5 | Business directory | 19 |
| Product page | 5 | Marketplace / listing | 8 |
| Press / news page | 2 | Press release (wire) | 4 |
| | | LinkedIn | 4 |
| | | Review / software platform | 2 |

**Notes:** Source-type counts among introductions verified to the specific item, by channel. Counts sum to the channel totals in Table C1.

by business directories, marketplace listings, wire press releases, and professional networking pages. Table C4 presents illustrative verified introductions from each channel, with links to the documenting pages.

**Table C4:** Examples of extracted introductions verified in public sources.

| Firm (country) | Extracted introduction (year) | Source |
|---|---|---|
| *Verified on the firm's own website:* | | |
| Siemens Limited (India) | Product: metro and rail signaling packages, CBTC and interlocking (2016) | press.siemens.com |
| Hewlett-Packard Norge AS (Norway) | Product: HP Spectre laptop family (2015) | hp.com |
| Blazeclan Europe (Belgium) | Process: FinOps cost governance (2020) | blazeclan.com |
| Brabender GmbH (Germany) | Product: liquid elastomeric waterproofing system (2023) | brabender-gmbh.de |
| *Verified in third-party sources:* | | |
| Motocaddy EMEA GmbH (Germany) | Product: M Series electric trolley range (2021) | golfmonthly.com |
| H&M Hennes & Mauritz S.A. (Greece) | Process: supplier sustainability audit program (2016) | sciencebasedtargets.org |
| UAB Nord Robotics (Lithuania) | Product: Nord Robotics palletizing system (2016) | linkedin.com |
| Advanced Accelerator Applications (Portugal) | Product: Lutetium Lu 177 peptide receptor radionuclide therapy (2018) | globenewswire.com |

**Notes:** Each row is an extracted introduction matched to a public page documenting the specific product or process; the source column links to the documenting page. The examples span the full firm-size range of the audit sample, from the 1–5 to the 101–1,000 employee band. Years are those returned by the extraction and are not independently corroborated in every case. Some linked sources document activity at the corporate-group level rather than establishing its attribution to the specific local legal entity.

We interpret the 9.7% verification rate as the share of audited introductions corroborated under the stated search protocol, rather than as an estimate of extraction accuracy. Failure to locate a supporting source does not by itself establish that an introduction is false. The audit does not establish the accuracy of the unverified records.

Three limitations should be noted. First, the sample is stratified by innovation direction and caps each country's share, so it is not a random draw from the firm population, and the country gradient in particular is shaped by that cap. Firm size, by contrast, is not balanced across cells, so the steep size gradient reflects real differences across the firms we study rather than an artifact of the sampling. Second, the audit does not include an unrestricted search arm: an earlier unrestricted run used a lighter search budget and is not depth-comparable to the two restricted arms, so the union of the two restricted channels is the appropriate benchmark for an unrestricted search. Third, verification measures public-web visibility rather than the veracity of the underlying introduction.

A final remark concerns the interpretation of the unverified majority. The audit searches the publicly accessible web and therefore cannot reach relevant reporting that is paywalled, no longer indexed, or otherwise unavailable through public search. Publisher agreements also show that model developers license non-public content. OpenAI, for example, has agreements with News Corp covering current and archived content from publications including the *Wall Street Journal*, *Barron's*, and *MarketWatch* (OpenAI, 2024b); with the *Financial Times* to provide attributed content in ChatGPT and help improve model usefulness (OpenAI, 2024a); and with Axel Springer covering publications including *Business Insider* and *Politico*, including otherwise paid content (OpenAI, 2023). These agreements do not establish whether any such material entered the model used in this study or identify the source of any particular extracted introduction. They do, however, reinforce the narrower point that failure to find a public-web source should not be interpreted as evidence that an introduction is false.

# D Production function regressions by firm size

The external-verification audit in Appendix C finds that public corroboration of the extracted introductions is scarcest among small firms. A natural concern is that the production-function evidence in Section 3 might likewise be driven by large, web-visible firms. This appendix addresses that concern by re-estimating the full specification of Table 1, column (4) — both innovation scores, both introduction-count regressors, the patent, size, and capital controls, and country$\times$NACE2 fixed effects — separately for firms below and at or above 20 employees. The innovation scores remain the cross-fitted TF-IDF predictions trained on the pooled sample and are held fixed across the size-specific regressions. The two reported size-specific estimation samples together contain 310 fewer observations than the pooled regression. Column (1) of Table D1 reproduces the pooled column-(4) estimates for reference.

The association between the extracted innovation scores and revenue is essentially invariant to firm size. The product-score coefficient is 0.053 in the pooled sample, 0.053 among firms with fewer than 20 employees, and 0.047 among firms with at least 20; the corresponding process-score coefficients are 0.024, 0.023, and 0.021, and all six estimates are significant at the

**Table D1:** Production function regressions by firm size.

| | (1) Pooled | (2) Employment $< 20$ | (3) Employment $\geq 20$ |
|---|---|---|---|
| Product innovation score (z) | 0.053*** | 0.053*** | 0.047*** |
| | (0.008) | (0.009) | (0.004) |
| Process innovation score (z) | 0.024*** | 0.023*** | 0.021*** |
| | (0.004) | (0.004) | (0.004) |
| Has any product text | 0.038 | 0.040 | 0.042 |
| | (0.039) | (0.041) | (0.037) |
| Has any process text | −0.010 | −0.002 | −0.011 |
| | (0.036) | (0.043) | (0.042) |
| log(1 + official patent count) | 0.033 | 0.101** | 0.059* |
| | (0.022) | (0.030) | (0.023) |
| log(1 + product introductions) | −0.043 | −0.038 | −0.103+ |
| | (0.056) | (0.057) | (0.057) |
| log(1 + process introductions) | 0.039 | 0.035 | 0.002 |
| | (0.049) | (0.057) | (0.068) |
| Log employment | 1.014*** | 0.997*** | 0.927*** |
| | (0.041) | (0.062) | (0.053) |
| Log capital | 0.152*** | 0.248*** | 0.237*** |
| | (0.032) | (0.050) | (0.047) |
| Capital missing | 1.571*** | 2.585*** | 3.657*** |
| | (0.346) | (0.507) | (0.752) |
| Num.Obs. | 309 222 | 249 627 | 59 285 |
| R2 | 0.861 | 0.806 | 0.907 |
| R2 Adj. | 0.860 | 0.804 | 0.904 |

+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001

Notes: The dependent variable is log revenue. The table uses shared full-sample scores. Column (1) reproduces column (4) of Table 1; columns (2) and (3) estimate the same specification on firms below and at-or-above 20 employees. Standard errors two-way clustered at country and NACE2 are reported in parentheses. The product and process innovation scores are the K=3 cross-fitted TF-IDF lasso predictions trained on the pooled sample (identical to Table 1); the two reported size-specific estimation samples together contain 310 fewer observations than the pooled regression.

0.1% level. Since firms below 20 employees account for roughly four fifths of the regression sample (249,627 of 309,222 firms), the pooled association cannot be attributed to a large-firm subsample. The patent coefficient, by contrast, is more sensitive to the split and is estimated imprecisely in the pooled specification. These estimates show that the association between the innovation scores and revenue is also present among small firms. They do not establish the accuracy of individual extracted introductions or explain why particular items could not be corroborated in the web audit.

# E Country-cap robustness

The country counts in the full snapshot are uneven, because the active-status and financial filters were applied after the initial per-country draw (Section 2). This appendix asks whether the direction results are driven by the over-represented countries. We re-run the analysis on two capped samples that keep at most 20,000 and at most 10,000 firms per country, drawing at random within countries above the cap and keeping smaller countries in full. The 20,000-firm cap binds for nine countries and retains 538,000 of the 605,000 firms. The 10,000-firm cap is more aggressive: it binds for 27 countries and retains 382,000 firms. Table E1 compares both caps with the baseline across sample composition, headline shares, the production function validation, the country- and sector-level relationships, and country rankings.

The results are stable under both caps. Headline firm-shares move by less than half a percentage point, country rankings are almost perfectly preserved, and the production function score coefficients stay positive and significant, shrinking modestly under the 20,000-firm cap and by about a fifth under the more aggressive 10,000-firm cap. The direction relationships keep their signs and significance throughout, with one exception: the country-level full-snapshot relationship between LLM greenness and green patenting is significant in the baseline and the 20,000-firm cap but not in the 10,000-firm cap. That same relationship remains significant in the regression sample and at the sector and country×sector levels, so the broader patent comparison is not fragile. We read the 20,000-firm cap as the more representative check and the 10,000-firm cap as a stress test.

**Table E1:** Country-cap robustness: baseline sample versus 20,000- and 10,000-firm per-country caps.

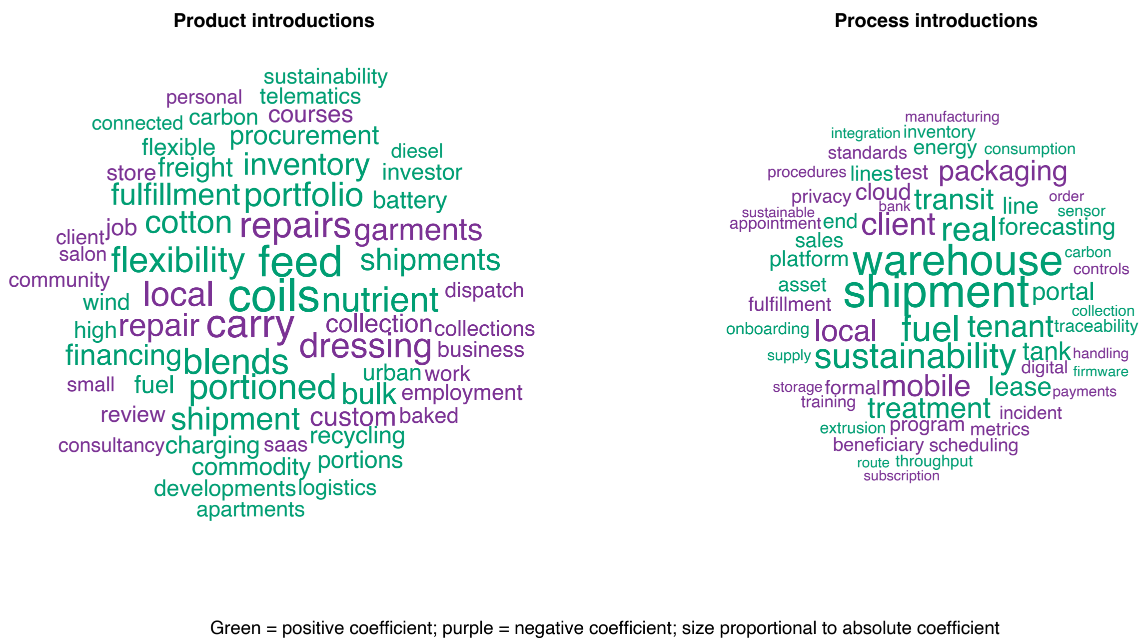
| | Baseline | 20k cap | 10k cap |
|---|---|---|---|
| *Panel A. Sample composition* | | | |
| Full snapshot (firms) | 605,830 | 537,940 | 381,914 |
| Countries above the cap | 0 | 9 | 27 |
| Largest single-country count | 51,877 | 20,000 | 10,000 |
| *Panel B. Headline firm-shares, full snapshot (%)* | | | |
| Product, any introduction | 43.67 | 43.80 | 43.71 |
| Product, augmenting | 24.60 | 24.68 | 24.68 |
| Product, labor-saving | 8.04 | 8.07 | 7.98 |
| Product, green | 14.11 | 14.20 | 13.95 |
| Process, any introduction | 26.29 | 26.36 | 26.28 |
| Process, augmenting | 15.46 | 15.51 | 15.42 |
| Process, labor-saving | 7.63 | 7.64 | 7.59 |
| Process, green | 11.63 | 11.66 | 11.57 |
| Any patent | 1.98 | 2.14 | 2.04 |
| Clean (Y02) patent | 0.40 | 0.45 | 0.42 |
| *Panel C. Production-function validation (both-score specification)* | | | |
| Product innovation score (z) | $0.053^{***}$ | $0.052^{***}$ | $0.041^{***}$ |
| | (0.008) | (0.009) | (0.008) |
| Process innovation score (z) | $0.024^{***}$ | $0.022^{***}$ | $0.019^{***}$ |
| | (0.004) | (0.004) | (0.003) |
| *Panel D. Country- and sector-level relationships (OLS slope)* | | | |
| Green vs. net-augmenting: process, full | $0.595^{***}$ | $0.592^{***}$ | $0.626^{***}$ |
| Augmenting vs. saving: process, full | $0.426^{***}$ | $0.427^{***}$ | $0.414^{***}$ |
| Green vs. green patent: country, regression | $0.278^{*}$ | $0.280^{*}$ | $0.299^{*}$ |
| Green vs. green patent: country, full | $0.322^{*}$ | $0.321^{*}$ | 0.281 |
| Green vs. green patent: sector, full | $0.596^{***}$ | $0.591^{***}$ | $0.594^{***}$ |
| Green vs. green patent: country×sector, full | $0.179^{***}$ | $0.193^{***}$ | $0.186^{**}$ |
| *Panel E. Country ranking stability (correlation with baseline)* | | | |
| Green product firm-share | — | 0.9999 | 0.9987 |
| Green process firm-share | — | 0.9998 | 0.9989 |
| Product net-augmenting share | — | 0.9996 | 0.9944 |
| Process net-augmenting share | — | 0.9990 | 0.9952 |

**Notes:** Each cap keeps at most the stated number of firms per country, drawing at random within countries above the cap (fixed seed) and keeping smaller countries in full. Panel C reports the product and process innovation-score coefficients from the both-score production-function specification (column (4) of Table 1), with standard errors two-way clustered on country and NACE2 in parentheses. Panel D reports the OLS slopes behind the country- and sector-level direction figures. Panel E reports correlations of country firm-shares between each capped sample and the baseline on the common country set. Across both caps the headline shares, the validation coefficients, and the direction and ranking results are stable; the single exception is the country-level full-snapshot green-versus-green-patent slope, which is significant in the baseline and the 20,000-firm cap but not in the more aggressive 10,000-firm cap, while the same relationship remains significant in the regression sample and at the sector and country×sector levels. Stars: $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

# F TF-IDF score diagnostics

The cross-fitted TF-IDF innovation score used in Section 3 is built from a lasso of extracted-text TF-IDF features on residualized log sales. Figure F1 reports the fold-averaged lasso coefficients as a word cloud. The terms that load most heavily are recognizable innovation vocabulary, supporting the validation claim that the score is picking up information beyond country-by-sector boilerplate.

**Figure F1:** Word cloud of fold-averaged TF-IDF lasso coefficients.

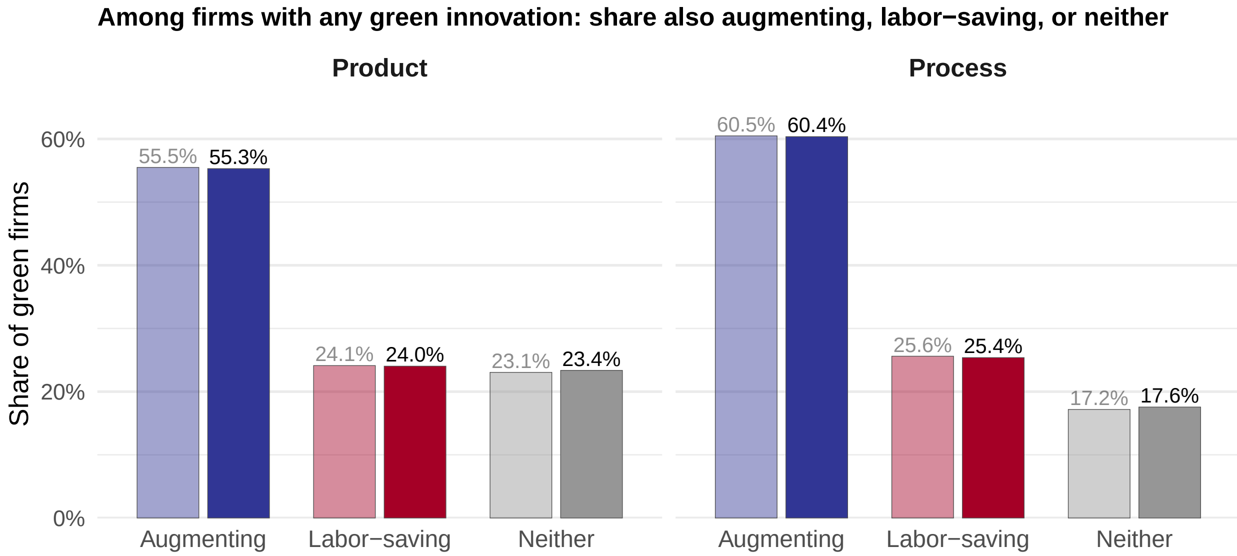



**Notes:** Product terms appear on the left and process terms on the right. Word size is proportional to coefficient magnitude; color encodes sign (green positive, purple negative). Coefficients are averaged across the $K = 3$ cross-fit folds; the top 60 terms by absolute coefficient are shown per panel. The lasso is fit at the firm level on the extracted product or process text after partialing out log employment, log capital, the capital-missing indicator, $\log(1 + \text{official patent count})$, and country$\times$NACE2 fixed effects. Negative-loading terms enter with a negative coefficient on residualized log sales: after partialing out size and country-by-sector fixed effects, firms whose extracted text is heavier on those terms have lower sales than their controls would predict. Construction details follow the cross-fit pipeline in Section 3.

# G Conditional shares among green firms

This appendix reports, conditional on a firm having any green innovation, the fraction that also has a labor-augmenting introduction, the fraction with a labor-saving introduction, and the fraction with neither. Product introductions are on the left and process introductions on the right; within each category the faded bar is the regression sample (309k firms) and the solid bar is the full snapshot (605k firms). The denominator is firms with at least one green product (left) or process (right) introduction, so the three categories are not mutually exclusive: a firm can have both an augmenting and a labor-saving introduction, in which case it contributes to both bars.

**Figure G1:** Labor direction among firms with green innovation.

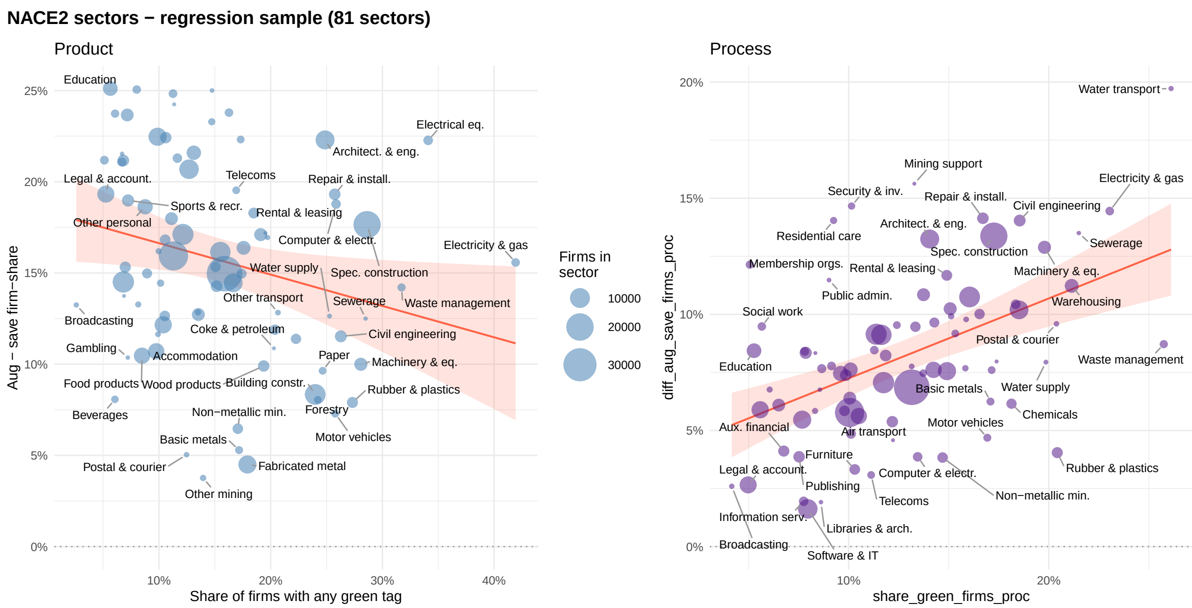



**Notes:** The figure reports, among firms with any green product (left) or process (right) introduction, the share that also has an augmenting introduction, a labor-saving introduction, or neither. Faded bars report the regression sample (309k firms); solid bars report the full 605k snapshot. The denominator within each panel is firms with the corresponding green tag. The augmenting and labor-saving categories overlap; "Neither" is firms with the green tag but neither a labor-augmenting nor a labor-saving introduction (typically because the labor effect was classified as "unclear").

# H NACE2 sector-level results

This appendix repeats the firm-level direction-tag analysis with each dot representing one NACE2 sector rather than one country. The sector set is the same in every figure: NACE2 codes with at least 100 firms in the regression sample. Dot labels are short verbal descriptors of the NACE Rev. 2 division (e.g. "Wholesale trade", "Software & IT", "Pharmaceuticals"). The aggregations otherwise follow the country-level construction in Sections 4–4.2.

**Figure H1:** Sector-level direction of innovation, regression sample.

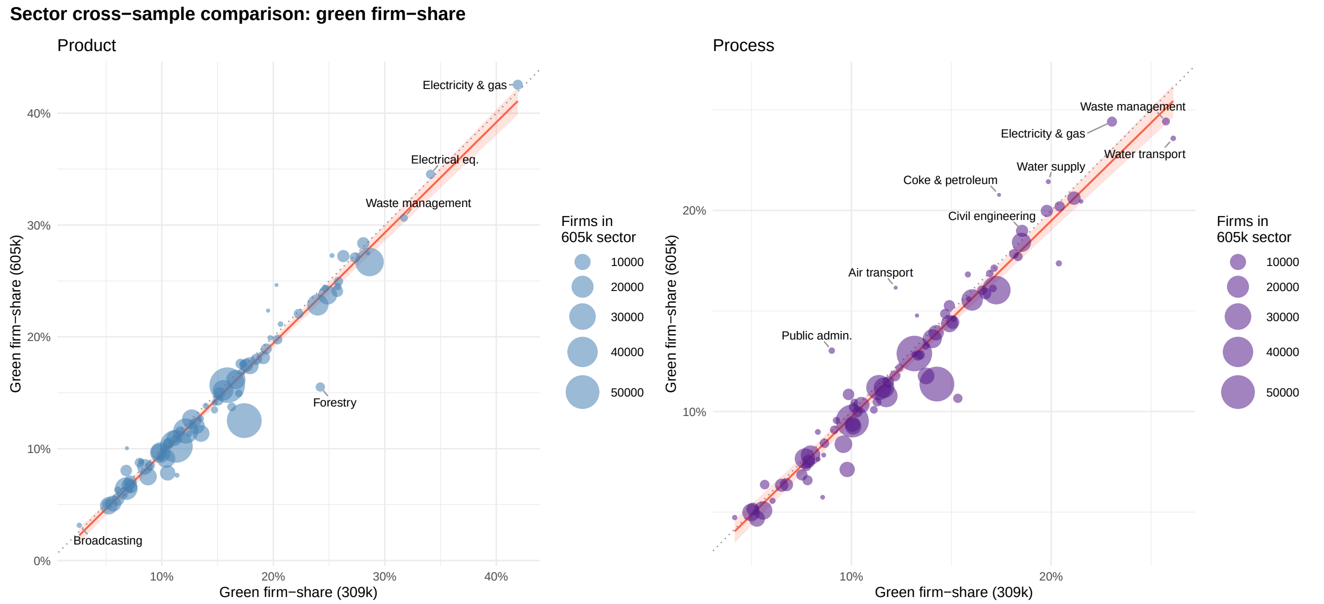



**Notes:** Each dot is one NACE2 sector with at least 100 firms in the regression sample. The x-axis reports the share of firms in the sector with any green product (left) or process (right) introduction. The y-axis reports the augmenting firm-share minus the labor-saving firm-share. The red line is an OLS fit with a 95% confidence band.

**Figure H2:** Sector-level direction of innovation, full snapshot.

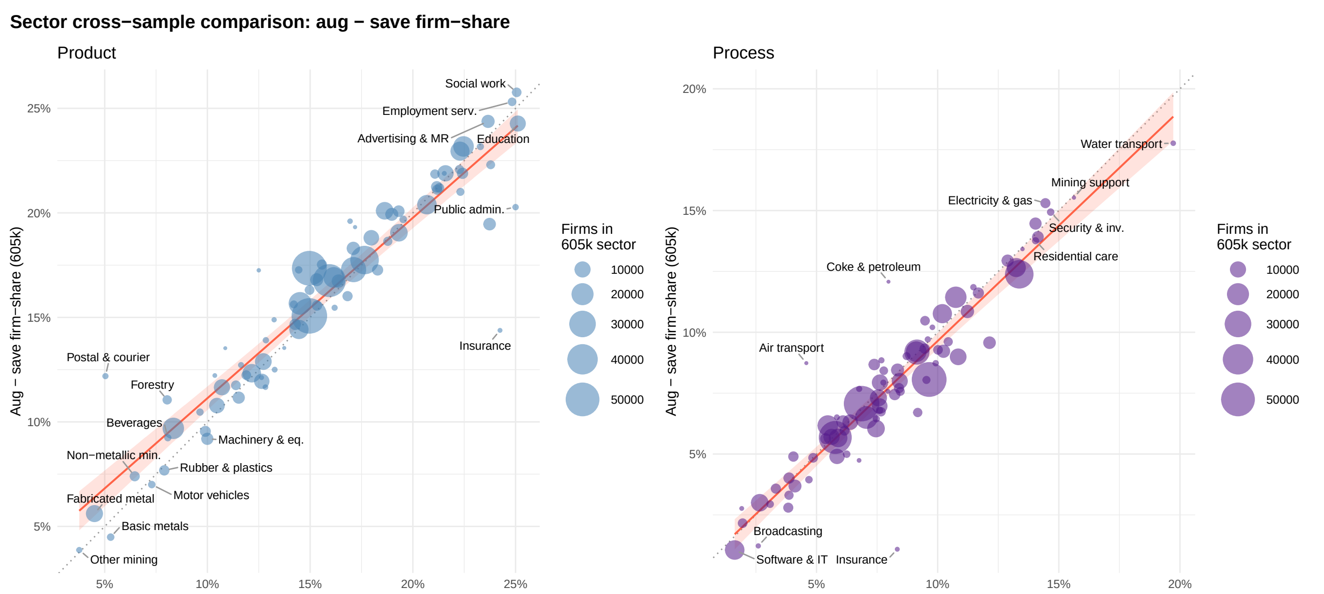

**Notes:** Each dot is one NACE2 sector, with product introductions shown on the left and process introductions on the right. The x-axis is the share of firms in the sector with at least one green introduction; the y-axis is the augmenting firm-share minus the labor-saving firm-share. The figure uses the full 605k snapshot and is restricted to sectors with at least 100 firms in the regression sample. Firms with no extracted text remain in the denominator and contribute zeros to the relevant numerators. Dot size is proportional to the number of firms in the sector; the red line is an OLS fit with a 95% confidence band.

**Figure H3:** Sector-level cross-sample comparison of green firm-shares.



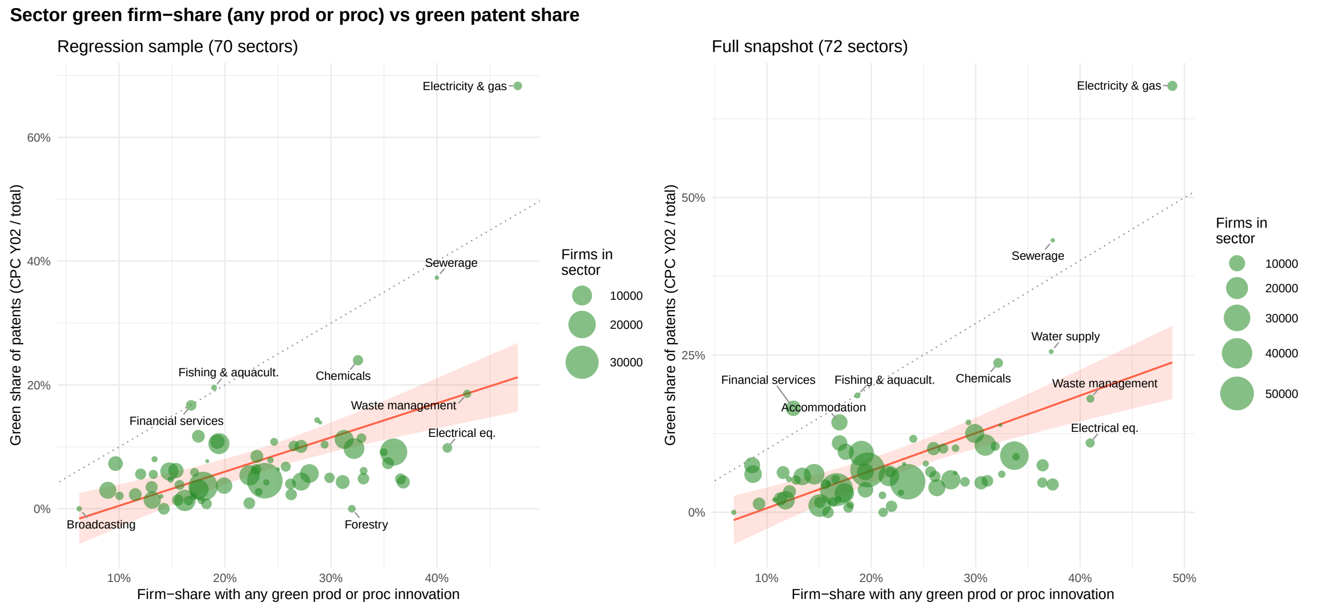

**Notes:** The x-axis reports the green firm-share in the 309k regression sample and the y-axis reports the green firm-share in the 605k full snapshot. Product innovation appears on the left and process innovation on the right. The dotted line is the 45-degree equality reference.

**Figure H4:** Sector-level cross-sample comparison of augmenting–saving firm-shares.




**Notes:** The x-axis reports the augmenting–saving firm-share in the 309k regression sample and the y-axis reports the same measure in the 605k full snapshot. Product innovation appears on the left and process innovation on the right.

**Figure H5:** Sector-level green firm-share vs green patent share.



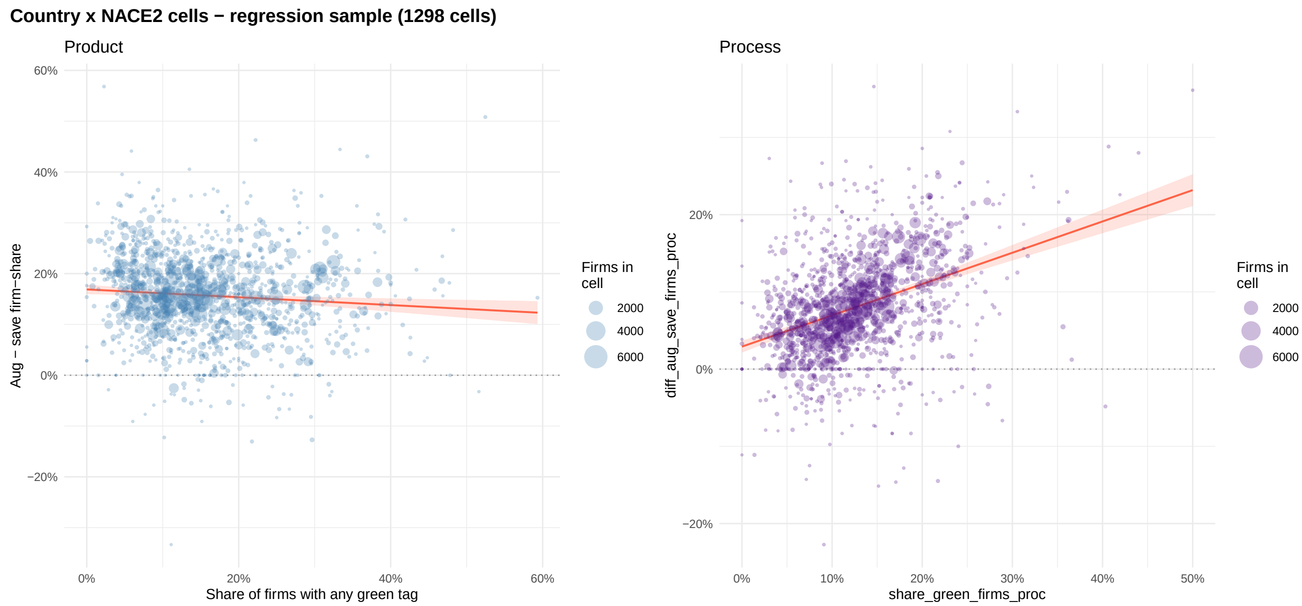

**Notes:** Green firm-share covers any green product or process introduction. The regression sample appears on the left and the full 605k snapshot on the right. The figure is restricted to sectors with at least 25 patents and at least 100 firms in the regression sample.

# I Country × NACE2 cell-level results

This appendix repeats the firm-level direction-tag analysis at the (country × NACE2) cell level. Each dot is now one country-sector cell rather than one country. With ∼2,700 such cells in the regression sample, individual labels would be unreadable, so we drop text labels and rely on dot density. The cell set is the same in every figure: cells with at least 25 firms in the regression sample.

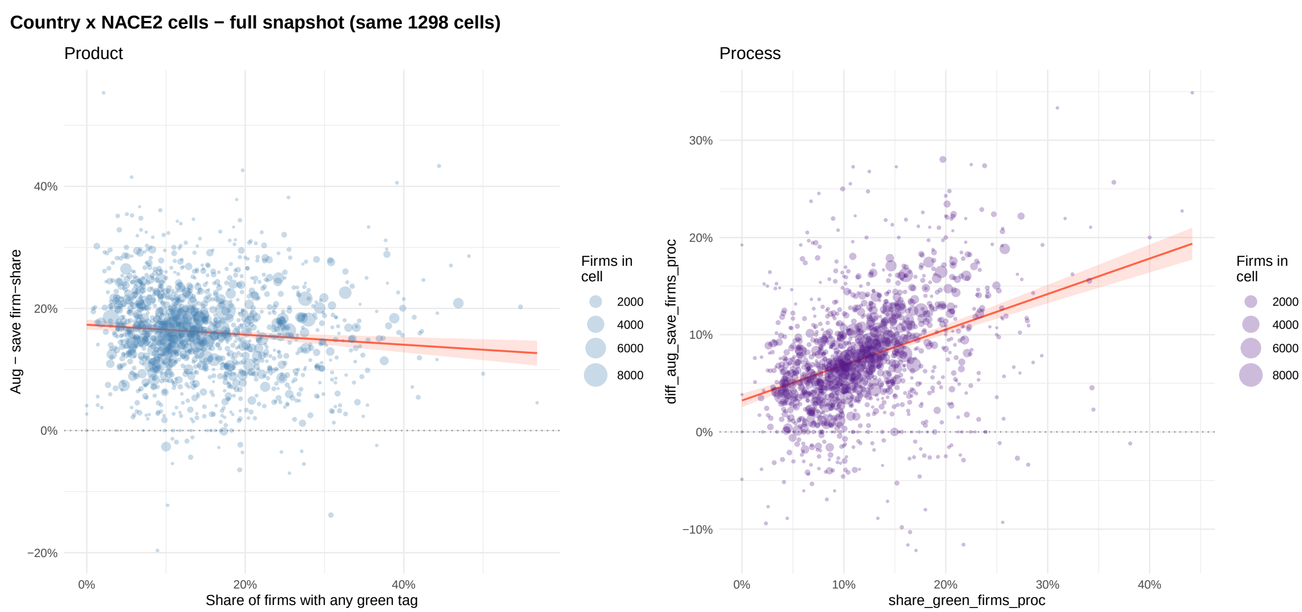
Figures I1 and I2 show that the cross-cell relationship between greenness and net labor direction differs between products and processes: the fitted relationship is negative on the product side and positive on the process side. The LLM-extracted green firm-share and the green patent share are positively associated at the cell level (Figure I5). As in the country-level comparison, these measures have different denominators and their levels are not directly comparable.

**Figure I1:** Cell-level direction of innovation, regression sample.



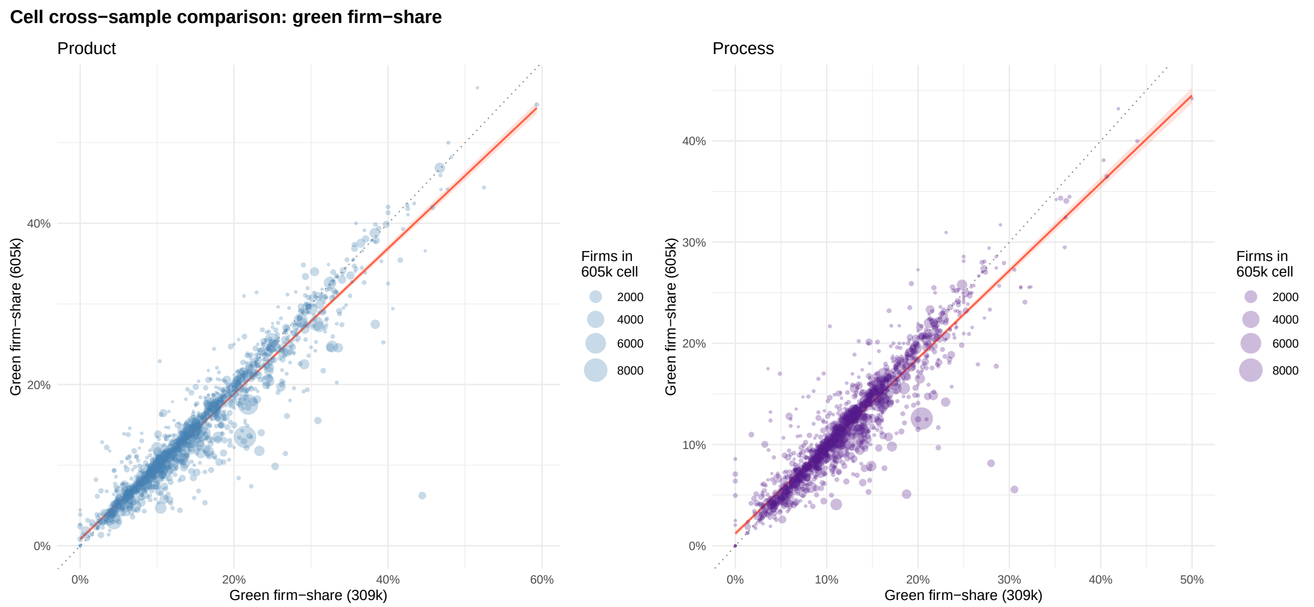

**Notes:** Each dot is one (country, NACE2) cell with at least 25 firms in the regression sample. The x-axis reports the share of firms in the cell with any green product (left) or process (right) introduction. The y-axis reports the augmenting firm-share minus the labor-saving firm-share. The red line is an OLS fit with a 95% confidence band.

**Figure I2:** Cell-level direction of innovation, full snapshot.



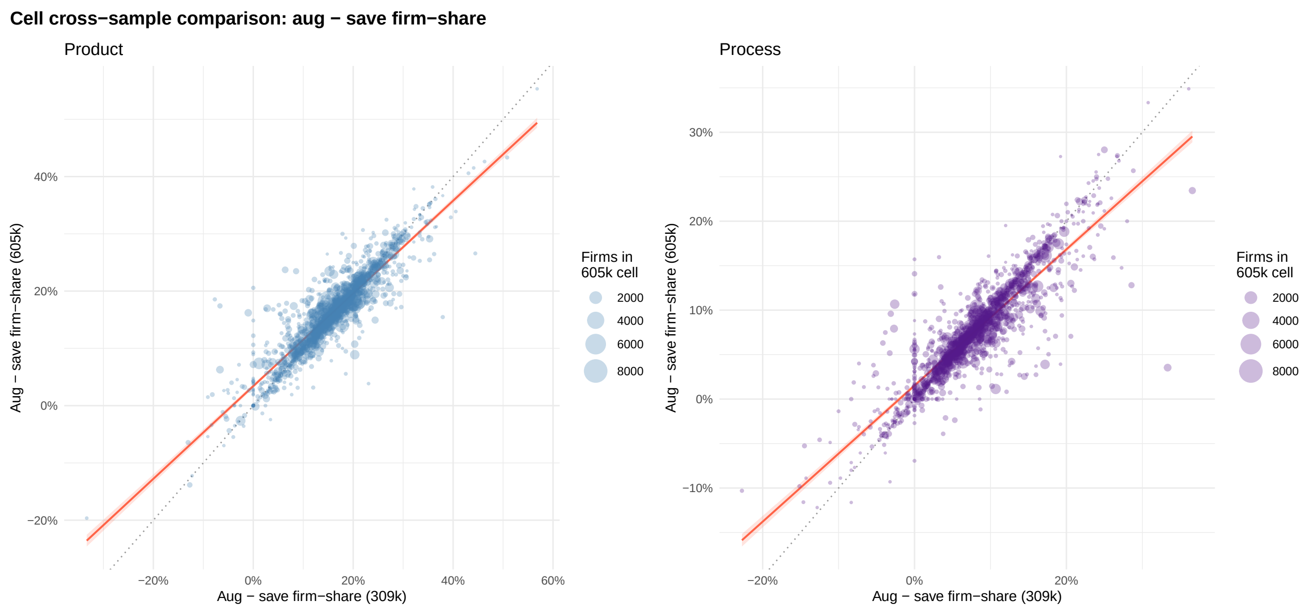

**Notes:** Each dot is one country–NACE2 cell, with product introductions shown on the left and process introductions on the right. The x-axis is the share of firms in the cell with at least one green introduction; the y-axis is the augmenting firm-share minus the labor-saving firm-share. The figure uses the full 605k snapshot and is restricted to cells with at least 25 firms in the regression sample. Firms with no extracted text remain in the denominator and contribute zeros to the relevant numerators. Dot size is proportional to the number of firms in the cell; the red line is an OLS fit with a 95% confidence band.

**Figure I3:** Cell-level cross-sample comparison of green firm-shares.



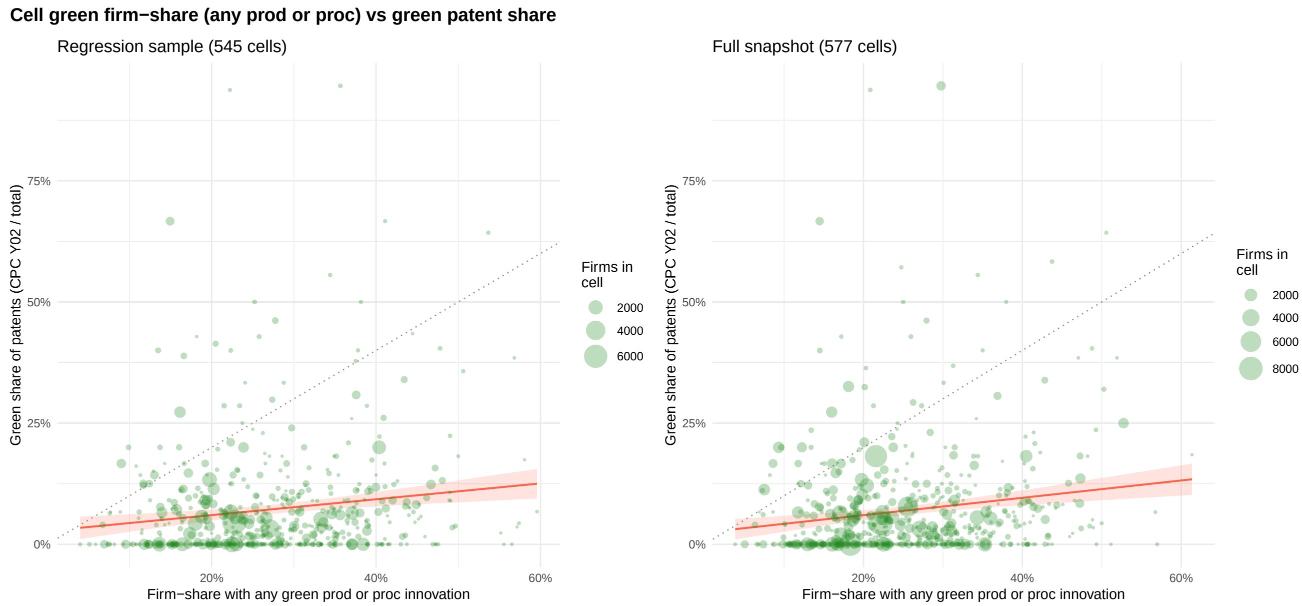

**Notes:** The x-axis reports the green firm-share in the 309k regression sample and the y-axis reports the green firm-share in the 605k full snapshot. Product innovation appears on the left and process innovation on the right. The dotted line is the 45-degree equality reference.

**Figure I4:** Cell-level cross-sample comparison of augmenting–saving firm-shares.




**Notes:** The x-axis reports the augmenting–saving firm-share in the 309k regression sample and the y-axis reports the same measure in the 605k full snapshot. Product innovation appears on the left and process innovation on the right.

**Figure I5:** Cell-level green firm-share vs green patent share.




**Notes:** Green firm-share covers any green product or process introduction. The regression sample appears on the left and the full 605k snapshot on the right. The figure is restricted to cells with at least 5 patents and at least 25 firms in the regression sample.

# J Sector vs country variance decomposition

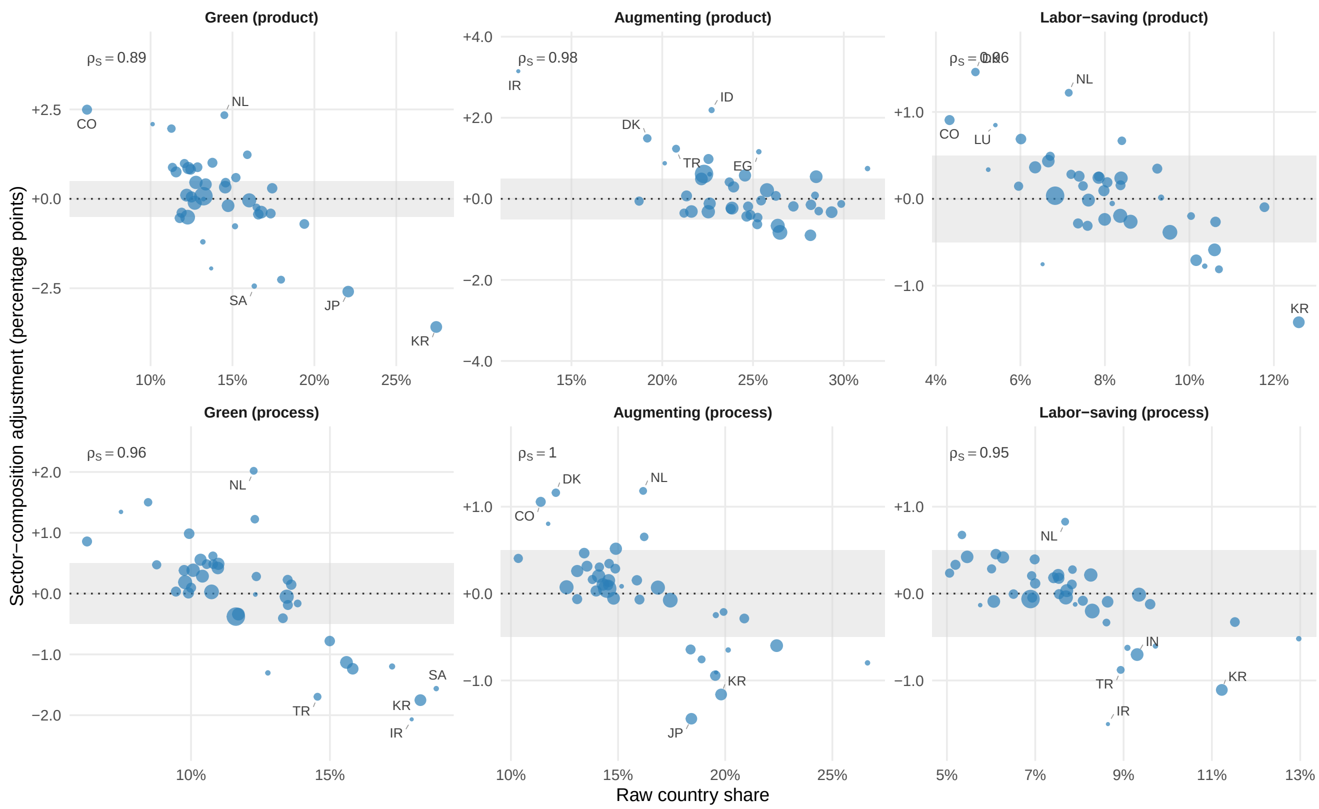
This appendix decomposes the cross-firm variation in the six direction tags into a country part and a NACE2 sector part. Table J1 reports $R^2$ shares; Figure J1 compares raw country shares to the corresponding sector-adjusted shares.

**Table J1:** Variance decomposition of firm-level direction tags.

| Tag | Country | Sector | Country + Sector | Interaction extra | Residual |
|---|---|---|---|---|---|
| Green (product) | 0.8 | 3.5 | 4.3 | 1.1 | 94.7 |
| Green (process) | 0.5 | 1.4 | 1.8 | 0.8 | 97.3 |
| Augmenting (product) | 0.4 | 0.8 | 1.2 | 0.8 | 98.0 |
| Augmenting (process) | 0.5 | 0.6 | 1.2 | 0.8 | 98.0 |
| Labor-saving (product) | 0.3 | 1.1 | 1.4 | 0.8 | 97.8 |
| Labor-saving (process) | 0.3 | 0.4 | 0.7 | 0.7 | 98.6 |

**Notes:** Rows are firm-level binary direction tags; entries are $R^2$ contributions in percent from a linear-probability regression of the tag on country and NACE2 sector fixed effects. The "Country" and "Sector" columns report the Shapley share of each fixed-effect block (the order-free average of the two sequential orderings). "Country + Sector" is the joint $R^2$ from the additive two-way fixed-effects specification. "Interaction extra" is the additional $R^2$ contributed by the full country$\times$sector interaction beyond the additive model. "Residual" is the remaining within-cell variation. Sample: 553,048 firms across 40 countries and 86 NACE2 sectors, restricted to countries with at least 25 firms in the regression sample.

**Figure J1:** Sector-composition adjustment to country direction-tag shares.




**Notes:** Each dot is a country; the $y$-axis is the sector-adjusted country share minus the raw country share, in percentage points. Positive values mean adjusting for sector mix raises the country's share; negative values mean the raw share was inflated by sector composition. The dotted reference line at zero marks no adjustment, and the gray band marks $\pm 0.5$ pp. The Spearman rank correlation between raw and adjusted shares appears in the top-left of each panel. Country labels mark the five countries with the largest absolute adjustment in each panel. The adjusted country share is constructed by taking the firm-level residual from a regression of the binary direction tag on NACE2 sector fixed effects, averaging within country, and adding back the grand mean. The construction isolates the within-sector cross-country contribution to the country share. Sample as in Table J1. Dot size is proportional to firm count in the country.