UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

DÉPARTEMENT DES SCIENCES ADMINISTRATIVES

WORKING PAPER NO. 38

The Economic Geography of Digital Production: Product Diversity, Regional Specialization, and the Hidden Structure of Quebec's Online Economy

Evidence from 390,108 product listings of 1,569 Quebec online stores across 17 administrative regions (Fabri-Ka snapshot, 1 November 2026)

Simon-Pierre Boucher

Département des sciences administratives
Université du Québec en Outaouais

simon-pierre.boucher@uqo.ca

Gatineau – Pavillon Alexandre-Taché
283, boulevard Alexandre-Taché
Gatineau, Québec, Canada J9A 1L8

November 2026

Version 1.0

Abstract

Abstract to be written after the final run.

Keywords: economic geography; online retail; product assortment; revealed comparative advantage; diversity; economic complexity; price dispersion; web-scraped data; Quebec

JEL Classification: R12, L81, O18, C81, D22, R32

1. introduction

TODO.

2. Related literature and positioning

The paper draws on six literatures. We summarise what each establishes, and where the present exercise sits relative to its closest empirical precedents. Every reference below was verified against Crossref or the publisher's page before citation; one critique of the method of reflections that could not be verified is not cited.

Economic geography and regional specialization. Industries are localised far beyond a random allocation of plants across space (Ellison and Glaeser, 1997; Duranton and Overman, 2005; Mori et al., 2005), with heterogeneity across industries and scales (Rosenthal and Strange, 2001) that also holds in Canadian plant microdata (Baldwin et al., 2008). The micro-foundations are sharing, matching and learning (Marshall, 1920; Duranton and Puga, 2004); density elasticities of productivity are modest and demanding to identify (Rosenthal and Strange, 2004; Gobillon, 2015; Duranton and Puga, 2020). Whether local specialisation or diversity favours growth remains contested (Glaeser et al., 1992; Henderson et al., 1995), and the related/unrelated-variety decomposition of entropy is the modern operationalisation (Frenken et al., 2007; Content and Frenken, 2016). Market size and love of variety (Krugman, 1980, 1991; Holmes et al., 2000) and the skill-ordered comparative advantage of cities (Davis and Dingel, 2020) motivate a link between regional scale and assortment breadth; urban scaling shows that the number of distinct business types grows roughly with the logarithm of population and that activities appear in a nested order across city sizes (Bettencourt et al., 2007; Youn et al., 2016; Hong et al., 2020). For Quebec and Canada, Polèse and Shearmur (2004, 2006) and Shearmur and Polèse (2007); Shearmur and Doloreux (2008) document a persistent centre–periphery gradient in employment location. None of this work uses the catalogues of online merchants as the object of measurement, and none covers Quebec's seventeen administrative regions in the 2020s. Our revealed assortment advantage (RAA) is the Balassa logic (Balassa, 1965) applied to listing counts, with the dartboard idea of Ellison and Glaeser (1997) supplying the null: a random allocation of *merchants*, not listings, to regions.

Economic complexity, relatedness and diversification. The Balassa index is skewed and unstable at the top, and symmetric or normalised variants exist (Hinloopen and Marrewijk, 2001; Yu et al., 2008; Laursen, 2015). The product space (Hidalgo et al., 2007) and the principle of relatedness (Hidalgo et al., 2018; Boschma, 2016) are robust regularities: regions diversify into activities

related to what they already do, documented with industries (Neffke et al., 2011), patents (Kogler et al., 2013; Rigby, 2013; Balland and Rigby, 2016), occupations (Muneepeerakul et al., 2013) and firms (Gao and Zhou, 2018), and combined with complexity in the smart-specialisation framework (Balland et al., 2018). Economic complexity and fitness indices summarise bipartite region–activity matrices (Hidalgo and Hausmann, 2009; Hausmann and Hidalgo, 2011; Hausmann et al., 2014; Tacchella et al., 2012; Cristelli et al., 2013; Servedio et al., 2018); the ECI is formally a spectral clustering of a region-similarity matrix (Mealy et al., 2019), rankings are sensitive to algorithm and sample (Morrison et al., 2017; Sciarra et al., 2020), and structure in bipartite matrices must be tested against degree-preserving null models (Saracco et al., 2015); see Hidalgo (2021) and Balland et al. (2022) for surveys. All applications we could verify use production-side bases (exports, patents, industries, occupations, firms). Nothing is known about the behaviour of these tools on a 17×20 matrix built from online listings, which is why we treat the complexity indicator as an exercise in validation rather than as a ranking (Section **??**).

Online retail, the long tail and the geography of online demand. Online variety created consumer surplus and shifted sales toward niche products (Brynjolfsson et al., 2003, 2011; Zentner et al., 2013; Waldfogel, 2017; Anderson, 2006), but hits remain dominant and the tail is thin in revenue (Elberse, 2008; Tan et al., 2017); search-cost theory predicts both long tails and superstars (Bar-Isaac et al., 2012). E-commerce reallocates retail share toward large firms and platform owners (Goldmanis et al., 2009; Zhu and Liu, 2018; Einav et al., 2016). The geography of online demand is systematic: consumers buy more online where local supply is thin and when they are preference minorities (Forman et al., 2009; Choi and Bell, 2011; Brynjolfsson et al., 2009), demand heterogeneity across markets limits the value of online variety for products with local tastes (Quan and Williams, 2018), distance still matters online (Hortaçsu et al., 2009; Blum and Goldfarb, 2006; Lendle et al., 2016), gains are larger for remote consumers (Dolfen et al., 2023; Couture et al., 2021), and e-commerce changes relative prices across space (Jo et al., 2019; Lieber and Syverson, 2012; Goldfarb and Tucker, 2019). Platform design shapes dispersion (Dinerstein et al., 2018; Ellison, 2009; Baye et al., 2004, 2006; Brynjolfsson and Smith, 2000; Clay et al., 2001). This stream is demand- or transaction-based. The supply-side question — what independent local merchants put online and how it varies across the regions of one province — is, to our knowledge, not addressed; Quan and Williams (2018) and Forman et al. (2009) are the closest mirrors, and our data cannot say whether regional assortments reflect regional tastes.

Regional inequality and digital adoption. Firms in large cities adopted the commercial internet first (Forman et al., 2005); the wage gains from internet investment accrued to already-advantaged places (Forman et al., 2012; Sinai and Waldfogel, 2004; Greenstein et al., 2018); broadband effects

on rural growth are positive but modest and hard to identify (Kolko, 2012; Whitacre et al., 2014; Haller and Lyons, 2015, 2019); adoption and usage divides differ (Goldfarb and Prince, 2008); Canadian household divides are documented with the Canadian Internet Use Survey (Haight et al., 2014; Statistics Canada, n.d.b, 2023) and business digital adoption with the SDTIU (Statistics Canada, n.d.c); the OECD measurement framework is OECD (2019) and the Quebec programme is Académie de la transformation numérique (n.d.). We found no peer-reviewed study of regional e-commerce *supply* in Canada or Quebec; the Quebec-specific evidence is grey literature (Institut de la statistique du Québec, 2024), and the paper is explicit that this stream is the thinnest in verified academic sources.

Product matching, hedonic prices and spatial price dispersion. Record linkage rests on Fellegi and Sunter (1969) and Christen (2012). Scraped online prices are a valid but selected source: online and offline prices coincide for most multi-channel retailers (Cavallo, 2017), scraped data have specific frequency and stockout properties (Cavallo, 2018; Cavallo and Kryvtsov, 2023), and the Billion Prices Project set the standards (Cavallo and Rigobon, 2016); online price-setting and cross-border integration are documented by Gorodnichenko and Talavera (2017); Gorodnichenko et al. (2018); Cavallo et al. (2014); web-scraped prices now enter official indices (Friedman, 2022). Dispersion is mostly relative and persistent (Kaplan and Menzio, 2015; Kaplan et al., 2019); US chains price almost uniformly across locations (DellaVigna and Gentzkow, 2019); spatial price gaps grow with distance (Atkin and Donaldson, 2015); variety and prices differ across cities (Handbury and Weinstein, 2014; Handbury, 2021); hedonic theory and practice follow Rosen (1974); Pakes (2003); Bajari and Benkard (2005). Cross-merchant dispersion among many *independent* small online stores in one province, with product fixed effects from matched titles and SKUs, has no direct precedent; the near-collinearity of merchant and region effects is specific to our setting.

Measurement error and representativeness of non-probability data. For non-probability big data the error is governed by the data-defect correlation rather than by sample size (Meng, 2018; Bradley et al., 2021; Lazer et al., 2014); inference requires explicit selection models (Elliott and Valliant, 2017; Buelens et al., 2018; Heckman, 1979; Groves, 2006; Little and Rubin, 2019), and statistical agencies are explicit about the conditions (Daas et al., 2015; Beaumont, 2020; Rao, 2020); measurement error in regressors and outcomes (Hausman, 2001; Bound et al., 2001) and the modifiable areal unit problem (Briant et al., 2010) are additional threats. We take these literatures literally: Section 4 reports a data-quality ledger, a category-label audit, a matching validation and a representativeness audit against Statistics Canada and ISQ counts.

What is and is not new here. The contribution is a documented, validated supply-side dataset of online product listings of Quebec-based merchants by administrative region; RAA and diversity measures with merchant-count-preserving permutation nulls and merchant-weighted variants; a region–category bipartite indicator whose eigenvector interpretation is tested against bootstraps and null models and reported as a family of components rather than a headline score; cross-merchant dispersion of matched products among independent stores with explicit collinearity diagnostics; and a representativeness audit of an aggregator against business counts. No method is new. We do not provide evidence on agglomeration economies, on diversification over time, on sales concentration, on gravity in online trade or on welfare, and we never call the indicator a production or economic complexity index: without verified product origin it describes observed online catalogues only.

3. Institutional setting and the Fabri-Ka platform

Quebec's administrative regions. Quebec is divided into seventeen administrative regions (*régions administratives*) defined by provincial decree and used by the Institut de la statistique du Québec (ISQ) for regional accounts, population estimates and business counts (Institut de la statistique du Québec, 2024; Gouvernement du Québec, n.d.a). Statistics Canada's economic regions for Quebec are by construction groupings of complete census divisions that coincide with the administrative regions (Statistics Canada, n.d.a); two census divisions (Brome-Missisquoi and La Haute-Yamaska) were transferred from Montérégie to Estrie on 28 July 2021, so Statistics Canada series on 2021 economic-region boundaries predate that change for those two regions while ISQ series and the current boundary files of the Ministère des Ressources naturelles et des Forêts (MRNF) reflect it. We use the MRNF boundaries (version 2026-10) and ISQ regional series throughout, and Statistics Canada series only where ISQ has no equivalent (Section 4.5).

Fabri-Ka. Fabri-Ka (`https://www.fabri-ka.com`) is a product-discovery aggregator built and operated by the author. It does not sell: each listing links to the merchant's own storefront. Its construction has two stages that matter for interpretation. First, a *discovery* pipeline assembles a registry of Quebec online stores from certification registries (Aliments du Québec, Les Produits du Québec, the CARTV organic register), sector associations (maple, cheese, cider, wine, distillery and microbrewery associations, the Conseil des métiers d'art), regional agri-food brands (Zone boréale, Créateurs de saveurs, Croquez Outaouais, Gaspésie gourmande and others), marketplaces and gift directories, and category-by-region search sweeps; each domain is visited, its e-commerce platform detected, its public catalogue endpoint tested, and Quebec location signals (postal codes, telephone area codes, wording) recorded. Second, *connectors* by e-commerce platform — Shopify's `products.json`, the WooCommerce Store API, Wix, Squarespace, Square, Votresite and a generic

JSON-LD reader — synchronise each store's public catalogue about every six hours, upserting listings by a content hash and retiring listings after two consecutive misses. Compliance per source (robots directives, terms, public endpoints only, removal on request) is documented in the platform's own compliance file and summarised in Appendix A.

Three features of the platform shape what the data can and cannot say. (i) The unit is a merchant's *published listing* at a point in time; the database keeps only the current price per listing, so there is no price history. (ii) Each store carries an *origin class* assigned from the discovery source (a maple producer listed by the producers' association is class A with confidence 0.95; a retailer listed by Les Produits du Québec is class C) and adjusted by site wording, with the textual evidence retained; this is a store-level prior, not product-level provenance. Les Produits du Québec's marks — "Produit du Québec" (at least 85% of direct production and manufacturing costs incurred in Quebec plus final substantial transformation), "Fabriqué au Québec" and "Conçu au Québec" — and the Aliments du Québec marks (Les Produits du Québec, n.d.; Gouvernement du Québec, n.d.b; Aliments du Québec, n.d.) are the only admissible origin evidence, and they attach to enterprises, not to listings. (iii) Discovery favours certified, association-member and search-indexed businesses and deliberately topped up Montréal through dedicated sources; it under-represents artisans without a web shop, industrial manufacturers selling direct, Etsy sellers (excluded by Etsy's terms) and Nord-du-Québec. The platform's public counters on 1 November 2026 reported 390,108 products, 1,569 live stores and 17 regions; the research sample derived below is smaller, and Section 4 traces every exclusion.

4. Data: snapshot, cleaning, provenance and representativeness

4.1. Snapshot and samples

The research copy is a read-only backup of the Fabri-Ka production database taken on 1 November 2026 at 17:38 UTC (application commit `689c7bb`); the only tables containing personal data (two user accounts, one self-registration) and the derived full-text index were dropped on the production node before the copy left it. The copy holds 465,388 product rows, of which 390,108 are active and published by the platform's own rules — gift cards, workshops, subscriptions, tickets and price-out-of-bounds quarantines are excluded upstream — in 1,569 enabled stores, plus 577,773 synchronisation events. Table 1 and Figure **??** trace the path to the two analysis samples. We drop one store flagged out of mission by the platform, keep one representative listing per (store, normalised title) because many merchants publish colour or format variants as separate products (19,245 non-representative listings), and remove eleven *mirror domains* — language or top-level-domain twins of the same business, detected by domain stem or by a title overlap of at least 90% of the smaller catalogue — which would double-count assortments and manufacture within-firm "cross-merchant" price pairs. The **assortment sample** has 370,054 listings in 1,557 stores and

twenty categories. The **price sample** further requires a Canadian-dollar store and listing, a positive posted price not "on request" and at most 50,000 CAD (9,635 listings have no posted price, 729 are in foreign currency, 78 exceed the cap): 359,612 listings. Within the assortment sample, 13.1% of listings are marked sold out by the merchant, 8.7% carry a strike-through price, 27.8% have several variants, 76.7% carry a merchant SKU and 75.6% a brand field.

Table 1. Sample flow from database rows to the analysis samples

Step	List
DB product rows (all statuses, all stores)	465
active = 1 and listing_status = 'published'	390
. . . of an enabled store with >= 1 published listing	390
. . . excluding store_kind = hors_mission	390
. . . keeping one representative per (store, normalised title) and non-empty title	370
. . . excluding mirror domains of the same business (sibling stores, non-representative) **[ASSORTMENT SAMPLE]**	370
. . . and CAD-priced store/listing	369
. . . and a positive listed price <= 50 000 CAD, not on request **[PRICE SAMPLE]**	359

Notes: Snapshot of 2026-11-01T17:38Z. Nothing is deleted: each step is a flag in the cleaned listing table. The assortment sample is used for H1–H4 and H7; the price sample for H5.

4.2. Merchant geography

The platform assigns a region from the discovery source (regional directories), from region mentions on the site, or from a map of forward sortation areas (FSA, the first three characters of the postal code); coordinates are FSA centroids obtained from a public geocoder, never civic addresses. At the snapshot 346 of the 1,569 live stores had no region, including the four largest catalogues (a book retailer, a souvenir printer, a trailer-parts seller and a pet-supply retailer). We recovered regions from official geography: for urban FSAs the sector label is matched to the MRNF municipality table (393 stores, 95.9% agreement with the point-in-polygon of the centroid where both exist), then the platform's own label is kept (865 stores), then the FSA centroid is located in the MRNF administrative-region polygons (82 stores). The order was amended after preregistration (A-001): a first attempt with the Statistics Canada 2021 economic-region layer produced 229 apparent conflicts between platform labels and polygons, of which 99 were a spelling artefact and most of the rest the Brome-Missisquoi/Haute-Yamaska transfer from Montérégie to Estrie that the 2021 layer predates; with the current MRNF layer the platform label and the centroid agree for 89.1% of stores, and the remaining disagreements (30 after the final rule) are rural FSAs whose centroids fall near a border, which is why the platform label is preferred to the centroid. After recovery, 229 stores (71,692 listings, 19% of the assortment sample) remain without a location; they come mainly from gift directories and search sweeps, are reported as a separate "Inconnue" row in descriptive tables, and

are excluded from all regional inference. One further caveat: the telephone field of 250 stores is the switchboard of the certification directory that listed them (212 carry the number of Les Produits du Québec); telephones are therefore never used for geography. Table 2 summarises the regional sample.

Table 2. Regional sample: stores, listings, certification and assortment diversity

Region	Stores	Listings	Median catalogue	Certified share	$\tilde{H}$ (M)	$\tilde{H}$ rarefied (15 stores) [2.5; 97.5
Montréal	237	61 419	65	0.59	0.834	0.779 [0.683; 0.857
Inconnue	226	71 692	77	0.07	–	-
Montérégie	190	57 945	32	0.59	0.853	0.786 [0.691; 0.864
Estrie	152	16 201	20	0.46	0.816	0.746 [0.643; 0.833
Capitale-Nationale	126	35 555	26	0.42	0.854	0.786 [0.674; 0.869
Chaudière-Appalaches	102	11 398	22	0.76	0.835	0.767 [0.664; 0.849
Laurentides	99	20 004	35	0.59	0.885	0.822 [0.739; 0.887
Saguenay–Lac-Saint-Jean	97	15 845	34	0.34	0.846	0.790 [0.696; 0.862
Lanaudière	66	8 828	34	0.56	0.831	0.774 [0.688; 0.845
Centre-du-Québec	61	9 356	19	0.41	0.824	0.767 [0.670; 0.848
Bas-Saint-Laurent	39	5 750	34	0.49	0.814	0.774 [0.691; 0.842
Outaouais	36	33 429	26	0.47	0.839	0.798 [0.725; 0.861
Mauricie	31	2 521	27	0.42	0.770	0.741 [0.679; 0.797
Gaspésie–Îles-de-la-Madeleine	30	7 514	28	0.30	0.837	0.799 [0.729; 0.860
Laval	25	6 528	33	0.84	0.805	0.785 [0.726; 0.839
Abitibi-Témiscamingue	21	5 206	36	0.38	0.870	0.845 [0.808; 0.879
Côte-Nord	16	587	18	0.38	0.791	0.788 [0.762; 0.802
Nord-du-Québec	3	276	27	0.00	0.655	-

Notes: Assortment sample. $\tilde{H}$ = normalised Shannon entropy of category shares (20 categories); M = merchant-equal weights. Rarefied entropy = mean over 2 000 draws of 15 stores without replacement; brackets give the 2.5th and 97.5th percentiles of the draws. 'Inconnue' = stores whose location could not be established (excluded from regional inference).

4.3. Category labels: audit and classifier (Amendment A-002)

The platform assigns one of twenty categories by keyword rules on the merchant's product type, tags, title and option names. We audited these labels on a stratified random sample of 500 listings (25 per category) annotated independently by two large language models (claude-opus-5-5 and gpt-5.5) given the platform's own taxonomy definitions; the two annotators agree with each other on 94.8% of listings but with the rule label on only 45.2% and 44.6% respectively, so the precision of the rule label against annotator consensus is 0.46 (Table 5). Errors are structured: the "other" bin is almost entirely resolvable into industrial, art or home goods; "art" absorbs industrial and craft supplies; beverages, alcohol, pets and outdoor goods are frequently mislabelled fashion or art when a merchant's tags are generic. Jewellery, garden and fashion are the most reliable rule labels.

4.4. Cross-merchant matching: candidates and validation

Tier-A (shared SKU) and tier-B (same brand, identical title) candidates number 1,837 and 457 pairs in different stores, of which 1,683 and 381 have a comparable variant structure; fuzzy tier C adds 446 (324 comparable) and identical-title-only tier D 2,713. On a validation sample of 60 pairs per tier, two independent annotators agree on 87.5% of pairs; among consensus pairs the share judged the same product is 0.92 for tier A, 0.90 for tier B, 0.77 for tier C and 0.17 for tier D (Table 3); restricted to comparable-variant pairs, tiers A and B reach 0.90 and 0.96. These are the precision figures we carry into the hedonic analysis; the tier-D figure confirms that identical titles across artisanal shops ("savon à la lavande") are not the same product.

Table 3. Validation of the cross-merchant matching tiers (60 pairs per tier)

Tier	Pairs with consensus	Same product	Different	Same product, comparable-variant pairs
A	56	0.92	0.02	0.90
B	54	0.90	0.00	0.96
C	52	0.77	0.10	0.77
D	48	0.17	0.60	0.19

Notes: Two independent LLM annotators (agreement 0.88; consensus on 88% of pairs). Tier A = shared merchant SKU; B = same brand and identical normalised title; C = same brand, fuzzy title; D = identical title without brand (never used for inference).

4.5. External data

Regional covariates and geography were downloaded from their official sources with a manifest of URLs, licences, vintages and checksums (Appendix A): ISQ population at 1 July 2025, ISQ business locations with employees in December 2025 (all industries, retail NAICS 44–45, manufacturing 31–33), ISQ disposable income per capita 2024 and regional GDP 2024, Statistics Canada Census 2021 land area and density, the MRNF administrative-region and municipality layers (version 2026-10) and the Statistics Canada 2021 FSA boundaries. No recent regional (as opposed to provincial) measure of household or business e-commerce adoption exists; the only regional split of household internet access is the 2012 ISQ survey, which we do not use.

4.6. Representativeness

Fabri-Ka's store counts track the regional business base closely in levels — Spearman correlations with ISQ business locations of 0.91 (all), 0.89 (retail) and 0.93 (manufacturing), and 0.84 with population — but coverage per thousand business locations varies six-fold across regions (1.8 to 11.1, coefficient of variation 0.43; Table 4), with the agri-food and craft regions covered by dedicated directories (Saguenay–Lac-Saint-Jean, Bas-Saint-Laurent, Gaspésie) over-represented

relative to Montréal and Laval. This is a discovery artefact, not a feature of the regional economies, and it is the main threat to every regional comparison below; Section **??** re-estimates the results on stores discovered through certification and association sources only and without Montréal.

Table 4. Coverage of the regional sample against external business and population counts

Region	Fabri-Ka stores	Business locations	Stores per 1 000 businesses	Population	Stores per 100 0
Bas-Saint-Laurent	39	7 441	5.24	204 755	
Saguenay–Lac-Saint-Jean	97	8 709	11.14	286 395	
Capitale-Nationale	126	24 239	5.20	814 004	
Mauricie	31	8 130	3.81	289 722	
Estrie	152	16 929	8.98	527 340	
Montréal	237	69 910	3.39	2 172 259	
Outaouais	36	8 566	4.20	428 751	
Abitibi-Témiscamingue	21	4 804	4.37	149 441	
Côte-Nord	16	2 672	5.99	89 621	
Nord-du-Québec	3	932	3.22	47 541	
Gaspésie–Îles-de-la-Madeleine	30	3 829	7.83	92 084	
Chaudière-Appalaches	102	15 715	6.49	459 158	
Laval	25	13 654	1.83	461 509	
Lanaudière	66	16 524	3.99	569 335	
Laurentides	99	22 002	4.50	679 190	
Montérégie	190	44 787	4.24	1 522 372	
Centre-du-Québec	61	9 270	6.58	264 820	

Notes: Business locations = ISQ, locations with employees, December 2025 (all industries; retail NAICS 44-45; manufacturing 31-33); population = ISQ 1 July 2025 provisional. Fabri-Ka stores are online storefronts of Quebec-based merchants discovered through certification registries, associations and search; coverage ratios are not sampling rates.

5. Conceptual framework and preregistered hypotheses

What a catalogue can and cannot reveal. Let s index merchants, $r(s)$ the administrative region of the merchant's business location, and c one of twenty product categories. A listing is a product a merchant chooses to publish online at the snapshot date. A region's online assortment is therefore the union of the publication decisions of the merchants located there, filtered by the aggregator's discovery process. Three forces shape it: the regional production and retail base (which merchants exist), the decision to sell online and to publish a catalogue (adoption), and the aggregator's coverage (discovery). Only the first is of economic interest; the other two are nuisance and are the object of the representativeness audit. Because the unit is a listing and not a sale, an employee or a shipment, all quantities below are *assortment* quantities, and because merchant location is not product origin, all regional statements concern merchants based in a region, not goods made there.

Hypotheses. The hypotheses, estimands, specifications, exclusions and decision rules were frozen in a preregistration document before any Phase 4 statistic was computed (SHA-256 recorded in the decision log; two amendments, A-001 on the region-assignment rule and A-002 on category labels, are reported in Section 4 and were decided before seeing results under the amended definitions).

H1 — Regional specialization (confirmatory).

Assortment composition differs across regions beyond what a random allocation of merchants to regions would produce. Statistic: Cramér's V of the region×category table; null: 2,000 permutations of region labels across stores. Decision: $p < 0.05$ (Holm-adjusted) under both merchant-equal and trimmed-listing weights.

H2 — Diversity and scale (confirmatory, low power).

Rarefied assortment diversity (entropy at fifteen stores) is associated with regional population. Statistic: Spearman ρ with a bootstrap interval; supported if the interval excludes zero.

H3 — Concentration (confirmatory).

Catalogue sizes are highly skewed: Gini ≥ 0.70 and top-decile share ≥ 0.50, with bootstrap intervals.

H4 — Related diversification (exploratory).

Within-store entry into categories absent at baseline increases with relatedness density; region-level entries are counted and declared untestable if fewer than ten.

H5 — Price dispersion (confirmatory, two tests).

For matched products, posted prices are (T1) geographically structured — joint wild-cluster-bootstrap test of region effects — and (T2) competitively structured — reseller versus producer effect.

H6 — Local-origin premium.

Declared *untestable* ex ante: the source has no product-level provenance. A store-certification contrast is reported descriptively and labelled as not being H6.

H7 — Digital-assortment complexity (validation).

A region–category complexity indicator is reported only if it passes bootstrap rank stability ($\rho \geq 0.8$), a store-reshuffle null, and a leave-one-region-out predictive check; otherwise it is reported as unstable and excluded from the index family.

The confirmatory family is $\{$H1, H2, H3, H5-T1, H5-T2$\}$ with Holm step-down at $\alpha = 0.05$; raw p-values are also reported. Every causal reading is excluded by design: the data are a single cross-section of posted listings from a non-probability discovery process.

6. Empirical methods and identification limits

6.1. Weighting schemes

Catalogue sizes are extremely skewed (Section **??**): the ten largest stores hold more than a third of all listings. Every regional statistic is therefore computed under three schemes. Under **L** each listing counts one. Under **M** (merchant-equal, the primary scheme for regional composition) each listing counts $1/n_s$, so that every store contributes one unit regardless of catalogue size. Under **T** listings count one after dropping the one per cent largest stores. The schemes answer different questions — L describes what a visitor sees, M describes the population of merchants — and we report where they disagree.

6.2. Revealed assortment advantage and the merchant-reshuffle null

For region r and category c let N_{rc} be the (weighted) count of listings. The revealed assortment advantage is

$$\mathrm{RAA}_{rc} = \frac{N_{rc}/\sum_{c} N_{rc}}{\sum_{r} N_{rc}/\sum_{rc} N_{rc}}, \tag{1}$$

the Balassa index on listings. Unlike export RCA it does not measure production, trade or competitiveness; it measures what the merchants of a region publish relative to what Quebec merchants publish overall. Dependence between region and category is summarised by Cramér's V of the $R \times C$ table and tested against a null that reshuffles region labels across *stores* while keeping every store's catalogue and every region's number of stores fixed; the p-value is the share of 2,000 permutations with V at least as large as observed. This is the dartboard logic of Ellison and Glaeser (1997) transposed to merchants: a region may look specialised simply because one large merchant is located there, and the null absorbs this. Uncertainty for each RAA_{rc} comes from 2,000 bootstrap resamples of stores; because resampling with replacement duplicates stores and biases V upward, bootstrap intervals are reported for the cells and not for V, whose inference is by permutation.

6.3. Diversity and concentration

Regional diversity is the normalised Shannon entropy of category shares, $\tilde{H}_r = -\sum_{c} p_{rc} \ln p_{rc} / \ln 20$, the effective number of categories $\exp(H_r)$ and the Herfindahl index, under L and M. Because entropy increases mechanically with the number of merchants, the primary H2 outcome is the *rarefied* entropy: the mean over 2,000 draws of fifteen stores without replacement, which puts regions with 15 and 235 stores on the same footing; regions with fewer than fifteen stores are excluded from that statistic. Concentration is the Gini coefficient of listings across stores, top-percentile shares, the rank–size slope, and a discrete power-law fit with $x_{\min}$ chosen by KS minimisation compared with a

log-normal on the same tail by a Vuong likelihood-ratio test (following Ellison and Glaeser, 1997).

6.4. Region–category network and complexity indicator

Let $M_{rc} = \mathbf{1}[\mathrm{RAA}_{rc} \geq 1]$ under scheme M. Diversity is the row sum, ubiquity the column sum; the ECI is the second eigenvector of $\tilde{M} = D_r^{-1} M D_c^{-1} M^\top$ (Hidalgo and Hausmann, 2009; Mealy et al., 2019), with the sign fixed so that it correlates positively with diversity; the fitness index is the fixed point of the non-linear iteration of Tacchella et al. (2012). Category relatedness is the Hidalgo-type proximity $\phi_{cc'} = \min\{P(c \mid c'), P(c' \mid c)\}$ computed from co-occurrence of categories across *stores*, and communities are obtained by Louvain over 200 seeds. Validation follows the preregistration: (V1) the Spearman correlation between the full-sample ECI and 500 store-bootstrap ECIs; (V2) the second eigenvalue of $\tilde{M}$ and the spectral radius of the bipartite adjacency matrix against 2,000 store-level region reshuffles (a merchant-preserving analogue of the null models in Saracco et al., 2015); (V3) the leave-one-region-out predictive correlation of the ECI with regional population, business locations and income; (V4) the normalised mutual information between community partitions across seeds.

6.5. Matched products and hedonic dispersion

No GTIN exists in the source. Cross-merchant identity is inferred in tiers — A: shared merchant SKU and token-set title similarity ≥ 70; B: same normalised brand and identical normalised title; C: same brand, same category and token-sort similarity ≥ 90; D: identical title without brand — restricted to listings in the price sample and to pairs of different stores that are not mirror domains of one business. Tiers A and B form the primary sample and C a sensitivity; D is never used for inference. Precision is estimated on a labelled sample of 240 pairs (Section 4). For listing i in match group g sold by store m in region $r(m)$,

$$\ln P_i = \alpha_g + \gamma_{r(m)} + \delta_{\mathrm{platform}(m)} + \beta_1 \ln n_m + \beta_2 \,\mathrm{Reseller}_m + \beta_3 \,\mathrm{Certified}_m + \varepsilon_i, \tag{2}$$

estimated by within-group least squares with standard errors clustered by store. Because each store belongs to one region, a store fixed effect cannot be included alongside γ_r: region effects are identified only through match groups that span regions, which is why the sample size in *stores* and the number of regions with at least five matched stores are reported, and why inference uses a wild cluster bootstrap (Rademacher weights by store, 9,999 draws, restricted residuals) for the joint test of γ_r and for β_2. Prices are posted minimum-variant prices at the snapshot; shipping, taxes and transactions are not observed.

6.6. Dynamics

The only temporal margin is listing entry and exit after store onboarding ended (24 August 2026). Weekly presence of each (store, category) pair is reconstructed from first and last observation dates; relatedness density is computed from the baseline weeks 36–39 and entry is measured over weeks 40–44, a temporal holdout that prevents the outcome from feeding the regressor. A linear probability model with store and category fixed effects and store-clustered standard errors is compared with a placebo that relabels categories in the proximity matrix. The exercise is exploratory by preregistration.

6.7. External data and the representativeness audit

Regional covariates come from the ISQ (population at 1 July 2025, business locations with employees in December 2025 by NAICS, disposable income per capita 2024, regional GDP 2024) and from Statistics Canada (Census 2021 land area and density on 2021 economic-region boundaries); vintages and concordance issues are recorded in the external-data manifest. Coverage is the number of Fabri-Ka stores per thousand business locations and per 100,000 inhabitants; the audit also tabulates discovery source and platform by region and category mix by platform, following the data-defect logic of Meng (2018).

7. results

TODO.

8. robustness

TODO.

9. discussion

TODO.

10. conclusion

TODO.

Reproducibility statement

The analysis is a deterministic pipeline (seed 20261101) run from a single `Makefile`: `make ingest clean match validate features models robust tables figures paper`, with `make test` running the invariant tests on the cleaned panel. The raw snapshot is an owner-authorised, read-only copy of the Fabri-Ka database taken with `sqlite3.Connection.backup()` on 2026-11-01T17:38:30Z (SHA-256 `8d963ba9...98fb`); the two tables containing personal data were dropped on the production node before transfer and nothing in production was modified. The snapshot is not redistributed (it contains merchant catalogues that belong to the merchants); the data dictionary, the sample-flow table, every intermediate aggregate, the labelled audit samples, the external-data manifest with URLs, licences, vintages and checksums, the run ledger with every API call and its cost, and the SHA-256 of every output are in the repository. LLM annotations (two independent models for the category audit and the matching validation; one model for the classifier training labels) are stored as CSV files so that the pipeline runs without API access; no empirical result depends on a live model call. Dependencies are pinned in `pyproject.toml` (Python 3.12).

References

Académie de la transformation numérique, "NETendances — enquêtes sur les usages numériques des Québécois," Technical Report, Académie de la transformation numérique, Université Laval, Québec n.d. Accessed 2026-11-01.

Aliments du Québec, "Aliments du Québec — site officiel," Technical Report, Aliments du Québec, Montréal n.d. Accessed 2026-11-01.

Anderson, Chris, *The Long Tail: Why the Future of Business Is Selling Less of More*, Hyperion, 2006. Accessed 2026-11-01.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs," *SSRN Electronic Journal*, 2015.

Bajari, Patrick and C. Lanier Benkard, "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, 2005, *113* (6), 1239–1276.

Balassa, Bela, "Trade Liberalisation and "Revealed" Comparative Advantage ¹," *The Manchester School*, 1965, *33* (2), 99–123.

Baldwin, John R., Desmond Beckstead, W. Mark Brown, and David L. Rigby, "Agglomeration and the Geography of Localization Economies in Canada," *Regional Studies*, 2008, *42* (1), 117–132.

Balland, Pierre-Alexandre and David Rigby, "The Geography of Complex Knowledge," *Economic Geography*, 2016, *93* (1), 1–23.

_ , Ron Boschma, Joan Crespo, and David L. Rigby, "Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification," *Regional Studies*, 2018, *53* (9), 1252–1268.

_ , Tom Broekel, Dario Diodato, Elisa Giuliani, Ricardo Hausmann, Neave O'Clery, and David Rigby, "The new paradigm of economic complexity," *Research Policy*, 2022, *51* (3), 104450.

Bar-Isaac, Heski, Guillermo Caruana, and Vicente Cuñat, "Search, Design, and Market Structure," *American Economic Review*, 2012, *102* (2), 1140–1160.

Baye, Michael R., John Morgan, and Patrick Scholten, "PRICE DISPERSION IN THE SMALL AND IN THE LARGE: EVIDENCE FROM AN INTERNET PRICE COMPARISON SITE," *The Journal of Industrial Economics*, 2004, *52* (4), 463–496.

_ , _ , and _ , "Chapter 6 Information, Search, and Price Dispersion," in "Handbooks in Information Systems Economics and Information Systems," Elsevier, 2006, pp. 323–375.

Beaumont, Jean-François, "Are probability surveys bound to disappear for the production of official statistics?," *Survey Methodology*, 2020, *46* (1), 1–28. Accessed 2026-11-01.

Bettencourt, Luís M. A., José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West, "Growth, innovation, scaling, and the pace of life in cities," *Proceedings of the National Academy of Sciences*, 2007, *104* (17), 7301–7306.

Blum, Bernardo S. and Avi Goldfarb, "Does the internet defy the law of gravity?," *Journal of International Economics*, 2006, *70* (2), 384–405.

Boschma, Ron, "Relatedness as driver of regional diversification: a research agenda," *Regional Studies*, 2016, *51* (3), 351–364.

Bound, John, Charles Brown, and Nancy Mathiowetz, "Measurement Error in Survey Data," in "Handbook of Econometrics," Elsevier, 2001, pp. 3705–3843.

Bradley, Valerie C., Shiro Kuriwaki, Michael Isakov, Dino Sejdinovic, Xiao-Li Meng, and Seth Flaxman, "Unrepresentative big surveys significantly overestimated US vaccine uptake," *Nature*, 2021, *600* (7890), 695–700.

Briant, A., P.-P. Combes, and M. Lafourcade, "Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?," *Journal of Urban Economics*, 2010, *67* (3), 287–302.

Brynjolfsson, Erik and Michael D. Smith, "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 2000, *46* (4), 563–585.

_ , Yu (Jeffrey) Hu, and Duncan Simester, "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales," *Management Science*, 2011, *57* (8), 1373–1386.

_ , _ , and Michael D. Smith, "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 2003, *49* (11), 1580–1596.

_ , _ , **and Mohammad S. Rahman**, "Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition," *Management Science*, 2009, *55* (11), 1755–1765.

Buelens, Bart, Joep Burger, and Jan A. van den Brakel, "Comparing Inference Methods for Non-probability Samples," *International Statistical Review*, 2018, *86* (2), 322–343.

Cavallo, Alberto, "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers," *American Economic Review*, 2017, *107* (1), 283–303.

_ , "Scraped Data and Sticky Prices," *The Review of Economics and Statistics*, 2018, *100* (1), 105–119.

_ **and Oleksiy Kryvtsov**, "What can stockouts tell us about inflation? Evidence from online micro data," *Journal of International Economics*, 2023, *146*, 103769.

_ **and Roberto Rigobon**, "The Billion Prices Project: Using Online Prices for Measurement and Research," *Journal of Economic Perspectives*, 2016, *30* (2), 151–178.

_ **, Brent Neiman, and Roberto Rigobon**, "Currency Unions, Product Introductions, and the Real Exchange Rate," *The Quarterly Journal of Economics*, 2014, *129* (2), 529–595.

Choi, Jeonghye and David R. Bell, "Preference Minorities and the Internet," *Journal of Marketing Research*, 2011, *48* (4), 670–682.

Christen, Peter, *Data Matching*, Springer Berlin Heidelberg, 2012.

Clay, Karen, Ramayya Krishnan, and Eric Wolff, "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry," *The Journal of Industrial Economics*, 2001, *49* (4), 521–539.

Content, Jeroen and Koen Frenken, "Related variety and economic development: a literature review," *European Planning Studies*, 2016, *24* (12), 2097–2112.

Couture, Victor, Benjamin Faber, Yizhen Gu, and Lizhi Liu, "Connecting the Countryside via E-Commerce: Evidence from China," *American Economic Review: Insights*, 2021, *3* (1), 35–50.

Cristelli, Matthieu, Andrea Gabrielli, Andrea Tacchella, Guido Caldarelli, and Luciano Pietronero, "Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products," *PLoS ONE*, 2013, *8* (8), e70726.

Daas, Piet J.H., Marco J. Puts, Bart Buelens, and Paul A.M. van den Hurk, "Big Data as a Source for Official Statistics," *Journal of Official Statistics*, 2015, *31* (2), 249–262.

Davis, Donald R. and Jonathan I. Dingel, "The comparative advantage of cities," *Journal of International Economics*, 2020, *123*, 103291.

DellaVigna, Stefano and Matthew Gentzkow, "Uniform Pricing in U.S. Retail Chains*," *The Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan, "Consumer Price Search and Platform Design in Internet Commerce," *American Economic Review*, 2018, *108* (7), 1820–1859.

Dolfen, Paul, Liran Einav, Peter J. Klenow, Benjamin Klopack, Jonathan D. Levin, Larry Levin, and Wayne Best, "Assessing the Gains from E-Commerce," *American Economic Journal: Macroeconomics*, 2023, *15* (1), 342–370.

Duranton, Gilles and Diego Puga, "Chapter 48 Micro-foundations of urban agglomeration economies," in "Handbook of Regional and Urban Economics Cities and Geography," Elsevier, 2004, pp. 2063–2117.

_ **and** _ , "The Economics of Urban Density," *Journal of Economic Perspectives*, 2020, *34* (3), 3–26.

_ **and Henry G. Overman**, "Testing for Localization Using Micro-Geographic Data," *The Review of Economic Studies*, 2005, *72* (4), 1077–1106.

Einav, Liran, Chiara Farronato, and Jonathan Levin, "Peer-to-Peer Markets," *Annual Review of Economics*, 2016, *8* (1), 615–635.

Elberse, Anita, "Should You Invest in the Long Tail?," *Harvard Business Review*, 2008, *86* (7/8), 88–96. Accessed 2026-11-01.

Elliott, Michael R. and Richard Valliant, "Inference for Nonprobability Samples," *Statistical Science*, 2017, *32* (2), 249–264.

Ellison, Ellison, "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 2009, *77* (2), 427–452.

Ellison, Glenn and Edward L. Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 1997, *105* (5), 889–927.

Fellegi, Ivan P. and Alan B. Sunter, "A Theory for Record Linkage," *Journal of the American Statistical Association*, 1969, *64* (328), 1183–1210.

Forman, Chris, Anindya Ghose, and Avi Goldfarb, "Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live," *Management Science*, 2009, *55* (1), 47–57.

_ , Avi Goldfarb, and Shane Greenstein, "How did location affect adoption of the commercial Internet? Global village vs. urban leadership," *Journal of Urban Economics*, 2005, *58* (3), 389–420.

_ , _ , and _ , "The Internet and Local Wages: A Puzzle," *American Economic Review*, 2012, *102* (1), 556–575.

Frenken, Koen, Frank Van Oort, and Thijs Verburg, "Related Variety, Unrelated Variety and Regional Economic Growth," *Regional Studies*, 2007, *41* (5), 685–697.

Friedman, Konny Williams, "Big Data in the U.S. Consumer Price Index," in "Big Data for Twenty-First-Century Economic Statistics," University of Chicago Press, 2022, pp. 69–98.

Gao, Jian and Tao Zhou, "Quantifying China's regional economic complexity," *Physica A: Statistical Mechanics and its Applications*, 2018, *492*, 1591–1603.

Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman, and Andrei Shleifer, "Growth in Cities," *Journal of Political Economy*, 1992, *100* (6), 1126–1152.

Gobillon, Combes, "The Empirics of Agglomeration Economies," in "Handbook of Regional and Urban Economics," Elsevier, 2015, pp. 247–348.

Goldfarb, Avi and Catherine Tucker, "Digital Economics," *Journal of Economic Literature*, 2019, *57* (1), 3–43.

_ and Jeff Prince, "Internet adoption and usage patterns are different: Implications for the digital divide," *Information Economics and Policy*, 2008, *20* (1), 2–15.

Goldmanis, Maris, Ali Hortaçsu, Chad Syverson, and Önsel Emre, "E-Commerce and the Market Structure of Retail Industries," *The Economic Journal*, 2009, *120* (545), 651–682.

Gorodnichenko, Yuriy and Oleksandr Talavera, "Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration," *American Economic Review*, 2017, *107* (1), 249–282.

_ , Viacheslav Sheremirov, and Oleksandr Talavera, "Price Setting in Online Markets: Does IT Click?," *Journal of the European Economic Association*, 2018, *16* (6), 1764–1811.

Gouvernement du Québec, "Portraits des régions administratives," Technical Report, Gouvernement du Québec, Québec n.d. Accessed 2026-11-01.

_ , "Reconnaître les aliments québécois," Technical Report, Gouvernement du Québec, Québec n.d. Accessed 2026-11-01.

Greenstein, Shane, Chris Forman, and Avi Goldfarb, "How Geography Shapes—and Is Shaped by—the Internet," in "Oxford Handbooks Online," Oxford University Press, 2018.

Groves, Robert M., "Nonresponse Rates and Nonresponse Bias in Household Surveys," *Public Opinion Quarterly*, 2006, *70* (5), 646–675.

Haight, Michael, Anabel Quan-Haase, and Bradley A Corbett, "Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet, level of online activity, and social networking site usage," *Information, Communication & Society*, 2014, *17* (4), 503–519.

Haller, Stefanie A. and Sean Lyons, "Effects of broadband availability on total factor productivity in service sector firms: Evidence from Ireland," *Telecommunications Policy*, 2019, *43* (1), 11–22.

_ **and Seán Lyons**, "Broadband adoption and firm productivity: Evidence from Irish manufacturing firms," *Telecommunications Policy*, 2015, *39* (1), 1–13.

Handbury, J. and D. E. Weinstein, "Goods Prices and Availability in Cities," *The Review of Economic Studies*, 2014, *82* (1), 258–296.

Handbury, Jessie, "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities," *Econometrica*, 2021, *89* (6), 2679–2715.

Hausman, Jerry, "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left," *Journal of Economic Perspectives*, 2001, *15* (4), 57–67.

Hausmann, Ricardo and César A. Hidalgo, "The network structure of economic output," *Journal of Economic Growth*, 2011, *16* (4), 309–342.

_ **,** _ **, Sebastián Bustos, Michele Coscia, Alexander Simoes, and Muhammed A. Yildirim**, *The Atlas of Economic Complexity*, The MIT Press, 2014.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153.

Henderson, Vernon, Ari Kuncoro, and Matt Turner, "Industrial Development in Cities," *Journal of Political Economy*, 1995, *103* (5), 1067–1090.

Hidalgo, C. A., B. Klinger, A.-L. Barabási, and R. Hausmann, "The Product Space Conditions the Development of Nations," *Science*, 2007, *317* (5837), 482–487.

Hidalgo, César A., "Economic complexity theory and applications," *Nature Reviews Physics*, 2021, *3* (2), 92–113.

_ **and Ricardo Hausmann**, "The building blocks of economic complexity," *Proceedings of the National Academy of Sciences*, 2009, *106* (26), 10570–10575.

_ **, Pierre-Alexandre Balland, Ron Boschma, Mercedes Delgado, Maryann Feldman, Koen Frenken, Edward Glaeser, Canfei He, Dieter F. Kogler, Andrea Morrison, Frank Neffke, David Rigby, Scott Stern, Siqi Zheng, and Shengjun Zhu**, "The Principle of Relatedness," in "Springer Proceedings in Complexity Unifying Themes in Complex Systems IX," Springer International Publishing, 2018, pp. 451–457.

Hinloopen, Jeroen and Charles Marrewijk, "On the empirical distribution of the Balassa index," *Weltwirtschaftliches Archiv*, 2001, *137* (1), 1–35.

Holmes, Thomas J., Masahisa Fujita, Paul Krugman, and Anthony J. Venables, "The Spatial Economy: Cities, Regions, and International Trade," *Southern Economic Journal*, 2000, *67* (2), 491.

Hong, Inho, Morgan R. Frank, Iyad Rahwan, Woo-Sung Jung, and Hyejin Youn, "The universal pathway to innovative urban economies," *Science Advances*, 2020, *6* (34).

Hortaçsu, Ali, F. Asís Martínez-Jerez, and Jason Douglas, "The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre," *American Economic Journal: Microeconomics*, 2009, *1* (1), 53–74.

Institut de la statistique du Québec, "Panorama des régions du Québec. Édition 2024," Technical Report, Institut de la statistique du Québec, Québec 2024. Accessed 2026-11-01.

Jo, Yoon, Misaki Matsumura, and David Weinstein, "The Impact of E-Commerce on Relative Prices and Consumer Welfare," NBER Working Paper 26506, National Bureau of Economic Research 2019.

Kaplan, Greg and Guido Menzio, "THE MORPHOLOGY OF PRICE DISPERSION," *International Economic Review*, 2015, *56* (4), 1165–1206.

_ **,** _ **, Leena Rudanko, and Nicholas Trachter**, "Relative Price Dispersion: Evidence and Theory," *American Economic Journal: Microeconomics*, 2019, *11* (3), 68–124.

Kogler, Dieter F., David L. Rigby, and Isaac Tucker, "Mapping Knowledge Space and Technological Relatedness in US Cities," *European Planning Studies*, 2013, *21* (9), 1374–1391.

Kolko, Jed, "Broadband and local growth," *Journal of Urban Economics*, 2012, *71* (1), 100–113.

Krugman, Paul, "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 1980, *70* (5), 950–959. Accessed 2026-11-01.

_ , "Increasing Returns and Economic Geography," *Journal of Political Economy*, 1991, *99* (3), 483–499.

Laursen, Keld, "Revealed comparative advantage and the alternatives as measures of international specialization," *Eurasian Business Review*, 2015, *5* (1), 99–115.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis," *Science*, 2014, *343* (6176), 1203–1205.

Lendle, Andreas, Marcelo Olarreaga, Simon Schropp, and Pierre-Louis Vézina, "There Goes Gravity: eBay and the Death of Distance," *The Economic Journal*, 2016, *126* (591), 406–441.

Les Produits du Québec, "Processus de vérification — marques de certification Produit du Québec, Fabriqué au Québec, Conçu au Québec," Technical Report, Les Produits du Québec n.d. Accessed 2026-11-01.

Lieber, Ethan and Chad Syverson, "Online versus Offline Competition," in "Oxford Handbooks Online," Oxford University Press, 2012.

Little, Roderick and Donald Rubin, *Statistical Analysis with Missing Data, Third Edition*, Wiley, 2019.

Marshall, Alfred, *Principles of Economics*, Macmillan (8th ed.; online edition, Library of Economics and Liberty), 1920. Accessed 2026-11-01.

Mealy, Penny, J. Doyne Farmer, and Alexander Teytelboym, "Interpreting economic complexity," *Science Advances*, 2019, *5* (1).

Meng, Xiao-Li, "Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election," *The Annals of Applied Statistics*, 2018, *12* (2), 685–726.

Mori, Tomoya, Koji Nishikimi, and Tony E. Smith, "A Divergence Statistic for Industrial Localization," *Review of Economics and Statistics*, 2005, *87* (4), 635–651.

Morrison, Greg, Sergey V. Buldyrev, Michele Imbruno, Omar Alonso Doria Arrieta, Armando Rungi, Massimo Riccaboni, and Fabio Pammolli, "On Economic Complexity and the Fitness of Nations," *Scientific Reports*, 2017, *7* (1).

Muneepeerakul, Rachata, José Lobo, Shade T. Shutters, Andrés Goméz-Liévano, and Murad R. Qubbaj, "Urban Economies and Occupation Space: Can They Get "There" from "Here"?," *PLoS ONE*, 2013, *8* (9), e73676.

Neffke, Frank, Martin Henning, and Ron Boschma, "How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions," *Economic Geography*, 2011, *87* (3), 237–265.

OECD, *Unpacking E-commerce*, OECD Publishing, 2019.

Pakes, Ariel, "A Reconsideration of Hedonic Price Indexes with an Application to PC's," *American Economic Review*, 2003, *93* (5), 1578–1596.

Polèse, Mario and Richard Shearmur, "Is Distance Really Dead? Comparing Industrial Location Patterns over Time in Canada," *International Regional Science Review*, 2004, *27* (4), 431–457.

_ **and** _ , "Why some regions will decline: A Canadian case study with thoughts on local development strategies," *Papers in Regional Science*, 2006, *85* (1), 23–46.

Quan, Thomas W. and Kevin R. Williams, "Product variety, across-market demand heterogeneity, and the value of online retail," *The RAND Journal of Economics*, 2018, *49* (4), 877–913.

Rao, J. N. K., "On Making Valid Inferences by Integrating Data from Surveys and Other Sources," *Sankhya B*, 2020, *83* (1), 242–272.

Rigby, David L., "Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes," *Regional Studies*, 2013, *49* (11), 1922–1937.

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 1974, *82* (1), 34–55.

Rosenthal, Stuart S. and William C. Strange, "The Determinants of Agglomeration," *Journal of Urban Economics*, 2001, *50* (2), 191–229.

_ **and** _ , "Chapter 49 Evidence on the nature and sources of agglomeration economies," in "Handbook of Regional and Urban Economics Cities and Geography," Elsevier, 2004, pp. 2119–2171.

Saracco, Fabio, Riccardo Di Clemente, Andrea Gabrielli, and Tiziano Squartini, "Randomizing bipartite networks: the case of the World Trade Web," *Scientific Reports*, 2015, *5* (1).

Sciarra, Carla, Guido Chiarotti, Luca Ridolfi, and Francesco Laio, "Reconciling contrasting views on economic complexity," *Nature Communications*, 2020, *11* (1).

Servedio, Vito D. P., Paolo Buttà, Dario Mazzilli, Andrea Tacchella, and Luciano Pietronero, "A New and Stable Estimation Method of Country Economic Fitness and Product Complexity," *Entropy*, 2018, *20* (10), 783.

Shearmur, Richard and David Doloreux, "Urban Hierarchy or Local Buzz? High-Order Producer Service and (or) Knowledge-Intensive Business Service Location in Canada, 1991-2001," *The Professional Geographer*, 2008, *60* (3), 333–355.

_ **and Mario Polèse**, "Do Local Factors Explain Local Employment Growth? Evidence from Canada, 1971–2001," *Regional Studies*, 2007, *41* (4), 453–471.

Sinai, Todd and Joel Waldfogel, "Geography and the Internet: is the Internet a substitute or a complement for cities?," *Journal of Urban Economics*, 2004, *56* (1), 1–24.

Statistics Canada, "Canadian Internet Use Survey, 2022," Technical Report, The Daily, Statistics Canada, Ottawa (released 2023-07-20) 2023. Accessed 2026-11-01.

_ , "Business Register (BR) — Integrated Metadata Base, record number 1105," Technical Report, Statistics Canada, Ottawa n.d. Accessed 2026-11-01.

_ , "Canadian Internet Use Survey (CIUS) — Integrated Metadata Base, record number 4432," Technical Report, Statistics Canada, Ottawa n.d. Accessed 2026-11-01.

_ , "Survey of Digital Technology and Internet Use (SDTIU) — Integrated Metadata Base, record number 4225," Technical Report, Statistics Canada, Ottawa n.d. Accessed 2026-11-01.

Tacchella, Andrea, Matthieu Cristelli, Guido Caldarelli, Andrea Gabrielli, and Luciano Pietronero, "A New Metrics for Countries' Fitness and Products' Complexity," *Scientific Reports*, 2012, *2* (1).

Tan, Tom Fangyun, Serguei Netessine, and Lorin Hitt, "Is Tom Cruise Threatened? An Empirical Study of the Impact of Product Variety on Demand Concentration," *Information Systems Research*, 2017, *28* (3), 643–660.

Waldfogel, Joel, "How Digitization Has Created a Golden Age of Music, Movies, Books, and Television," *Journal of Economic Perspectives*, 2017, *31* (3), 195–214.

Whitacre, Brian, Roberto Gallardo, and Sharon Strover, "Broadband's contribution to economic growth in rural areas: Moving towards a causal relationship," *Telecommunications Policy*, 2014, *38* (11), 1011–1023.

Youn, Hyejin, Luís M. A. Bettencourt, José Lobo, Deborah Strumsky, Horacio Samaniego, and Geoffrey B. West, "Scaling and universality in urban economic diversification," *Journal of The Royal Society Interface*, 2016, *13* (114), 20150937.

Yu, Run, Junning Cai, and PingSun Leung, "The normalized revealed comparative advantage index," *The Annals of Regional Science*, 2008, *43* (1), 267–282.

Zentner, Alejandro, Michael Smith, and Cuneyd Kaya, "How Video Rental Patterns Change as Consumers Move Online," *Management Science*, 2013, *59* (11), 2622–2634.

Zhu, Feng and Qihong Liu, "Competing with complementors: An empirical look at Amazon.com," *Strategic Management Journal*, 2018, *39* (10), 2618–2642.

Appendices

A. Source audit and compliance

The Phase 0 source audit (`reports/00_source_audit.md`) documents the three tables kept from the snapshot (`stores`, 4,082 rows; `products`, 465,388; `sync_log`, 577,773), their fields, missingness and the platform's own compliance review per source: public catalogue endpoints only, polite rates, nothing behind authentication, retained attribution and removal on request. The research used no scraping of the platform and a single control request to its public statistics endpoint to date-stamp the counters.

B. Data dictionary

The full field-level dictionary of the cleaned tables (`listings_clean`, `stores_master`, `sibling_stores`, `match_pairs`, `match_groups`, `sync_log`) is generated by the pipeline (`reports/data_dictionary.md`). Key derived fields: `title_norm` (ASCII-folded, lower-cased title used for duplicate detection and matching); `f_analysis_assort` and `f_analysis_price` (sample membership flags); `region_final` and `region_source` (assigned region and its provenance: municipality lookup, platform label, FSA-centroid point-in-polygon, or unknown); `category_model` (classifier label, Amendment A-002); `comparable` (variant-structure comparability of a matched pair).

C. Category taxonomy and audit details

Table 5. Audit of the rule-based category labels (500-listing stratified sample)

Rule label	Listings with annotator consensus	Precision of the rule label
Maple & honey	24	0.50
Fine grocery	24	0.42
Coffee & tea	23	0.35
Chocolate	24	0.62
Bakery	25	0.40
Meat & fish	22	0.32
Cheese & dairy	24	0.58
Beverages	24	0.17
Alcohol	25	0.28
Beauty & care	24	0.62
Home & decor	22	0.36
Fashion	25	0.72
Jewellery	25	0.80
Art & craft	25	0.52
Children	21	0.62
Pets	24	0.29
Outdoor & sport	19	0.32
Garden	25	0.84
Industrial	24	0.42
Other	25	0.00
All	474	**0.46**

Notes: 25 random listings per category annotated independently by two LLMs (claude-opus-5-5, gpt-5.5) with the platform's taxonomy; agreement rule vs A 0.45, rule vs B 0.45, A vs B 0.95; consensus exists for 95% of listings. Precision = share of consensus labels equal to the rule label. Disagreements listed in the replication package.

D. Within-region concentration

E. Region-level indicators

F. Preregistration, decision log and amendments

The frozen preregistration (`research/preregistration.md`), the decision log (`research/decisions.md`, D-001 to D-007) and the two amendments are reproduced in the repository. A-001 (region assignment) replaced "platform label first" by "official geography first" after a first attempt with the Statistics Canada 2021 economic-region layer produced spurious Estrie/Montérégie conflicts that predate the 2021 boundary change; the final rule uses the MRNF municipality table for urban forward sortation areas, then the platform label, then the FSA-centroid point-in-polygon. A-002 (category labels) introduced a supervised classifier after the audit measured the keyword rules at 46% precision, with the primacy rule stated before any result was computed under the new labels.

Table 6. Within-region concentration of catalogue sizes

Region	Stores	Gini [95% CI]	Top-10% share
Outaouais	36	0.922 [0.632; 0.937]	0.946
Capitale-Nationale	126	0.891 [0.751; 0.929]	0.866
Gaspésie–Îles-de-la-Madeleine	30	0.891 [0.472; 0.912]	0.876
Montérégie	190	0.890 [0.724; 0.925]	0.854
Centre-du-Québec	61	0.885 [0.519; 0.921]	0.861
Laval	25	0.830 [0.492; 0.868]	0.809
Estrie	152	0.823 [0.680; 0.882]	0.756
Chaudière-Appalaches	102	0.815 [0.661; 0.857]	0.776
Laurentides	99	0.810 [0.652; 0.871]	0.725
Abitibi-Témiscamingue	21	0.806 [0.587; 0.850]	0.795
Bas-Saint-Laurent	39	0.806 [0.447; 0.854]	0.776
Saguenay–Lac-Saint-Jean	97	0.803 [0.618; 0.846]	0.736
Montréal	237	0.782 [0.675; 0.843]	0.689
Lanaudière	66	0.765 [0.645; 0.809]	0.679
Mauricie	31	0.730 [0.495; 0.792]	0.701
Côte-Nord	16	0.517 [0.318; 0.565]	0.455
Nord-du-Québec	3	–	–

Notes: 500 store-bootstrap resamples per region; regions with fewer than five stores omitted.

Table 7. Region-level indicators (reported for transparency; unstable under bootstrap)

Region	Stores	Diversity	ECI	Fitness
Montréal	237	9	1.70	1.15
Abitibi-Témiscamingue	21	10	1.27	1.15
Côte-Nord	16	8	1.25	0.98
Laval	25	7	1.16	0.79
Outaouais	36	7	1.13	0.82
Capitale-Nationale	126	9	0.32	1.02
Montérégie	190	9	0.16	1.08
Gaspésie–Îles-de-la-Madeleine	30	8	0.05	0.88
Estrie	152	7	0.03	0.73
Laurentides	99	11	-0.07	1.70
Lanaudière	66	8	-0.38	0.92
Saguenay–Lac-Saint-Jean	97	11	-0.70	1.66
Nord-du-Québec	3	4	-0.72	0.45
Chaudière-Appalaches	102	8	-0.96	0.89
Centre-du-Québec	61	11	-1.19	1.54
Bas-Saint-Laurent	39	6	-1.49	0.63
Mauricie	31	6	-1.56	0.60

Notes: Diversity = number of categories with RAA ≥ 1. ECI standardised; sign fixed so that corr(ECI, diversity) ≥ 0.

Table 8. QDACI components by region

Region	diversity_rarefied	diversity_lo	diversity_hi	specialization	specialization_lo	specialization_hi	concentration_gini
Abitibi-Témiscamingue	0.85	0.81	0.88	0.74	0.62	1.63	0.81
Bas-Saint-Laurent	0.77	0.69	0.84	0.60	0.53	1.14	0.81
Capitale-Nationale	0.79	0.67	0.87	0.29	0.28	0.48	0.89
Centre-du-Québec	0.77	0.67	0.85	0.58	0.56	0.93	0.89
Chaudière-Appalaches	0.77	0.66	0.85	0.57	0.51	0.79	0.81
Côte-Nord	0.79	0.76	0.80	0.82	0.68	1.14	0.52
Estrie	0.75	0.64	0.83	0.38	0.34	0.61	0.82
Gaspésie–Îles-de-la-Madeleine	0.80	0.73	0.86	0.68	0.65	1.22	0.89
Lanaudière	0.77	0.69	0.85	0.51	0.50	0.76	0.76
Laurentides	0.82	0.74	0.89	0.32	0.29	0.51	0.81
Laval	0.79	0.73	0.84	0.44	0.42	0.94	0.83
Mauricie	0.74	0.68	0.80	0.89	0.72	1.30	0.73
Montréal	0.78	0.68	0.86	0.37	0.34	0.47	0.78
Montérégie	0.79	0.69	0.86	0.21	0.18	0.34	0.89
Nord-du-Québec	–	–	–	1.08	0.77	1.48	–
Outaouais	0.80	0.73	0.86	0.43	0.38	0.93	0.92
Saguenay–Lac-Saint-Jean	0.79	0.70	0.86	0.34	0.33	0.52	0.80

Notes: Components describe observed online catalogues of Quebec-based merchants, not production or output. Complexity omitted because the ECI failed the preregistered stability criterion.