Local Bias in Global Platforms: Evidence from a Cross-Market Merger

Sai Chand Chintala* Jūra Liaukonytė†
Tommaso Bondi‡ Dominykas Šeputis§

October 8, 2025

Abstract

As digital platforms expand globally, understanding heterogeneous local market preferences becomes essential for optimizing market integration strategies. The rapid growth of the "Global Localization" industry highlights the increasing demand for companies to tailor their content and products to resonate with local audiences. This paper theoretically and empirically examines the impact of international cross-market integration on consumer behavior within a major global Consumer-to-Consumer (C2C) online marketplace, which merged neighboring geographic markets. We focus on a specific merger case between two markets characterized by asymmetrical market sizes and income levels. We leverage this quasi-exogenous change in market thickness to evaluate the effects of this merger on key platform outcomes such as prices and volume of transactions, utilizing language-agnostic embeddings to identify similar items within and across markets with different languages. We find that despite the substantial increase in market efficiency due to the merger—evidenced by a reduction in price differences from 46% to 19% between the two markets—there remains a strong preference for local trade, which we term "local bias." Local bias is in part explained by localized and idiosyncratic consumer preferences for various products across different markets. The perceived degree of commoditization of items within categories predicts the extent of local bias. We investigate other potential underlying mechanisms of this persistent local bias, providing insights into the complexities of market integration strategies and their implications for digital platforms.

Keywords: Platform Mergers, C2C Platforms; Market Efficiency, Local Bias
JEL codes: L14, L81, D12, F14, M31, O33

*Cornell University, SC Johnson College of Business. Email: sc2947@cornell.edu.
†Cornell University, SC Johnson College of Business. Email: jurate@cornell.edu.
‡Cornell University, SC Johnson College of Business. Email: tbondi@cornell.edu.
§University of Amsterdam

The authors thank the anonymous company for providing the data. Some details in the paper have been obfuscated at the request of the partner company. The authors especially thank Emaad Manzoor for his significant insights and feedback on the early draft of the paper, as well as Omid Rafieian, Nathan Yang, Rakesh Allu, Muzeeb Shaik, and Chris Forman for their helpful comments and discussions.

1 Introduction

With digital commerce growing at 18.2% annually, the global digital platforms market is projected to rise from $9.61 billion in 2023 to $51.16 billion by 2033 (Future Market Insights 2023). In this dynamic environment, online platforms continually seek innovative strategies to enhance their value propositions and to grow. One such emerging strategy for global platforms is the integration of different geographic markets to create a more cohesive and efficient ecosystem. Cross-market integration allows platforms to offer a wider variety of products and more competitive prices, increasing user adoption and sales. Conversely, international boundaries appear to hinder trade (McCallum 1995, Anderson and Van Wincoop 2003). The advantages of a broader market come with the challenges of addressing and accommodating local market preferences.

A prime example of an ongoing cross-market integration is Amazon's European Expansion Accelerator (EEA). This program simplifies the process for sellers to expand their operations across nine European countries with just a few clicks. By automating account setup, translations, listings, shipping arrangements, and compliance checks, the EEA enables sellers to quickly and efficiently access customers in markets across borders. This integration strategy has proven particularly beneficial for small and medium-sized businesses, allowing them to diversify their revenue streams and significantly increase their market reach across Europe. Participants in the EEA have reported substantial growth, with year-on-year sales increasing by 10% (Amazon 2022). Other platforms, such as eBay and Airbnb, also utilize automatic translation to display listings in local languages, further facilitating cross-border commerce and access to a bigger consumer base.

The integration of distinct international markets by digital platforms increases market thickness, potentially enhancing market efficiency. However, consumers in different markets might exhibit a bias or preference for local trade due to various factors. As digital platforms expand beyond national boundaries, these local biases can limit the expected convergence towards market efficiency and pose challenges to achieving a fully integrated market. Understanding and addressing these nuances is important for platforms aiming to optimize their international market integration strategies. Indeed, in an era where global reach is becoming more accessible, the key question surfacing right now is how platforms can adapt their content and products to resonate better with local audiences. Evidence of this is seen in the rapidly growing *"Global Localization"* industry, which reached $67.9 billion in 2023, including 27 global companies with annual revenues over $100 million (Nimdzi Insights 2024).

To examine the above-mentioned issues, we study the impact of integration of international markets on consumer behavior within a leading global Consumer-to-Consumer (C2C) platform. We leverage the variation caused by the quasi-exogenous change in market thickness resulting from the integration of two adjacent geographic markets—referred to as Markets A and B—with differing market sizes and income levels. Market B is significantly larger than Market A and has more price-sensitive consumers. Before the integration, each market-specific platform operated in the local language with its own distinct website.

Our first research question investigates equilibrium price changes post-merger. To compare prices of similar items across markets, we use an approach that considers both structured and unstructured product attributes to identify comparable products. Initially, we show that, consistent with the lower price sensitivity of Market A compared to Market B, prices were 46% higher in Market A prior to the merger after controlling for item compositional differences across markets. After the merger, the listing and sale prices in Market A decreased significantly, moving in the opposite direction of price changes in Market B and Market C, the latter of which did not experience any merger and is used as a control. This equilibrium price adjustment aligns with the prediction of our proposed simple theoretical framework.

Our second question examines the degree of convergence towards a fully efficient market by assessing whether price differences between comparable products in both markets disappear. We find a post-merger price difference of 19.75% even after accounting for shipping cost differences, indicating a preference for local trade, which we refer to as "local bias". We observe heterogeneity of local bias across different product categories. We explore other mechanisms that may contribute to preferences for local trade, such as shipping time differences, frictions due to communication and trust, financial transaction frictions etc.

This paper proceeds as follows: Section 2 reviews the related literature on the home bias, market thickness in platforms, and the impact of international trade on digital platforms. Section 3 provides a description of our empirical setting and Section 4 presents the relevant descriptive statistics. Section 5 outlines our approach to identifying comparable products within and across different markets. Section 6 develops a simple theoretical framework to understand market efficiency and local bias. Section 7 provides empirical results on the effects of the merger on prices, buyer outcomes, seller outcomes, and transaction efficiency. Section 8 discusses potential mechanisms and alternative explanations for the observed local bias. Finally, Section 9 concludes with the implications of our findings and suggestions for future research.

2 Related Literature

Our work contributes to the interrelated streams of literature about home bias and local tastes, the impact of market thickness on platforms, and the role of international trade in digital platforms.

Home bias and local tastes: Research on home bias (Wolf 2000, Hillberry and Hummels 2003, 2008, Yi 2010, McCallum 1995) and local tastes highlights their persistent impact on market behavior. Consumers have a product bias based on regional preference (Schooler 1965), and this bias extends to the digital markets (Blum and Goldfarb 2006). Hortaçsu et al. (2009) demonstrate that physical distance significantly impacts trade volumes on eBay and MercadoLibre, highlighting a notable home-state bias that persists despite the digital nature of transactions. This bias, however, is driven by goods that need to be locally consumed, such as event tickets, or cultural factors, such as affinity for similar sports teams via trading sports memorabilia. Burtch et al. (2014) and Elfenbein et al. (2023) extend this by showing that both geographic proximity and cultural similarity are important determinants of trade patterns. Lin and Viswanathan (2016) further reveal that home bias significantly influences online investments, as lenders tend to favor local borrowers. Chintagunta and Chu (2021) demonstrate that seller locations act as branding cues, creating asymmetric geographic preferences where one Chinese province's preference for another is not always reciprocated. Seller locations influence buyer preferences by serving as quality indicators and reputation mechanisms, with overall trade patterns shaped by trust, cultural similarities, and institutional factors. Together, these studies underscore the persistent role of geographic and cultural factors in shaping online economic interactions.

Three studies most closely related to our work—Hortaçsu et al. (2009), Chintagunta and Chu (2021), and Elfenbein et al. (2023)—focus on the determinants of aggregate online trade flows among various city, state, or province pairs. Our work contributes to this literature by examining how individual customers transact when comparable, non-location-specific items are available both within and across markets. In addition to documenting asymmetric cross-market trade flows after the merger of two markets, we quantify a price premium that local consumers are willing to accept to avoid cross-market trade. We also demonstrate that this local bias is heterogeneous across categories, providing insights into another important dimension of what drives local bias.

Platforms and Market Thickness: Existing research on platforms explores the influence of network effects and the enhanced value that users derive from the platform (Katz and Shapiro 1985, Rysman 2004, Chu and Manchanda 2016), the optimal market thickness (Loertscher et al. 2022), the platform's ability to manage market thickness (Bimpikis et al. 2020). Previous studies on the impact of increased market thickness on platform outcomes present mixed findings. Li and Netessine (2020) investigated the effects of increased market thickness in online platforms, discovering that higher market thickness reduces matching rates due to increased competition among listings, which complicates the search process for users. Fong (2024) explored market size and competition in online dating platforms, finding that larger market sizes increase user participation and competition. However, increased market thickness was found to decrease user participation, aligning with the findings on market thickness by Li and Netessine (2020). Farronato et al. (2023) examined the merger of two major pet-sitting platforms, revealing that network effects benefit users of the acquiring platform, but the lack of differentiation harms users of the acquired platform, leading to negligible net benefits at the market level. Reshef (2023) studied the impact of entry in platform markets, showing that while new entrants expand the market, low-quality incumbent firms experience a decline in individual business outcomes due to intensified competition. Collectively, these studies highlight the complex interplay between market size, competition, and user outcomes in platform markets. Previous studies have largely focused on the increase in market thickness following integration or mergers within the same region or market. Our research expands on this by examining the effects of the increase in market thickness resulting from the integration of international markets that are asymmetric in size and price sensitivity. In addition, the geographic or cultural dissimilarities between these markets could lead to outcomes that differ from those observed in previous studies that focus only on a single market.

Platforms and International Trade: Several recent studies emphasize the significant impact of digital platforms on international trade, facilitating cross-border transactions and reducing trade barriers. Lendle et al. (2016) examine how eBay reduces the impact of geographical distance on international trade. Brynjolfsson et al. (2019) provide empirical evidence that the introduction of Machine Translation on a large digital platform significantly increased international trade by reducing language barriers among users and improving market efficiency. Hui (2020) find that the integration of an existing logistics service increases international trade. These studies look at the impact of different strategies that firms adopt, such as online trade, machine translation, and special shipping programs. We contribute to this body of literature by examining a new angle: the integration across markets on a digital platform and how it affects international trade.

3 Empirical Setting

3.1 Institutional Background

The data for this study comes from one of the leading global Consumer-to-Consumer online marketplaces, which connects buyers and sellers across various categories. The platform earns revenue from fees on each transaction and offers a shipping service. Initially, the platform operated within individual countries.

Figure 1 displays the transaction volumes between ten selected markets before and after the mergers. Initially, only diagonal cells were populated, representing 100% of trade within each market. After the mergers, various cross-market trades became available. In the figure, buyer markets are on the Y-axis, and seller markets are on the X-axis. The size of each bubble represents the proportion of transaction volume for each market pair combination, with each row totaling 100% of the buyer market's transactions.

Figure 1: Cross-Market Transactions

Notes: The figure displays 10 buyer markets on the Y-axis and 10 seller markets on the X-axis. The size of each bubble represents the transaction volume between any two markets following their respective mergers. The sum of transactions across each row totals 100% of the buyer market's transactions. For instance, A (row) - A (column) is 62%, meaning that post-merger, 62% of market A's transactions are local. Similarly, A (row) - B (column) is 38%, indicating that 38% of market A's buyer volume is purchased from market B. Markets A, B, and C are the focus of this paper. Seven additional markets that underwent mergers were randomly selected and presented for illustrative purposes only. Market C never experienced any mergers and is used as a control market.

In this paper, we focus on markets A, B, and C (the 3 × 3 matrix highlighted in yellow). The merger between markets A and B is unique among all existing mergers because both markets had well-established pre-existing marketplaces before the merger, allowing us to evaluate platform outcomes before and after the merger. In contrast, other mergers occurred simultaneously with the opening of new markets. Market C did not undergo any merger and still operates separately, serving as a control group in our analysis. We obtained data on a representative sample of product listings and transactions from these three markets: A, B, and C. Seven additional markets that underwent mergers were randomly selected and presented for illustrative purposes only.

3.2 Data

For the analysis, we use the transactions data from a randomly drawn representative sample of the buyers on the platform, along with a randomly drawn sample of the listings across the three markets A, B, and C. We implemented a randomization check to ensure that the provided subsamples are representative of the activity on the entire platform across the three markets. The data contains details of characteristics of the products sold. Due to data privacy regulations, we encoded text descriptions with privacy-preserving embeddings to be able to transfer the data outside the company's internal data infrastructure. Below we summarize the nature and the characteristics of the embeddings used.

Specifically, we work with embeddings produced by the `clip-ViT-B-32-multilingual-v1` model (Please see Appendix D for a detailed description of the embeddings). This model is a variant of OpenAI's CLIP model (specifically, `clip-ViT-B-32`) (Radford et al. 2021), which produces unified embeddings of text and images in the same 512-dimensional latent space. The `clip-ViT-B-32-multilingual-v1` model is derived from `clip-ViT-B-32` via multilingual knowledge distillation[1]. The embeddings produced by `clip-ViT-B-32-multilingual-v1` are thus language-agnostic: the same textual descriptions in a different languages will be mapped to the same embeddings. This model is trained on a vast, diverse internet dataset that includes multiple languages, enabling it to capture the semantic essence of text as well as to understand and represent the semantic meanings of descriptions rather than just the literal words, facilitating accurate representation and comparison of items across languages. For example, an item described in French can be matched with its equivalent described in German, based on the semantic similarity of their descriptions. Overall, the language agnosticity of the embeddings is key for our analysis, as it allows for effective comparison of item descriptions across different languages.

[1]Specifically, `clip-ViT-B-32` serves as the teacher and mutlilingual DistilBERT (Reimers and Gurevych 2020a) as the student, trained on parallel corpora from the following languages (enumerated as ISO 639 language codes): ar, bg, ca, cs, da, de, el, es, et, fa, fi, fr, fr-ca, gl, gu, he, hi, hr, hu, hy, id, it, ja, ka, ko, ku, lt, lv, mk, mn, mr, ms, my, nb, nl, pl, pt, pt, pt-br, ro, ru, sk, sl, sq, sr, sv, th, tr, uk, ur, vi, zh-cn, zh-tw.

4 Descriptive Statistics

The merger between markets A and B occurred during the last several years.[2] Our data covers a 16-month period: 6 months before and 10 months after the merger. We have the data from three markets— two markets that eventually merged (A and B) and a third market (C) that did not undergo any merger. As mentioned above, prior to the integration, the company managed the online platform as independent entities in markets A and B. Post integration, the sellers and buyers from both markets could view and transact across listings on a single website in their local language (i.e the same website on which they used to transact in the pre-integration period). Markets A and B are geographical neighbors with different per capita income levels, with market A having a higher per capita income than market B. Additionally, country A is substantially smaller both geographically and by population. Therefore, it is expected that the merger effects will be asymmetric: country B's presence will significantly affect market A's outcomes, while the reverse is less likely. Indeed, Figure 1 confirms that, from the perspective of market B, purchases from market A are minimal. Therefore, this paper focuses on analyzing the platform outcomes from the perspective of market A after its merger with market B.

We observe transaction details with the characteristics of the products. We exclude the long tail of categories with less than 1% sales. Table 1 presents a summary of the platform-level outcomes in three markets - A, B, and C. To preserve the confidentiality of the data when reporting platform-level outcomes, we standardize the values (of the first five rows) by using the values in the pre-merger period of Market A as the base. Market A shows a higher average amount per transaction compared to Market B. This relationship holds even after adjusting for inflation using Consumer Price Index (CPI) values. On average, buyers in Market B conduct 10 to 15 times more transactions per month than those in Market A. Additionally, the monthly transaction volume increases in the post-merger period compared to the pre-merger period. Market B has approximately 10 times the number of unique buyers and sellers compared to Market A in both periods. This greater number of participants in Market B is also evident in the percentage of cross-market transactions. For both the seller

[2] We keep the exact date anonymous at the request of the company.

side and buyer side, cross-market transactions represent a significant proportion of total transactions in Market A compared to Market B.

Table 1: Descriptive Statistics: Outcomes in Markets A, B, and C Pre and Post Merger

	Market A		Market B		Market C	
	Pre	Post	Pre	Post	Pre	Post
Avg. Transaction Amount (Price, Indexed)	1.00	0.92	1.00	1.02	1.00	1.03
	(2.35)	(2.21)	(4.06)	(3.19)	(2.23)	(2.25)
Monthly Transaction Volume (Indexed)	1.00	1.79	1.00	1.20	1.00	2.50
	(0.29)	(0.70)	(0.30)	(0.49)	(0.39)	(0.30)
N of Unique Monthly Buyers (Indexed)	1.00	1.63	1.00	1.20	1.00	2.52
	(0.25)	(0.61)	(0.26)	(0.46)	(0.35)	(1.16)
N of Unique Monthly Sellers (Indexed)	1.00	1.30	1.00	1.20	1.00	2.39
	(0.25)	(0.48)	(0.24)	(0.45)	(0.34)	(1.06)
% of Cross-Market Transactions (Buyer Side)	0.00	18.87	0.00	0.60	0.00	0.00
% of Cross-Market Transactions (Seller Side)	0.00	8.35	0.00	1.78	0.00	0.00

Notes: The table shows the mean values with SD in parenthesis. To protect the confidentiality of the data, for the first five rows, for each market, the values in the column pre-merger are normalized to 1, with all other columns rescaled accordingly. The last two rows in the table are divided by a constant to obfuscate the actual values.

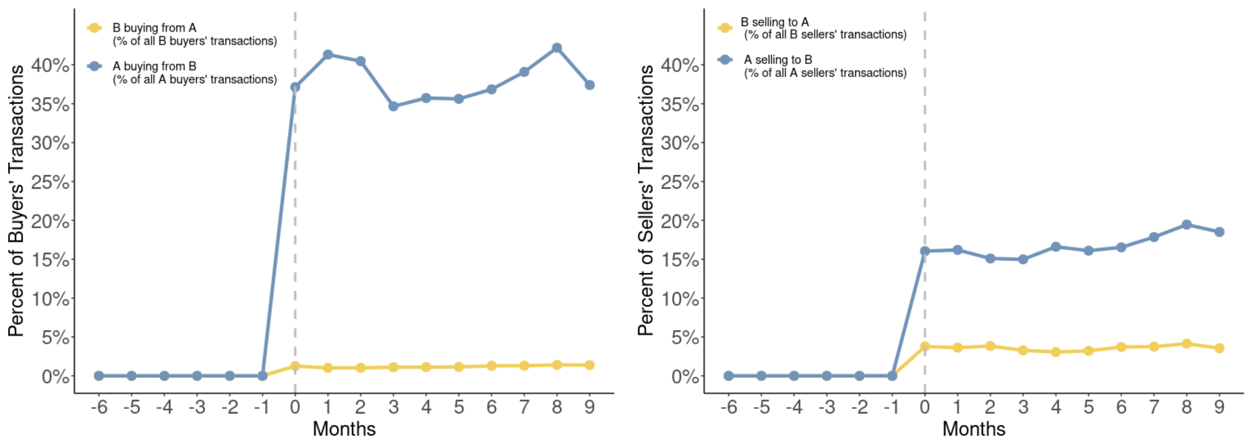
We use market C as the control market to analyze the effects of the integration. An important concern when using Market C as a control to understand the outcomes of Market A is the much higher growth rate of Market C compared to Market A. In terms of unique monthly buyers, Market C grew faster than Market A, with similar asymmetry in the number of monthly sellers. Theoretically, the entry of new buyers and sellers could contribute to downward pressure on prices and improve efficiency in the market due to increased market thickness (Focarelli and Panetta 2003). Therefore, any resulting comparison of the post-merger outcomes of Market A with those of Market C could be interpreted as a conservative estimate of the merger effect.

Figure 2 panel (a) illustrates the immediate increase in cross-market transactions in Market A. From the buyers' perspective, even during the first month post-integration, cross-market purchases in Market A surged to around 40% of the total buyers' transactions in A and remained around the same range throughout post-period. As anticipated, given that Market B's items are less expensive compared to those in Market A, and that Market B is larger, the share of cross-market purchases originating from Market B were minimal.

Figure 2: Discontinuity in Cross-Market Transactions

(a) Buyers' Transactions

(b) Sellers' Transactions



Notes: Panel A shows the percentage of Cross-Market Transactions from the Buyers' perspective as part of the total Buyers' transactions. Panel B shows the percentage of Cross-Market Transactions from the Sellers' perspective as part of the total Sellers' transactions. The percentages are scaled by a constant to obfuscate the actual values.

Figure 2 panel (b) illustrates cross-market transactions from the sellers' perspective: a significant increase is also evident in both markets. Notably, of all items sold by the sellers in Market A, around 16% of the sales originate from buyers in Market B. This occurs due to the much larger size of Market B and despite the fact that buyers from Market B overwhelmingly buy from Market B sellers, since products in Market B are on average cheaper than in Market A. Compared to Market B buyers, Market B sellers account for a slightly higher percentage of the total cross-country trade, driven by their lower prices relative to Market A sellers. This results in a relatively high proportion of Market A buyers engaging in cross-market purchases.

Both panels of the graph also demonstrate a sharp discontinuity in trade following the merger, with the timing of the merger serving as a quasi-natural experiment in our analysis.

5 Identifying Comparable Products

One challenge with data from the listings, and thus all subsequent analyses relying on comparisons across products within and across different markets, is the inherent heterogeneity in the products listed. Specifically, a key technical challenge is constructing unique identifiers for each product. In settings similar to the current platform, sellers typically do not report stock keeping unit (SKU) codes or other product identifiers. Therefore, product identifiers must be reconstructed from other observable product descriptors.

We develop and implement several approaches to construct product identifiers—which we term "pseudo-product IDs"—using seller-provided structured and unstructured descriptors of each product. One of our approaches leverages text embeddings to create these pseudo-product IDs. Due to privacy concerns, the company does not store images beyond a certain period, and thus image embeddings are unavailable during the sample period. However, we obtained image embeddings for a month outside the sample period and conducted a robustness test by reconstructing pseudo-product IDs using these image embeddings (see section 7.2). For the main sample analysis to evaluate different approaches, we consider deriving pseudo-product IDs from the product's category nest (a combination of category and subcategory), attribute 1, attribute 2, attribute 3, and text description. Our approaches are summarized in Table 2.

Table 2: Summary of Approaches to Create Pseudo-Product IDs

Pseudo-ID method	Selected Characteristics
1	category nest × attribute 1
2	category nest × attribute 1 × attribute 2
3	category nest × attribute 1 × attribute 2 × attribute 3
4	category nest × attribute 1 × attribute 2 × attribute 3 + text description

Notes: The categories of the products sold on the platform are nested in hierarchical order with different category levels. The attributes refer to the different characteristics of the products.

In each approach, all products having the same category nest and size are first grouped under a single *candidate* pseudo-product ID. From a consumer perspective, this initial grouping aligns with how consumers typically search the platform by specifying a category and additional attributes through filters, forming their initial consideration sets. In approach 1, the candidate pseudo-product ID is not fragmented further (i.e., products having the same category nests and sizes are assigned to the same pseudo-product-ID). Hence, approach 1 is likely to produce pseudo-IDs that are too coarse-grained, and will serve as an illustrative baseline.

In approach 2, each candidate pseudo-product ID is fragmented further based on attribute 1. This approach considers that platform's consumers often refine their searches and preferences by brand within their selected category and attribute 1, leading to more precise consideration sets.

In approach 3, each candidate pseudo-product ID is fragmented further based on attribute 3 (i.e., products with the same category nest, attribute 1, attribute 2, and attribute 3 are assigned the same pseudo-product ID). Approach 3 is thus a finer-grained variant of approach 1, accounting for the fact that consumers may also consider the attribute 3 of the item when making purchase decisions.

In approach 4, we further fragment the pseudo-product IDs produced by approach 3 by clustering the text descriptions of products. Specifically, we cluster the products within each candidate pseudo-ID based on the CLIP embeddings of their text descriptions (using the k-medoids algorithm configured with the cosine distance metric) and assign the same pseudo-ID to all products within the same cluster. We use silhouette coefficients (Rousseeuw 1987) to select the optimal k for each candidate pseudo-ID. Approach 4 uses all the structured and unstructured product descriptors available in our dataset. This method leverages detailed text descriptions to reflect the nuanced preferences consumers may have, ensuring the resulting pseudo-product-IDs consist of the most similar products possible.

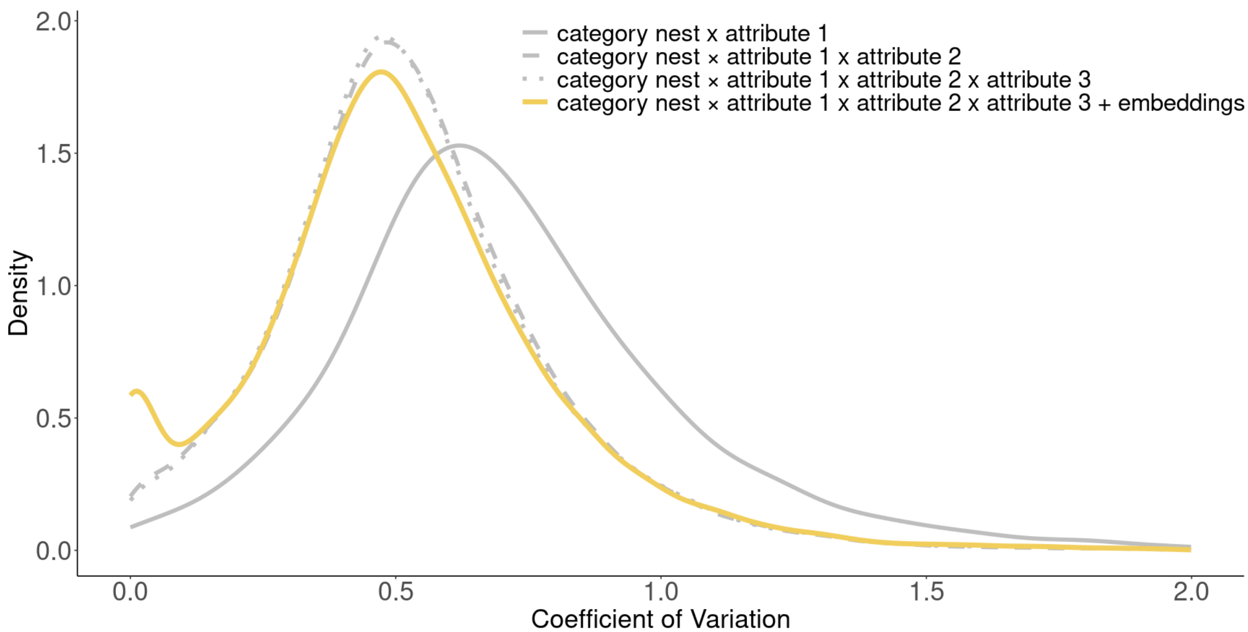
The validity of our pseudo-product IDs is key to accurately measuring platform outcomes, as they form the basis for comparing product listings within and across markets. If our pseudo-product IDs are too fine (erring towards assigning different pseudo product-IDs to products with the same true SKU), the consideration sets for some products will be underestimated. If our pseudo product-IDs are too coarse (erring towards assigning the same pseudo product-ID to products with different true SKUs), the consideration set for some products will be overestimated. Hence, we provide empirical support for the validity of our pseudo-IDs with a validation test, detailed further below.

In our validation test, we compute the coefficient of variation[3] (CoV) of the prices of products having the same pseudo product-ID. The assumption underlying this test is that the transacted prices of pre-owned products with the same true SKU should converge towards equilibrium prices. Consequently, homogeneous items should exhibit negligible price

[3]The coefficient of variation of a quantity is the sample standard deviation of the quantity divided by its mean. As such, it is a mean-standardized measure of dispersion of a quantity in a sample.

variation, resulting in a low within-SKU price CoV. However, there could be a variation in how the sellers price the products on the platform resulting in a non-zero CoV. Different pricing strategies, product conditions, and seller motivations can contribute to this variability. Overly coarse-grained pseudo-IDs, in particular, will have large CoVs.

Figure 3: Coefficient of Variation comparison by Pseudo-ID methods



Notes: This figure presents the density plots for Coefficients of Variation (CoV) of price obtained from different constructions of pseudo-product IDs. The yellow plot represents the density of CoV using the combination of category nest × attribute 1 × attribute 2 × attribute 3 with embeddings while other plots in grey indicate the other combinations used to construct pseudo-product IDs. For a homogeneous group of products, the density of the CoV of the price should ideally be close to zero. Among the different combinations, the fourth approach (category nest × attribute 1 × attribute 2 × attribute 3) with embeddings performs the best, making it our preferred method for creating pseudo-product IDs.

In Figure 3, for each pseudo-ID construction approach, we plot the distributions of the price CoVs across all pseudo-IDs. The pseudo-IDs produced by approach 1 (that we expect to be overly coarse) have a median price CoV of 0.67: this is significantly higher[4] than the median price CoVs of the pseudo-IDs produced by the other approaches. Including products' status and textual descriptions reduces the price CoVs of the resulting pseudo-IDs, with approach 4 having the lowest median CoV. Hence, we adopt approach 4 for subsequent analyses.

[4]We measure the statistical significance of the difference of medians with a one-tailed Wilcoxon signed-rank test.

6 Market Efficiency and Local Bias: Simple Theoretical Framework

6.1 Model Set Up

To answer our first research question about the equilibrium prices after the integration, we develop and solve a simple theoretical model. The model provides a framework to understand how the integration affects market dynamics and pricing strategies. More specifically, the objective of this model is to derive a proposition for the equilibrium prices that can be empirically tested.

There are two countries, A and B. There is a continuum of mass 1 of consumers in each country, indexed by j, of measure $\gamma > 0$ and 1 respectively. We do not restrict γ to be above 1. Sellers compete on prices, $\mathcal{P}_A$ and $\mathcal{P}_B$. Consumers from the two countries differ in their price sensitivity: we define as $\alpha > 0$ the price sensitivity of country A consumers, and normalize the price sensitivity of country B consumers to 1. We do not restrict α to be above 1.

Each consumer j has a home bias $\chi \geq 0$ for the seller of their country. For simplicity (and in line with our empirical application), we assume that consumers in both countries are equally biased; relaxing this assumption renders the algebra cumbersome while not affecting the spirit of the model in a major way.

On top of being locally biased, each consumer $j \in [0, 1]$ also has an idiosyncratic *relative* preference, or taste, ζ_j for the good sold by seller A. For analytical tractability, we assume that $\zeta_j \sim U(-\bar{\zeta}, \bar{\zeta})$. Notice how the relative taste for product A follows the same distribution for consumers of country A and B: in other words, χ is solely responsible for consumer's preference for the same country good, whereas ζ_j captures country-independent, product-specific consumer taste. In practical terms, a consumer in Market A might prefer the product sold by the seller in Market B ($\zeta_j < 0$), and trade this off with her own home bias $\chi \geq 0$, as well as the prices charged by the two sellers, $\mathcal{P}_A$ and $\mathcal{P}_B$.

In what follows, we employ the following

Assumption 1. $\chi < \bar{\zeta}$.

Simply put, this assumption requires that home bias is not strong enough to prevent all cross-country trade (when prices are equal). This is in line with our data, and ensures that market integration has non-trivial effects on prices and market shares, and thus on sellers' and buyers' welfare. Without loss of generality, we restrict $\bar{\zeta} = 1$, so that the previous assumption can be rewritten as $\chi < 1$.

The aforementioned positive market efficiency effect is either compounded or attenuated by a price effect: how does market integration affect optimal pricing by each seller? To this end, we find the Nash equilibria in prices of three separate games: *i*) price competition between a country A and a country B seller (post integration scenario); *ii*) price competition between two country A sellers; *iii*) price competition between two country B sellers.

6.1.1 A Brief Discussion of the Model's Assumptions

While our model captures the main forces at play pre- and post-integration, it is worth emphasizing some important assumptions (and how they could be relaxed in future work).

1. **Duopoly competition both pre- and post-integration.** Alternatively, one can start with duopolies in both countries A and B, and then characterize the price equilibrium when all four sellers compete post-integration. However, solving for the equilibrium with four firms is cumbersome. Similarly, one could start with two monopolistic sellers in A and B, and study their duopoly competition post-integration. However, this would overestimate the benefits of integration for (all) consumers. Thus, we elected to go with two sellers in each scenario; while this doesn't capture increased competition (and thus more refined taste matches), it delivers some key and sensible predictions on price dynamics.

2. **Sellers fully internalize local bias in pricing decisions.** In reality, sellers might learn the extent of local bias over time. In this sense, our model is best seen as describing the long-run convergence of the pricing equilibrium, after the sellers have fully learned the amount of bias present in the market.

3. **Fixed product offering by sellers.** We treat sellers' product offerings as fixed and do not consider the possibility of sellers adapting the quality of their items in response to a change in their buyers' pool. This is in line with our empirical context.

4. **Symmetric local bias.** This assumption is made purely for analytical simplicity. It is not hard to modify the model to account for the case of $\chi_A > \chi_B \geq 0$.

5. **No stock-outs.** Our model implicitly (and realistically) assumes a deep supply of products. Consumers can be hurt by market integration because of higher prices, but not due to missing out on a product to another buyer (either same-country or cross-country).

6. **No Tariffs.** Neither country is imposing tariffs on products bought from foreign sellers. While χ could partly reflect tariffs on foreign products, it will be interesting to study how incorporating both tariffs and heterogeneous price sensitivities between countries influence the results.

6.2 Equilibrium Prices: Post-Integration

Given our specification, post-integration profits for seller A are given by

$$\begin{aligned}
\pi_A &= \pi_{AA} + \pi_{AB} \\
&= \mathcal{P}_A\Big(\gamma\Big(Prob(\zeta_A - \alpha\mathcal{P}_A + \chi \geq -\alpha\mathcal{P}_B)\Big) + Prob(\zeta_A - \mathcal{P}_A \geq -\mathcal{P}_B + \chi)\Big) \qquad (1)\\
&= \mathcal{P}_A\Big(\gamma\Big(\frac{1-\alpha(\mathcal{P}_A - \mathcal{P}_B) + \chi}{2}\Big) + \Big(\frac{1-(\mathcal{P}_A - \mathcal{P}_B) - \chi}{2}\Big)\Big)
\end{aligned}$$

and similarly for B:

$$\begin{aligned}
\pi_B &= \pi_{BA} + \pi_{BB} \\
&= \mathcal{P}_B\Big(\gamma\Big(Prob(\zeta_A - \alpha\mathcal{P}_A + \chi \leq -\alpha\mathcal{P}_B)\Big) + Prob(\zeta_A - \mathcal{P}_A \leq -\mathcal{P}_B + \chi)\Big) \\
&= \mathcal{P}_B\Big(\gamma\Big(\frac{1+\alpha(\mathcal{P}_A - \mathcal{P}_B) - \chi}{2}\Big) + \Big(\frac{1+(\mathcal{P}_A - \mathcal{P}_B) + \chi}{2}\Big)\Big).
\end{aligned} \tag{2}$$

Notice that to go from the second-to-last to the last line in both profits expressions we have used straightforward properties of the uniform distribution's CDF.

We are now ready to present our first result, characterizing prices as a function of home bias (χ), relative market size (γ), and relative price sensitivity (α).

Proposition 1 (Post-Integration Optimal Prices)**.** *Equilibrium prices are given by*

$$\mathcal{P}_A^{Post} = \frac{1+\gamma-\chi(1-\gamma)}{2(1+\alpha\gamma)}, \qquad \mathcal{P}_B^{Post} = \frac{1+\gamma+\chi(1-\gamma)}{2(1+\alpha\gamma)}. \tag{3}$$

A few remarks are in order. First, when either $\chi = 0$ or $\gamma = 1$, $\mathcal{P}_A^{Post} = \mathcal{P}_B^{Post}$. To understand why this is the case, first notice that, whenever $\chi = 0$, the two sellers are fully symmetric and are competing over the same market (of size $\gamma + 1$) by setting symmetric prices, namely

$$\mathcal{P}_A^{Post} = \mathcal{P}_B^{Post} = \frac{1+\gamma}{2(1+\alpha\gamma)}. \tag{4}$$

Conversely, when $\gamma = 1$, we have

$$\mathcal{P}_A^{Post} = \mathcal{P}_B^{Post} = \frac{1}{1+\alpha}. \tag{5}$$

Provided that $\chi > 0$, $\mathcal{P}_A^{Post} > \mathcal{P}_B^{Post}$ whenever $\gamma > 1$. Perhaps surprisingly, this is true independently on α – although it is easy to verify that the *magnitude* of this difference is, intuitively, decreasing in α. In other words, seller A charges a higher price whenever country A is larger, even if *i*) home bias is so strong that each seller essentially only sells to their own country and *ii*) $\alpha > 1$, that is, country A buyers have a lower willingness to pay compared to their country B peers.

We now turn to our central question: what are the effects of market integration? More specifically: How do the post-integration prices compare to the pre-integration ones?

6.3 Prices Pre vs. Post Integration

We start with the following:

Lemma 1 (Pre-Integration Optimal Prices)**.** *Before market integration, optimal prices are given by*

$$\mathcal{P}_A^{Pre} = \frac{1}{\alpha}, \qquad \mathcal{P}_B^{Pre} = 1. \tag{6}$$

Thus, we obtain the following:

Proposition 2 (Price Contamination)**.** *Assume that country A's consumers price sensitivity is lower than country B's: $\alpha < 1$. Then, following market integration, prices in country A decrease for every $\gamma > 0$ and $0 < \chi < 1$:*

$$\mathcal{P}_A^{Pre}(\gamma,\chi) - \mathcal{P}_A^{Post}(\gamma,\chi) > 0. \tag{7}$$

The opposite is true for country B.

Proposition 2 highlights how pre-integration optimal prices adjust to the merging of the two markets: seller A optimal prices decrease to reflect seller's A attempts to tap into market B – as well as to defend its market shares in market A – while the opposite is true for seller B, who is now optimally increasing prices to capitalize on the newly accessible wealthier market.

Importantly, $\mathcal{P}_A^{Pre}(\gamma,\chi) > \mathcal{P}_A^{Post}(\gamma,\chi)$ and $\mathcal{P}_B^{Pre}(\gamma,\chi) < \mathcal{P}_B^{Post}(\gamma,\chi)$ jointly imply that $\mathcal{P}_A^{Post}(\chi) - \mathcal{P}_B^{Post}(\chi) < \mathcal{P}_A^{Pre}(\chi) - \mathcal{P}_B^{Pre}(\chi)$, that is, price differences between the two countries are reduced. And yet, we once again emphasize that prices do not fully converge across the two countries – unless $\chi = 0$ or $\gamma = 1$, neither of which, however, is the case in our empirical setting.

7 Market Efficiency and Local Bias: Empirical Results

7.1 Prices

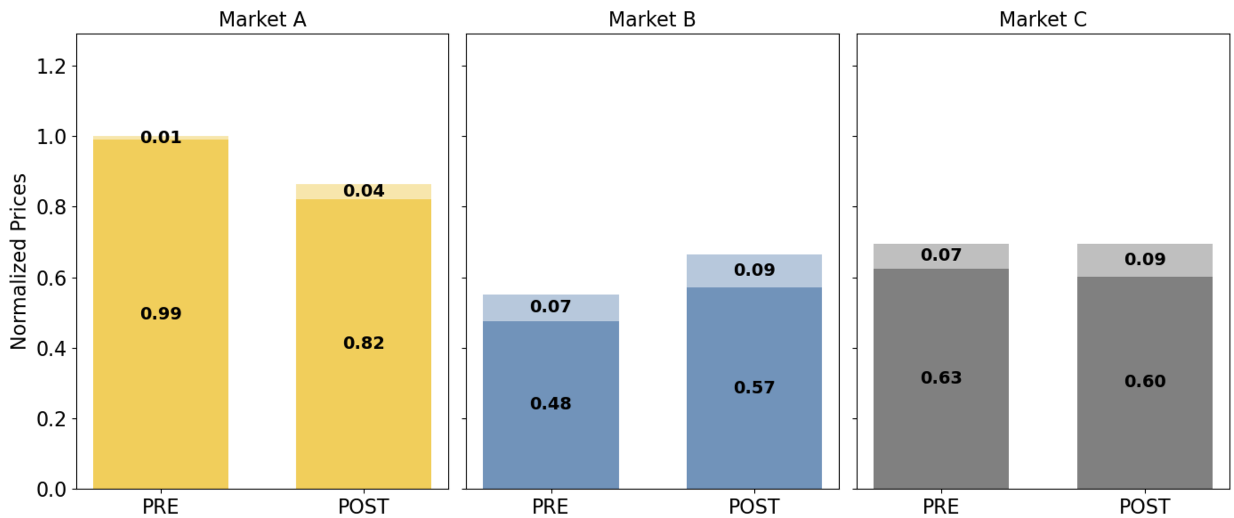
Proposition 2 in the theoretical model outlined in Section 6 predicts that after the merger, prices in Market A will decline, but the decline will be constrained by preferences for local trade.

We evaluate these predictions by comparing average prices before and after the merger across all three markets. While prices are inherently endogenous, our goal is to evaluate the changes in equilibrium levels—comparative statistics with respect to changes in market thickness as predicted by our theoretical model—while remaining agnostic about the exact processes of adjustment (however, we provide some informal discussion on these potential processes below).

In this evaluation, we control for compositional differences in item types by including pseudo-product ID fixed effects (see Section 5 for details on our approach to evaluating product similarity). In our main analysis, we use prices not adjusted for inflation because consumers, over relatively short periods and high levels of inflation, exhibit inattention to inflation and do not swiftly adjust their price perceptions (Weber et al. 2023). Consequently, inflation-adjusted prices could potentially lead to misinterpretations of market dynamics. However, in Appendix B, we demonstrate that the main results hold even when prices are adjusted for inflation. Figure 4 illustrates the results of the price differences controlling for item similarity using both sale prices and listing prices of the products. Several key takeaways emerge from this descriptive evidence. First, the listing and sale prices in Market A, which were higher before the merger, significantly decrease after the merger. This is directly consistent with Proposition 2 in Section 6. Second, average prices in Market B increase.

Finally, the post-merger selling prices in Markets A and B move in opposite directions, suggesting that the relative change is attributable to the merger. In contrast, the changes in prices in Market C remain much more stable compared to Markets A and B.

Figure 4: Average Prices Controlling for Pseudo-Product IDs Pre- and Post- Merger



Notes: The figure compares the average prices in three markets—A, B, and C—before and after the merger of markets A and B, after controlling for pseudo-product ID fixed effects. The total height of each bar reflects the listed prices. Darker shaded areas within the bars represent the sale prices of products, while lighter shaded areas represent negotiated discounts. To protect the confidentiality of the data, the listing price of market A pre-merger is normalized to 1, with all other prices rescaled accordingly.

Interestingly, Figure 4 also illustrates that in market A, negotiations on the platform were barely existent prior to the merger, averaging a tiny percentage of the listed price. After the merger, the negotiated amount increased almost five-fold. Thus, one of the plausible mechanisms facilitating the observed price convergence after the merger can be partly attributed to changes in negotiation behaviors among consumers.

7.2 Buyer Outcomes and Local Bias

We now analyze the price differences at the platform level from the perspective of buyers in Market A. To do so, we need to compare the price paid for a specific pseudo-product ID in Market A with the prices of the same pseudo-product IDs in Market B that are sold during the same week. The rationale for using the same week restriction is that it approximates keeping supply factors for a given pseudo-product ID constant by minimizing temporal variability in availability. In such markets, the inventory for some items can be highly variable and dependent on individual sellers, making short-term comparisons more reliable.

Table 3: Shipping Costs

	A → A		**A → B**		**B → A**		**B → B**	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Avg. Order Cost (Indexed)	1.00	0.94	-	1.10	-	0.94	0.59	0.60
	(1.87)	(1.89)	-	(2.31)	-	(2.50)	(1.48)	(1.55)
Avg. Shipping Cost (Indexed)	0.06	0.07	-	0.18	-	0.18	0.07	0.09
	(0.00)	(0.00)	-	(0.01)	-	(0.06)	(0.01)	(0.01)

Notes: The table shows the average order amounts and shipping costs. To protect the confidentiality of the data, the values in the column of A to A pre-merger are normalized to 1, with all other values rescaled accordingly. The shipping prices are indicated from seller market to buyer market. For example, B to A means the shipping price paid on a sale from Market B to Market A.

We use the pseudo-product IDs generated using the preferred approach 4 (from Section 5) which uses the grouping of category nest × attribute 1 × attribute 2 × attribute 3 characteristics along with CLIP embeddings of the products' textual descriptions.

In this exercise, we compare the average price differentials for comparable items sold across the two markets, incorporating shipping costs which can be a main source of friction for trade flow (Anderson and Van Wincoop 2004). To ensure a conservative comparison, for all items in available in Market B, we include the cross-market shipping costs. This allows us to estimate the potential cost if an item were bought by a Market A buyer from a Market B seller and shipped to Market A. Since this specific counterfactual is not directly observable, we incorporate a median observed shipping cost from all such cross-market trades. For items sold by Market A sellers to Market A buyers, we add the actual observed shipping costs. To preserve anonymity, the indexed shipping costs relative to the average order transaction amount across the two markets, both before and after the merger, are presented in Table 3.

To illustrate, consider a Market A buyer looking for a specific product in a specific week. There is a listing in Market B priced at \$80 [5] and a corresponding listing in Market A for \$100. It costs \$3 to ship the shoes locally within Market A, and \$4 to ship from Market B

[5] The currency \$ is only used for illustrative purpose

to Market A. From the perspective of a buyer in Market A, the resulting price comparison would be $84 if bought from Market B, compared to $103 if bought from Market A. This implies a 22.6% price differential in Market A compared to Market B for the specific product in that specific week.

We systematically analyze these comparisons at the platform level between Markets A and B by evaluating price differences for products sold within the same week across both markets, using the following specification, which controls for week × pseudo-product ID fixed effects:

$$total\_price_{it} = \beta \times \mathrm{A}_{it} + \gamma_{it} + \varepsilon_{it} \tag{8}$$

Here, $total\_price_{it}$ is the sale price of an item with a pseudo-product ID i, with the respective shipping cost included, depending on whether the product is listed in Market A or B. A_{it} is a dummy variable (1 or 0) indicating whether the item is available from sellers in Market A. γ_{it} represents the double interaction of *Pseudo-Product ID*× *Week* fixed effects, used to compare the price differences of pseudo-product IDs available in both markets within the same week. Our estimates of β can be interpreted as the average sale price difference between markets A and B conditional on the pseudo-product ID and the week of sale.

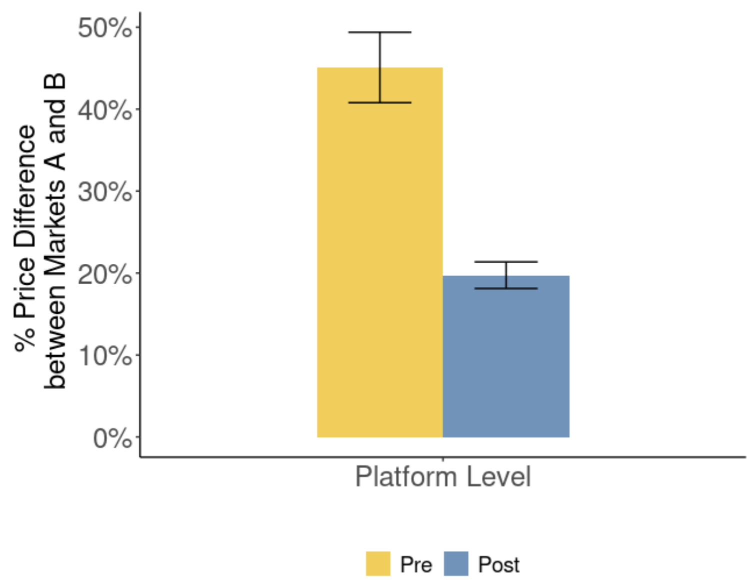
Table 4: Price Differences from the Perspective of Buyers in Market A

	(1)		(2)		(3)	
	Pre	**Post**	**Pre**	**Post**	**Pre**	**Post**
Price Difference - A (β)	0.451***	0.197***	0.433***	0.178 ***	0.416***	0.104***
	(0.040)	(0.015)	(0.039)	(0.015)	(0.039)	(0.016)
Num.Obs.	1.00	7.13	1.13	8.92	1.13	8.92
R^2	0.061	0.438	0.137	0.415	0.139	0.415
Pseudo-Product ID× *Year_Week* f.e.	yes	yes	no	no	no	no
Pseudo-Product ID f.e.	no	no	yes	yes	yes	yes
Week f.e.	no	no	no	no	yes	yes

Notes: This table reports the price differences at the platform level for buyers in market A in the pre and post-periods. To ensure confidentiality, we mask the estimates and standard errors by dividing them with the mean values of the total price in the period, in Market A. The number of observations is normalized with the value from the pre-period from the model (1).*** p<0.001, ** p<0.01, * p<0.05.

Table 4 presents the estimation results for price differences from the perspective of buyers in Market A at the platform level for all categories. To preserve the anonymity of price levels, we rescale the estimates using the mean of prices in Market A, allowing the reported coefficients to be interpreted as a percentage price differential—indicating how much, on average, products available locally in Market A are more expensive than those in Market B (controlling for shipping cost differential). For comparison, we also include specifications that feature only pseudo-product ID fixed effects, as well as combinations of pseudo-product ID fixed effects and week fixed effects. The estimates are similar across the specifications. Figure 5 visually depicts the price differences from the preferred specification in the column (1) of Table 4.

Figure 5: Price Differences from the Perspective of Buyers in Market A



Notes: This figure shows the changes in average price differences at the platform level for buyers in market A before and after the merger.

Overall, our findings indicate that at the platform level, the percentage price difference prior to the merger was 45.12%, which decreased significantly to 19.75% following the merger. This reduction in the percentage price difference between the pre- and post-merger periods supports the argument for increased market efficiency. However, if the market were fully efficient, we would expect the percentage price difference to be zero. In the current scenario, we observe a persistent 19.75% difference in the post-merger period, indicating the presence of friction that limits the expected convergence towards a fully efficient market. We refer to this friction as a preference for local trade, or "local bias". In Section 8, we explore the potential underlying reasons for this preference and persistent bias.

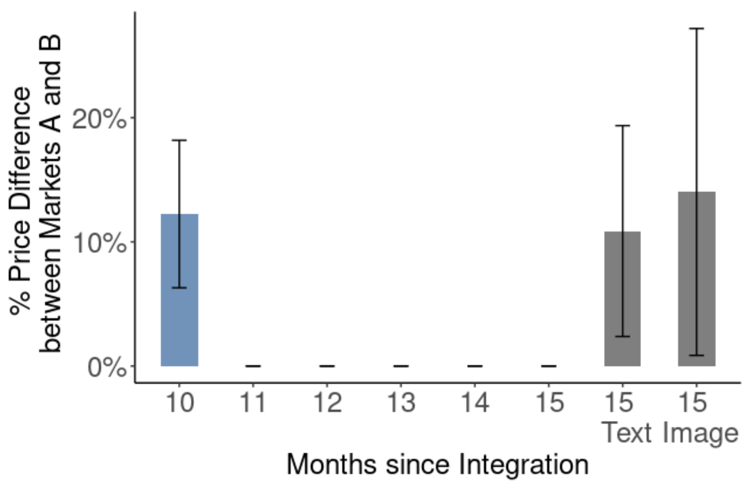
Robustness Test with Image Embeddings. One concern with the current method of relying solely on CLIP embeddings from textual descriptions to identify comparable products is the potential omission of other unobserved characteristics such as the information from the images associated with the listing. Information from images can be valuable in predicting return rates (Dzyabura et al. 2023), review helpfulness (Ceylan et al. 2024), demand estimation (Compiani et al. 2023) as they often capture nuances that text descriptions may miss. CLIP embeddings are particularly valuable for this task because they are trained on a vast dataset of images paired with text, allowing the model to learn a shared representation space where both visual and textual information are aligned. This alignment facilitates the model's ability to capture subtle visual cues and semantic relationships that are often missed by traditional models. Specifically, we utilized the `clip-ViT-B-32-laion2B-s34B-b79K` model, which is a Vision Transformer-based architecture fine-tuned on the LAION-2B dataset[6], known for its ability to encode rich visual semantics into compact embeddings. Using this model, we processed and obtained CLIP embeddings of the photos of listings and transactions for a one month period outside of the sample period.

To test the robustness of the results, we modify the preferred approach 4 (from Section 5) by substituting text embeddings with image embeddings to generate pseudo-product IDs. The underlying rationale for this approach is that the estimates using the pseudo-product IDs—whether generated via text or image embeddings should show minimal variation. We first estimate the price differential for the last month of the sample period using the specification in Equation 8 from Section 7.2. The estimate of the price differential is 12.89% for the last month using text embeddings and is represented by the blue bar (month 10) in Figure 6. Next, we estimate the price differential for the one month outside the sample period using the pseudo-product IDs generated from the CLIP embeddings of the text, finding the

[6]LAION-2B dataset consists of 2 billion CLIP-filtered image-text pairs

price differential to be 10.17% (significant at the 5% level). Finally, we estimate the price differential for the same month using CLIP embeddings of the images [7], finding this estimate to be 14.07% (significant at the 5% level). These two price differences are represented by the two gray bars (month 15) on the right side in Figure 6. The variation between the estimates using the classification based on CLIP embeddings of text and CLIP embeddings of images seems to be low. Hence, this mitigates the concern regarding the potential impact of the absence of image embeddings on identifying comparable products in the baseline analysis.

Figure 6: Price Differences using out-of-sample image embeddings



Notes: This figure shows the changes in average price differences at the platform level for buyers in market A before and after the merger using the text and image embeddings to construct the pseudo-product IDs. The blue bar shows the price difference in the last month of the sample period using pseudo-product IDs constructed from text embeddings. The gray bars on the right show the price difference using pseudo-product IDs constructed from text and image embeddings separately, based on data from five months outside the sample period.

7.3 Seller Outcomes

To evaluate the effect of integration on Market A sellers, we compare two outcomes for incumbent sellers in Market A—–transaction frequency and sales–before and after the merger, with those of control Market C using a two-way fixed effects (TWFE) Difference-in-Differences specification:

[7] For the out-of-sample period, CLIP embeddings of text and CLIP embeddings of images are generated separately

$$Y_{ijt} = \delta[I_{j(i)} \times Post_t] + \gamma_t + \theta_i + \varepsilon_{ijt} \tag{9}$$

where i denotes the incumbent seller, j denotes the Market (Treatment(A) or Control(C)), and t denotes month. Y_{ijt} is the outcome variable *log(1+ Transactions)* or *log(1+ Sales)*. I_j is an indicator variable that takes a value of one for items in Market A and zero otherwise, $Post_t$ is a post-treatment indicator that equals one for months in the post-merger period, γ_t is the month fixed effects and θ_i is the seller fixed effects. The results of this estimation are presented in Panel (A) of Table 5. We see a decrease in both the sales and transactions for incumbent sellers in Market A. However, we notice that the treatment and control groups are not comparable as they are not balanced (please see Figure B1) with respect to the covariates in the pre-treatment period and the TWFE does not inherently adjust for imbalances in covariates leading to biased estimates.

We proceed to address the issue with TWFE by using Inverse Probability Weights (IPW) (Robins et al. 2000) in Equation 9. To make sure that the sellers in markets A and B exhibit comparable parallel trends, we use matching using Inverse Probability Weights from propensity scores generated via Entropy Balancing (Hainmueller 2012). The matching approach effectively eliminates virtually all differences between the treated sellers in Market A and the control group sellers in Market C. Figure B1 illustrates this balance.

To enhance the comparability between treated and untreated sellers, we incorporate an inverse probability of treatment weighting into Equation 9.

$$\Pr(A_i = 1) = \Pr(\alpha_0 + Z_i\alpha + \eta_i > 0) \tag{10}$$

We include the outcomes from the six pre-treatment periods as the covariates and calculate the weights for Equation 9 using the propensity score $\Pr(A_i = 1)$, which represents the likelihood of a seller being in the treated group. The weights for the computation of ATT are given by:

$$w_i = A_i + \frac{Pr(A_i = 1)}{(1 - Pr(A_i = 1))}(1 - A_i) \tag{11}$$

For the estimation, we use data from all sellers in Market A who had at least one transaction every month in the pre-period. We keep all sellers in Market C for potential donor pool. The results of the estimation in Panel A of Table 5 show that the sellers in Market A experienced a 9.1% decrease in frequency of transactions and a 11.4% decrease in sales relative to the sellers in control Market C. When we limit the transactions of the sellers in Market A to only buyers in Market A, we find that the sellers in Market A have a larger magnitude of decrease in transactions and sales, at 20.5% and 32.6% respectively.

We plot the monthly treatment effects by modifying the Equation 9 and present the plots for seller transactions in both markets and for seller transactions in Market A, in Figure 7 and in Figure 8 respectively. The zero estimates in the pre-periods for the plots provide support for the plausibility of the parallel trends assumption. Both quantitative estimates and visual plots consistently indicate a sharper decline within Market A when only intra-market transactions are considered. Interestingly, when including cross-market sales, the magnitude of the decrease lessens, suggesting that cross-market transactions partially offset the losses from within-market sales. This is further confirmed by the observation that the average sale price of cross-market transactions exceeds that of within-market transactions in the post-period, underscoring a strategic adjustment in seller behavior to mitigate losses. This adjustment likely reflects a shift towards higher-value cross-market transactions as a compensatory mechanism for the declines observed within the local market.

As a test of robustness, we estimate Synthetic Difference-in-Differences (Arkhangelsky et al. 2021) using *synthdid* package in R. We use the Jackknife estimator option to calculate the standard errors. The results from Synthetic DiD estimation, presented in Panel D of Table 5, indicate that the percentage decrease in both transactions and sales is within the same range as the estimates from DiD using IPW.

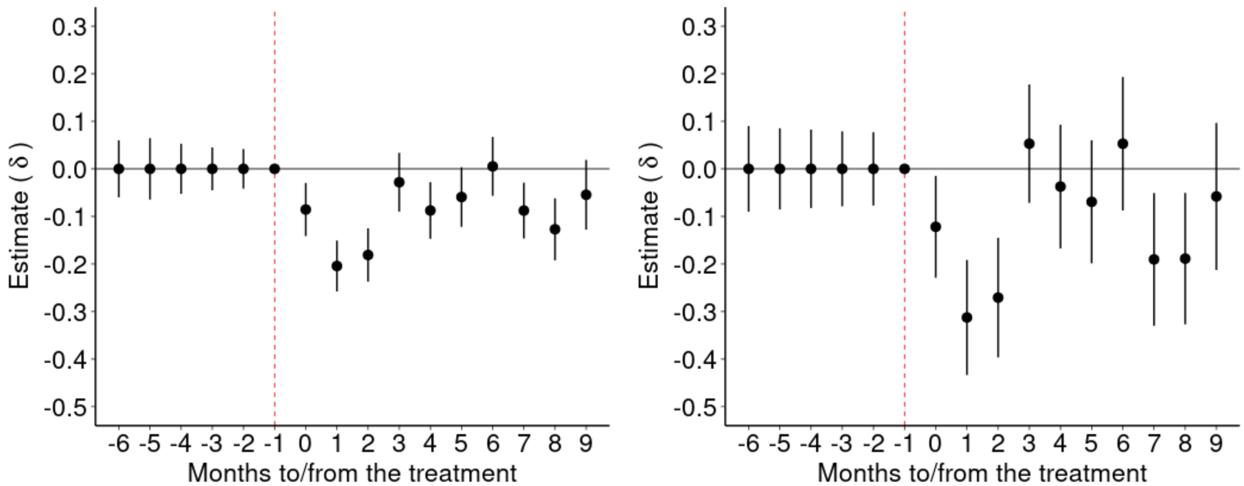
Table 5: Change in Outcomes for Sellers in A

	Transactions	Sales
(A) *Difference-in-Differences*		
treated $\times$ *post*	-0.027***	-0.109***
	(0.001)	(0.002)
Num.Obs.	196,069	196,069
R^2	0.348	0.267
(B) *Difference-in-Differences - IPW*		
treated $\times$ *post*	-0.091***	-0.114**
	(0.021)	(0.040)
Num.Obs.	145,666	145,666
R^2	0.815	0.647
(C) *DiD - IPW - Only Market A Transactions*		
treated $\times$ *post*	-0.205***	-0.326***
	(0.021)	(0.040)
Num.Obs.	145,666	145,666
R^2	0.815	0.647
(D) *Synthetic Difference-in-Differences*		
treated $\times$ *post*	-0.102***	-0.140***
	(0.013)	(0.030)
Num.Obs.	145,666	145,666
Seller f.e.	yes	yes
Time f.e.	yes	yes

Notes: Panel A reports the results of TWFE estimation, for incumbent sellers - existing in both pre and post-periods in Market A, with the incumbent sellers in Market C as the control group. Panel B reports results for a IPW DID. Panel B reports results using IPW DID for sales in only Market A. Panel D reports the results for a Synthetic DID. The dependent variables are *log(1+ Transactions)* and *log(1+ Sales)*. *** p<0.001, ** p<0.01, * p<0.05; standard errors reported in parentheses. The number of observations is divided by a constant to obfuscate the actual values.

Figure 7: Monthly Treatment Effects for the Sellers in Market A

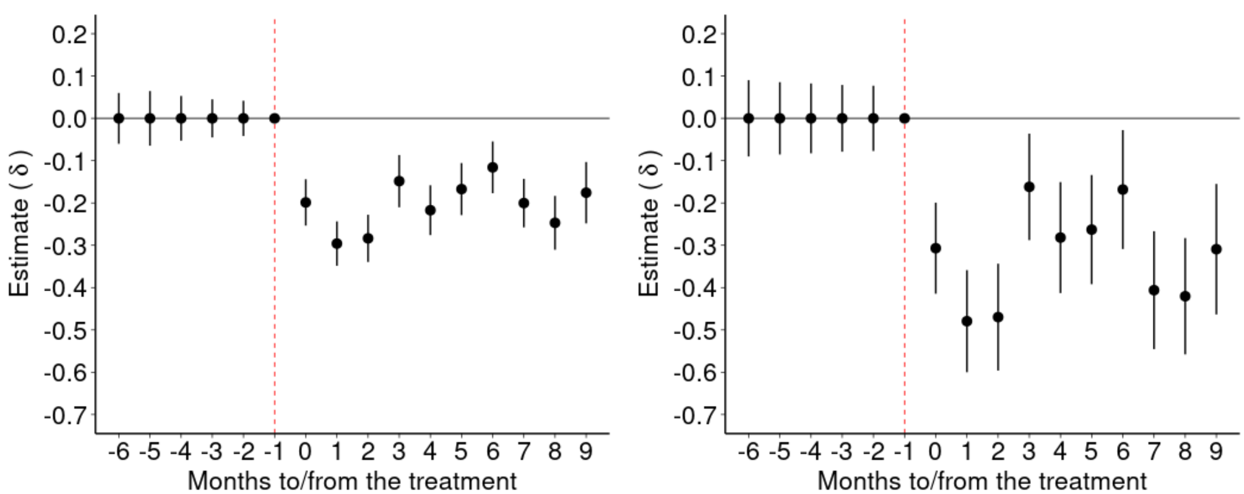
(a) Transactions (b) Sales



Notes: Panel (a) shows the monthly treatment effects for the Transactions and Panel (b) shows the monthly treatment effects for the Sales. The estimation uses the data from all (within and cross-market) transactions for sellers in Market A.

Figure 8: Monthly Treatment Effects for the Sellers in Market A Limited to Transactions in Market A

(a) Transactions (b) Sales



Notes: Panel (a) shows the monthly treatment effects for the Transactions and Panel (b) shows the monthly treatment effects for the Sales. The estimation uses the data from the within transactions for sellers in Market A.

8 Mechanisms and Alternative Explanations

So far, we have shown two important results: (1) the price differential has substantially decreased in Market A, but (2) the remaining price differential is still significant, around

20%. Why might this be the case? While we do not attempt to exhaustively and definitively document and identify the underlying mechanisms behind our results, we discuss potential explanations for the main findings. Table 6 lists the different potential mechanisms and their likelihood of explaining the price difference due to local bias.

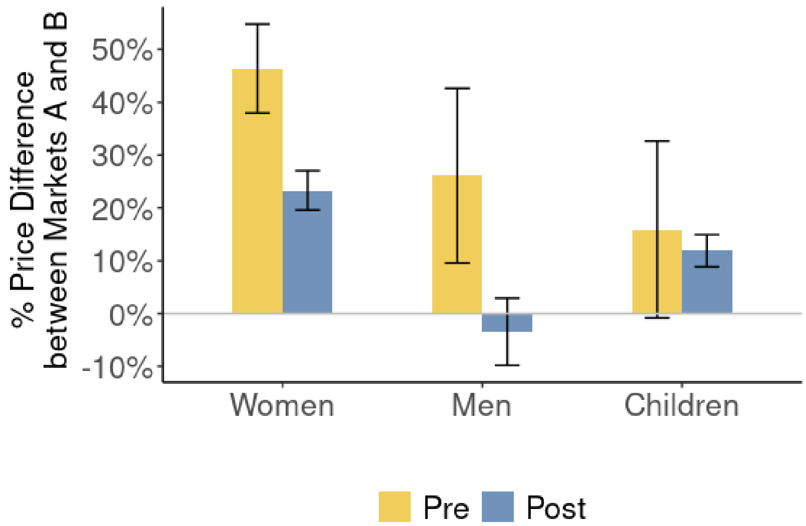
Table 6: Summary of Mechanisms Contributing to Local Bias and their Likelihood

Mechanism	**Description**	**Likelihood of Explaining Local Bias**
Preference Heterogeneity	Specificity of preferences in certain product categories, driving consumers to favor local sellers	High (Section 8.1)
Shipping Costs & Times	Differences in shipping costs and times between local and non-local transactions	Low (Section 8.2)
Frictions in Communication and Trust	Reduced trust and communication barriers in non-local transactions	Medium to High (Section 8.3)
Other Frictions	Financial transaction issues, return process complexity and delays	Low (Section 8.4)

8.1 Preference Heterogeneity

Using the same methodology as described in section 7, we estimate price differentials before and after the merger across major product categories: Women, Men, and Children. Figure 9 shows price differences for these categories from the perspective of buyers in Market A. In all three categories — Women, Men, and Children — the price difference significantly decreases from the pre-merger to the post-merger period, as predicted by our theoretical framework. However, the remaining price differential varies substantially across categories. Overall, we see that the platform-level local bias reported in Figure 5 is driven by transactions within Women's category. The local bias for Men's category products is positive and statistically significantly different from zero, and it completely disappears for Children's category products.

Figure 9: Price Differences Across Categories from the Perspective of Buyers in Market A



Notes: This figure shows the changes in average price differences across three different product categories for buyers in market A before and after the merger of markets A and B, after controlling for pseudo-product ID fixed effects.

The observed variance in local bias post-merger across different categories—specifically the pronounced bias in Women's category items, the lack of bias in Men's category items, and the minimal bias in Children's category items—can be attributed to several factors.

First, the heightened local bias for Women's category items may stem from the greater diversity and specificity of preferences in this category. The markets for Women's category often exhibit a higher degree of trend sensitivity and regional variation in style preferences, leading consumers to favor local sellers who are more attuned to these localized trends. This aligns with the theory of localized preference structures, where consumers exhibit a stronger affinity for products that reflect their immediate cultural and social milieu.

Second, the disappearance of local bias for Men's category items might be explained by the relatively homogenous nature of Men's category fashion across different regions. Items in Men's category tends to be less varied and subject to fewer regional trends compared to other categories. Consequently, the standardization and universal appeal of Men's category items reduce the need for localized purchasing, as consumers are more inclined to perceive products as substitutable regardless of the seller's location.

Third, the relatively small local bias observed for Children's category items can be further understood through the lens of the literature on buying for others (i.e. the items being bought by the parents). Research such as Gillison and Reynolds (2016) and Babin et al. (1994) suggests that when individuals purchase items for others, their purchasing decisions are more utilitarian, driven more by functional and practical considerations than by personal preferences or localized trends. The buyers often prioritize factors like price, quality, and convenience, which might be less influenced by regional variations. Thus, even though the same individuals may be buying both Women's category and Children's category items, their purchasing criteria differ significantly, resulting in a weaker inclination toward local sellers and a prioritization of better deals.

8.2 Shipping Time Differences

While shipping costs for cross-market purchases are higher compared to local purchases, our local bias estimation already accounts for these differences, therefore shipping cost differences do not explain the persistent local bias. However, another significant factor related to shipping that can contribute to local bias is the time taken for delivery. Although we do not observe the delivery times in the data, we know from the platform that cross-market deliveries take, on average, one day longer. While minimal, longer shipping times can theoretically increase uncertainty and perceived risk associated with cross-market transactions. This uncertainty could potentially deter buyers from choosing cross-market options and reinforce a preference for local sellers who can fulfill orders more quickly and reliably.

To investigate whether differences in shipping times influence price premiums, we leverage the fact that intra-city transactions typically involve shorter shipping times. This allows us to assess whether buyers are willing to pay a premium for faster deliveries by examining the systematic price differences between intra-city and inter-city purchases. Specifically, we compare the prices of products purchased within the same city to those from different cities by buyers in Market A. Our estimation includes a specification that controls for week $\times$

pseudo-product ID fixed effects, as well as seller city fixed effects:

$$total\_price_{it} = \beta \times \text{same\_city}_{it} + \gamma_{it} + \theta_{it} + \varepsilon_{it} \tag{12}$$

Here, $total\_price_{it}$ is the sale price of an item with a pseudo-product ID i. $same\_city_{it}$ is a dummy variable (1 or 0) indicating whether the item was bought from a seller in the same city or not. γ_{it} represents the double interaction of *Pseudo-Product ID*$\times$*Week* fixed effects, used to compare the price differences of pseudo-product IDs available within the same week. θ_{it} represents the seller city fixed effects. Our estimate of β can be interpreted as the average sale price difference between products bought from the same city and a different city. If the buyers in Market A pay a higher price for the products they buy from the same city, it means that they pay a price premium for the shorter shipping times. The result of the estimation of Equation 12 indicates that there is no statistically significant difference in prices between same-city transactions and cross-city transactions, meaning that the buyers in Market A do not seem to pay higher prices for shorter shipping times.

In addition, shipping time differences alone are unlikely to fully account for the local bias we observe. If they were the sole driver, we would expect to see consistent price differences across product categories post-merger. However, as discussed in Section 8.1, local bias completely disappears for Men's category items. This suggests that other factors, beyond shipping times, must be influencing buyer preferences and driving the persistence of local bias in other categories.

8.3 Frictions in Communication and Trust

Another potential mechanism contributing to the observed local bias is the friction in communication between buyers and sellers from different markets. A meta-analysis of previous research shows that sharing a common language increases trade flow by 44% (Egger and Lassmann 2012). While the platform translates product listings into the user's local language, reducing friction in listings, the nature and extent of pre-sale communication may still influence transactions. On this platform, the buyers can chat with sellers during the purchase process, and detailed communication may be easier in their native language (when compared to the machine-translated chats), potentially leading to a preference for local sellers over those from non-native markets.

Trust is another factor that plays an important role in the bilateral trade. The buyers and sellers in different markets can have different levels of dyadic trust between them. Reduced bilateral trust results in decreased trade between two countries (Guiso et al. 2009). Hortaçsu et al. (2009) suggest that trust may play a role in reinforcing the home bias. However, Guiso et al. (2009) finds that to the extent that the goods are homogeneous, the effect of trust on trade appears to be both economically and statistically insignificant. As we use comparable products in our analysis of price differences due to local bias, this approach may help alleviate some of the concerns related to trust. However, the concerns due to the difference in the bilateral trust remain.

8.4 Other Frictions

8.4.1 Friction due to Returns

The ease of returns could influence buyers' preference for local sellers. The extra time required to ship a product back to a non-local seller, along with the subsequent refund processing, may lead buyers to favor local sellers. Additionally, concerns about potential disputes during the return process could play a role. On this platform, buyers can use an integrated shipping option, and sellers process refunds upon receiving returned products. In the event of disagreements, the platform acts as an arbitrator and issues a final decision. Given that the difference in shipping times between markets is between one to two days, return-related frictions are unlikely to be a significant factor contributing to local bias.

8.4.2 Frictions in Financial Transactions

The effectiveness of online payment systems plays a vital role in driving online trade (Gomez-Herrera et al. 2014). Potential frictions in financial transactions could plausibly contribute to local bias in online marketplaces operating across different platforms. Financial transaction frictions could include currency exchange fees, differences in payment methods, and varying levels of trust in cross-border payment systems. These frictions can deter buyers from purchasing from foreign sellers, leading to a preference for local transactions where such issues are minimized. Assessing the likelihood of these frictions affecting market A buyers on a platform purchasing from market B sellers involves understanding the specific financial and regulatory environment. We argue that financial frictions are unlikely to explain the observed local bias post-merger as the payments are processed on the platform without any additional charge.

9 Concluding Remarks

This study investigates the impact of international cross-market integration on consumer behavior in a major C2C online marketplace. By analyzing the effects of market integration on key platform outcomes, including sales, prices, and transaction frequency, we provide a comprehensive assessment of how market efficiency is influenced by the integration of diverse linguistic markets. Our approach utilized language-agnostic embeddings to identify similar items across markets, revealing that while price differences decreased significantly post-integration, local biases persisted.

Despite significant improvements in market efficiency post-merger, local bias remains a substantial barrier to full integration. The findings underscore the need for digital platforms to adopt strategies that address not only economic factors but also psychological and logistical barriers to cross-market transactions. First, the platform could mitigate the effect of local bias by subsidizing cross-market shipping costs. By lowering the cost of cross-market transactions, platforms can encourage consumers to make purchases from neighboring markets, thereby reducing the preference for the items in the local market. Next, to the extent that price is seen as a signal for quality (Wolinsky 1983), the observed local bias in pricing analysis can inform the platform's pricing recommendations to sellers. In addition, the platform could modify the search and discovery process by adjusting the weights for items from local and neighboring markets. This adjustment can enhance the visibility of cross-market items, encouraging consumers to consider products from neighboring regions and fostering a more integrated market. Additionally, the platform could implement features such as improved translation services and standardized information requests to reduce communication barriers. By addressing these frictions, the platform can enhance trust and facilitate smoother transactions between buyers and sellers from different markets.

References

Amazon (2022) 2022 Amazon SME Report. URL https://assets.aboutamazon.com/51/f6/01cca00a4447830799f028a0f0b4/amazon-sme-report-2022-eu.pdf, accessed: 2024-06-07.

Anderson JE, Van Wincoop E (2003) Gravity with Gravitas: A Solution to the Border Puzzle. *The American Economic Review* 93(1):170–192.

Anderson JE, Van Wincoop E (2004) Trade Costs. *Journal of Economic Literature* 42(3):691–751.

Arkhangelsky D, Athey S, Hirshberg DA, Imbens GW, Wager S (2021) Synthetic Difference-in-Differences. *American Economic Review* 111(12):4088–4118.

Babin BJ, Darden WR, Griffin M (1994) Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value. *Journal of Consumer Research* 20(4):644–656.

Bimpikis K, Elmaghraby WJ, Moon K, Zhang W (2020) Managing Market Thickness in Online Business-to-Business Markets. *Management Science* 66(12):5783–5822.

Blum BS, Goldfarb A (2006) Does the Internet Defy the Law of Gravity? *Journal of International Economics* 70(2):384–405.

Brynjolfsson E, Hui X, Liu M (2019) Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform. *Management Science* 65(12):5449–5460.

Burtch G, Ghose A, Wattal S (2014) Cultural Differences and Geography as Determinants of Online Prosocial Lending. *MIS Quarterly* 38(3):773–794.

Ceylan G, Diehl K, Proserpio D (2024) Words Meet Photos: When and Why Photos Increase Review Helpfulness. *Journal of Marketing Research* 61(1):5–26.

Chintagunta PK, Chu J (2021) Geography as Branding: Descriptive Evidence from Taobao. *Quantitative Marketing and Economics* 19:53–92.

Chu J, Manchanda P (2016) Quantifying Cross and Direct Network Effects in Online Consumer-to-Consumer Platforms. *Marketing Science* 35(6):870–893.

Compiani G, Morozov I, Seiler S (2023) Demand estimation with text and image data .

Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. Burstein J, Doran C, Solorio T, eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186 (Minneapolis, Minnesota: Association for Computational Linguistics), URL http://dx.doi.org/10.18653/v1/N19-1423.

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. (2020) An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations.*

Dzyabura D, El Kihal S, Hauser JR, Ibragimov M (2023) Leveraging the Power of Images in Managing Product Return Rates. *Marketing Science* 42(6):1125–1142.

Egger PH, Lassmann A (2012) The Language Effect in International Trade: A Meta-Analysis. *Economics Letters* 116(2):221–224.

Elfenbein DW, Fisman R, McManus B (2023) The Impact of Socioeconomic and Cultural Differences on Online Trade. *Management Science* 69(10):6181–6201.

Farronato C, Fong J, Fradkin A (2023) Dog Eat Dog: Balancing Network Effects and Differentiation in a Digital Platform Merger. *Management Science* .

Focarelli D, Panetta F (2003) Are Mergers Beneficial to Consumers? Evidence from the Market for Bank Deposits. *American Economic Review* 93(4):1152–1172.

Fong J (2024) Effects of Market Size and Competition in Two-Sided Markets: Evidence from Online Dating. *Marketing Science* .

Future Market Insights (2023) Digital Commerce Platform Market Report. URL https://www.futuremarketinsights.com/reports/digital-commerce-platform-market, accessed: 2024-06-20.

Gillison S, Reynolds K (2016) Shopping for Yourself versus Shopping for Someone Else. *Journal of Consumer Marketing* 33(4):225–234.

Gomez-Herrera E, Martens B, Turlea G (2014) The Drivers and Impediments for Cross-border e-commerce in the EU. *Information Economics and Policy* 28:83–96.

Guiso L, Sapienza P, Zingales L (2009) Cultural Biases in Economic Exchange? *The quarterly journal of economics* 124(3):1095–1131.

Hainmueller J (2012) Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political analysis* 20(1):25–46.

Hillberry R, Hummels D (2003) Intranational Home Bias: Some Explanations. *Review of Economics and Statistics* 85(4):1089–1092.

Hillberry R, Hummels D (2008) Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data. *European Economic Review* 52(3):527–550.

Hortaçsu A, Martínez-Jerez FA, Douglas J (2009) The Geography of Trade in Online Transactions: Evidence from eBay and Mercadolibre. *American Economic Journal: Microeconomics* 1(1):53–74.

Hui X (2020) Facilitating Inclusive Global Trade: Evidence from a Field Experiment. *Management Science* 66(4):1737–1755.

Katz ML, Shapiro C (1985) Network Externalities, Competition, and Compatibility. *The American Economic Review* 75(3):424–440.

Lendle A, Olarreaga M, Schropp S, Vézina PL (2016) There Goes Gravity: eBay and the Death of Distance. *The Economic Journal* 126(591):406–441.

Li J, Netessine S (2020) Higher Market Thickness Reduces Matching Rate in Online Platforms: Evidence from a Quasiexperiment. *Management Science* 66(1):271–289.

Lin M, Viswanathan S (2016) Home Bias in Online Investments: An Empirical Study of an Online Crowdfunding Market. *Management Science* 62(5):1393–1414.

Loertscher S, Muir EV, Taylor PG (2022) Optimal Market Thickness. *Journal of Economic Theory* 200:105383.

McCallum J (1995) National Borders Matter: Canada-US Regional Trade Patterns. *The American Economic Review* 85(3):615–623.

Nimdzi Insights (2024) The 2024 Nimdzi 100. URL https://www.nimdzi.com/wp-content/uploads/2024/03/The-2024-Nimdzi-100.pdf, accessed: 2024-03-22.

Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, et al. (2021) Learning Transferable Visual Models from Natural Language Supervision. *International Conference on Machine Learning*, 8748–8763 (PMLR).

Reimers N, Gurevych I (2019) Sentence-bert: Sentence embeddings using siamese bert-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* (Association for Computational Linguistics), URL http://arxiv.org/abs/1908.10084.

Reimers N, Gurevych I (2020a) Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4512–4525.

Reimers N, Gurevych I (2020b) Making monolingual sentence embeddings multilingual using knowledge distillation. Webber B, Cohn T, He Y, Liu Y, eds., *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4512–4525 (Online: Association for Computational Linguistics), URL http://dx.doi.org/10.18653/v1/2020.emnlp-main.365.

Reshef O (2023) Smaller Slices of a Growing Pie: The Effects of Entry in Platform Markets. *American Economic Journal: Microeconomics* 15(4):183–207.

Robins JM, Hernan MA, Brumback B (2000) Marginal Structural models and Causal Inference in Epidemiology. *Epidemiology* 11(5):550–560.

Rousseeuw PJ (1987) Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics* 20:53–65.

Rysman M (2004) Competition between Networks: A Study of the Market for Yellow Pages. *The Review of Economic Studies* 71(2):483–512.

Sanh V, Debut L, Chaumond J, Wolf T (2019) Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv* abs/1910.01108.

Schooler RD (1965) Product Bias in the Central American Common Market. *Journal of Marketing Research* 2(4):394–397.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. *Advances in neural information processing systems* 30.

Weber M, Candia B, Afrouzi H, Ropele T, Lluberas R, Frache S, Meyer BH, Kumar S, Gorodnichenko Y, Georgarakos D, et al. (2023) Tell Me Something I Don't Already Know: Learning in Low and High-Inflation Settings. Technical report, National Bureau of Economic Research.

Wolf HC (2000) Intranational Home Bias in Trade. *Review of Rconomics and Statistics* 82(4):555–563.

Wolinsky A (1983) Prices as Signals of Product Quality. *The Review of Economic Studies* 50(4):647–658.

Yi KM (2010) Can Multistage Production Explain the Home Bias in Trade? *American Economic Review* 100(1):364–393.

Appendix

Table of Contents

A Proofs

Proof of Proposition 1

Proof. We begin with the profit functions for both sellers, A and B, after integration. The profit functions are given by:

$$\pi_A = \mathcal{P}_A\left(\gamma\left(\frac{1-\alpha(\mathcal{P}_A-\mathcal{P}_B)+\chi}{2}\right)+\frac{1-(\mathcal{P}_A-\mathcal{P}_B)-\chi}{2}\right),$$

and

$$\pi_B = \mathcal{P}_B\left(\gamma\left(\frac{1+\alpha(\mathcal{P}_A-\mathcal{P}_B)-\chi}{2}\right)+\frac{1+(\mathcal{P}_A-\mathcal{P}_B)+\chi}{2}\right).$$

To find the equilibrium price for A, we first take the derivative of π_A with respect to $\mathcal{P}_A$:

$$\frac{\partial \pi_A}{\partial \mathcal{P}_A} = \frac{1+\gamma-2\mathcal{P}_A(\gamma\alpha+1)+\mathcal{P}_B(\gamma\alpha+1)+(\gamma-1)\chi}{2}.$$

Setting this equal to zero:

$$1+\gamma-2\mathcal{P}_A(\gamma\alpha+1)+\mathcal{P}_B(\gamma\alpha+1)+(\gamma-1)\chi = 0.$$

Rearranging for $\mathcal{P}_A$, we get:

$$\mathcal{P}_A = \frac{1+\gamma+\mathcal{P}_B(\gamma\alpha+1)+(\gamma-1)\chi}{2(\gamma\alpha+1)}. \tag{13}$$

Similarly, for B:

$$\frac{\partial \pi_B}{\partial \mathcal{P}_B} = \frac{1+\gamma-2\mathcal{P}_B(\gamma\alpha+1)+\mathcal{P}_A(\gamma\alpha+1)-(\gamma-1)\chi}{2}.$$

Setting this equal to zero:

$$1+\gamma-2\mathcal{P}_B(\gamma\alpha+1)+\mathcal{P}_A(\gamma\alpha+1)-(\gamma-1)\chi = 0.$$

Rearranging for $\mathcal{P}_B$, we get:

$$\mathcal{P}_B = \frac{1+\gamma+\mathcal{P}_A(\gamma\alpha+1)-(\gamma-1)\chi}{2(\gamma\alpha+1)}.$$

By substituting this expression for $\mathcal{P}_B$ into Eq. (13), we obtain:

$$\mathcal{P}_A = \frac{1+\gamma+\frac{1+\gamma+\mathcal{P}_A(\gamma\alpha+1)-(\gamma-1)\chi}{2(\gamma\alpha+1)}(\gamma\alpha+1)+(\gamma-1)\chi}{2(\gamma\alpha+1)}.$$

After simplifying the resulting expression, we get:

$$\mathcal{P}_A = \frac{1+\gamma-\chi(1-\gamma)}{2(1+\alpha\gamma)}.$$

Similarly, substituting $\mathcal{P}_A$ into the equation for $\mathcal{P}_B$, we obtain:

$$\mathcal{P}_B = \frac{1+\gamma+\chi(1-\gamma)}{2(1+\alpha\gamma)}.$$

This concludes the proof.

□

Proof of Lemma 1

Proof. Before integration, seller A profits are given by

$$\pi_A = \mathcal{P}_A\Big(\frac{1-\alpha(\mathcal{P}_A-\mathcal{P}'_A)}{2}\Big). \tag{14}$$

Taking the first order conditions yields

$$\frac{\pi_A}{\mathcal{P}_A} = \Big(1-\alpha(\mathcal{P}_A-\mathcal{P}'_A)\Big) - \alpha\mathcal{P}_A = 0, \tag{15}$$

or $\mathcal{P}_A = \frac{1+\alpha\mathcal{P}_{A'}}{2\alpha}$. Imposing $\mathcal{P}_A = \mathcal{P}'_A$ (which is easily verified to be the only possibility in equilibrium) yields the result: $\mathcal{P}_A = \frac{1}{\alpha}$.

The proof for $\mathcal{P}_B = 1$ follows exactly the same steps and corresponds to the case of $\alpha = 1$. This concludes the proof.

□

Proof of Proposition 2

Proof. Given the expressions we find for the optimal prices of firm A in the previous two propositions, it is enough to show that

$$\frac{1}{\alpha} - \left(\frac{1+\gamma-\chi(1-\gamma)}{2(1+\alpha\gamma)}\right) > 0 \;\; \forall\alpha < 1. \tag{16}$$

To see this, notice that upon simple algebraic manipulation Eq. (16) simplifies to

$$\alpha(\gamma-1)(\chi-1) < 2. \tag{17}$$

But this is obviously satisfied since $\alpha < 1$, $\chi - 1 \in (-1, 0)$ and $\gamma - 1 \in (-1, \infty)$. Thus, the LHS is either negative or, when positive, bounded above by α.

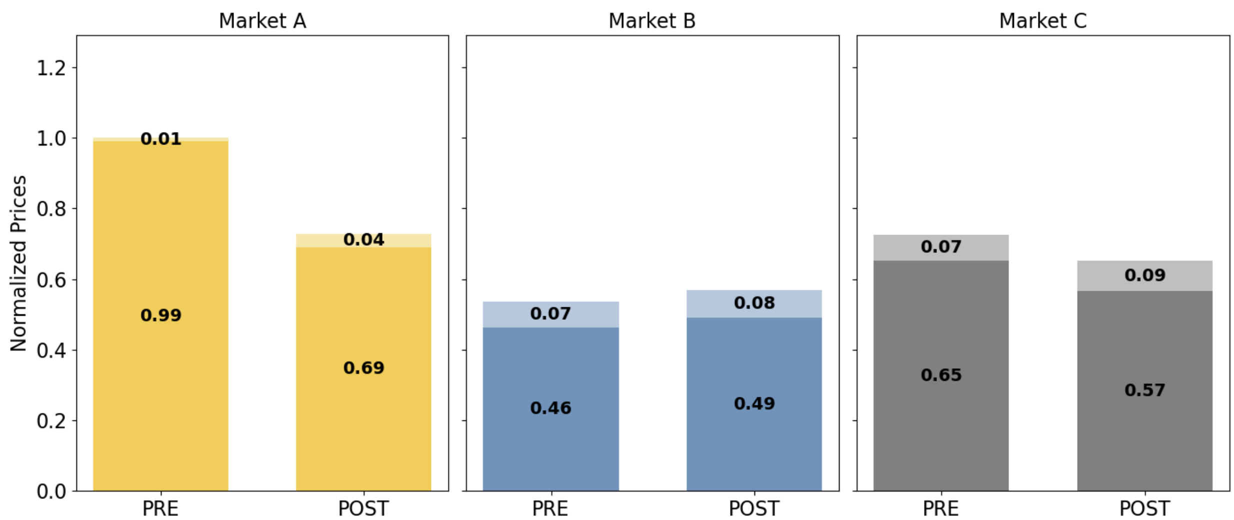
This concludes the proof.

□

B Robustness for Price Difference

The observed price changes could be attributed to inflation rather than the merger. To account for inflation, we adjusted the prices using an Inflation Adjustment factor, calculated as the ratio of the CPI of the base month to the CPI of the month in which the transaction occurred. The price differences across markets, adjusted for inflation, are presented in Figure A1. These adjusted price differences follow the same patterns as those shown in Figure 4.

Figure A1: Average Prices Controlling for Item Similarity Pre- and Post- Merger of Markets A and B

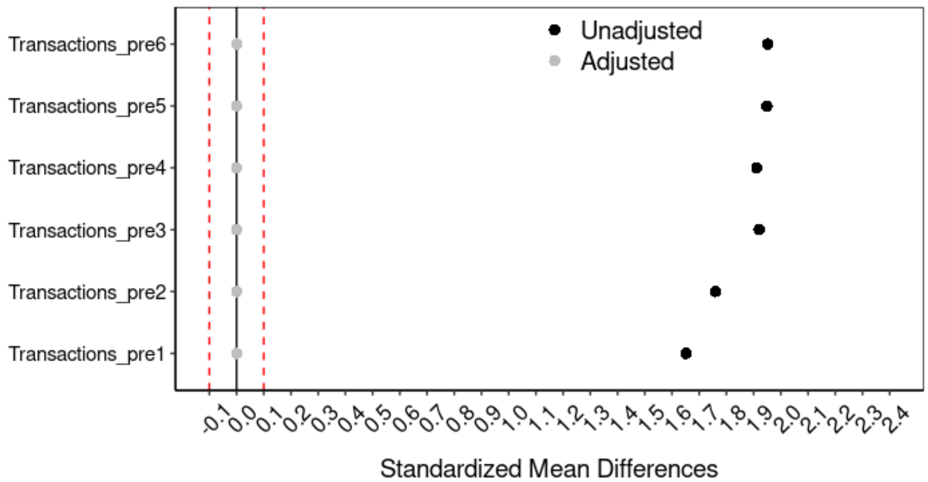


Notes: The figure compares the average prices in three markets—A, B, and C—before and after the merger of markets A and B. In the figure, darker shades represent the sale prices of products, while lighter shades represent the listing prices. Prices in B increase post the merger by x%. The listing prices and sale prices in A are higher before the merger and decrease after the merger due to higher negotiated discounts. To protect the confidentiality of the data, the prices are normalized.

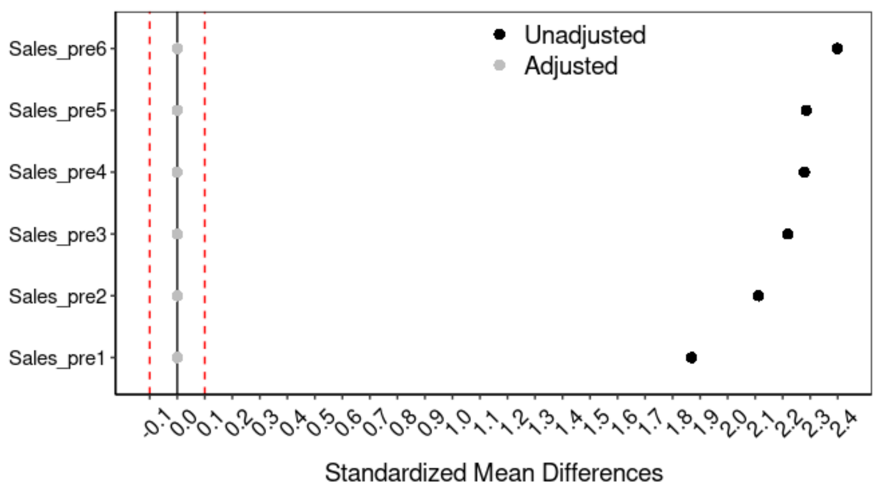
C Covariate Balance for Seller Outcomes

Figure B1: Covariate Balance

(A) Transactions



(B) Sales



Notes: Panel A shows the Covariate balance for Transactions and Panel B shows the Covriate balance for Sales in the pre-treatment periods.

D Details about the Multi-lingual Image/Text Embeddings

`clip-ViT-B-32-multilingual-v1` embeddings are vector representations of images and text that exhibit several desirable properties. First, they are *sentence* embeddings; in contrast with mean-pooled token embeddings from transformer models like BERT (Devlin et al. 2019), sentence embeddings are explicitly designed to capture semantic textual similarity between whole sentences (Reimers and Gurevych 2019).

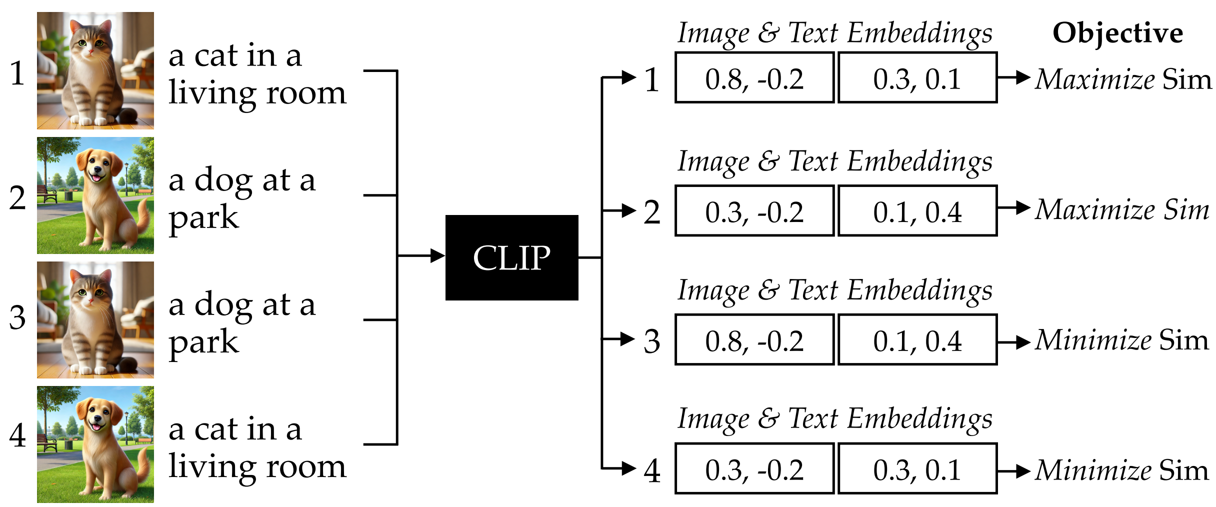
Second, they are *multilingual* embeddings; sentences in different languages having a similar semantic meaning will have similar embeddings.

Third, they are also *image* embeddings; images are embedded in the same vector space as text such that semantically similar image-text pairs will have similar embeddings. Each of these desirable properties are engineered using a sequence of fine-tuning steps, which we detail below.

D.1 Jointly Embedding Images and Text

The originating embedding model for `clip-ViT-B-32-multilingual-v1` is CLIP (Radford et al. 2021). CLIP embeddings are derived via contrastive representation learning on a corpus of 400 million image-caption pairs.

Figure C1: CLIP Embeddings



Notes: For each batch of $N \times N$ image-text pairs (here $N = 2$), CLIP (Reimers and Gurevych 2020b) jointly learns embeddings of images and text such that (i) the cosine similarity of the embeddings of image-text pairs that are image-caption pairs (here, pairs 1 and 2) is maximized, and (ii) the cosine similarity of the embeddings of image-text pairs that are not image-caption pairs (here, pairs 3 and 4) is minimized.

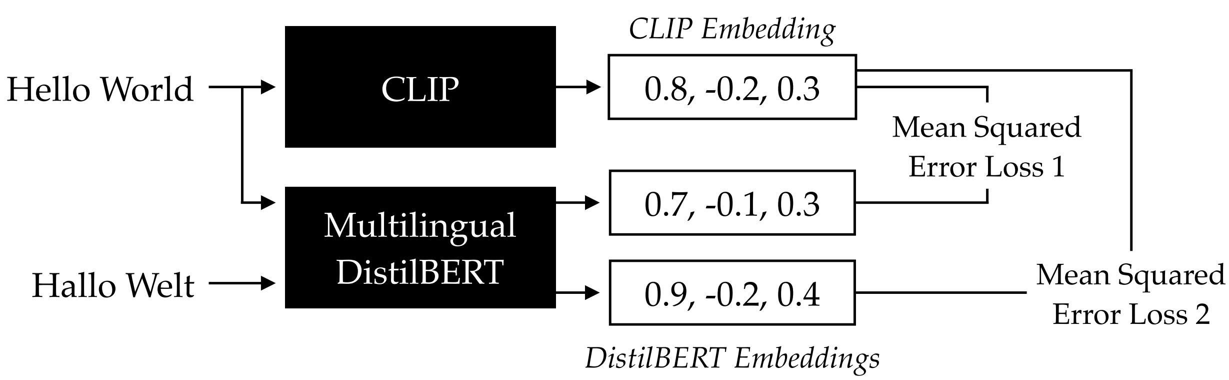
Specifically, for each training batch of $N \times N$ image-text pairs (where the text may or may not be the image's caption), CLIP learns image and text embeddings such that (i) the cosine similarity of the image and text embeddings for image-caption pairs is high, and (ii) the cosine similarity of the image and text embeddings of image-text pairs that are not image-caption pairs is low.

The initial image and text embeddings are derived from image and text transformers respectively (Vaswani et al. 2017, Dosovitskiy et al. 2020), which are in turn fine-tuned during the representation learning process.

D.2 Learning Multilingual Embeddings

CLIP embeddings are monolingual. `clip-ViT-B-32-multilingual-v1` adapts CLIP to over 50 different languages using multilingual knowledge distillation (Reimers and Gurevych 2020b).

Figure C2: Multilingual Embeddings



Notes: During the process of multilingual knowledge distillation (Reimers and Gurevych 2020b), multilingual DistilBERT (Sanh et al. 2019) is fine-tuned to produce embeddings of parallel source-translation sentence pairs (here, "Hello World" in English and "Hallo Welt" in German) that are each similar to the CLIP embedding of the source sentence with respect to the mean squared error.

Specifically, CLIP is used to fine-tune multilingual DistilBERT (Sanh et al. 2019) on a corpus of parallel (i.e., source and translated) sentences $(s_1, t_1), (s_2, t_2), \ldots$ in 50+ languages such that DistilBERT's embeddings of the source and translated sentences are similar to CLIP's embedding of the source sentence. This results in similar embeddings for semantically similar sentences agnostic to their language.