



AEA Publication Dynamics and Supportive Literature Management

Final Project for GLBL 6060 Applied Python for Global Affairs

Rosemary Zhang and Yirui Zhao

April 25, 2024

Yale University

AEA Publication Analysis:

- **Field Popularity:** Identify emerging research areas.
- **Collaboration Networks:** Analyze patterns in author collaborations.
- **Theme Discovery:** Uncover prevalent research topics.
- **Geographical Trends:** Examine focus of research across different countries.

Literature Search:

- Locate research that aligns with your specific interests.

Literature Review Support:

- Facilitate paper categorization with automated CSV uploads.

Source of Data:

We collect paper information from the American Economic Association website.

Included Journals:

1. **American Economic Review (AER)** - Volumes 89 to present.
2. **AER: Insights** - Volume 1 to present.
3. **AEJ: Applied Economics** - Volume 1 to present.
4. **AEJ: Economic Policy** - Volume 1 to present.
5. **AEJ: Macroeconomics** - Volume 1 to present.
6. **AEJ: Microeconomics** - Volume 1 to present.

Database Size:

Our sample includes **7,375 papers**.

Database Variables

Ghosting the Tax Authority: Fake Firms and Tax Fraud in Ecuador

Title

Paul Carrillo
Dave Donaldson
Dina Pomeranz
Monica Singhal

Authors

AMERICAN ECONOMIC REVIEW: INSIGHTS
VOL. 5, NO. 4, DECEMBER 2023
(pp. 427-44)

Journal

Issue

Download Full Text PDF

Article Information

Abstract
An important but poorly understood form of firm tax evasion arises from "ghost firms"—fake firms that issue fraudulent receipts so that their clients can claim false deductions. We provide a unique window into this global phenomenon using transaction-level tax data from Ecuador. Five percent of firms use ghost invoices annually. Among these firms, ghost transactions comprise 14 percent of purchases. Ghost transactions are prevalent among large firms and firms with high-income owners and exhibit suspicious patterns, such as bunching below financial system thresholds. An innovative enforcement intervention targeting ghost clients rather than ghosts themselves led to substantial tax recovery.

Abstract

Citation
Carrillo, Paul, Dave Donaldson, Dina Pomeranz, and Monica Singhal. 2023. "Ghosting the Tax Authority: Fake Firms and Tax Fraud in Ecuador." *American Economic Review: Insights*, 5 (4): 427-44.
DOI: 10.1257/aeri.20220321

Choose Format: ▾

Additional Materials
Replication Package
Online Appendix (3.17 MB)
Author Disclosure Statement(s) (20.89 KB)

JEL Classification
D22 Firm Behavior: Empirical Analysis
H25 Business Taxes and Subsidies Including sales and value-added (VAT)
H26 Tax Evasion and Avoidance
K34 Tax Law
L25 Firm Performance: Size, Diversification, and Scope
O14 Industrialization: Manufacturing and Service Industries; Choice of Technology

JEL Codes

On the left, we show the information that we webscraped for each paper ^a.

^aThe JEL classification system categorizes scholarly literature in economics.

1. Field Popularity:

- Use the year slider and journal dropdown to explore trends, hide fields by clicking their names, and use the toolbar for zooming and saving graphs.

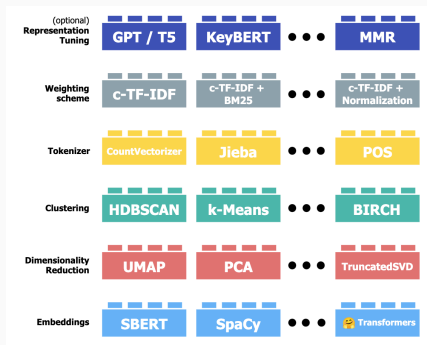
2. Collaboration Networks:

- Search an author to view their network, highlight connections by clicking nodes, and use the toolbar to zoom and save network sections.

3. Geographical Trends:

- Use the year slider to track annually popular countries, ranked by number of works involved.

4. Theme Discovery:



- We use **BER topic modeling** as an unsupervised technique to identify prevalent topics based on all literature abstracts in our database.

4. Theme Discovery:

- **Step 0: Preprocessing Abstract**

- To ensure clearer results, apply lemmatization to standardize words. The project's focus is on identifying themes rather than authors' arguments.

- **Step 1: Converting Documents to Numerical Representations**

- Utilizing **Sentence-Transformers** ('all-MiniLM-L6-v2') for effective capture of semantic similarity between documents.

- **Step 2: Reducing Dimensionality of Embeddings to Manageable Space**

- Employing **UMAP (Uniform Manifold Approximation and Projection)** for superior performance in handling complex, nonlinear datasets compared to PCA.

4. Theme Discovery:

- **Step 3: Clustering Reduced Embeddings into Topics**

- Utilizing **K-Means** and silhouette scores to determine the optimal number of clusters, assessing within-cluster similarity and out-of-cluster discrimination.

- **Step 4: Tokenizing Topics and Generating Topic Representations**

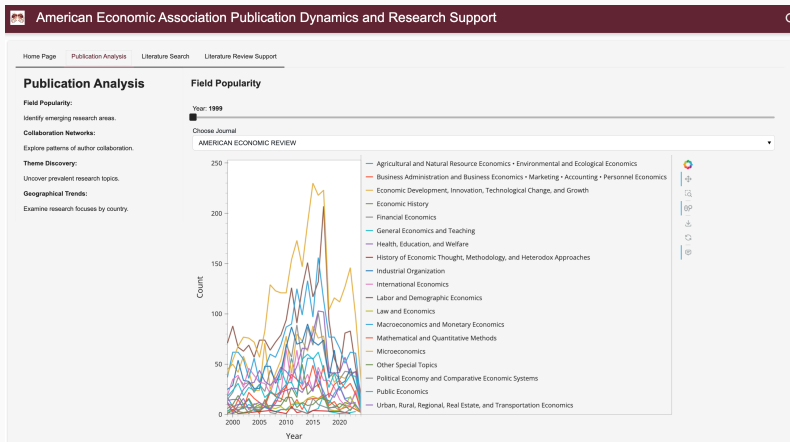
- Using **CountVectorizer** to extract unigrams (single words), bigrams (sequences of two adjacent words), and trigrams from the text.

- **Step 5: Token Weighting for Accurate Topic Representation**

- Employing **c-TF-IDF (Cluster-Term Frequency-Inverse Document Frequency)** to emphasize the distinctiveness of terms within each cluster, assigning higher weights to terms prevalent within a specific cluster but rare in others.

AEA Publication Analysis - Video Demo

[Link to the Demo Video](#)



Objective: Assist in finding the most relevant papers.

1. Filter Database:

- Search literature by criteria such as years, journals, JEL codes¹, authors, and keywords. Dynamic updates displayed as filters are applied.

2. Articles Recommendation:

- Enter potential research questions to find similar articles.
- Utilize natural language processing (NLP) and vector similarity calculations to rank articles based on relevance to specified interests.

3. Download Data:

- Download filtered literature data in CSV format.

¹JEL codes options sourced from

<https://cran.r-project.org/web/classifications/JEL.html>.

Link to the Demo Video

10

Objective: Streamline literature reviews.

1. Upload CSV Containing Paper Titles and Abstracts:

- Enables the inclusion of papers from various journals, independent of the CSV file created in the previous section.

2. Automated File Validation:


- Ensures correct format adherence.
- Verifies completeness of columns.
- Enforces a record count limit of up to 50 records.

3. Analysis and Grouping:

- Utilizes the OpenAI API to categorize papers by content.
- Sorts papers into thematic groups based on analysis.

Literature Review Support - Video Demo

[Link to the Demo Video](#)

 American Economic Association Publication Dynamics and Research Support

[Home Page](#) [Publication Analysis](#) [Literature Search](#) [Literature Review Support](#)

Literature Review Support

Upload Dataset:

Upload your dataset in CSV format containing paper titles and abstracts.

Update Categorization:

After uploading, press the button to categorize the literature based on similarity.

Please ensure your dataset contains the necessary columns: 'Title' and 'Abstract'.

Choose File [demo.csv](#)

Congratulations! Your input information is shown on the right.

Consider using the Literature Search Section to generate the .csv file. Users can also upload papers from other journals they are interested in. This tab uses the model 'gpt-3.5-turbo-0125'.

Update Categorization

	Title	Abstract
Mobilizing the Masses for Genocide	Do political elites use armed groups to foster civilian participation in genocidal violence? Are armed groups employed strategically? How do they mobilize civilians? I investigate these questions using data from the Rwandan Genocide. To establish causality, I exploit exogenous variation in armed groups' transport costs induced by weather fluctuations: the shortest distance of each village to the main road intersected with rainfall along the dirt tracks between main road and village. I find (i) 1 additional armed-group member resulted in 7.3 more civilian perpetrators; (ii) armed-group leaders employed their men strategically; and (iii) armed groups invoked civilians' obedience.	
Subsidy Policies and Insurance Demand	Using data from a two-year pricing experiment, we study the impact of subsidy policies on weather insurance take-up. Results show that subsidies increase future insurance take-up through their influence on payout experiences. Exploring mechanisms of the payout effect, we find that for households that randomly benefited from financial education, receiving a payout provides a one-time learning experience that improves take-up permanently. In contrast, households with poor insurance knowledge continuously update take-up decisions based on recent experiences with disasters and payouts. Combining subsidy policies with financial education can thus be effective in promoting long-run insurance adoption.	
Validating Migration Responses to Flooding Using Satellite and Vital Registration Data	Rainfall measures may be imperfect proxies for floods, given factors such as upstream water balance, proximity to rivers, and topography. We check the robustness of flooding-migration relationships by combining nationally-representative survey data with measures of flooding derived from weather stations, gridded products, and remote sensing tools. Linear probability models reveal that extreme flooding is negatively associated with out-migration. Rainfall-based proxies produce results qualitatively similar to those using the satellite-based measure of inundation, but only the latter is able to discern non-monotonic effects throughout the distribution. Moreover, estimates differ widely across areas, suggesting that households respond differently to rainfall and flooding.	
Agricultural Productivity and Structural Transformation: Evidence from Brazil	We study the effects of the adoption of new agricultural technologies on structural transformation. To guide empirical work, we present a simple model where the effect of agricultural productivity on structural development depends on the factor-bias of technical change. We test the predictions of the model by studying the introduction of genetically engineered soybean seeds in Brazil, which had heterogeneous effects on agricultural productivity across areas with different soil and weather characteristics. We find that technical change in soy production was strongly labor-saving and led to industrial growth, as predicted by the model.	
Has the Increased Attachment of Women to the Labor Market Changed a Family's Ability to Smooth Income Shocks?	An increase in a married woman's attachment to the labor market affected her family's ability to smooth unexpected income shocks. Between 1970 and 1990, the sharp rise in labor market attachment provided an increasingly important channel for smoothing shocks to spousal income. As the participation rate stabilized, this contribution to smoothing evaned out. In the Great Recession, both spouses received negative income shocks, and access to transfer income became the main insurance mechanism. Volatility of consumption followed volatility of family income trends but at a lower magnitude. Families' ability to weather income shocks didn't change during the 1970-2010 period.	

Group: Impacts of Weather on Economic Outcomes

Titles:

- Validating Migration Responses to Flooding Using Satellite and Vital Registration Data
- Rainfall Forecasts, Weather, and Wages over the Agricultural Production Cycle
- Temperature, Laboratory Reallocation, and Industrial Production: Evidence from India
- Communication Infrastructure and Stabilizing Food Prices: Evidence from the Telegraph Network in China
- Projection Bias in Catalog Orders
- The Accelerated Benefits Demonstration: Impacts on the Employment of Disability Insurance Beneficiaries

Explanation:

This group explores the effects of weather variations on various economic outcomes. The papers within this group investigate how weather conditions influence migration responses to flooding, wage levels over the agricultural production cycle, labor reallocation in response to temperature changes, the stabilization of food prices through improved communication infrastructure, consumer behavior in relation to catalog orders based on weather, and the impacts of weather on employment outcomes for disability insurance beneficiaries.

Group: Socioeconomic Impacts of Weather Shocks

Titles:

- Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall
- The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather
- Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US
- Every Day is Earth Day: Evidence on the Long-Term Impact of Environmental Activism
- Mortality, Temperature, and Public Health Provision: Evidence from Mexico
- The Fiscal Cost of Hurricanes: Disaster Aid versus Social Insurance

Explanation:

This group focuses on the socioeconomic implications of weather shocks on various aspects such as health, education, mortality rates, environmental activism, and public health provision. The papers within this group analyze how weather variability impacts different socioeconomic outcomes and discuss the responses and adaptations to weather-related challenges.

Group: Policy Responses to Weather-Related Challenges