Advancing Job Classification Systems: Leveraging AI to enhance historical mapping between ISIC revisions *

Gavin Lukanc † Matthew Hernandez ‡ Benjamin Nhan § Caleb Jensen ¶

June 18, 2024

Abstract

One key challenge in studying labor market trends over time is classifying job types and standardizing these classifications when they change. Traditional methods assume equal correspondence among codes, which often fails to capture complex relationships. We propose a new method using AI tools and provided descriptions to better translate between new and old classification systems. Our algorithm generates a probability mapping from newer to older versions. We test this method on the Indian Employment Unemployment Survey (EUS) data from 2009 and 2011, during the transition from the International Standard Industrial Classification (ISIC) version 3.1 to 4. Assuming economic conditions were similar, we back-coded the 2011 data to ISIC version 3.1 using our tool and compared the demographic makeup of job codes in EUS 2011 to EUS 2009. The results were positive, showing our method provides a more accurate framework for translating between classification systems, considering their complexities.

Keywords: Job Classification, ISIC 3.1, ISIC 4, International Standard Industrial Classifications, AI, Probability Mapping

1 Introduction

Classifying job types accurately is essential for understanding labor market trends. These classifications help policymakers tailor policies to support sectors most in need. As economies grow and new industries appear, job roles become more complex, making precise classifications crucial for tracking employment changes and addressing sector-specific challenges. For example, rapid technological advances and the digital transformation of commerce require updates to job categories. However, job classifications often lag behind industry innovations, leading to discrepancies in labor data that can affect policy decisions. Accurate and current job classifications enable policymakers to allocate resources effectively and design initiatives that help the labor market adapt.

*We thank the World Bank, Mario Gonert, and Professor Vasilaky
†Cal Poly MSQE,(e-mail: glukanc@calpoly.edu)
‡Cal Pol MSQE,(e-mail: mhern371@calpoly.edu)
§Cal Poly MSQE,(e-mail: bmnhan@calpoly.edu)
¶Cal Poly MSQE,(e-mail: cijensen@calpoly.edu)

The main challenge with job classifications is their frequent updates to reflect economic changes. This disrupts the continuity of labor market statistics, as new job codes may not align with old ones. For instance, in the technology sector, roles like data scientists and software engineers have emerged rapidly, replacing older roles such as 'computer programmers and 'network administrators.' If classification systems are not updated promptly, it can lead to misleading statistics that do not accurately represent the current job market. This misalignment complicates efforts to track job growth, as new job titles can obscure real employment trends. It also affects wage data, making it difficult to compare salaries over time due to job reclassification discrepancies. Moreover, this impacts the identification of educational needs for emerging fields. If new job roles are not accurately classified and tracked, developing appropriate training and educational programs becomes challenging. As a result, policymakers and educators may struggle to allocate resources effectively, leading to skill gaps in the workforce. Addressing these issues requires a dynamic and responsive job classification system that can keep pace with rapid economic changes, ensuring that labor market statistics remain accurate and relevant.

In this study, we develop a tool using Google's Gemini Aplication Program Interface (API) and advanced AI techniques to accurately map old job classifications to new ones. Our methodology leverages machine learning algorithms to analyze patterns and trends over time, resulting in precise predictions of historical job codes. We apply this tool to data from the Indian Employment Unemployment Survey (EUS) for the years 2009 and 2011 to validate its effectiveness and reliability. Our findings show that this approach improves the accuracy of labor market statistics, provides deeper insights into how job roles evolve, and helps policymakers make more informed decisions.

Additionally, changes in job classifications every few years, such as the shift from ISIC 3.1 to ISIC 4, pose significant challenges. This shift included major updates like reclassifying 'telecommunications' to encompass new digital services, reflecting the growing importance of technology and the digital economy. While these updates are necessary to accurately represent emerging industries in labor statistics, they complicate the creation of consistent long-term data. Revising classification systems makes it difficult to maintain continuity in data series because the criteria and categories for classification change. Historical comparisons become less reliable unless the mappings between old and new systems are precise and continuous. This is particularly problematic for longitudinal studies that track trends over time, as inaccurate mappings can lead to data misinterpretation, obscuring true economic trends and potentially misleading policy decisions. These discrepancies affect researchers, policymakers, and businesses, all of whom rely on reliable labor market data to make informed decisions. Therefore, ensuring that classification updates are integrated smoothly and that historical data is accurately mapped to new categories will provide better labor market analyses over time.

The data we use for this study are from the Indian Employment Unemployment Survey (EUS) for the years 2009 and 2011, which correspond to ISIC revisions 3.1 and 4. These data sets provide a robust foundation for validating our methodology due to their coverage of employment information and the opportunity to examine the effects of classification system changes on labor market statistics. However, using these datasets involves several ethical considerations and potential biases, including selection bias in survey participation, biases in how questions are framed, and inaccuracies in mapping old job codes to new ones. Additionally, the "black box" nature of AI models can obscure the reasoning behind certain outcomes, potentially reinforcing existing biases if not managed carefully. To address these challenges, our solution leverages Google's API to develop a replicable and repeatable methodology for mapping old and new job classifications, offering more nuanced predictions of historical job codes. This refined mapping enhances the accuracy of historical labor market analysis and provides deeper insights into job evolution. By applying this methodology to the Indian data, we test its effectiveness and reliability, ensuring its applicability in various economic contexts and future classification updates. Sharing our findings through detailed reports and presentations to policymakers will ensure they are accurately understood and used, maximizing the impact of our research on policy decisions.

Our findings contribute to three main areas of literature: the evolution and impact of industry classification systems on labor statistics and economic analysis, the effect of industry innovation on labor market adaptability, and the relevance of traditional industry classifications in analyzing productivity.

Firstly, we engage with studies on the transition from the SIC to the NAICS system and its implications for maintaining consistent and comparable economic data over time by Yuskavage [8] and Bayard and Klimek [1] as well as the recent NAICS updates in response to digital commerce from Haley and Keller [5]. The evolution of industry classification systems, such as the transition from the Standard Industrial Classification (SIC) to the North American Industry Classification System (NAICS), reflects the significant changes in the economy over time. These changes have shifted from a predominantly industrial focus to one that is increasingly driven by technology and services. Yuskavage [8] undertook the challenging task of converting historical data from the SIC to NAICS, ensuring that the data remains consistent and comparable across time periods. This effort allows for a more accurate understanding of the past and how it influences the present economic landscape. Similarly, Bayard and Klimek [1] addressed the complexities of maintaining data integrity while industry standards evolve. Their work ensures that economic analyses remain relevant and reliable, even as the definitions and categories of industries change over time. More recently, Haley and Keller [5] examined the impact of the NAICS 2022 update on the retail sector, highlighting how classification systems must adapt to the growing influence of e-commerce and digital transformation. Taken together, these studies demonstrate the critical role that data adaptation plays in understanding economic shifts. Industry classifications constantly change, reflecting the evolving economy and its impact on jobs. The combined work of these researchers emphasizes their continuous effort to track and analyze the economy's growth and transformation.

Our research also contributes to the ongoing discussion about how industry innovation, as measured by Research and Development investment, affects the adaptability of the labor market. This includes examining job market fluidity Moscarini and Thomsson [7] and the unexpected impact of minimum wage policies on employment across different industry sectors Card and Krueger [2]. The OECD's classification system, which categorizes industries based on their R&D investment, is crucial for understanding the role of different economic sectors in driving innovation and their overall impact Galindo-Rueda and Verger [3]. This system identifies sectors that are at the forefront of technological advancement and provides a foundation for analyzing how the labor market adapts to industry changes driven by R&D investment. Moscarini and Thomsson [7] highlights the dynamic nature of the U.S. job market, demonstrating how changes in industry innovation directly influence job mobility and employment patterns. This research underscores the importance of understanding the relationship between industry-level innovation and labor market outcomes. Furthermore, Card and Krueger[2] present evidence that contradicts the traditional view by demonstrating that higher minimum wages do not necessarily lead to job losses, even in industries like fast food, which typically invest less in research and development. This finding suggests that the economic responses to wage policies are complex and closely tied to the structure of the industry in question. Collectively, these studies by Galindo-Rueda and Verger[3] Moscarini and Thomsson[7], and Card and Krueger[2] challenge traditional economic models and offer a more detailed view and understanding of the complex connections between industry classification, worker mobility, and minimum wage policies in the field of labor economics.

Finally, our study speaks to the work that critiques the usefulness of traditional industry classifications in analyzing productivity, especially regarding how these classifications correspond with production functions and market categories as explored by Jack E. Triplett and Gollop[6]. Additionally, we examine the complexities involved in measuring overall productivity trends and the possible biases that may arise as discussed by Gullickson and Harper[4]. Jack E. Triplett and Gollop[6] critically evaluated the U.S. Standard Industrial Classification (SIC) system and found that only 20 percent of the industries were classified in a way that was consistent with production theory. This finding raises concerns about the applicability of traditional industrial classifications for productivity research. Their empirical analysis employed a diversification index to demonstrate the variability of production functions within industries, suggesting that many industry classifications might not effectively capture the true economic dynamics of production processes or market categories. In a related study, Gullickson and Harper[4] revisited potential biases in aggregate productivity trends, particularly addressing earlier issues with measuring service outputs and incorporating updated data. Although they observed negative multifactor productivity trends in some industries, the broader implications for bias were inconclusive. Nevertheless, they highlighted that the measurement challenges identified in prior research still persist. Together, these studies Jack E. Triplett and Gollop[6] Gullickson and Harper[4] emphasize the ongoing challenge of accurately capturing the true scale and scope of productivity across diverse industries. They also highlight the need for continuous refinement of industry classification systems to better align with evolving economic structures and output measurement methodologies.

2 Data

The data used in this study follow two parts. In Part I, we collect tables of ISIC codes and their corresponding descriptions across multiple years. To prepare the data, we web scraped the ISIC 3.1 and ISIC 4 revisions into a PDF. We then performed a quality check on the ISIC codes and their descriptions. In Part II, we apply our algorithm to re-categorize ISIC job codes in the past using "current" categorizations. We then test how this re-categorization affects the summary statistics of individuals in each job category. This helps us understand if and when re-coding job categories creates problems of comparability within an ISIC sector across years.

In Part I, we summarize the data collection for the ISIC 3.1 and ISIC 4 classifications, an extensive process undertaken by the United Nations Statistics Division, resulting in detailed datasets that reflect the structure and nature of the global economy. In a survey, people are asked to describe the main output of the company they work in (or what they do in their self-employment). This information is then recorded and converted into ISIC codes. The ISIC 3.1 data was collected from a broad range of economic activities and organized in a clear, systematic format. It includes explanatory notes and correspondence tables, which offer detailed guidelines for classifying economic activities. Similarly, ISIC 4 continues this detailed categorization, allowing for an accurate reflection of changes in the economy.

In Part II, we use labor market data from India to apply our job code mapping algorithm. India's labor market is noted for its diverse and rapidly evolving industries, making it an ideal candidate for our study. We use data from the Employment Unemployment Survey (EUS) conducted in India in 2009 and 2011, a period of significant economic transformation marked by changes in occupational structures and employment patterns. The EUS is a comprehensive national labor force survey conducted by the National Sample Survey Office (NSSO) under the Ministry of Statistics and Programme Implementation (MoSPI). The survey employs a stratified two-stage sampling procedure, with census villages and urban blocks serving as primary sampling units (PSUs) and households as secondary sampling units (SSUs). This methodology ensures representative coverage of the workforce. We examine the transition from the International Standard Industrial Classification (ISIC) 3.1 to ISIC 4 and its implications for employment outcomes and demographics by job type in these data.

There are several potential biases and ethical considerations in our data. For the development of the recoding from ISIC 3.1 to ISIC 4, we use the Google Gemini API to analyze job code transitions. This approach raises ethical concerns, particularly about AI transparency and possible biases in our data sources, such as the ISIC and EUS datasets. "Flawed data" refers to potential inaccuracies in job descriptions or misclassifications within these datasets. If the AI makes decisions based on such flawed data, it could misrepresent job market trends. Such errors could lead to uneven funding or investments, misguided educational and training programs, and poor job creation strategies, severely affecting those reliant on industries shown as declining or inaccurately booming, like workers in unstable job markets or recent graduates entering supposed growth fields. Furthermore, these inaccuracies can skew labor market analyses and result in ineffective policy measures. The "black box" nature of AI, which makes it hard to see where biases start, whether from societal inequalities in the data or the AI's processing methods, adds to the complexity. Those most affected are often those with the least ability to adapt, such as low-income workers, small-scale entrepreneurs, and communities dependent on single industries, potentially trapping them in economic hardship due to these misrepresentations.
2011 data

3 Problem and Estimation Procedure

Matching job codes across years poses a challenge because job codes are typically increasingly disaggregated over time. For example, ISIC 4 codes have several potential matches with ISIC 3.1. Code 0119 in ISIC 4, for example, could be either 0111 or 0112 in ISIC 3.1. Both of the options involve farming, but 0111 deals with sugar beets and related food items, while 0112 involves vegetable seeds and mushrooms. Currently, mappings between ISIC editions are done via brute force "1/n" probabilities. The table simply shows an ISIC 4 code and a list of n possible 3.1 codes could map to. This simplistic method does not provide an effective continuity between classification systems, so policymakers have difficulty analyzing trends and job information over time. Accurately differentiating between codes like the above farming example is crucial to analyzing job market statistics, which is why an inefficient mapping system can lead to incorrect summary statistics and,

Table 1: summary stats for 2009 in ISIC 3.1

	Category	Freq	urban	males	literacy	avg_age	avg_hsize	scl
1	Activities of private households	1637	0.73	0.39	0.60	37.09	4.61	
2	Agriculture, hunting and forestry	67134	0.09	0.61	0.66	38.94	5.81	
3	Construction	18827	0.33	0.85	0.69	35.29	5.19	
4	Education	8109	0.45	0.63	0.98	38.08	4.90	
5	Electricity, gas and water supply	857	0.53	0.94	0.95	41.00	4.67	
6	Extraterritorial organizations and bodies	1	1.00	1.00	1.00	36.00	3.00	
7	Financial intermediation	1815	0.71	0.86	0.99	39.08	4.61	
8	Fishing	627	0.46	0.94	0.80	37.56	6.04	
9	Health and social work	2127	0.57	0.62	0.97	39.06	4.59	
10	Hotels and restaurants	3115	0.61	0.81	0.82	36.50	4.99	
11	Manufacturing	20452	0.54	0.73	0.80	35.29	5.23	
12	Mining and quarrying	1221	0.41	0.89	0.76	38.17	5.10	
13	Other community, social and personal service activities	4336	0.52	0.74	0.74	38.52	5.08	
14	Public administration and defence	7701	0.60	0.87	0.96	41.91	4.84	
15	Real estate, renting and business activities	2420	0.77	0.90	0.97	35.93	4.81	
16	Transport, storage and communications	9292	0.51	0.98	0.85	36.01	5.20	
17	Unemployed	286450	0.41	0.38	0.71	22.44	5.87	
18	Wholesale and retail trade	23663	0.58	0.88	0.88	37.84	5.45	

therefore, policy decisions based on wrong statistics. Using Google Gemini's API, we ameliorate the problem of blunt probability mappings. We have created a tool that searches through ISIC 3.1 and has the AI evaluate the actual probability of the ISIC 4 code mapping to it.

Our tool has two main variables, an ISIC 4 code and an ISIC 3.1 code. Our goal is to map from ISIC 4 to its corresponding 3.1 code(s). The ISIC 4 code is what we initially input into our model, and it will return a list of potential 3.1 codes, along with proportions for each 3.1 code. The algorithm is visualized below:

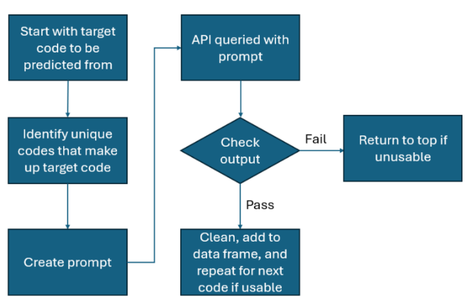

Figure 1: ISIC Algorithm

After finding those potential codes, the algorithm plugs them into Gemini, which completes the process by ensuring it makes sense, then adding it to a correspondence table, or restarting the

Table 2: Summary Stats for 2011 in ISIC 4

	Category
1	Accommodation and food service activities
2	Activities of extraterritorial organizations and bodies
3	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
4	Administrative and support service activities
5	Agriculture, forestry, and fishing
6	Arts, entertainment and recreation
7	Construction
8	Education
9	Electricity, gas, steam and air conditioning supply
10	Financial and insurance activities
11	Human health and social work activities
12	Information and communication
13	Manufacturing
14	Mining and Quarrying
15	Other service activities
16	Professional, scientific, and technical activities
17	Public administration and defence; compulsory social security
18	Real estate activities
19	Transportation and storage
20	Unemployed
21	Water supply; sewerage, waste management and remediation activities
22	Wholesale and retail trade; repair of motor vehicles and motorcycles

process if it returns an output that does not make sense. The final output is a complete correspondence table, featuring ISIC 4 codes, its potential ISIC 3.1 codes, and the proportion of those 3.1 codes. The table is accurate and detailed.

The Google Gemini API is a powerful tool that is essentially a collection of tools and processes that can easily be applied using scripted programming. The Google Gemini API utilizes AI models and allows users to integrate AI into their pipeline. Our use of Gemini involves prompting it to analyze the different versions of ISIC and predict the proportion of jobs now coded in ISIC 4 that were previously coded in ISIC 3.1. This method uses real-world data, making it significantly more accurate than blunt approaches. Once Gemini generates the results, the data is cleaned and the code outputs a small data frame that is added to a larger correspondence table and then moves on to the next code batch in the pipeline. Although this process can be time-consuming, it ensures higher accuracy, precision, and clarity. For real world analysis, this tool can be applied to translate years from one ISIC classification system to another, allowing for continuity over time. While the codes are predicted at the granular, 4 digit level, they can be aggregated into 3, 2 or even single digit level after, depending on the desires of the researcher. To show this, we applied the tool's predictions to data from the Indian Employment Unemployment Survey (EUS) for the years 2009 and 2011 to evaluate its effectiveness and reliability. The methodology involves feeding the ISIC 4 codes into Gemini, which then maps each ISIC 4 code to a probability of each ISIC 3.1 code it was previously made up of. We can then reclassify the 2011 data from an ISIC 4 code system back to ISIC 3.1 and look for demographic or other trends within industries over time. This method gives researchers the ability to study changes in the labor market over time with a more precise knowledge of the makeup of that market.

While we do not believe there is any literature directly related to the mapping between ISIC editions, one paper touches on the difficulty of changing classification systems in the context of the North American Industry Classification System (NAICS). NAICS is a U.S. specific job code classification system, separate from international systems. The differences can present difficulties in the global job market and time series analysis. In Bayard and Klimek [1] the authors discuss the Standard Industry Classification (SIC) compared to NAICS. In their paper, the authors develop an algorithm that maps from SIC to NAICS classifications, using data from 1992 and 1997. Its approach to non-unique codes, meaning ones that could map to a few different options, is to randomly assign a code based on the proportion of that code's appearance.

4 Results

4.1 Main Findings

We begin by evaluating our tool for back prediction. We choose Google's Gemini model[1] as it is open-source for up to 60 queries per minute and performs well in accuracy compared to competitors like ChatGPT. We build an algorithm to generate a new correspondence table that back-predicts new codes. We find that compared to older methods, our approach generates more nuanced predictions of how job classifications change over time. (Place holder to discuss prediction of Indian Data). One example of this can be seen in ISIC 4 code 2670 which is "Manufacture of optical instruments and photographic equipment". This code corresponds to 3 in version 3.1 which are 3230 "Manufacture of television and radio receivers, sound or video recording or reproducing apparatus, and associated goods", 3320 "Manufacture of optical instruments and photographic equipment" and 3312 "Manufacture of instruments and appliances for measuring, checking, testing, navigating and other purposes, except industrial process control equipment". Previously, because these codes are all different until the 2 digit level, we would have just assigned a weight of 0.33 to each. However, under our new prediction method, 3320, which should have the most overlap with 2670, is assigned a weight of 0.5 while the other 2 are given 0.25 each. This is evidence of the tool showing that 3320 should be the highest weighted, but previously we would have had weighted back predicted categories with equal probability. This shows that this method of classification is a step up from previous methods and can be used to create correspondence tables in the future.

The main issue we address is predicting proportions without data on job creation and adjustment over time. There is no validation dataset that shows, for example, jobs 1 through 1000 in country A, how they were coded under ISICS 3.1 and then how those same jobs are coded in ISICS 4. This makes prediction challenging as there is no labeled training set to train a back prediction model on. To address this we use Google's widely available deep learning algorithm Gemini.

We first acquire the codes and descriptions by scraping the ISIC revision documentation. This provides us with each 1, 2, 3, and 4 digit code along with the description associated with each. We

[1]https://ai.google.dev/gemini-api/docs/api-overview

Table 3: Variance Summary Statistics

Min	0
Q1	0
Median	.01
Q3	.04
Max	.13

then clean this data in Python, creating Excel files that have only the 4-digit codes and descriptions for both versions of 3.1 and 4. This allows us to prepare standardized prompts for Gemini, which returns a table in textual form. The text file is then read into Python as a .csv file and converted into a dataframe. The problem is that asking Gemini to predict takes time with each additional query adding several seconds. To optimize the process, the algorithm checks each predicted code to see if it corresponds to a single or multiple codes in a different version. If only one, it is immediately assigned a correspondence value of 1, significantly reducing runtime.

Mechanisms to explain findings One question that remains with this method is reproducibility. Since we are feeding prompts to an AI model we will not always achieve exactly the same result. To discover the true magnitude of the variance we run the model 100 times and measure the variance of each code correspondence pair. Below are summary statistics for the variance as well as a histogram of their distribution.

This table shows that the variances across runs are quite small, even close to zero with some higher outliers causing slight concern. However, overall we can be confident that while each running of the model will produce slightly different results, they will be nearly with very few large moves. This gives us confidence that using such models can provide a stable estimate of the correspondence table we desire.

We also we need to test the model's accuracy. We do so below by looking a two years of the Indian Employment Unemployment Survey (EUS) discussed in the data section. These give us a look at two very similar years where the primary difference should only be in how the jobs were classified, granting us insight to how our tool is preforming. Below is a discussion of the results when applying the tool to a real world scenario.

4.2 Application to India Data

We apply our solution to the 2011 India dataset described in the data section. We take the prediction table created by our tool and apply it to the 2011 data, backcoding it into ISIC 3.1. We then sort the newly changed codes into the ISIC 3.1 categories to create a new table of summary statistics. To test how well our tool makes the 2011 data resemble the 2009, we subtract the backcoded summary statistics from the 2009 summary statistics, then square the differences in the case of proportional calculations, or take the absolute value in the case of whole numbers. The closer these values are to zero, the smaller the difference compared to 2009, indicating the accuracy of the backcoding. We test this against the blunt case that simply picks a random option from possible ISIC 3.1 values for each ISIC 4 code. Our goal is to see how well our tool predicts various outcomes and

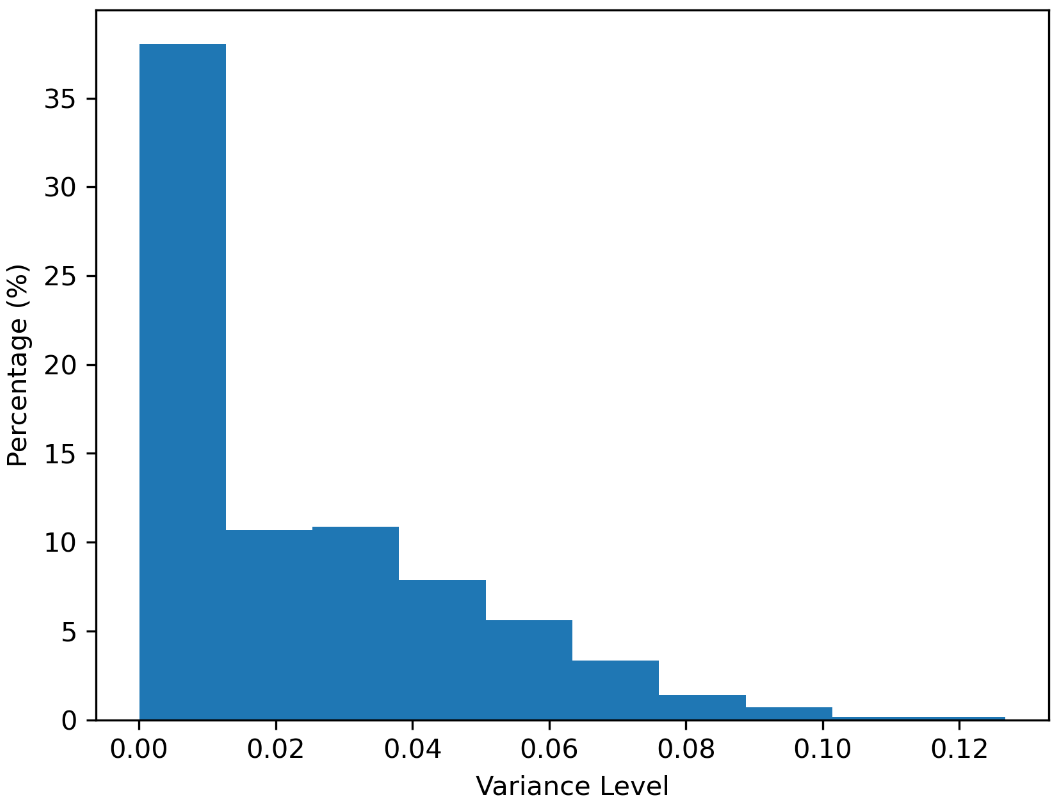

Figure 2: Variance Histogram

demographic information in different job categories. For example, by looking at the proportion of literate workers in construction, we can get a sense of what India was like in 2009. Then, after backcoding, we examine what the tool predicts the literacy rate to be in 2011 if the same job categories from 2009 were still in use. Since the country's demographic and economic conditions changed between 2009 and 2011, we know we cannot make the differences actually equal to zero, no matter how well we predict the codes. Instead, our aim is to produce a vision of India in 2011 as if it was still using ISIC 3.1, This should lead to most differences being close to, but not actually zero as, although there are differences in the Indian economy from 2009 to 2011, we believe in general they are not vast. This approach allows us to understand how the tool performs in predicting and maintaining continuity in labor market statistics. The summary statistic tables are below:

Table 4: Comparison of differences between ISIC 3.1 and 2011 ISIC 4 data backcoded via the blunt method

	Category	Freq	urban	males	literacy	avg_age	h_size	school	marital	labor	vocation	wage	hours
1	Activities of private households	76	0.00	0.00	0.00	1.21	0.15	0.00	0.00	0.17	0.92	285.38	0.16
2	Agriculture, hunting and forestry	5553	0.00	0.00	0.00	0.97	0.10	0.00	0.00	0.01	0.90	288.08	0.39
3	Construction	1407	0.00	0.00	0.00	0.62	0.12	0.00	0.00	0.03	0.83	350.60	0.29
4	Education	1041	0.00	0.00	0.00	0.82	0.06	0.00	0.00	0.03	0.76	982.45	0.23
5	Electricity, gas and water supply	140	0.00	0.00	0.00	0.96	0.04	0.00	0.00	0.01	0.58	1025.04	0.11
6	Extraterritorial organizations and bodies	0	0.00	1.00	0.00	12.00	1.00		1.00	1.00	1.00	1200.00	0.00
7	Financial intermediation	122	0.00	0.00	0.00	0.42	0.02	0.00	0.00	0.04	0.77	1035.87	0.38
8	Fishing	123	0.00	0.00	0.00	0.44	0.40	0.00	0.00	0.03	0.61	333.44	0.37
9	Health and social work	313	0.00	0.00	0.00	0.16	0.11	0.00	0.00	0.04	0.52	814.88	0.16
10	Hotels and restaurants	428	0.00	0.00	0.00	1.39	0.17	0.00	0.00	0.06	0.85	560.48	0.62
11	Manufacturing	1291	0.00	0.00	0.00	0.74	0.09	0.00	0.00	0.06	0.52	388.80	0.36
12	Mining and quarrying	252	0.01	0.00	0.00	0.16	0.30	0.00	0.00	0.03	0.77	1528.60	0.16
13	Other community, social and personal service activities	1599	0.00	0.00	0.00	0.75	0.15	0.00	0.00	0.05	0.64	374.70	0.29
14	Public administration and defence	1077	0.00	0.00	0.00	0.14	0.02	0.00	0.00	0.02	0.84	1121.21	0.04
15	Real estate, renting and business activities	555	0.00	0.00	0.00	1.36	0.04	0.00	0.00	0.07	0.61	1024.94	0.17
16	Transport, storage and communications	955	0.00	0.00	0.00	0.88	0.01	0.00	0.00	0.03	0.63	648.58	0.11
17	Unemployed	1732	0.00	0.00	0.00	0.42	0.10	0.00	0.00	0.09	0.91	286.90	1.48
18	Wholesale and retail trade	87	0.00	0.00	0.00	0.64	0.12	0.00	0.00	0.05	0.84	395.98	0.30

Table 5: Comparison of differences between ISIC 3.1 and 2011 ISIC 4 data backcoded via the AI tool method

	Category	Freq	urban	males	literacy	avg_age	h_size	school	marital	labor	vocation	wage	hours
1	Activities of private households	76	0.00	0.00	0.00	0.97	0.14	0.00	0.00	0.17	0.92	292.61	0.38
2	Agriculture, hunting and forestry	5561	0.00	0.00	0.00	1.21	0.08	0.00	0.00	0.01	0.90	287.08	0.22
3	Construction	1400	0.00	0.00	0.00	0.61	0.09	0.00	0.00	0.03	0.83	328.46	0.07
4	Education	1061	0.00	0.00	0.00	0.69	0.07	0.00	0.00	0.04	0.76	963.19	0.02
5	Electricity, gas and water supply	140	0.00	0.00	0.00	0.94	0.04	0.00	0.00	0.01	0.59	1014.15	0.13
6	Extraterritorial organizations and bodies	0	0.73	0.33	0.02	0.71	3.00		0.08	0.08	1.00	1601.43	14.29
7	Financial intermediation	128	0.00	0.00	0.00	0.25	0.02	0.00	0.00	0.04	0.77	1026.51	0.26
8	Fishing	128	0.00	0.00	0.00	0.14	0.30	0.00	0.00	0.03	0.60	361.52	0.36
9	Health and social work	313	0.00	0.00	0.00	0.05	0.16	0.00	0.00	0.03	0.52	897.10	0.11
10	Hotels and restaurants	428	0.00	0.00	0.00	1.39	0.18	0.00	0.00	0.06	0.86	585.40	0.42
11	Manufacturing	1283	0.00	0.00	0.00	0.77	0.09	0.00	0.00	0.06	0.52	425.20	0.25
12	Mining and quarrying	249	0.00	0.00	0.00	0.19	0.30	0.00	0.00	0.03	0.77	1321.60	0.06
13	Other community, social and personal service activities	1563	0.00	0.00	0.00	0.04	0.14	0.00	0.00	0.05	0.65	318.44	0.51
14	Public administration and defence	1074	0.00	0.00	0.00	0.19	0.02	0.00	0.00	0.02	0.84	1122.98	0.07
15	Real estate, renting and business activities	563	0.00	0.00	0.00	1.53	0.04	0.00	0.00	0.06	0.62	891.32	0.22
16	Transport, storage and communications	931	0.00	0.00	0.00	0.74	0.01	0.00	0.00	0.03	0.63	591.21	0.21
17	Unemployed	1732	0.00	0.00	0.00	0.87	0.12	0.00	0.00	0.10	0.91	1624.00	18.83
18	Wholesale and retail trade	87	0.00	0.00	0.00	0.69	0.11	0.00	0.00	0.05	0.84	388.08	0.21

An important note here is that India changed from 2009 to 2011. We discussed above that we wan the differences to be as close to 0 as possible, but that is not entirely possible. As seen in the data section, some categories are going to change drastically across time. For example, "vocation" increased drastically across these years. In 2009, not a lot of people had vocational training, while in 2011 many had. Similarly, wages increased by a significant margin between the two years. The mean wage in 2009 was 1507; in 2011 it was 2129. Generally the differences were close to 0, but it is still valuable to compare the methods of backcoding to eachother with their differences. Below is a table that shows, in each category, which method was smaller.

Table 6: Which tool is better in each category

	Category	Freq	urban	males	literacy	avg_age	h_size	school	marital	labor	vocation	wage	hours
1	Activities of private households	tie	blunt	AI Tool	blunt	AI Tool	AI Tool	blunt	AI Tool	AI Tool	blunt	blunt	blunt
2	Agriculture, hunting and forestry	blunt	AI Tool	blunt	blunt	blunt	AI Tool	blunt	blunt	AI Tool	blunt	AI Tool	AI Tool
3	Construction	AI Tool	AI Tool	blunt	blunt	AI Tool	AI Tool	blunt	AI Tool	blunt	AI Tool	AI Tool	AI Tool
4	Education	blunt	AI Tool	blunt	blunt	AI Tool	blunt	AI Tool	AI Tool	blunt	AI Tool	AI Tool	AI Tool
5	Electricity, gas and water supply	tie	blunt	blunt	blunt	AI Tool	blunt	blunt	AI Tool	blunt	blunt	AI Tool	blunt
6	Extraterritorial organizations and bodies	tie	blunt	AI Tool	blunt	AI Tool	blunt		AI Tool	AI Tool	tie	blunt	blunt
7	Financial intermediation	blunt	AI Tool	AI Tool	blunt	AI Tool	AI Tool	blunt	AI Tool	AI Tool	AI Tool	AI Tool	AI Tool
8	Fishing	blunt	blunt	AI Tool	blunt	AI Tool	AI Tool	blunt	AI Tool	blunt	AI Tool	blunt	AI Tool
9	Health and social work	tie	blunt	blunt	AI Tool	AI Tool	blunt	blunt	blunt	AI Tool	blunt	blunt	AI Tool
10	Hotels and restaurants	tie	blunt	blunt	AI Tool	AI Tool	blunt	blunt	AI Tool	blunt	blunt	blunt	AI Tool
11	Manufacturing	AI Tool	AI Tool	AI Tool	blunt	blunt	AI Tool	AI Tool	AI Tool	blunt	AI Tool	blunt	AI Tool
12	Mining and quarrying	AI Tool	AI Tool	AI Tool	blunt	blunt	AI Tool	AI Tool	blunt	AI Tool	blunt	AI Tool	AI Tool
13	Other community, social and personal service activities	AI Tool	blunt	AI Tool	AI Tool	AI Tool	AI Tool	AI Tool	AI Tool	blunt	blunt	AI Tool	blunt
14	Public administration and defence	AI Tool	AI Tool	blunt	blunt	blunt	AI Tool	blunt	blunt	AI Tool	AI Tool	blunt	blunt
15	Real estate, renting and business activities	blunt	blunt	blunt	AI Tool	blunt	AI Tool	AI Tool	blunt	AI Tool	blunt	AI Tool	blunt
16	Transport, storage and communications	AI Tool	blunt	blunt	blunt	AI Tool	AI Tool	blunt	blunt	AI Tool	AI Tool	AI Tool	blunt
17	Unemployed	tie	blunt	blunt	AI Tool	blunt	blunt	AI Tool	blunt	blunt	blunt	blunt	blunt
18	Wholesale and retail trade	tie	AI Tool	AI Tool	blunt	blunt	AI Tool	blunt	blunt	AI Tool	blunt	AI Tool	AI Tool

This table says that the changes between 2009 data and back-coded 2011 data are smaller in the blunt case in 82 categories, while it is smaller in the AI case in 81 categories. However, as mentioned above, the AI tool represents India's accurate of information in 2011. The tool shows changes in wages and vocation as higher compared to the blunt version. Therefore, the AI tool is stronger compared to the blunt method and can provide a more precise display of countries in a different ISIC system.

Checking the accuracy of the tool In Table 4, we used Indian labor outcomes categorized under ISIC 4.0 and 3.1 to evaluate the accuracy of our prediction tool when applied to real-world data. The graph of the probability differences visualizes the distribution of probability differences across various intervals of 0.10. The results show that many probability differences cluster around 0, suggesting that the tool's predictions are closely aligned with actual labor outcomes. As we move further from 0, fewer occupations exhibit high variance in probability differences, indicating that the tool effectively predicts labor outcomes in India, with larger deviations being less frequent.

Concerns and Limitations of the Mapping Methodology While our new method using Google Geminiâs API significantly improves the accuracy of mapping job codes between different ISIC versions, it is not without challenges and limitations. The success of our approach heavily relies on the accuracy and relevance of the data given to the AI model. If the underlying data or its initial categorization are flawed, it could lead to incorrect mappings. Additionally, while the AI enhances the mapping process, computational errors or misinterpretations by the AI can occur, especially in complex cases. This method also does not entirely eliminate biases that may have existed in previous job categorizations. For example, our methodology may unintentionally reinforce past job classification biases toward particular industries or ethnicities.

Applicability of the Findings The validity of our findings is promising but should be applied cautiously. The specific AI-driven approach we developed is tailored to the ISIC coding system, which may not directly translate to other classification systems without modifications. Therefore, while our results are robust within the context of ISIC codes, applying this method to entirely different systems, like NAICS or company-specific frameworks, would require additional calibration and testing. This approach is best suited for economies or sectors that closely follow or are compatible with the ISIC classification principles. Extending these findings to any person, firm, or country without considering local economic structures and industry classification standards might lead to less accurate or relevant outcomes.

5 Conclusion

The problem we studied is how frequently changing job classifications make it difficult to maintain accurate labor market statistics. Policymakers need precise job categories to create effective policies across job sectors. Traditional methods of job categorization often do not capture the nuances of job changes over time. To address this, we developed a tool using Google's AI to match old and new job categories more accurately. We applied our tool to the Indian Employment Unemployment Survey (EUS) for 2009 and 2011, which covers significant changes in job classifications, and compared demographic statistics by job code using our back-predicted job categories to the actual categories.

Our method can produce detailed correspondences between old and new job classifications. For example, job code 113 in the new system corresponds to about three-fifths of jobs previously classified under code 111 and two-fifths under code 112, while job code 112 in the new system matches entirely to job code 111 from the old system. When back-predicting job codes using the EUS, we find that our tool provides more accurate and consistent job classification mappings compared to traditional methods, enhancing the reliability of labor market data.

These findings provide a clearer picture of how job roles evolve, aiding in a better understanding and tracking of employment trends. By offering a more nuanced view of job classifications, policymakers can design better-targeted policies and interventions that align with the actual job market dynamics. This improved understanding helps address sector-specific challenges and supports more effective resource allocation.

There are several caveats to our findings. AI models can be hard to understand and trust because they often work like a "black box," making it difficult to see how predictions are made. This opacity can be a problem for policymakers who need to understand the basis of the data they use. Additionally, our dataset is specific to India, so our findings might not apply universally. There are also technical limitations and potential biases in the AI models we use, such as selection bias in survey participation and inaccuracies in job code mappings. Addressing these issues is important to make our approach more applicable and reliable. Furthermore, job roles and technologies change quickly, necessitating continuous updates to classification systems. These challenges underscore the need for ongoing improvements and testing of our methods.

Given our work, future research should test this AI method with different datasets from other countries to see how well it works elsewhere. Future studies should also aim to make AI predictions more transparent and understandable to build trust in the results. It is crucial to continually update job classification systems to keep pace with economic and technological changes. Additionally, future research should explore using real-time data to make job classifications more responsive to current economic trends. By doing so, we can ensure that the occupation code system remains robust and relevant, supporting better policy decisions and economic planning worldwide.

6 Attribution

Matthew worked on creating the correspondence tables using the India data to check the accuracy of the tool at the 4-digit level with their probability differences. He also worked on creating the summary stats for the binary variables, along with some averages. Also, worked on portions of writing the paper and the layout of the presentation. Gavin worked on summary statistics for India in 2009 and 2011. He used the prediction table made by the AI to backcode from ISIC 4 to ISIC 3.1 and made summary maps of India. Ben worked on the writing of the paper, along with necessary edits. He did the results tables and slides. He collaborated with Matthew on the code of the probability differences and summary statistics. Caleb worked on creating the code for the tool and method of the algorithm to classify codes. He also helped in writing and editing parts of the introduction, results and methods sections as well as drafts of the abstract.

7 Appendix

Table 7: Mapping of Job Codes Between Versions and Their Proportions

version_4	version_3.1	Proportion of Jobs
0111	0111	1
0112	0111	1
0113	0111	0.6
0113	0112	0.4
0114	0111	1
0115	0111	1
0116	0111	1
0119	0111	0.6
0119	0112	0.4
0121	0113	1
0122	0113	1
0123	0113	1
0124	0113	1
0125	0112	0.5
0125	0113	0.5
0126	0111	0.75
0126	0113	0.25
0127	0113	1
0128	0111	0.2
0128	0112	0.3
0128	0113	0.5
0129	0111	0.6
0129	0112	0.2
0129	0200	0.2
0130	0112	0.75
0130	0200	0.25
0141	0121	1
0142	0121	1
0143	0122	1
0144	0121	1
0145	0122	1
0146	0122	1
0149	0122	1
0150	0130	1
0161	0140	1
0162	0140	0.75
0162	2892	0.25
0163	0111	0.75
0163	0140	0.25
0164	0140	1

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
0170	0150	1
0210	0200	1
0220	0200	1
0230	0112	0.25
0230	0113	0.25
0230	0200	0.5
0240	0200	1
0311	0501	1
0312	0501	1
0321	0122	0.3
0321	0502	0.7
0322	0122	0.5
0322	0502	0.5
0510	1010	1
0520	1020	1
0610	1110	1
0620	1110	1
0710	1310	1
0721	1200	1
0729	1320	1
0810	1410	0.9
0810	1429	0.1
0891	1421	1
0892	1030	1
0893	1422	1
0899	1429	1
0910	1110	0.6
0910	1120	0.3
0910	7421	0.1
0990	1010	0.0833
0990	1020	0.0833
0990	1030	0.0833
0990	1200	0.0833
0990	1310	0.0833
0990	1320	0.0833
0990	1410	0.0833
0990	1421	0.0833
0990	1422	0.0833
0990	1429	0.0833
0990	4510	0.0833
0990	7421	0.0833
1010	1511	1
1020	1512	1

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
1030	1513	0.75
1030	1549	0.25
1040	1514	1
1050	1520	1
1061	1531	1
1062	1532	1
1071	1541	0.8
1071	1549	0.2
1072	1542	1
1073	1543	1
1074	1544	1
1075	1512	0.25
1075	1513	0.25
1075	1544	0.25
1075	1549	0.25
1079	1549	0.8
1079	2429	0.2
1080	1533	1
1101	1551	1
1102	0113	0.5
1102	1552	0.5
1103	1553	1
1104	1554	1
1200	1600	1
1311	1711	1
1312	1711	0.8
1312	2699	0.2
1313	1712	0.9
1313	1729	0.05
1313	1810	0.05
1391	1730	1
1392	1721	0.8
1392	3430	0.2
1393	1722	1
1394	1723	1
1399	1729	0.95
1399	3699	0.05
1410	1810	1
1420	1820	1
1511	1820	0.5
1511	1911	0.5
1512	1912	0.75
1512	1920	0.15

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
1512	3699	0.1
1520	1920	1
1610	2010	1
1621	2021	1
1622	2022	1
1623	2023	1
1629	1920	0.25
1629	2029	0.75
1629	3699	0
1701	2101	1
1702	2102	1
1709	1729	0.25
1709	2109	0.5
1709	2221	0.25
1709	3699	0
1811	2109	0.5
1811	2221	0.4
1811	2892	0.1
1812	2222	1
1820	2230	1
1910	2310	0.9
1910	2411	0.1
1920	1010	0.25
1920	1020	0.25
1920	2320	0.5
2011	1551	0.25
2011	2330	0.25
2011	2411	0.25
2011	2429	0.25
2012	2412	1
2013	2413	1
2021	2421	1
2022	2422	1
2023	2424	1
2029	2429	1
2029	3699	0
2030	2430	1
2211	2511	1
2219	1920	0.25
2219	2519	0.5
2219	3610	0.1
2219	3699	0.15
2220	1920	0.05

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
2220	2109	0.05
2220	2519	0.05
2220	2520	0.7
2220	3610	0.05
2220	3699	0.1
2310	2610	1
2391	2692	1
2392	2691	0.65
2392	2693	0.35
2393	2691	1
2394	2694	1
2395	2695	1
2396	2696	1
2399	2699	1
2410	2710	1
2420	2330	0.1
2420	2720	0.9
2431	2710	0.5
2431	2731	0.5
2432	2732	1
2511	2811	1
2512	2812	1
2513	2813	1
2520	2927	1
2591	2891	1
2592	2892	1
2593	2893	0.8
2593	2929	0.2
2599	2899	0.7
2599	3190	0.2
2599	3699	0.1
2610	2429	0.0588
2610	2520	0.0588
2610	3000	0.0588
2610	3110	0.0588
2610	3120	0.0588
2610	3130	0.0588
2610	3210	0.5882
2610	3230	0.0588
2620	3000	1
2630	3190	0.4
2630	3220	0.5
2630	3230	0.1

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
2640	3230	0.8
2640	3694	0.2
2651	3190	0.25
2651	3220	0.25
2651	3312	0.25
2651	3313	0.25
2652	3330	1
2660	3311	1
2670	3230	0.25
2670	3312	0.25
2670	3320	0.5
2680	2429	1
2710	3110	0.75
2710	3120	0.25
2720	3140	1
2731	3130	0.8
2731	3320	0.2
2732	3130	1
2733	2520	0.25
2733	3120	0.5
2733	3190	0.25
2740	3150	0.8
2740	3190	0.2
2750	2930	1
2790	2922	0.2
2790	2929	0.1
2790	3110	0.3
2790	3120	0.1
2790	3130	0.05
2790	3150	0.05
2790	3190	0.1
2790	3210	0.1
2811	2911	0.6
2811	2912	0.05
2811	3110	0.25
2811	3430	0.08
2811	3530	0.01
2811	3591	0.01
2812	2519	0.25
2812	2912	0.5
2812	2913	0.25
2813	2912	1
2814	2913	1

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
2815	2914	0.8
2815	2930	0.2
2816	2915	0.9
2816	3599	0.1
2817	2429	0.2
2817	2899	0.1
2817	3000	0.5
2817	3230	0.1
2817	3610	0.1
2818	2922	1
2819	2919	0.6
2819	2922	0.2
2819	2930	0.1
2819	3312	0.1
2821	2921	1
2822	2922	0.8
2822	3190	0.2
2823	2923	1
2824	2924	1
2825	2925	1
2826	2926	0.8
2826	2929	0.2
2910	3410	1
2920	3420	1
2930	3190	0.2
2930	3430	0.7
2930	3610	0.1
3011	3511	0.9
3011	3610	0.1
3012	3512	1
3020	3190	0.3
3020	3520	0.6
3020	3610	0.1
3030	2927	0.25
3030	3530	0.7
3030	3610	0.05
3040	2927	1
3091	3591	1
3092	3592	0.8
3092	3699	0.2
3100	3599	0
3100	3610	1
3211	3330	0.25

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
3211	3691	0.75
3212	3330	0.25
3212	3699	0.75
3220	3692	1
3230	1920	0.25
3230	3693	0.75
3240	3694	1
3250	1721	0.0455
3250	2423	0.0455
3250	2919	0.0909
3250	3311	0.6364
3250	3312	0.0909
3250	3320	0.0909
3290	1810	0.1
3290	1912	0.05
3290	2029	0.05
3290	2211	0.05
3290	2519	0.05
3290	2520	0.1
3290	2699	0.05
3290	2899	0.1
3290	3311	0.1
3290	3693	0.05
3290	3699	0.1
3311	2811	0.2273
3311	2812	0.2273
3311	2813	0.0909
3311	2892	0.0909
3311	2893	0.0909
3311	2899	0.0909
3311	2927	0.0455
3311	2929	0.0455
3311	3420	0.0909
3312	2911	0.0952
3312	2912	0.0476
3312	2913	0.0476
3312	2914	0.0476
3312	2915	0.0476
3312	2919	0.0952
3312	2921	0.0476
3312	2922	0.0476
3312	2923	0.0476
3312	2924	0.0476

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
3312	2925	0.0476
3312	2926	0.0476
3312	2929	0.0476
3312	3110	0.0476
3312	3230	0.0476
3312	3599	0.0476
3312	3694	0.0476
3312	3699	0.0476
3312	7250	0.0476
3313	3190	0.0789
3313	3220	0.0526
3313	3230	0.0789
3313	3311	0.0789
3313	3312	0.0789
3313	3313	0.5263
3313	3320	0.0526
3313	5260	0.0526
3314	2520	0.05
3314	2922	0.05
3314	2929	0.05
3314	3110	0.25
3314	3120	0.25
3314	3130	0.05
3314	3140	0.05
3314	3150	0.05
3314	3190	0.15
3314	3210	0.05
3315	3511	0.4
3315	3512	0.1
3315	3520	0.1
3315	3530	0.2
3315	3599	0.1
3315	6303	0.1
3319	1721	0.05
3319	1723	0.02
3319	2023	0.03
3319	2029	0.01
3319	2519	0.1
3319	2520	0.05
3319	2610	0.02
3319	2699	0.01
3319	3311	0.2
3319	3312	0.05

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
3319	3330	0.02
3319	3692	0.01
3319	3694	0.01
3320	2813	0.05
3320	2911	0.05
3320	2912	0.05
3320	2914	0.05
3320	2915	0.05
3320	2919	0.05
3320	2921	0.05
3320	2922	0.05
3320	2923	0.05
3320	2924	0.05
3320	2925	0.05
3320	2926	0.05
3320	2929	0.05
3320	3000	0.05
3320	3110	0.05
3320	3220	0.05
3320	3311	0.05
3320	3313	0.05
3320	4540	0.05
3510	4010	1
3520	4020	1
3530	1549	0.5
3530	4030	0.5
3600	4100	1
3700	9000	1
3811	9000	1
3812	2330	0.25
3812	9000	0.75
3821	2412	0.1
3821	9000	0.9
3822	2330	0.25
3822	9000	0.75
3830	3710	0.7
3830	3720	0.3
3900	4510	0.5
3900	9000	0.5
4100	4520	1
4210	4520	1
4220	4520	1
4311	4510	1

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
4312	4510	0.8
4312	4550	0.2
4321	4530	1
4322	4530	1
4329	4530	0.75
4329	4540	0.25
4330	4530	0.5
4330	4540	0.5
4390	4520	0.4
4390	4530	0.2
4390	4540	0.2
4390	4550	0.2
4510	5010	1
4520	5020	1
4530	5030	1
4540	5040	1
4610	5110	1
4620	5121	1
4630	5122	1
4641	5131	1
4649	5139	1
4651	5151	1
4652	5152	1
4653	5159	1
4659	5159	1
4661	5141	1
4662	5142	1
4663	5139	0.25
4663	5143	0.75
4669	5149	1
4690	5190	1
4711	5211	1
4719	5219	1
4721	5220	1
4722	5220	1
4723	5220	1
4730	5050	1
4741	5239	1
4742	5233	1
4751	5232	1
4752	5234	1
4753	5233	0.6
4753	5239	0.4

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
4759	5233	1
4761	5239	1
4762	5233	1
4763	5239	1
4764	5239	1
4771	5232	1
4772	5231	1
4773	5239	1
4774	5240	1
4781	5252	1
4782	5252	1
4789	5252	1
4791	5251	0.75
4791	5259	0.25
4799	5259	1
4911	6010	1
4912	6010	1
4921	6021	1
4922	6021	0.75
4922	6022	0.2
4922	9241	0.05
4923	6023	1
4930	6030	1
5011	6110	1
5012	6110	1
5021	6120	1
5022	6120	1
5110	6210	0.75
5110	6220	0.25
5210	6302	1
5221	5020	0.6
5221	6010	0.2
5221	6303	0.2
5222	6303	1
5223	6303	1
5224	6301	1
5229	6309	1
5310	6411	1
5320	6412	1
5510	5510	1
5520	5510	1
5590	5510	1
5610	5520	1

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
5621	5520	1
5629	5520	1
5630	5520	1
5811	2211	0.75
5811	7240	0.25
5812	2211	0.95
5812	7240	0.05
5813	2212	0.8
5813	2219	0.15
5813	7240	0.05
5819	2219	0.6
5819	7240	0.4
5820	7221	0.8
5820	7240	0.2
5911	9211	0.7
5911	9213	0.3
5912	9211	0.75
5912	9231	0.25
5913	9211	1
5914	9212	1
5920	2213	0.25
5920	7240	0.25
5920	9211	0.25
5920	9213	0.25
6010	7240	0.6
6010	9213	0.4
6020	7240	0.55
6020	9213	0.45
6110	6420	1
6120	6420	1
6130	6420	1
6190	6420	1
6201	7229	1
6202	7210	0.2
6202	7229	0.7
6202	7230	0.1
6209	7290	1
6311	7230	1
6312	7240	1
6391	9220	1
6399	7499	1
6411	6511	1
6419	6519	0.8

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
6419	6719	0.2
6420	6599	1
6430	6599	1
6491	6591	1
6492	6592	1
6499	6592	0.8
6499	6599	0.2
6511	6601	0.85
6511	6603	0.15
6512	6603	1
6520	6601	0.8
6520	6603	0.2
6530	6602	1
6611	6711	1
6612	6712	0.75
6612	6719	0.25
6619	6599	0.8
6619	6719	0.2
6621	6720	1
6622	6720	1
6629	6720	1
6630	6602	0.6
6630	6712	0.4
6810	7010	0.8
6810	7514	0.2
6820	7020	0.8
6820	7514	0.2
6910	7411	1
6920	7412	1
7010	7414	1
7020	7414	1
7110	7421	1
7120	7422	0.8
7120	7523	0.2
7210	7310	1
7220	7310	0.5
7220	7320	0.5
7310	7430	0.9
7310	7499	0.1
7320	7413	1
7410	7421	0.6
7410	7499	0.4
7420	7494	0.8

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
7420	9220	0.2
7490	6309	0.2
7490	7414	0.3
7490	7421	0.2
7490	7492	0.1
7490	7499	0.2
7500	8520	1
7710	7111	1
7721	7130	1
7722	7130	1
7729	7130	1
7730	7111	0.0909
7730	7112	0.0909
7730	7113	0.0909
7730	7121	0.1818
7730	7122	0.1818
7730	7123	0.1818
7730	7129	0.0909
7730	7130	0.0909
7740	6599	1
7810	7491	0.75
7810	9249	0.25
7820	7491	1
7830	7491	1
7911	6304	1
7912	6304	1
7990	6304	0.8
7990	7513	0.1
7990	9214	0.05
7990	9219	0.02
7990	9241	0.03
8010	7492	1
8020	5260	0.5
8020	7492	0.5
8030	7492	1
8110	7020	1
8121	7493	1
8129	7493	0.75
8129	9000	0.25
8130	0140	0.8
8130	9000	0.2
8211	7499	1
8219	6411	0.25

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
8219	7499	0.75
8220	7499	1
8230	7499	1
8291	7499	1
8292	7495	1
8299	7499	1
8411	7511	0.7
8411	7514	0.3
8412	7512	1
8413	7513	1
8421	7521	1
8422	7522	1
8423	7523	1
8430	7530	1
8510	8010	1
8521	8021	1
8522	8022	0.95
8522	8090	0.05
8530	8030	1
8541	9241	0.9
8541	9309	0.1
8542	8090	0.75
8542	9219	0.25
8549	8090	0.8
8549	9241	0.2
8550	7414	0.5
8550	7499	0.3
8550	8532	0.2
8610	8511	1
8620	8512	1
8690	8519	1
8710	8519	1
8720	8519	0.6
8720	8531	0.4
8730	8519	0.5
8730	8531	0.5
8790	8531	1
8810	8532	1
9000	9214	0.7
9000	9219	0.2
9000	9220	0.1
9101	7514	0.5
9101	9231	0.5

Table 7 continued from previous page

version_4	version_3.1	Proportion of Jobs
9102	9232	1
9103	9233	1
9200	9249	1
9311	9241	1
9312	9241	1
9319	9241	1
9321	9219	1
9329	9219	0.5
9329	9241	0.2
9329	9249	0.3
9411	9111	1
9412	9112	1
9420	9120	1
9491	9191	1
9492	9192	1
9499	0150	0.25
9499	6599	0.5
9499	9199	0.25
9511	7250	1
9512	3220	0.75
9512	5260	0.25
9521	3230	0.5
9521	5260	0.5
9522	5260	1
9523	5260	1
9524	3610	0.6
9524	5260	0.4
9529	5260	1
9601	9301	1
9602	9302	1
9603	9303	1
9609	9309	1
9700	9500	1
9810	9600	1
9820	9700	1
9900	9900	1

Table 8: Job Codes Prediction Table Results

Codes	Variance
0113-0111	0

Table 8 continued from previous page

Codes	Variance
0113-0112	4.62223E-33
0119-0111	0.020833333
0119-0112	0.020833333
0125-0112	0.020833333
0125-0113	0.020833333
0126-0111	0.010833333
0126-0113	0.010833333
0128-0111	0.025833333
0128-0112	0.000833333
0128-0113	0.020833333
0129-0111	0.04
0129-0112	0.013333333
0129-0200	0.013333333
0130-0112	0.005833333
0130-0200	0.005833333
0162-0140	0.0025
0162-2892	0.0025
0163-0111	0.063333333
0163-0140	0.063333333
0230-0112	0
0230-0113	0
0230-0200	0
0321-0122	0.003333333
0321-0502	0.003333333
0322-0122	0
0322-0502	0
0810-1410	0.000833333
0810-1429	0.000833333
0910-1110	0.03
0910-1120	0.0037
0910-7421	0.0127
0990-1010	0.005069917
0990-1020	0.00287415
0990-1030	0.00027808
0990-1200	0.00027808
0990-1310	0.002507559
0990-1320	0.003112245
0990-1410	7.55858E-06
0990-1421	1.88964E-06
0990-1422	1.88964E-06
0990-1429	1.88964E-06
0990-4510	0.00011935
0990-7421	0.00011935

Table 8 continued from previous page

Codes	Variance
1030-1513	0.010833333
1030-1549	0.010833333
1071-1541	0
1071-1549	2.88889E-34
1075-1512	0.00226875
1075-1513	0.000352083
1075-1544	0.000352083
1075-1549	0.000102083
1079-1549	0.01
1079-2429	0.01
1102-0113	0.0325
1102-1552	0.0325
1312-1711	0.010258333
1312-2699	0.010258333
1313-1712	0.010833333
1313-1729	0.010833333
1313-1810	2.88889E-34
1392-1721	0.000833333
1392-3430	0.000833333
1399-1729	0.0025
1399-3699	0.0025
1511-1820	0.050833333
1511-1911	0.050833333
1512-1912	0.023333333
1512-1920	0.003333333
1512-3699	0.013333333
1629-1920	0.003333333
1629-2029	0.013333333
1629-3699	0.003333333
1709-1729	0
1709-2109	0
1709-2221	0.000833333
1709-3699	0.000833333
1811-2109	0.0453
1811-2221	0.092633333
1811-2892	0.025833333
1910-2310	0.000833333
1910-2411	0.000833333
1920-1010	0.01
1920-1020	0.003333333
1920-2320	0.023333333
2011-1551	0.000833333
2011-2330	0.013333333

Table 8 continued from previous page

Codes	Variance
2011-2411	0.040833333
2011-2429	0.003333333
2029-2429	0.000833333
2029-3699	0.000833333
2219-1920	0
2219-2519	0.020833333
2219-3610	0.005833333
2219-3699	0.005833333
2220-1920	0.005833333
2220-2109	0.000833333
2220-2519	2.88889E-34
2220-2520	0.000833333
2220-3610	0.000833333
2220-3699	0.000833333
2392-2691	1.84889E-32
2392-2693	0
2420-2330	0.003333333
2420-2720	0.003333333
2431-2710	0.013333333
2431-2731	0.013333333
2593-2893	0.040833333
2593-2929	0.040833333
2599-2899	0.01
2599-3190	0.005833333
2599-3699	0.000833333
2610-2429	0.0075
2610-2520	0.000833333
2610-3000	0.000833333
2610-3110	2.88889E-34
2610-3120	0.000833333
2610-3130	2.88889E-34
2610-3210	0.0075
2610-3230	0.000833333
2630-3190	0.011633333
2630-3220	0.011633333
2630-3230	0.005633333
2640-3230	0.000833333
2640-3694	0.000833333
2651-3190	0
2651-3220	0
2651-3312	0
2651-3313	0
2670-3230	0.0075

Table 8 continued from previous page

Codes	Variance
2670-3312	0.000833333
2670-3320	0.013333333
2710-3110	0
2710-3120	0
2731-3130	0.000833333
2731-3320	0.000833333
2733-2520	0.0025
2733-3120	0.003333333
2733-3190	0.005833333
2740-3150	0.000833333
2740-3190	0.000833333
2790-2922	0.000709366
2790-2929	2.75482E-05
2790-3110	0.002837466
2790-3120	0.001701102
2790-3130	2.75482E-05
2790-3150	2.75482E-05
2790-3190	0.002837466
2790-3210	2.75482E-05
2811-2911	0.016639118
2811-2912	2.75482E-05
2811-3110	2.75482E-05
2811-3430	0.002754821
2811-3530	0.002066116
2811-3591	0.002066116
2812-2519	0.000833333
2812-2912	0.01
2812-2913	0.005833333
2815-2914	0.0075
2815-2930	0.0075
2816-2915	0
2816-3599	2.88889E-34
2817-2429	0.003333333
2817-2899	0.003333333
2817-3000	0.003333333
2817-3230	2.88889E-34
2817-3610	2.88889E-34
2819-2919	0.0175
2819-2922	0.000833333
2819-2930	0.000833333
2819-3312	0.010833333
2822-2922	0.0075
2822-3190	0.0075

Table 8 continued from previous page

Codes	Variance
2826-2926	0.000833333
2826-2929	0.000833333
2930-3190	0.000833333
2930-3430	0.010833333
2930-3610	0.0075
3011-3511	0.000833333
3011-3610	0.000833333
3020-3190	0.040833333
3020-3520	0.043333333
3020-3610	0.000833333
3030-2927	0.005833333
3030-3530	0.0075
3030-3610	0.000833333
3092-3592	0.000833333
3092-3699	0.000833333
3100-3599	0.000833333
3100-3610	0.000833333
3211-3330	0.0075
3211-3691	0.0075
3212-3330	0.020833333
3212-3699	0.020833333
3230-1920	0.075833333
3230-3693	0.075833333
3250-1721	0.000709366
3250-2423	0.003663912
3250-2919	0.000709366
3250-3311	0.022506887
3250-3312	0.000709366
3250-3320	0.000709366
1810-1810	0.000833333
1912-1912	7.22224E-35
2029-2029	7.22224E-35
2211-2211	7.22224E-35
2519-2519	0.000833333
2520-2520	0.000833333
2699-2699	7.22224E-35
2899-2899	0.003333333
3311-3311	0.000833333
3693-3693	7.22224E-35
3699-3699	0.003333333
3311-2811	0.012458678
3311-2812	0.001246556
3311-2813	2.75482E-05

Table 8 continued from previous page

Codes	Variance
3311-2892	0.001012397
3311-2893	0.000557851
3311-2899	0.004428375
3311-2927	0.000261433
3311-2929	0.000261433
3311-3420	0.001636639
3312-2911	0.001264792
3312-2912	0.000252446
3312-2913	0.000306767
3312-2914	4.61363E-05
3312-2915	4.61363E-05
3312-2919	0.000252446
3312-2921	0.000306767
3312-2922	0.000306767
3312-2923	4.61363E-05
3312-2924	4.61363E-05
3312-2925	4.61363E-05
3312-2926	4.61363E-05
3312-2929	0.000252446
3312-3110	4.61363E-05
3312-3230	4.61363E-05
3312-3599	4.61363E-05
3312-3694	4.61363E-05
3312-3699	4.61363E-05
3312-7250	0.000252446
3313-3190	0.000120489
3313-3220	0.001553873
3313-3230	0.001553873
3313-3311	0.002430858
3313-3312	0.000759298
3313-3313	0.000297689
3313-3320	0.000365872
3313-5260	0.000507954
3314-2520	0.00037037
3314-2922	0.00037037
3314-2929	0.00037037
3314-3110	0.00037037
3314-3120	0.00037037
3314-3130	0.000648148
3314-3140	0.000648148
3314-3150	0.000648148
3314-3190	0.003981481
3314-3210	0.00037037

Table 8 continued from previous page

Codes	Variance
3315-3511	1.15556E-33
3315-3512	2.88889E-34
3315-3520	1.54074E-33
3315-3530	0.003333333
3315-3599	0.000833333
3315-6303	0.000833333
3319-1721	0.00425994
3319-1723	0.00011935
3319-2023	0.003217007
3319-2029	0.001117082
3319-2519	0.000564701
3319-2520	1.88964E-06
3319-2610	0.00027808
3319-2699	0.00011935
3319-3311	0.000731293
3319-3312	0.000761527
3319-3330	1.88964E-06
3319-3692	0.00011935
3319-3694	2.72865E-05
3320-2813	0.000119897
3320-2911	8.02616E-05
3320-2912	8.02616E-05
3320-2914	8.02616E-05
3320-2915	8.02616E-05
3320-2919	0.028974435
3320-2921	8.02616E-05
3320-2922	8.02616E-05
3320-2923	8.02616E-05
3320-2924	8.02616E-05
3320-2925	8.02616E-05
3320-2926	8.02616E-05
3320-2929	8.02616E-05
3320-3000	8.02616E-05
3320-3110	8.02616E-05
3320-3220	8.02616E-05
3320-3311	8.02616E-05
3320-3313	8.02616E-05
3320-4540	8.02616E-05
3530-1549	0.015833333
3530-4030	0.015833333
3812-2330	0.010833333
3812-9000	0.010833333
3821-2412	0.000833333

Table 8 continued from previous page

Codes	Variance
3821-9000	0.000833333
3822-2330	0
3822-9000	0
3830-3710	0.015833333
3830-3720	0.015833333
3900-4510	0
3900-9000	0
4312-4510	0.01
4312-4550	0.01
4329-4530	0.000833333
4329-4540	0.000833333
4330-4530	0.013333333
4330-4540	0.013333333
4390-4520	0.003333333
4390-4530	0.000833333
4390-4540	0.000833333
4390-4550	2.88889E-34
4663-5139	0
4663-5143	0
4753-5233	0.0075
4753-5239	0.0075
4791-5251	0.000833333
4791-5259	0.000833333
4922-6021	0.0057
4922-6022	0.005633333
4922-9241	0.010033333
5110-6210	0.003333333
5110-6220	0.003333333
5120-6210	0.0075
5120-6220	0.0075
5221-5020	0.01
5221-6010	0.0025
5221-6303	0.0025
5811-2211	0.0025
5811-7240	0.0025
5812-2211	0.000833333
5812-7240	0.000833333
5813-2212	0.000833333
5813-2219	0.003333333
5813-7240	0.000833333
5819-2219	0.0075
5819-7240	0.0075
5820-7221	0.005833333

Table 8 continued from previous page

Codes	Variance
5820-7240	0.005833333
5911-9211	0.0025
5911-9213	0.0025
5912-9211	0.000833333
5912-9231	0.000833333
5920-2213	0
5920-7240	0
5920-9211	0
5920-9213	0
6010-7240	0.003333333
6010-9213	0.003333333
6020-7240	0.003333333
6020-9213	0.003333333
6202-7210	0.000833333
6202-7229	0
6202-7230	0.000833333
6419-6519	0.000833333
6419-6719	0.000833333
6499-6592	0
6499-6599	0
6511-6601	0.010833333
6511-6603	0.010833333
6520-6601	0.003333333
6520-6603	0.003333333
6612-6712	0
6612-6719	0
6619-6599	0.173333333
6619-6719	0.173333333
6630-6602	0
6630-6712	0
6810-7010	0.003333333
6810-7514	0.003333333
6820-7020	0.000833333
6820-7514	0.000833333
7120-7422	0.005833333
7120-7523	0.005833333
7220-7310	0.003333333
7220-7320	0.003333333
7310-7430	0.005833333
7310-7499	0.005833333
7410-7421	0.023333333
7410-7499	0.023333333
7420-7494	0.003333333

Table 8 continued from previous page

Codes	Variance
7420-9220	0.003333333
7490-6309	0.003333333
7490-7414	0.0025
7490-7421	0.003333333
7490-7492	0.000833333
7490-7499	0.01
7730-7111	0.000731293
7730-7112	1.88964E-06
7730-7113	1.88964E-06
7730-7121	0.002517007
7730-7122	0.005969388
7730-7123	0.001088435
7730-7129	0.004449282
7730-7130	0.008459864
7810-7491	0.003333333
7810-9249	0.003333333
7990-6304	0.043333333
7990-7513	2.88889E-34
7990-9214	0.000833333
7990-9219	0.001033333
7990-9241	0.0247
8020-5260	0.050833333
8020-7492	0.050833333
8129-7493	1.84889E-32
8129-9000	1.15556E-33
8130-0140	0.0025
8130-9000	0.0025
8219-6411	0.005833333
8219-7499	0.005833333
8411-7511	0.01
8411-7514	0.01
8522-8022	1.84889E-32
8522-8090	1.15556E-33
8541-9241	0.000833333
8541-9309	0.000833333
8542-8090	0.03
8542-9219	0.03
8549-8090	0.000833333
8549-9241	0.000833333
8550-7414	0.0325
8550-7499	0.000833333
8550-8532	0.040833333
8720-8519	0.013333333

Table 8 continued from previous page

Codes	Variance
8720-8531	0.013333333
8730-8519	0.020833333
8730-8531	0.020833333
9000-9214	0.01
9000-9219	0.003333333
9000-9220	0.003333333
9101-7514	0.020833333
9101-9231	0.020833333
9329-9219	0
9329-9241	1.15556E-33
9329-9249	1.15556E-33
9499-0150	0.001633333
9499-6599	0.005633333
9499-9199	0.0112
9512-3220	0.000833333
9512-5260	0.000833333
9521-3230	0.0075
9521-5260	0.0075
9524-3610	0.0925
9524-5260	0.0925

References

[1] Bayard, K. N. and S. D. Klimek (2003). Creating a historical bridge for manufacturing between the standard industrial classification system and the north american industry classification system.

[2] Card, D. and A. B. Krueger (1995). Time-series minimum-wage studies: A meta-analysis. *The American Economic Review 85*, 238–243.

[3] Galindo-Rueda, F. and F. Verger. Oecd taxonomy of economic activities based on rd intensity.

[4] Gullickson, W. and M. J. Harper (2002). Bias in aggregate productivity trends revisited.

[5] Haley, S. and K. Keller (2023). The naics 2022 update and its effect on bls employment estimates in the retail trade sector.

[6] Jack E. Triplett, D. Mark Kennet, R. J. and F. M. Gollop (1995). Proceedings of the 6th asis sigicr classificanon research workshopÂ· do industrial classifications need re-inventing? an analysis of the relevance of the u.s. sic system for productivity research. pp. 145–172.

[7] Moscarini, G. and K. Thomsson (2007, 12). Occupational and job mobility in the us. *Scandinavian Journal of Economics 109*, 807–836.

[8] Yuskavage, R. E. (2007). Converting historical industry time series data from sic to naics.