Correlation between Unemployment Rates and Migrant Children to Non-Parent Sponsors

Team 080: Narges Bassirzadeh, Benjamin V Falby, MinYoung Kim,
Guysnove M Lutumba, Vakula Mallapally, and Clinton R Raye

Introduction & Problem Definition

This paper grapples with the problem of child migrant labor in the U.S. as highlighted by recent investigations in the New York Times (Murray et al., 2023). This reporting has further served to illustrate the inadequacy of federal resources at departments such as DOL (Department of Labor) for identifying places where child labor might be spiking (Drier, 2023b). Utilizing Health & Human Services (HHS) data secured by the Times and data from the US Bureau of Labor Services, we have analyzed the correlation between child migrant release destinations/dates and local unemployment rates. Our aim was to create an interactive visualization tool featuring a user-friendly interface, facilitating exploration of release patterns crossed with unemployment statistics. If such a tool revealed strong patterns, interested parties, whether government auditors, migrant advocates or reporters, would have a means for more quickly identifying potential child labor hot spots.

Literature Review

Our effort's novelty is crucial to its importance, especially as the HHS-released data set shows a recent surge in unaccompanied child migration (Murray et al., 2023), a phenomenon that has attracted attention from policymakers and advocates. The long-understood urgency surrounding child labor trafficking (Huijsmans & Baker, 2012) underscores the significance of our exploration of the HHS dataset, as explicitly encouraged by The New York Times (Murray et al., 2023), Dreier (2023a & 2023b). Rosenblum's (2015) examination of policy challenges from earlier surges in unaccompanied child migration provided valuable insights into the tension between humanitarian protection and immigration law enforcement, enhancing our team's insight into child placement pressures. Chavez & Menjivar (2010) provided a contextual understanding of the child migration process and broadly enhanced our domain understanding.

Existing literature on child migrant labor often focuses on three main areas: migration drivers, life outcomes, and the qualitative experience of migrants.

To the first area, Hamilton & Bylander (2021) examine migration motivations among Mexican children, emphasizing age, family dynamics, and community norms. While insightful for understanding migration patterns and the U.S. economic landscape, their research's specificity may limit its applicability to current trends. Reynolds (2013) discusses historical perspectives, suggesting many child migrants arrive with the intention to work. However, neither study quantitatively addresses local labor-demand contexts upon arrival. Wong's (2014) work significantly contributes to analytics-oriented literature by exploring the drivers of child migration, particularly violence in home countries. While not directly addressing U.S. labor market pressures, his analytical approach aligns with our project's methodology. However, he used basic tools compared to our advanced approach. On the other hand, Eggert et al. (2010) develops a two-region model for labor migrants, offering valuable insights but limited by its focus on only two regions and failure to consider multiple locations' interplay. In the limited quantitative literature on employment and child migration Eggert et al., 2010 and Chow et al., 2023 discuss Southeast Asia, exploring household networks, gender dynamics, and relay migration. However, this regional focus may not fully align with Central and South American contexts of child migration to the U.S. Work-adjacent studies, such as those by Crea et al., (2017 & 2022) document the work motivation to migrate but are not relevant to our problem statement.

As for the second category of migrant labor research, Evans et al., (2021), and Hasson et al. (2021), focus on life outcomes for child migrants to the U.S., often considering predictors for long-term self-sufficiency and positive outcomes, particularly related to non-parent sponsors leading to foster care. Li & Sun (2020) explore developmental outcomes for children affected by migration in the context of parent migration and left-behind children in China. In contrast, our research homes in on the immediate U.S. labor environment upon children's release.

The remainder of our sources fall in that third general category, addressing the experience that characterizes the child migrant journey. These sources tend to provide strong domain instruction, but in a different manner than the structural bent of Rosenblum and Chavez et al. Gimeno et al., (2021), Canizales (2023), Vega et al., (2023) and Iglesias et al., (2024) explore the lived experiences of child migrants in various facets, sometimes employment-related, but often dedicated to health and emotional well-being. These studies collectively underscore the complexity of child migration, suggesting that economic conditions, policy environments, and family dynamics are intertwined. However, a dedicated analysis of how the logistics of child migration may interact with U.S. unemployment rates and non-parent sponsor capacity is currently unavailable.

The literature reveals a gap in quantitative, demand-side exploration of labor and child migration to the U.S., presenting a new opportunity. Despite the risk of no meaningful correlation, our study's unique methodology was worth pursuing for its potential insights and the possibility of providing federal agencies and advocates with a predictor for child labor surges. Our study's necessity hinges on its novelty, as it explores the interaction between unemployment and migration patterns through innovative data integration. Notably, the absence of existing visuals representing this relationship underscores the uniqueness of our approach. Additionally, our algorithm's direct correlation of these distinct data types represents a pioneering effort in this field, contributing to the success of our work.

Methodology

Our method integrated three datasets: HHS release temporal data (comprising over 500,000 observations), Bureau of Labor Services county-level employment data, and a ZIP code to county linking dataset. Data was wrangled with Pandas, visualization with Dash and machine learning tools handled large datasets efficiently. We used regression analysis and other factor-based models to explore whether employment rates influence release destination and complemented that analytical exploration with visualization and user interface design. Our time parameters dictated the scope of our work, an exploration of the problem statement in a single state, Iowa, and the eventual visual extension to the full fifty.

We started our study by merging three distinct datasets to create a comprehensive resource for analysis. This consolidated dataset underwent preprocessing to ensure data quality and consistency, setting the stage for robust analysis. Combination of HHS data and zip code/county linking data culminated in the removal of 255 observations. Nearly all removals were related to zip code reconciliation challenges or total omissions. This still resulted in the preservation of more than 99.9% of the data. Engineered factors such as “release_year” and “release_month” were incorporated for analysis purposes.

An initial exploratory analysis was performed to examine the release of child migrants to sponsors across counties in Iowa. This step was pivotal in searching out patterns and informing the direction of our subsequent analysis and intended visualization. Our analysis's core involved estimating the correlation between the unemployment rate and the prevalence of migrant child release (particularly for non-parent sponsorships) both at a state level and specifically within individual counties. Statistical methods were rigorously applied to ensure the validity of our findings. The approach and results would be used to substantiate relevant visualizations.

The visualization interface was concurrently developed using Dash, a Python web application framework ideal for building analytical web applications. Dash leverages Flask, Plotly.js, and React.js, enabling the creation of highly interactive, web-based data visualizations that are both user-friendly and aesthetically pleasing.

A bar chart and a map were prototyped to represent the data. These charts included interactive elements such as dropdown menus to select specific states, thereby dynamically updating to display only the counties within the chosen state. The interactivity was achieved using Dash’s callback functionality, which allows components of the web application to communicate in real-time and update asynchronously.

Experiments

The correlation between unemployment rates and release destination (particularly to non-parent sponsors) held many nuances when aggregate data was compared to localized county-level data. Figure 1 below depicts the Pearson correlation coefficient across the entire data set.

The correlation was calculated with the number of migrants released to non-parental sponsors and the unemployment rate on a per county basis for each month present in the data set. The resulting coefficient is very near zero at -0.03, indicating no significant relationship between unemployment rates and non-parent sponsorships.

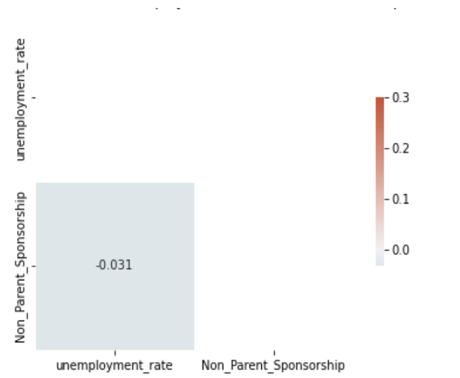

Figure 1: Correlation Matrix between Unemployment and Non-parent Sponsors

Figure 2 below shows the same correlation coefficients calculated for each county in Iowa individually. This perspective highlighted a wide range of relationship strength between unemployment and migrant counts. Although not depicted in this report, correlation factor shows a similar theme for all US states.

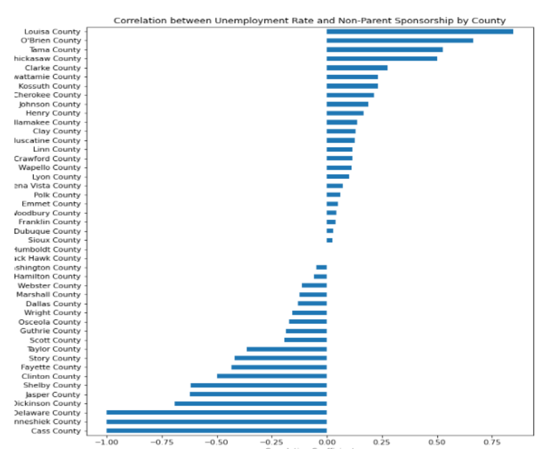

Figure 2: Correlation Matrix between Unemployment and Non-parent Sponsors per County

When examined at a high level, for localized data sets there can exist a strong positive, or negative, correlation between the two factors.

Linear Regression Model

To increase the fidelity of analysis compared to simple correlation, a series of linear regression models were trained on a subset of the overall data, namely all observations for Iowa. All models applied to Iowa initially assumed the dependent variable as the unemployment rate data. Each linear regression model was trained on different independent variables. Preliminary models using only child release date as independent showed limited significance, and a highly negative coefficient suggesting that an increase in migrating children corresponded to a decrease in the unemployment rate.

To capture any potential reactive impacts of the unemployment rate, the following models incorporated time series techniques, resulting in slightly improved performance. Additional variables utilized variables in further models resulted in more significant variance attributable to the selected factors, as indicated by the R-squared. However, the highest R-squared value was only 43%, indicating there was not a significant portion of the variance in migrant release data attributable to unemployment data. Table 1 below lists the variables used and the resulting R-squared value for each linear regression model.

Table 1: Linear Regression Results

Linear Regression Type	Independent Variables	Significant Variables	R^2
Simple	-Date of release	-Date of Release	1.80%
Time Series	-Date of release	-Date of release	5.80%
Expanded	-Date of Release -County -Month	-Date of Release -County -Month	43%
Time Series (with lag)	-Date of release (1-month lag) -County -Month -Gender -Sponsor Category	-Date of Release (1-month lag) -County -Month	43%
Time Series (with lag)	-Date of release (1-month lag) - Date of release (2-month lag) -County -Month -Gender -Sponsor Category	-Month -County	43%

Cross-Correlation Function (CCF) Analysis

We employed CCF analysis to uncover potential correlations in temporal dynamics. CCF analysis indicated a low negative correlation between the unemployment rate and the number of migrant children, with the strongest correlation observed at a lag of 0.75. The analysis highlighted mild inverse relationships between the variables during specific time lags.

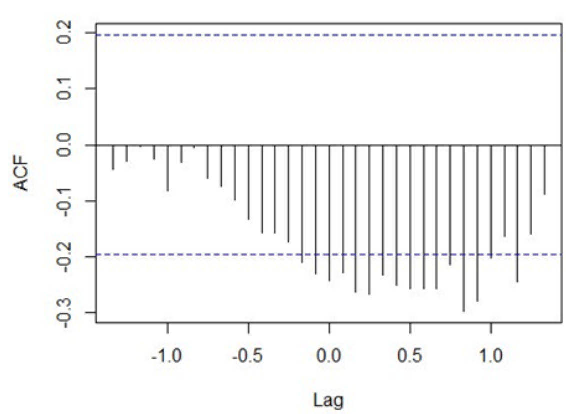

Figure 3: Cross-Correlation Function

ARIMA Correlation Analysis

We employed the ARIMA (Auto-Regressive Integrated Moving Average) model to assess potential correlations between the residuals of child arrival patterns and fluctuations in the unemployment rate, aiming to further capture any additional temporal patterns not accounted for by traditional linear regression techniques. ARIMA correlation analysis revealed a very weak linear relationship between the ARIMA residuals of child release and the unemployment rate. The analysis suggested limited to absent linear association between changes in child release patterns and fluctuations in the unemployment rate.

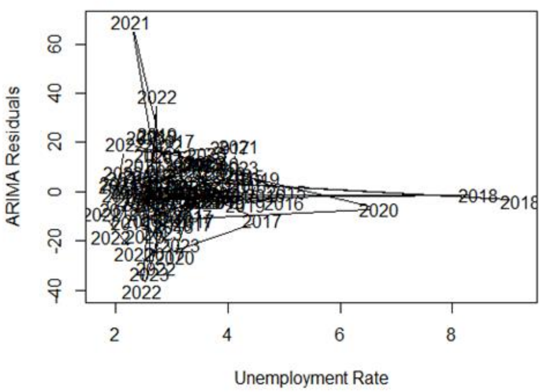

Figure 4: ARIMA Analysis

Machine Learning Models

The final analytical approach, also applied solely to data from Iowa, used machine learning to find correlations between unemployment rates and a child's release date. The machine learning model used 80% of the data for training and 20% for testing. The initial model only considered unemployment rates and a child's release date. Table 2 lists results for 3 machine learning models trained on the Iowa data.

Table 2: Machine Learning Results

ML Model Results	ML – Child's Release Date Only	ML – Four Factors	5_fold Cross-Validation
Mean Square Error (MSE)	0.89	0.89	0.88
Mean Absolute Error (MAE)	0.71	0.72	0.70
Root Mean Squared Error (RMSE)	0.94	0.94	0.93
R^2 Score	0.02	0.07	0.08

The results did not indicate any significant relationship between these variables. In a second model, more features were added, including the child's release date, month, county, and sponsor category. Although the addition of these features slightly improved the R-squared value, the relationship between the variables remained weak. We then applied a 5-fold cross-validation model to our dataset. The results continued to exhibit no significant change.

Visualization

It was then decided the visualization would be developed to allow independent insight into the Pearson correlation coefficient between the unemployment rate and the number of migrants released to a county on a per month basis. By way of selecting the desired filter, a user can evaluate the correlation coefficient for every county in the United States. Users can filter by state, county, the migrant's gender, the category of sponsor, as well as by start and end time. Figure 5 below demonstrates the choropleth resulting from a user's filter selection. Correlation coefficients are color coded with blue representing a coefficient of one and red representing a coefficient of negative one.

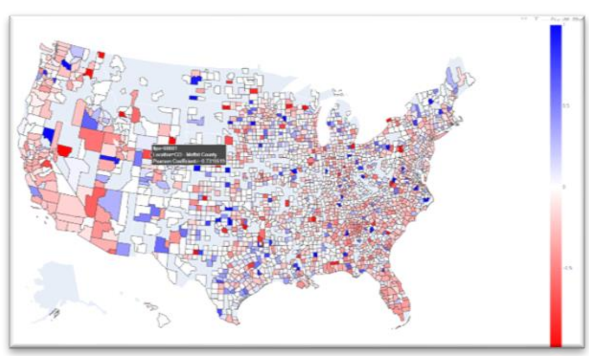

Figure 5: Pearson Correlation Choropleth

Additionally, all counties selected by the filters have their unique correlation coefficient plotted as a bar chart. Figure 6 below illustrates an example of this visualization.

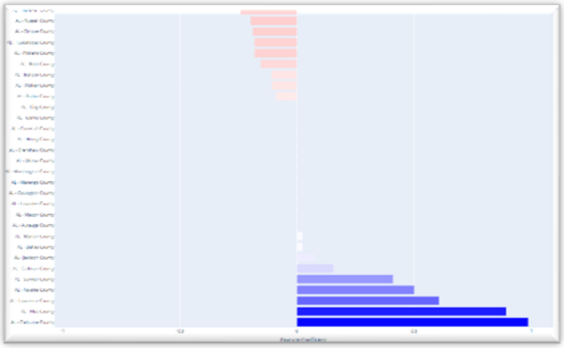

Figure 6: Pearson Correlation Bar Chart

Evaluation

While it was unclear that a definitive, general connection exists between unemployment rate and migrant release location, there are two proposed methods for enhancing the future utility of the developed analytical tool.

First, the tailored models created for a set of areas will be used to predict future release of child migrants to said places based on newly released unemployment data. An error signal comparing the predicted value to the actual number of migrants to that area can be calculated for each period of interest, say the error in prediction for each month of the next year. Low error signals would indicate stability in the correlation coefficients between the two factors over time, building confidence in the strength of unemployment data as a predictor of migrant placement.

Secondly, since this analysis will be of most interest to activists, politicians, government agents, etc., a user study will be performed to determine the user friendliness of the tool. Questions asked of users enlisted to experiment with the visualization will include those such as: is the tool intuitive to navigate? Were you able to efficiently encounter information for areas and times? Does the tool provide unique insight into the relationship between unemployment and migration? Users would be asked to answer each question on a 1-5 scale and with short form responses to provide quantitative and qualitative observations into the tool's quality.

Conclusion

All models ran in R studio version 4.3.1. No single method conclusively determined the relationship's nature for the Iowa subset. Linear regression showed limited dependency while altogether neglecting temporal dependencies. CCF did provide some insight into specific time lags. On the state level, ARIMA showed very weak correlation. Finally, in our machine learning approach, an initial model using only unemployment rates and a child's release date did not reveal a significant relationship. Despite the addition of more features such as the month, county, and sponsor category, the relationship remained weak. And subsequent 5-fold cross-validation did not yield significant changes. Collectively, these findings in Iowa suggested that the initial premise, the possibility that unemployment rate might mark migrant release location, was likely false. We saw value in continuing to develop our code for visualizing the relationship, as it might quickly corroborate our analysis in Iowa. That said, we decided a Pearson correlation would work no better nor worse than the complex alternatives we had entertained.

Scaling correlation with a Pearson coefficient to the full national dataset did not highlight any further, general insight into the relationship between the two variables. While a global insight may not be provided, localized results by county can provide more value to interested observers.

*__Equal Contribution Credit__: All tasks were performed by all team members equally. Authors are listed in alphabetical order.

References

Canizales, S. L. (2023). Work primacy and the social incorporation of unaccompanied, undocumented Latinx youth in the United States. *Social Forces, 101*(3), 1372–1395. https://doi.org/10.1093/sf/soab152

Chavez, L., & Menjívar, C. (2010). Children without borders: A mapping of the literature on unaccompanied migrant children to the United States. *Migraciones internacionales*, *5*(3),71111.https://ssrn.com/abstract=2333016

Chow, C., Zhou, X., Fu, Y., Jampaklay, A., & Jordan, L. P. (2023). From Left-Behind Children to Youth Labor Migrants: The Impact of Household Networks, Gendered Migration, and Relay Migration in Southeast Asia. *Social Sciences, 12*(135). https://doi.org/10.3390/socsci12030135.

Crea, T. M., Evans, K., Lopez, A., Hasson, R. G., Palleschi, C., & Sittley, L. (2022). Unaccompanied immigrant children in long-term foster care: Identifying and operationalizing child welfare outcomes. *Child & Family Social Work*. https://doi.org/10.1111/cfs.12902

Crea, T. M., Lopez, A., Taylor, T., & Underwood, D. (2017). Unaccompanied migrant children in the United States: Predictors of placement stability in long term foster care. *Children and Youth Services Review*, 73, 93-99. https://doi.org/10.1016/j.childyouth.2016.12.009

Dreier, H. (2023, February 25). Alone and Exploited, Migrant Children Work Brutal Jobs Across the U.S. *The New York Times*. https://www.nytimes.com/2023/02/25/us/unaccompanied-migrant-child-workers-exploitation.html

Dreier, H. (2023, December 28). They're Paid Billions to Root Out Child Labor in the U.S. Why Do They Fail? *The New York Times*. https://www.nytimes.com/2023/12/28/us/migrant-child-labor-audits.html

Eggert, W., Krieger, T., Meier, V. (2010). Education, Unemployment and Migration. *Journal of Public @Economics*, 94, 5-6 (repec.org)

Evans, K., Crea, T. M., Chu, Y., Salas-Wright, C. P., Takeuchi, D., Egmont, W., & Todo-Bom-Mehta, C. (2021). Paths to Self-Sufficiency for Youth Served Through the Unaccompanied Refugee Minor Foster Care Program in the United States. *Journal of International Migration and Integration*. https://doi.org/10.1007/s12134-021-00900-7

Gimeno-Monterde, C., Gómez-Quintero, J. D., & Aguerri, J. C. (2021). Unaccompanied young people and transition to adulthood: Challenges for childcare services. *Children and Youth Services Review*, 121, 105858. DOI: 10.1016/j.childyouth.2020.10585

Hamilton, E. R., & Bylander, M. (2021). The Migration of Children from Mexico to the USA in the Early 2000s. *Population Research and Policy Review, 40*(2), 337–361. https://doi.org/10.1007/s11113-020-09591-x.

Hasson, R. G., Crea, T. M., Easton, S. D., Evans, K., & Underwood, D. (2021). Clocking in: Employment outcomes for unaccompanied refugee minors leaving foster care in the United States. *Children and Youth Services Review*, 126, 106040. https://doi.org/10.1016/j.childyouth.2021.106040

Huijsmans, R. and Baker, S. (2012), Child Trafficking: 'Worst Form' of Child Labour, or Worst Approach to Young Migrants?. *Development and Change*, 43: 919-946. https://doi.org/10.1111/j.1467-7660.2012.01786.x

Iglesias, E., Montserrat, C., Llosada-Gistau, J., & Gallart, J. (2024). The well-being of unaccompanied migrant children: Between dreams and reality. Children and Youth Services

Review, 157, 107389. https://doi.org/10.1016/j.childyouth.2023.107389

Li, X., Sun X. (2020). Child Development in Rural China: Does Parental Migration Matter? *Children and Youth Services Review*, 119, 105654 DOI: 10.1016/j.childyouth.2020.105654

Menjívar, C., & Perreira, K. M. (2017). Undocumented and unaccompanied: children of migration in the European Union and the United States. *Journal of Ethnic and Migration Studies*, *45*(2), 197–217. https://doi.org/10.1080/1369183x.2017.1404255

Murray, E., Dreier, H., & Lai, K. K. R. (2023, December 28). Where Migrant Children Are Living, and Often Working, in the U.S. *The New York Times*. https://www.nytimes.com/interactive/2023/12/28/us/migrants-children-data.html

Reynolds, J. F. (2013). (Be)laboring Childhoods in Postville, Iowa. *Anthropological Quarterly*, *86*(3), 851–889. http://www.jstor.org/stable/43652639

Rosenblum, M. R. (2015). Unaccompanied Child Migration to the United States: The Tension between Protection and Prevention. Washington DC: Migration Policy Institute. https://www.migrationpolicy.org/research/unaccompanied-child-migration-united-states-tension-between-protection-and-prevention

Vega Potler, N. J., Zhang, J., Hackley, B., Choi, J., Xie, X., Punsky, B., Pineda, L., Shapiro, A., & Authors Affiliations. (2023). Persistence of Emotional Distress in Unaccompanied Migrant Children and Adolescents Primarily From the Northern Triangle of Central America. *JAMA Network Open, 6*(6), e2318977. https://doi.org/10.1001/jamanetworkopen.2023.18977

Wong, T. K. (2014). Statistical analysis shows that violence, not U.S. immigration policies, is behind the surge of unaccompanied children crossing the border. Retrieved from Center for American Progress. 11wong-uacs2.pdf (ucsd.edu)