Uncovering Migrant Fertility Patterns Through Consumer Data

Francesco Rampazzo[1]
Micol Morellini, Ridhi Kashyap, Douglas Leasure, and Melinda Mills[2]

Introduction and Background

Migration and fertility are key processes shaping population change but remain difficult to measure accurately because of inconsistent definitions and limited data across countries (Bilsborrow 2016). These limitations obscure how migration influences family formation, union dynamics, and fertility.

Research shows a tight link between migration and fertility: migration can reshape fertility behaviours, and family formation can motivate migration (Kulu 2005; Mussino and Strozza 2012). Five hypotheses are commonly used to explain these patterns: socialisation, selection, adaptation, disruption, and interrelation of events (Kulu 2005; Milewski 2010; Kulu and González-Ferrer 2014; Adserà and Ferrer 2015). Together they emphasise that reproductive behaviour reflects both origin norms and destination contexts.

Evidence from Europe confirms that both origin and destination matter (Mussino and Cantalini 2024; Kulu and González-Ferrer 2014), and that selective emigration can affect fertility at origin (Anelli and Balbo 2021). Advances in computational social science now allow near real-time demographic insight from novel sources, including consumer transactions, while raising concerns about representativeness and inference (Breen and Feehan 2024; Lazer et al. 2020). Building on this literature, we use large-scale consumer data to study how migration relates to fertility among Italian migrants across Europe.

Italy and Hypotheses

Italy combines persistently low fertility with high emigration of young and educated adults, especially toward Northern and Western Europe. Constraints in welfare support, labour markets, and childcare costs limit family formation in Italy, while family-friendly policies abroad make childbearing more feasible.

H1 (Consumer data as a fertility proxy). Purchase behaviour, especially on childcare and household goods, can indicate fertility patterns and complement traditional data.

H2 (Migration and fertility abroad). Italian migrants are more likely to have children abroad where social and policy contexts are more supportive (Adserà and Ferrer 2015; Anelli and Balbo 2021).

H3 (Geographic selection in fertility). Migrants self-select into destinations with higher fertility norms and stronger welfare systems, shaping reproductive outcomes (Kulu and González-Ferrer 2014; Mussino and Cantalini 2024).

---

[1] University of Manchester

[2] University of Oxford

Methodology

We combine consumer transaction data and survey data from *MammaPack*, an e-commerce platform serving Italian migrants in 21 European countries. The dataset includes more than 22,000 customers and 1.2 million orders with product categories, purchase frequency, and delivery locations. A survey of 700 respondents provides verified demographic outcomes.

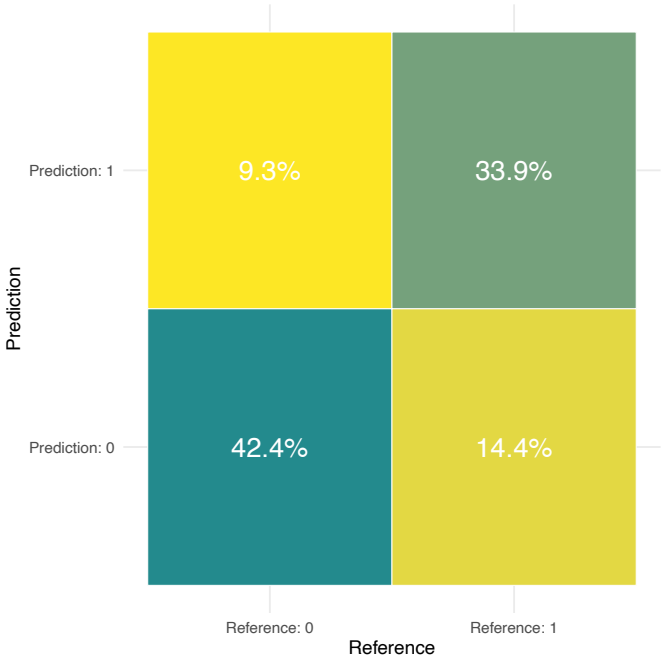
Products were grouped into 21 consumption categories using k-means and variable selection (VSURF and Boruta) (Genuer et al. 2017; Kursa and Rudnicki 2010). A Generalized Linear Mixed Model predicts the probability that a household includes children based on spending patterns, with random intercepts by country. Predictor selection used LASSO, and performance was assessed with k-fold cross-validation. The model achieves 76% accuracy, indicating that consumer spending, particularly on childcare goods, is a reliable behavioural proxy for fertility.

Results

We estimate the probability that a household includes children using a GLMM with spending on childcare, personal care, and snack foods as key predictors, and a country random intercept:

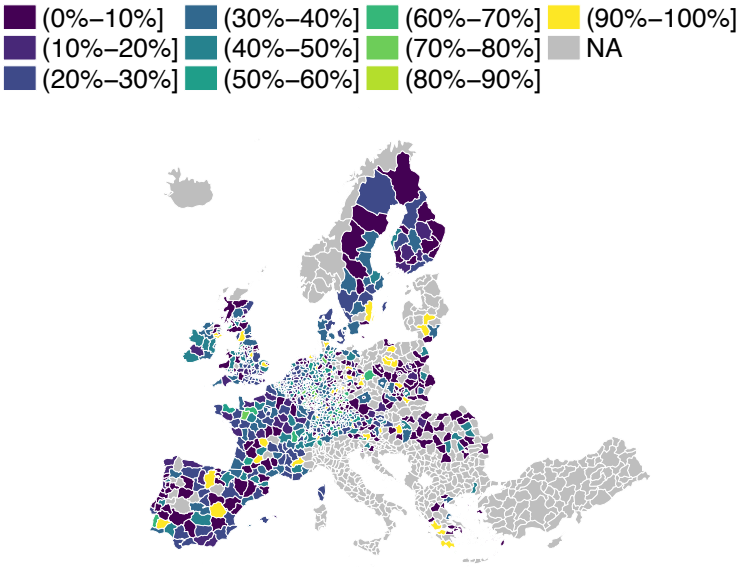
$$\begin{aligned}&\operatorname{logit}\big(P(\text{household\_minors} = 1)\big)\\ &= \beta_0 + \beta_1 \text{spent\_child} + \beta_2 \text{spent\_personal\_care} + \beta_3 \text{spent\_snack\_food} + b_{\text{country}}.\end{aligned}$$

Childcare spending is strongly and positively associated with having children in the household ($\beta_1 \approx 1.22, p < 0.001$). Personal care spending is negatively associated ($\beta_2 \approx -0.31, p < 0.01$), and snack foods are positively associated ($\beta_3 \approx 0.41, p < 0.01$). Model fit is good (AIC = 548.0), and country-level variance indicates moderate heterogeneity (variance 0.0847, SD 0.291). The model correctly classifies roughly three in four cases: accuracy 76.3 per cent compared with a 51.7 per cent baseline, Kappa 0.52, sensitivity 82 per cent, specificity 70 per cent, positive predictive value 75 per cent, and negative predictive value 78 per cent.



Confusion matrix of predicted versus observed classes.

Origin-region differences within Italy are small and not statistically significant once spending behaviour and destination effects are considered. Across Europe, predicted probabilities show a clear north-west to south-east gradient. The likelihood of having children is highest for Italians residing in Germany, Ireland, and Belgium, and lowest in Portugal, Romania, and Greece. At a finer scale, high values appear in East-Central and South-West Germany, the Aegean Islands, Silesia, and Lithuania, while very low values appear in Central Greece, Portuguese Centro and Algarve, and Southern and Northern Romania.



Predicted probability of households with children across European destinations.

These results support all three hypotheses. H1 is corroborated because consumer purchases, especially of childcare products, provide a valid behavioural proxy for fertility. H2 is consistent with higher predicted fertility abroad, which aligns with the view that migration can be both an economic and a reproductive strategy (Adserà and Ferrer 2015; Anelli and Balbo 2021). H3 is supported by the strong cross-country gradient, indicating selection into destinations with higher fertility norms and stronger family policy regimes (Kulu and González-Ferrer 2014; Mussino and Cantalini 2024).

Conclusions

Consumer transaction data can meaningfully capture links between migration and fertility for populations that are hard to observe in traditional sources. Behavioural data from *MammaPack* show that spending patterns, particularly on childcare, predict the presence of children with 76 per cent accuracy, which supports H1. Cross-national patterns indicate higher predicted fertility for Italians in Germany, Ireland, and Belgium, supporting H2 and H3 and reinforcing selection and adaptation mechanisms.

Several limitations remain. Consumer data may be affected by selection into platform use, and purchases are a probabilistic rather than causal indicator of family status. Future work should triangulate with administrative or survey records, extend to other migrant groups, and track temporal change.

Overall, computational and behavioural data provide scalable and timely insight into how migration and fertility interact in contemporary Europe.

References

Adserà, Alicia, and Ana Ferrer. 2015. "Immigrants and Fertility Adaptation: Evidence from Canada." *Journal of Population Economics* 28 (4): 965–1004. https://doi.org/10.1007/s00148-015-0564-0.

Anelli, Massimo, and Nicoletta Balbo. 2021. "Fertility Drain or Fertility Gain? Emigration and Fertility in Italy During the Great Recession." *European Journal of Population* 37 (3): 679–708. https://doi.org/10.1007/s10680-020-09577-0.

Bilsborrow, Richard E. 2016. *Migration, Population Structure, and Development: Implications for Policy and Planning*. Dordrecht: Springer.

Breen, Michael, and Dennis Feehan. 2024. "New Data Sources for Demographic Research." *Population and Development Review* 50 (1): 5–35. https://doi.org/10.1111/padr.12680.

Genuer, Robin, Jean-Michel Poggi, Christine Tuleau-Malot, and Nathalie Villa-Vialaneix. 2017. "Random Forests for Big Data." *Big Data Research* 9: 28–46.

Kulu, Hill. 2005. "Migration and Fertility: Competing Hypotheses Re-Examined." *European Journal of Population* 21 (1): 51–87. https://doi.org/10.1007/s10680-004-6793-4.

Kulu, Hill, and Amparo González-Ferrer. 2014. "Family Dynamics Among Immigrants and Their Descendants in Europe: Current Research and Opportunities." *European Journal of Population* 30 (4): 411–35. https://doi.org/10.1007/s10680-014-9322-0.

Kursa, Miron B, and Witold R Rudnicki. 2010. "Feature Selection with the Boruta Package." *Journal of Statistical Software* 36: 1–13.

Lazer, David, Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, et al. 2020. "Computational Social Science: Obstacles and Opportunities." *Science* 369 (6507): 1060–62. https://doi.org/10.1126/science.aaz8170.

Milewski, Nadja. 2010. "Immigrant Fertility in the Context of Assimilation and Socialization: A Case Study of Germany." *European Journal of Population* 26 (3): 297–323. https://doi.org/10.1007/s10680-010-9211-0.

Mussino, Eleonora, and Stefano Cantalini. 2024. "Multiple Origins and Multiple Destinations: The Fertility of Immigrant Women in Europe." *European Journal of Population*. https://doi.org/10.1007/s10680-024-01121-4.

Mussino, Eleonora, and Salvatore Strozza. 2012. "The Fertility of Immigrants After Arrival: The Italian Case." *Demographic Research* 26 (4): 99–130. https://doi.org/10.4054/DemRes.2012.26.4.