Online Appendix

The long road to economic independence of German women, 1973-2011

A. Income information

The scientific use-files of the German Microcensus ask respondents about the their personal income per month, including *any* income source net of income tax and social security contributions. The data do not include information about the amount of different income sources.

Respondents are not obliged to answer on the income question (in contrast to other questions). As a consequence, the income variable contains about 25% missings. We do not impute missing values, because we assume them to be not missing at random (Pepinsky 2018).

Respondents answer the income question using income brackets. The ranges and the number of income brackets change over time (Lengerer et al. 2010). Lengerer et al. (2010) provide an algorithm that assigns the inflation-adjusted median income within each bracket to a person. On that basis, we calculate the woman's income contribution C_W as

$$C_W = \frac{I_W}{I_W + I_M} \cdot 100$$

.

A possible consequence of responding to income brackets are higher shares on specific contributions than we would expect given a continuous measure. A woman's contribution of 50% does not mean that both partners contribute *exactly* the same income but that they claim to be within the same income range defined by the bracket.

B. Identification of couples

Prior to 1996, a couple always referred to married partners in the data. Large changes of the survey in 1996 shifted the focus from marriages to partnerships. We thus need to identify unmarried couples in the data between 1973-1995 using plausible assumptions about cohabitation and the age structure within couples. We use an algorithm provided by Lengerer (2007) to identify unmarried couples for the waves 1973-1995. It is possible that despite this algorithm, our sample prior to 1996 under-represents unmarried couples.

C. Notes on the kernel estimates

The kernel density estimates provided here rest on the assumption that all values are part of a Gaussian probability distribution ,which implies that there should be (as many) values above and below each data point. This is also the case for values at the end of the scales. This assumption causes two problems for the visualizations.

First, there are no values below zero and above 100. Thus, the probability of such values should be zero. However, the kernel density estimation does not account for that. The R-package *ggridges* offers a solution to that problem with the truncating option. By using this option, I fixed the range for the kernel density estimation for the interval [0,100]. The truncated kernel density estimator puts all density masses outside the interval into it. Thus, the truncated densities transform into a cumulative distribution that sums up to 1.

The second problem is strongly connected to the first. The assumption of the scale end points as points out of a probability distribution can result in an overemphasis of the densities for x-values in the ranges near the end points *within* the interval. There *are* observations near the end points for each year, but we need to interpret the results, especially the density changes over time, for the areas close to zero and to 100 with caution.

References

Lengerer, Andrea. 2007. *Zur Abgrenzung nichtehelicher Lebensgemeinschaften im Mikrozensus*, volume 2007/04. Mannheim.

Lengerer, Andrea, Julia Schroedter, Mara Boehle, Tobias Hubert, and Christof Wolf. 2010. "Datenhandbuch GESIS-Mikrozensus-Trendfile: Harmonisierung der Mikrozensen 1962 bis 2006.".

Pepinsky, Thomas B. 2018. "A note on listwise deletion versus multiple imputation." *Political Analysis* 26:480–488.