The clock is ticking. How much longer until you get divorced?

Lukas Hackl [*1] **Christopher Prinz** [*2] **Fabian Rostomily** [*3]

Abstract

This study investigated the validity of the "seven-year itch" phenomenon using German divorce data spanning from 1997 to 2022. Findings indicated a statistically significant peak in divorce probabilities around the fifth to sixth year of marriage, lending support to the notion of the seventh-year itch. Additionally, couples with children exhibited a higher likelihood of divorce, particularly in later marriage years. However, the study acknowledged limitations in fully exploring the impact of the Covid-19 pandemic on divorce rates due to data constraints. Methodologically dimension reductions, employing bootstrap methods for validation and artifact correction for refining estimators were used, revealing nuances in divorce dynamics. Despite the provided findings, the study recognized limitations, such as the exclusive focus on married couples and incomplete post-pandemic data. The source code can be found in our git-repository.

1. Introduction

The seven-year itch is to the society a well known saying. It describes the situation that relationships as well a marriages tend to be getting more complicated and raise problems which are likely to end in affairs, breakups and divorces. But is this thesis actually true or just superstition? What is the actual most probable year one will get separated? To deal with these questions we looked into the data of divorces in Germany. The used dataset consists of yearly divorces from 1997 until 2022, the marriage duration until divorce and the number of children present in the marriage 1. To find scientific reasoning on which the saying could be based on, we first calculated the probability of getting divorced giving a marriage duration 1. Further the data was analyzed in regards to the number of children present in the marriage and how children affect divorces in Germany. For this topic we took a second dataset into account which consists of the yearly number of married couples with and without children. We also took into account a temporal analysis regarding the Covid-19 pandemic and how it could have influenced it. The data showed statistical evidence on which the seven-year itch can be based on. Further we reached to the conclusion that relatively couples with children are more likely to get divorced, especially the more children are involved in a marriage, the more divorces tend to appear in the marriages with a higher duration. The influence of the pandemic could not be fully explored with this dataset, since more post-pandemic data is needed.

2. Data and Methods

The seven-year itch is known to be applicable for relationships in general including marriages and also non-married relationships. Since there is no way to have scientific records of breakups in general, we decided to concentrate only on data for divorces. The table consists of the number of yearly legally valid divorces in Germany from 1997 to 2022 with additional information about the marriage duration as well as the number of children present in the marriages. Thus, the number of in total over 4.5 million divorces are resolved along the three dimensions time (years), number of children and marriage duration. Regarding the involved children the numbers are divided into four categories: 0 children, 1 child, 2 children and ≥ 3 children. The 26 categories for the marriage duration's are reaching from 2 years to 25 years and also < 2 years and ≥ 26 years. In figure 1 the data is visualized according to its three dimensions. An implemented data loader loads and prepossesses the csv-file, resulting in an numpy-array that we further used to perform the dimension reductions for the specific analytical questions. During our analysis regarding the effects of the pandemic we discovered an error in the used dataset. The sum of all divorces in 1997 are about 40000 divorces off compared to other datasets, where as the total number of divorces in all years from 1998 to 2022 align perfectly with the other data. In our analyses we therefore decided to omit the year 1997 from the dataset, since we could not merge

[*]Equal contribution [1]Matrikelnummer 5379585, lukas.hackl@student.uni-tuebingen.de, BSc Computer Science [2]Matrikelnummer 4157933, christopher.prinz@student.uni-tuebingen.de, MSc Computer Science [3]Matrikelnummer 3947830, Fabian.Rostomily@student.uni-tuebingen.de, MSc Computer Science.

Project report for the "Data Literacy" course at the University of Tübingen, Winter 2023/24 (Module ML4201). Style template based on the ICML style files 2023.

Figure 1. 3D-visualization of the yearly number of divorces in Germany resolved along the marriage duration and the number of minors involved. The size of the points is scaled with the number of divorces that occurred in year x, with marriage duration y and involved minors z. For the purpose of a better visualization the last category of the marriage duration ($\geq$ 26 years), was omitted, since the data points in this category describe a sum across multiple marriage durations. Thus, the values are too high to allow an equally scaled plot in size and color.

the missing data from another source.

For the divorce probabilities we assumed every married person gets divorced at some point, at latest at the point of death. This assumption allowed us to calculate the probability $P(divorce|duration)$ with the given data, by summing up the divorces along the number-of-children and along the year axes:

$$P(divorce|duration) = \frac{n_{div,i}}{n_{div}}, \tag{1}$$

with the number of divorces with duration i $n_{div,i}$ and the number of total divorces n_{div}. Further, we derived the mean, median and most-likely duration to get divorced. To validate these three estimators, we implemented a bootstrap to simulate 100 datasets with 10000 samplings. After performing the bootstrap on the data, we discorvered an artefact, that leads to deviations in the analysis and the estimators. Therefore, a second bootstrap was performed on the corrected data. The outcomes are described more detailed in 3 2 in 3. To answer the question how children influence the likelihood of divorce we decided to incorporate a weighting factor for divorces per child, aligning with the number of married couples with and without children in Germany. Consequently, we opted to integrate a second dataset into our analysis. This supplementary dataset, provided by spans from 1996 to 2019 and withholds information on the count of married couples with and without children in Germany. To align the two datasets, we truncated the second dataset by 2 years at the beginning and trimmed our data by the last 4 years, to ensure both datasets have data starting at 1998 and concluding at 2019. In addition, we were required to reorganize our initial dataset, condensing it along the third axis from four distinct groups into two broader categories — divorces with and divorces without children. After both datasets had the same dimensions, we derived the fractions divorces with and without children per married couples and respectively divorces with and without children per marriage with children and without. The resulting plot is shown in 3. As figure 1 already intuitively seems to show, are the number of children affecting the marriage duration at which the divorces take place. We further analysed this by summing up the data across the year axis to get the divorces for every child-category across the marriage durations. We further calculated the probability similar to 1, but for the four groups of involved children individually. Therefore, we got the frequencies of divorces across the different marriage

durations along the z-axis of the dataset. The results are given in figure 4 in 3.

For the analysis on the effect of Covid-19 on divorces we reduced the data across the y- and z-axis, resulting in the number of all divorces per year. We further had a look at the summed up data eight years before and during Covid-19 to show any trends before and during the pandemic. In the process of further research we found out that in order to get a legal divorce you have to have a mandatory one-year separation (Bundesministerium), during which a married couple has to live apart from each other. As Covid-19 began in early 2020 and our dataset only has data until 2022, where the virus and lockdowns were still present, an effect would therefore only show in the next 2-3 years if there was one to begin with. Therefore, we ultimately decided against a further analysis of this topic, since we do not have sufficient data to make any concrete statements. This work is not the first one that has a look into data regarding divorces. There are for example other studies that examines divorces in regard to consequences Brüggmann (2023) or broader analysis including factors like happiness and cohabitation Zimmermann & Easterlin (2006), but nothing regarding the specific stereotypical seven-year itch.

3. Results

To estimate if there is statistic evidence to support the so called seven-year-itch we calculated the probability of getting divorced, given a marriage duration in years according to 1. After estimating the mean, median and most-likely duration year we performed two bootstraps for validation. In figure 2 the first bootstrap shows that the estimation for the most-likely duration to get divorced is off by one year. We figured this is, because all recorded divorces with a duration ≥ 26 were summed up to one count value. Therefore, the value influences the estimators and especially leads to a low-quality estimator for the most-likely marriage duration to get divorced. After correcting for this deviation in the data, the second bootstrap shows that the estimators describe the underlying distribution well. The mode of the sampled distribution and therefore the most-likely marriage duration to get divorced is at six years. The median with 10 years is only distributed across two different durations. The mean with 11.49 years has the least abbreviations in the simulated datasets and is distributed only over one bucket.

In addressing our inquiry about the potential negative correlation between the likelihood of divorce and the number of children, we utilized figure 3, which clearly illustrates that the number of couples in Germany with children is lower than the number of couples without children. Notwithstanding the observed trend, it is noteworthy that a higher incidence of divorces is evident among couples with children as opposed to those without. Overall, the total number of divorces is on a decline for both groups — married couples with children and those without.

In Figure 4, the divorce frequency is depicted across 26 durations, ranging from less than two years of marriage to more or equal to 26 years. As evident from the data, a higher frequency of divorces occurs within the initial five years of marriage for couples with less than two children.

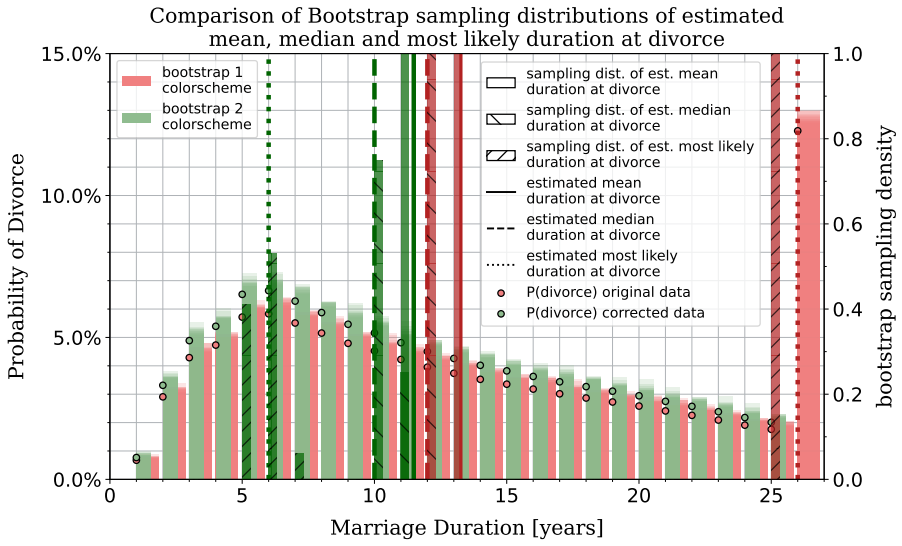

Figure 2. Comparison of the two performed bootstraps. The first bootstrap was performed on the original data. The second one was performed on the corrected data where the marriage duration category ≥ 26 was omitted. The three estimators (mean, median and most-likely) on the corrected dataset seem to better describe the distribution of the divorces according to the performed bootstrap, since the identified artifact is no longer influencing the estimators.

Furthermore the divorce frequency shows an increase for couples with three or more children after 12 years of marriage. In total, the data indicates a correlation: the more children are involved in a marriage, the more the higher divorce-frequencies tend to shift towards marriage durations. Additionally we can see that divorces in marriages of ≥ 26 duration without children are we more frequent. This is probably due to the fact that the dataset only takes minors into account and it is logical that less to none minors are present in marriages with this high duration. As mentioned

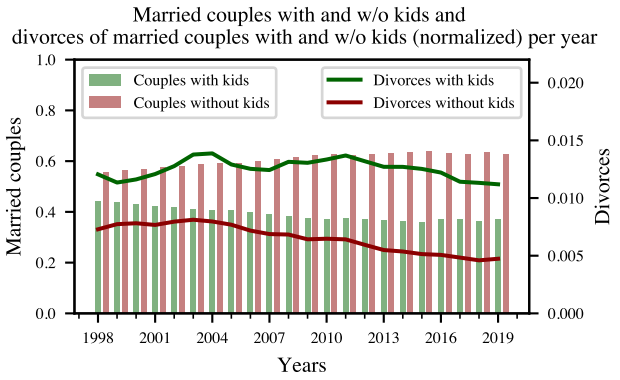

Figure 3. Plot with two y-axis to highlight the number of married couples with and w/o kids and the weighted probability of divorce

in 2 we did not further investigate the Covid-19 effect on divorces. Nevertheless it is noteworthy that during the start of the analysis of this topic, we figured a downwards trend for divorces in the last 15 years.

4. Discussion & Conclusion

As mentioned before, we could find a statistical evidence to support the seven-year-itch theory. The used data of divorces in Germany shows that the most probable marriage year to get divorced is at a marriage duration of six years. Thus, the seventh year of marriage is the one that most probably ends in a divorce. Nevertheless the data is not complete. For calculating the probability we could only use data until a marriage duration of 25 years. Even though the trend of the data seems to make it plausible that there is no other year with a higher probability, we can not make a final statement. Our examination of the data reveals that the likelihood of divorce is in total higher among couples with children compared to marriages without. This finding is on the one hand counter-intuitive to the assumption couples with children are more likely to stay together. On the other hand and according to our data, having children in a marriage is challenging and can lead to a higher divorce probability. Several factors may contribute to the observed outcomes. Firstly, due to constraints within the dataset, compromises were necessary. Reducing the initial four groups of children counts in marriages into a binary domain led to a loss of granularity. Secondly, by summing durations across all groups, we forfeit the ability to differentiate between varying periods within each category. This means we cannot discern, for instance, whether the probability of divorce differs significantly during the first 15 years of marriage while raising children, as opposed to being together for 15 years without children. This lack of temporal specificity may impact our understanding of the nuanced dynamics influencing marital stability over time. Interestingly, our findings revealed a contrasting trend when the divorces take place and children are involved. Specifically, it is less likely for couples with more than one child to experience divorce in the initial five to eight years of marriage compared to those without children. Intuitively, the reason for this could be that couples marriages with more children have a stronger desire to stay together for the kids. Also one could speculate that there are less problems created by the presence of children or less initial problems that become more apparent with children. Although until now there were no results to make a supported statement regarding the effect of Covid-19, we can state a downwards trend in divorces overall. Reasons for this can be multi-factorial as also highlighted by Destatis. To reason further about this we have to take topics like migration and emigration into account, which ultimately demands extending research.

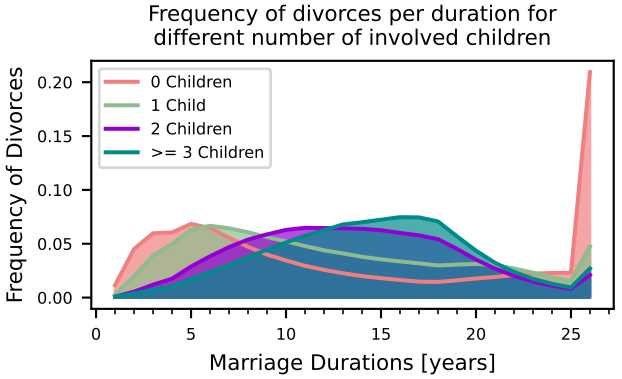

Figure 4. Visualization of divorce frequencies given the marriage duration for 0, 1, 2 and $\geq$ children involved in the marriage. The data shows a correlation of present children and the point in marriage duration, when divorces appear.

Contribution Statement

Christopher Prinz examined divorce regarding the influence of children. Fabian Rostomily analyzed the seven-year itch saying and build the visualization for the data intuition. Lukas Hackl analyzed the effect of the pandemic and engaged in research and the most part of the writing.

References

Brüggmann, D. *Consequences of Divorce in Germany*. PhD thesis, Hertie School, 2023.

Bundesministerium. § 1565 scheitern der ehe. URL https://www.gesetze-im-internet.de/bgb/__1565.html.

Destatis. Destatis - ehen im wandel. URL https://www.destatis.de/DE/Themen/Querschnitt/Demografischer-Wandel/Hintergruende-Auswirkungen/demografie-ehen.html#:~:text=Nach%20der%20deutschen%20Vereinigung%20ging,(einschlie%C3%9Flich%20der%20gleichgeschlechtlichen%20Eheschlie%C3%9Fungen).

Zimmermann, A. C. and Easterlin, R. A. Happily ever after? cohabitation, marriage, divorce, and happiness in germany. *Population and Development Review*, 32 (3):511–528, 2006. ISSN 00987921, 17284457. URL http://www.jstor.org/stable/20058902.