Marriage and Longevity

Collin Kennedy

December 10th 2020

1 Introduction

Whether or not marriage is linked to greater longevity is by no means a nuanced question; in fact, the ostensible advantages in life expectancy enjoyed by married individuals is relatively well documented. However, very few studies (if any) have had the luxury of having such an extensive, multi-generational database of over 400,00 observations at their disposal, nor have any really been able to determine if marriage truly leads to greater life expectancy. It was my intention to capitalize on the availability of this data to see if I could glean any new insight; Is the link between marriage and prolonged life expectancy a causal one?

For my project, I considered three regression models in order to evaluate the effect of marriage on life expectancy. I first ran a relatively straightforward multiple linear regression with a time fixed effect in order to get a general sense of the relationship between marriage and longevity, while controlling for a handful of other predictors.

Next, I add a layer of depth to my model by considering differences in longevity between brothers. By comparing life spans between brothers, a luxury afforded by the incredible scale of the families of england database, I was essentially able to control for the environment in which individuals grew up in (as well as hereditary factors). While I had initially suspected this (differencing) would alleviate the effect I found in the first model, the effect of marriage on longevity remained relatively persistent; the brother who married lived a little over 2 and a half years longer than the brother who didn't, on average.

Lastly, I sought to explore whether it's the act of getting married, or the length of marriage that matters most for predicting longevity. I expanded on my differenced brothers' model, but where the parameter(s) of interest are dummy variables to capture the effect of differences in marriage length (between brothers) on the difference in life expectancy. In plain English, I explore how much longer one brother lives based on how much longer his marriage his, compared to his other brothers. I find that the longer the difference in brothers' respective marriage lengths, the greater the return.

It is important to point out that this effect is almost surely non-linear. While I find that a greater difference in marriage is associated with a greater return to longevity for the brother experiencing the longer marriage, it would be appropriate to say that this effect has *decreasing marginal returns* on longevity. However, due to time constraints I didn't explore this non-linearity to the length it deserves (or really, at all). but nevertheless is important to be mindful of.

2 Methods and Results

Prior to any actual analysis, I first had to construct a subset of the FOE dataset that included variables of immediate interest. In R, I created 4 different subsets of the main database, one subset for "sons," "brothers," "fathers," and "wives," respectively, and then merged them to create my family dataframe. Each subset contained attributes related to educational achievement, occupational status, birth year and death age (which I use as a measure of longevity), marriage year, region of birth, and inherited wealth.

As part of my preliminary data exploration, I also wanted to determine if there was any sort of a linear relationship between whether or not an individual was married (*dmarried*) and death age (*dage*):

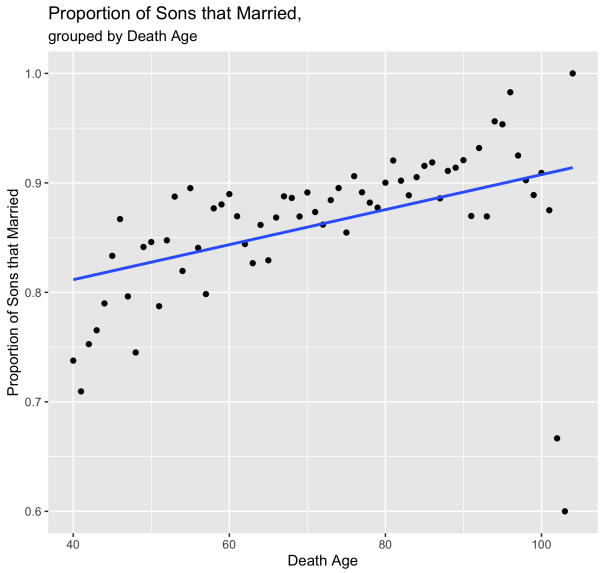

Figure 1:

Note here that I restrict the visualization to only include individuals living to at least 40. I find that when I do so, the relationship between death age and whether or not an individual married is relatively linear. This ends up being my justification for restricting my dataset(s) that I use to build the following linear models to only include sons (and eventually brothers) that lived to at least age 40.

2.1 Model 1: Marriage and Longevity

First, I considered a regression model where *dmarried_son* was the parameter of interest:

$$Y_{longevity} = \beta_0 + \beta_1 X_{dmarried\_son} + \beta_2 X_{ded} + \beta_3 X_{Occrank} + \lambda_t + \beta_4 X_{regbirth} + \beta_5 X_{inherited\_lnwealth} \quad (1)$$

note: where λ_t is a time fixed effect

Hypothesis:

$H_0 : X_{dmarried\_son} = 0$
$H_a : X_{dmarried\_son} \neq 0$

Table 1: Regression Table: Model 1

	Dependent variable:
	dage_son
dmarried_son	3.613***
	(0.242)
ded_son	0.699
	(0.426)
Occrank_son	4.875***
	(0.852)
regbirth_son	−0.005
	(0.029)
inherited_lnwealth	0.074
	(0.197)
Observations	26,807
R^2	0.056
Adjusted R^2	0.049
Residual Std. Error	12.431 (df = 26603)
Note:	*p<0.1; **p<0.05; ***p<0.01

Most notably, when I restrict my sample to only consider brothers who a) lived to at least 40, and b) got married before 40 (or did not marry at all), and after controlling for occupational and educational achievement, region of birth, and inherited (log) wealth, and including a time fixed effect, I find that for individuals that lived to at least 40, those who married live about 3.6 years longer on average than individuals who do not get married ($p < 0.01$). That's a decent chunk of extra life, but this result may not be impervious to endogeneity concerns. For example, perhaps *dmarried_son* is simply picking up the additional years of life that the type of people that get married normally get, regardless of whether or not they get married. For example, perhaps more fit and physically healthy individuals are more likely to get married. Because I can't directly control for this attribute, since it is not included in the database, I had to come up with a different solution to hedge this potential issue of omitted variable bias.

2.2 Model 2: Using brothers to compare (differences) in longevity

As a workaround, I considered looking at the *difference* in longevity between brothers. The advantage of such method is that I am able to indirectly control for a multitude of unobserved factors that I am otherwise unable to control for. This theoretically results in a much more precise and unbiased estimate of the difference in longevity between the average brother who does get married and the average brother who does not.

Consider the regression model I estimated and my results:

$$Y_{\Delta longevity} = \beta_0 + \beta_1 X_{\Delta dmarried} + \beta_2 X_{\Delta ded} + \beta_3 X_{\Delta Occrank} + \beta_4 X_{sameRegbirth} \tag{2}$$

Variables:

$X_{\Delta dmarried} = X_{dmarriedSon} - X_{dmarriedBrother}$

$X_{\Delta ded} = X_{dedSon} - X_{dedBrother}$

$X_{\Delta Occrank} = X_{OccrankSon} - X_{OccrankBrother}$

$$X_{sameRegbirth} = \begin{cases} 1, & \text{if born in same region} \\ 0, & \text{otherwise} \end{cases}$$

Hypothesis:

$H_0 : X_{\Delta dmarried} = 0$

$H_a : X_{\Delta dmarried} \neq 0$

Table 2: Regression Table: Model 2

	Dependent variable:
	delta_dage
delta_dmarried	2.686^{***}
	(0.364)
delta_ded	-1.071
	(0.716)
delta_Occrank	8.231^{***}
	(1.653)
same_regbirth	-0.929
	(0.657)
Constant	0.843
	(0.635)
Observations	10,743
R^2	0.008
Adjusted R^2	0.008
Residual Std. Error	16.927 (df = 10738)
Note:	*p<0.1; **p<0.05; ***p<0.01

Again, using similar restrictions on the data that I used to construct my first linear model, I find that the son who gets married lives about 2.6 years longer than the brother who doesn't on average, after controlling for differences in educational achievement (Δded), occupational rank ($\Delta Occrank$), and whether or not the brothers were born in the same region ($same\_regbirth$) (p <.01).

2.3 Model 3: Length of Marriage, or Marriage in General?

Lastly, I wanted to consider the effect of marriage on longevity from a different perspective. In particular, I wanted to evaluate to what extent *the difference* in *length* of marriage, (between brothers) matters. To do so, I defined a new variable, $X_{\Delta marriage\_length}$, which is calculated as the (vector) difference between the marriage length of the "son" and the marriage length of the "brother". It is important to point out that marriage length itself is defined as the difference between the year of marriage (for both the son's and brother's marriage) and the year in which their respective spouse died. I also only restricted my sample to consider sons and brothers who were only married once, as to remove the possibility of picking up the effect of additional marriages.

Having created this new $X_{\Delta marriage\_length}$ variable, it enabled me to create 3 dummy variables to capture the effect of the difference in marriage length between brothers over three, 15 year intervals:

Variables:

$$X_{\Delta marriage\_length0\_14} = \begin{cases} 1, & 0 \leq \text{difference in marriage length} < 15 \\ 0, & \text{otherwise} \end{cases}$$

$$X_{\Delta marriage\_length15\_29} = \begin{cases} 1, & 15 \leq \text{difference in marriage length} < 30 \\ 0, & \text{otherwise} \end{cases}$$

$$X_{\Delta marriage\_length30\_plus} = \begin{cases} 1, & \text{difference in marriage length} \geq 30 \\ 0, & \text{otherwise} \end{cases}$$

I then considered the following regression model, including $\Delta Occrank$ as a control:

$$Y_{\Delta longevity} = \beta_1 X_{\Delta marriage\_length0\_14} + \beta_2 X_{\Delta marriage\_length15\_29} + \beta_3 X_{\Delta marriage\_length30\_plus} + \beta_4 X_{\Delta Occrank} \tag{3}$$

Table 3: Model 3: Regression Table

	Dependent variable:
	delta_dage
delta_Occrank	3.189
	(1.944)
delta_marriage_length0_14	0.500
	(0.305)
delta_marriage_length15_29	2.214***
	(0.420)
delta_marriage_length30_plus	2.820***
	(0.642)
Observations	5,191
R^2	0.004
Adjusted R^2	0.003
Residual Std. Error	16.589 (df = 5187)
Note:	*p<0.1; **p<0.05; ***p<0.01

I also visualized these three binary variables:

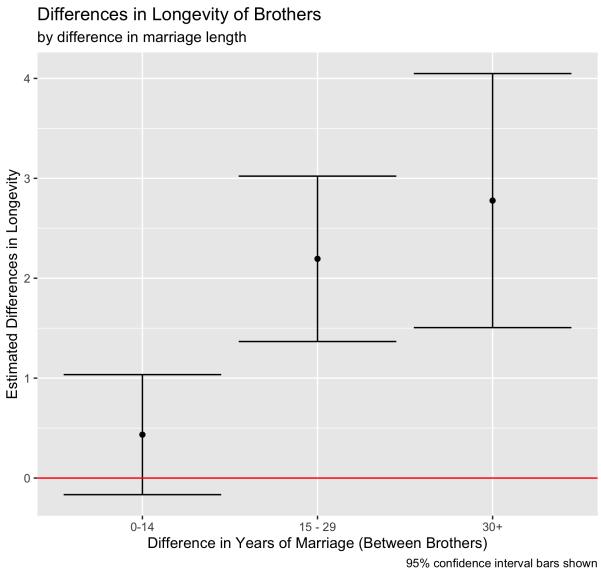

Figure 2:

Based on the estimated coefficients in Table 3, and visually confirmed by the plot in Figure 2, it is evident that the length of marriage does matter. For example, I find that when the difference in length of marriage experienced by brothers is between 15 and 29 years inclusive (e.g., the son's marriage lasts 30 years but the other brother's marriage lasts only 6 years), the son who experienced the longer marriage lives about 2.2 years longer on average. When the difference in marriage length is even greater (over 30 years), so is the return to longevity; the son who experiences the (significantly) longer marriage lives about 2.8 years longer than his brother, on average.

Notice however that when the length of marriage between brothers is not very different (between 0 and 14 years inclusive), there is no statistically discernible effect on the difference in life expectancy. Visually, this can be seen again in Figure 1, where the 95% confidence interval for the $delta\_marriage\_length0\_14$ estimate contains 0.

It appears that the *length* of marriage is what truly matters. If the simple act of getting married offered such considerable returns to life expectancy, there would likely be a statistically significant effect for a coefficient that captures the effect of a shorter difference in marriage length.

2.3.1 Model 3: Robustness Check

Endogeneity is often a concern in econometric settings, and this is no different. To determine if this model (3) was robust to potential identification concerns, I considered a hypothetical way in which this model could be invalidated and then constructed a new variable related to said concern to include as a control, and then evaluated how the results changed (if at all).

For example, I considered a scenario, or rather an argument, which goes something like this: perhaps this result is being driven by a son marrying a younger wife, and that (assuming the wife caring for their husband is at least part of the reason for the husband living longer) they are able to care for their husband for a greater amount of time, relative to a couple where the wife is older at the time of marriage.

To determine if this is a legitimate issue, I defined a new variable called ΔX_{spouse_age}, which represents the difference in age between the wife of the of son at marriage, and the wife of the brother(s) at marriage, and incorporate it in my model as a control:

New Variable:
$X_{\Delta spouse\_age} = X_{spouse\_age\_son} - X_{spouse\_age\_broth}$

(Robust) Model 3:

$$Y_{\Delta longevity} = \beta_1 X_{\Delta marriage\_length0\_14} + \beta_2 X_{\Delta marriage\_length15\_29} + \beta_3 X_{\Delta marriage\_length30\_plus} + \beta_4 X_{\Delta Occrank} + \beta_5 X_{\Delta\_spouse\_age} \quad (4)$$

Table 4: Robustness Check

	Dependent variable:	
	delta_dage	
	(1)	(2)
delta_Occrank	3.189	2.921
	(1.944)	(1.970)
delta_marriage_length0_14	0.500	0.435
	(0.305)	(0.307)
delta_marriage_length15_29	2.214***	2.194***
	(0.420)	(0.422)
delta_marriage_length30_plus	2.820***	2.777***
	(0.642)	(0.649)
delta_spouse_age		0.017
		(0.034)
Observations	5,191	5,127
R^2	0.004	0.004
Adjusted R^2	0.003	0.003
Residual Std. Error	16.589 (df = 5187)	16.585 (df = 5122)
Note:		*p<0.1; **p<0.05; ***p<0.01

The main thing to note here is how little the estimates for $\Delta marriage\_length$ change. While the estimates do change ever so slightly, the new estimates are still well within the 95% confidence interval of the first estimate(s) in the original 3rd regression model. The difference in marriage length between brothers appears to be robust to concerns of this effect on marriage being heavily influenced by the spouse's age (youth in particular).

3 Conclusion

Marriage is undoubtedly associated with greater life expectancy. Even after comparing differences in longevity between brothers, it is clear that marriage is linked to greater longevity.

Of course, as is often the issue with regression models in economics, the exogeneity assumption may be violated here. While I make a solid effort to control for certain major factors that are also heavily tied to longevity(e.g., educational achievement), it's important to consider the multitude of ways that omitted variable bias could be leading to poor and misleading estimates of the effect of marriage on longevity.

Healthier individuals who are more likely to live longer may be more likely to get married. Restricting my sample to only include people that live to at least 40 is an attempt to combat this issue, but it is by no means a perfect solution to the problem.

Granted, this may cut the other way: The act of marriage, and to a greater extent being married, may promote greater mental and physical health, which of course leads to greater life expectancy. My findings in my 3rd model somewhat support this idea: There is a considerable difference in longevity between brothers who experience marriages of variable length, suggesting that there is something about *being married* that is driving this difference. As always, the truth may lie somewhere in between. With all that said, this is still not a causal estimate, but it is still interesting enough.

Going forward, it would be great to be able to explore the impact of marriage on women's lives; while I restricted my focus to men due to the greater availability of data, it would be interesting to see if the results are similar across the board, or if women are affected by marriage differently than man are. It would also be interesting to see if there are any interaction effects. Does the effect of marriage on longevity depend on wealth of the couple?...
