Supplement to "Where did it go wrong? Marriage and divorce in Malawi"

(*Quantitative Economics*, Vol. 12, No. 2, May 2021, 505–545)

Laurens Cherchye
Department of Economics, University of Leuven (KU Leuven)

Bram De Rock
ECARES, Université Libre de Bruxelles and Department of Economics, University of Leuven (KU Leuven)

Frederic Vermeulen
Department of Economics, University of Leuven (KU Leuven)

Selma Walther
Department of Economics, University of Sussex

Appendix OA1 provides more details on our data construction. Appendix OA2 gives a concrete example of an instability network. Appendices OA3 and OA4 present additional empirical results.

Appendix OA1: Data construction

All values used in our empirical application were converted to real terms using the spatial and temporal price index provided in the IHS. In some cases we recoded outliers, namely the top 1% of values, to be equal to the value at the 99th percentile.

Bounds on wages and land prices

Wages We calculated the median observed wage per hour of hired workers in the district, separately for males and females. Where there were insufficient observations, we used the regional median instead. The bounds were zero and two times this median.

Laurens Cherchye: laurens.cherchye@kuleuven.be
Bram De Rock: bderock@ulb.ac.be
Frederic Vermeulen: frederic.vermeulen@kuleuven.be
Selma Walther: s.walther@sussex.ac.uk
We thank Veerle Hennebel and Yuki Kimura for excellent research assistance. Laurens Cherchye gratefully acknowledges the European Research Council (ERC) for his Consolidator Grant 614221. Part of this research is also funded by the Research Fund of the KU Leuven (BOF) and the EOS project 30544469 of FWO/FNRS. Bram De Rock gratefully acknowledges FWO and FNRS for their financial support (including the EOS project 30544469). Frederic Vermeulen gratefully acknowledges financial support from the Research Fund KU Leuven through the grant STRT/12/001, from the FWO through the grant G057314N, and from FWO/FNRS through the EOS project 30544469.

Available at http://qeconomics.org. https://doi.org/10.3982/QE1272

Land price per acre For each plot of owned land, households were asked how much they could earn if they rented it out for 1 year. We regressed this value on plot characteristics: the size of the plot, the soil type of the plot, the soil quality of the plot, whether the plot is swamp or wetland, and how the household acquired the plot. We then used the predicted values of this regression to estimate the rental income for those plots where the reported rental income was missing. The rental income was summed for each household and divided by the total acres of land, giving an average rental income per acre for each household. We then obtained the median rental income per acre for each village and for each district. We used the median rental income per acre for the village where there were at least seven observations per village; where there were fewer, we used the median rental income per acre for the district. The bounds on the land price were zero and two times this median.

Production

Inputs We calculated the cost of inputs into production as the total of direct inputs, such as the costs of fertilizer, seeds and transport, the cost of indirect inputs, namely machinery, and the cost of hired labor. For machinery, we calculated the use value of each item by first calculating the remaining age of the item as twice the mean age of this item in the sample minus its current age, with a minimum of 2 years. The annual consumption stream from each item was the amount of money the item could be sold for, if sold today, divided by the remaining age of the item. The cost of hired labor was calculated as the number of days this labor was used times the average daily wage for these laborers, as reported by the household. The survey distinguished between male, female and child laborers, providing a more accurate measure of the total cost. Free labor was also valued at these rates and included as a costly input.

Revenue The revenue was calculated as the sum of all crop sales during the rainy and dry seasons and the value of all own agricultural production that was consumed by the household. The latter value originates from the survey itself, where households were asked how much of each consumed food they had grown themselves. This was then valued at local prices by the World Bank Living Standards Measurement Study team.[1]

Consumption

Consumption was split into four categories: public consumption, private nonassignable consumption, private consumption of the man and private consumption of the woman.

Public consumption This included expenditure on children's education and health, expenditure on the education and health of other household members (not the husband or wife), expenditure on children's clothing, expenditure on durables (which was calculated as a use value or consumption stream, using the same method described for ma-

[1] Many thanks to Talip Kilic for sharing his Stata code that allowed us to separately identify consumption from own production and consumption from purchases.

chinery above), expenditure on public nondurables (such as candles, light bulbs, and books), expenditure on rent and expenditure on public bills (such as firewood and the landline telephone).

Private nonassignable consumption The largest component of private nonassignable consumption was food, consisting of food purchased, the value of food from own production, and the value of food received as a gift. This category also included private bills (such as the mobile telephone) and private nondurables (such as cigarettes, tickets for public transport, soap, and stationery items).

Private consumption of the man and woman This consisted of the health, education and clothing expenses of the man or woman.

Time

The model requires two time variables: agricultural labor and leisure.

Agricultural labor Agricultural labor was calculated as the total number of hours of agricultural work on the household's plots in the rainy and dry seasons of the past year, reported by the husband or wife. Where certain information was missing, such as the individual reported the number of days worked but not the number of hours per day, we used the village median for this information, where there were at least seven observations in the village. Otherwise, we used the district median.

Leisure In order to calculate leisure hours, we first required a measure of total available hours. As reported working hours are fairly low, leading to likely overestimates of true leisure time, we calculated total time available as the number of hours worked by the hardest working man or woman in the sample in the past year. This included both agricultural and wage labor and resulted in a value of 6120 hours. We assumed that this hardest worker works full-time and has zero leisure. We then calculated leisure for each individual as 6120 minus the annual hours of agricultural and wage labor of each individual.

Landholdings

In order to accurately measure the land income of individuals on divorce, we required exact information on the amount of land owned by each spouse. We defined land to be owned if it was inherited, granted by local leaders, part of a bride price, purchased with a title or purchased without a title. Land that was owned either solely by the spouse or owned by the spouse jointly with someone outside the household was assumed to accrue to that spouse on divorce. Land not owned by either spouse was assumed to disappear after divorce, while land owned jointly by the spouses was allowed to be endogenously split in the simulations.

Covariates in regressions

Here, we explain how the covariates in the regressions were defined. All covariates from the data are from the 2010 wave. The 2013 wave was only used to see whether the couple had divorced.

 # children This is the number of own or adopted children living in the household.

Age of man/woman This is the age of the man or woman in 2010.

Head education level This is a series of dummy variables that define the highest education level of the head, which ranges from no education to tertiary education.

Same age This is a dummy variable that equals one if the spouses have the same age ± 2 years, and equals zero otherwise.

Same education This is a dummy variable that equals one if the spouses have the same education level, and equals zero otherwise.

Negative income shocks The number of negative income shocks (events like unusually high prices for agricultural inputs, unusually high prices for food purchased, and unusually low prices for food sold) that a household was exposed to in the past year. It ranges from 0 to 11.

N churches This is the number of churches in the marriage market, as reported by village informants.

Distance to road, urban center This is the average distance to the nearest road or nearest urban center (Lilongwe, Zomba, or Blantyre) in kilometers, in the marriage market.

Sex ratio This is the ratio of men to women at the village level in the IHS sample, calculated based on the heads of household. Single-headed households count as one male or one female, while married households count as one male and one female.

Land This is the total number of acres of land owned by the household.

N households in marriage market This is the total number of households in that particular household's marriage market.

Marriage market literacy This is the fraction of adults who can read the local language.

Marriage market female primary education This is the fraction of female adults who have at least primary education.

Marriage market children in school This is the fraction of children who are currently enrolled in school.

Public/private share of consumption This is the share of public or private consumption in total consumption.

Nonlabor income (NLI) This is an output of the structural model and is the difference between total consumption and other inputs on the one hand and labor and land income on the other hand.

Land income This is an output of the structural model and gives the total number of acres of land owned by the spouse multiplied by the shadow price of land. It measures the annual rental yield on the land.

Wage This is an output of the structural model and gives the hourly shadow wage of agricultural labor of the husband or wife.

Appendix OA2: Example of instability network

In Figure 1, we illustrate the instability network of one particular cluster. Women are indexed *Wi* and men are indexed *Mi*, and we only display men and women who have blocking pairs in the cluster. Arrows depict these blocking pairs. In this cluster, $M18$, $M26$, and $W13$ are popular. $M18$ has a blocking pair with 27 women, meaning that he could be better off with any of these women than in his marriage, and each of these women would be better off with him. Similarly, $M26$ has a blocking pair with 13 women. $W13$ is the only woman with more than two blocking pairs: she has six. She can form a profitable blocking pair with $M1$, $M2$, $M17$, and $M30$, in addition to $M18$ and $M26$. However, she is best off with $M18$ (measured by the associated *BP* index). Similarly, $M18$ is best off with $W13$. The thick arrow depicts the fact that these two individuals are each others' favorite blocking pair: hence, they would both be best off divorcing their partners and marrying each other. The instability in this cluster is driven by $M18$, $M26$, and $W13$: if these three individuals were removed from the cluster, all marriages would be stable. The most likely explanation for the fact that these individuals have a large number of blocking pairs is that they are highly productive.

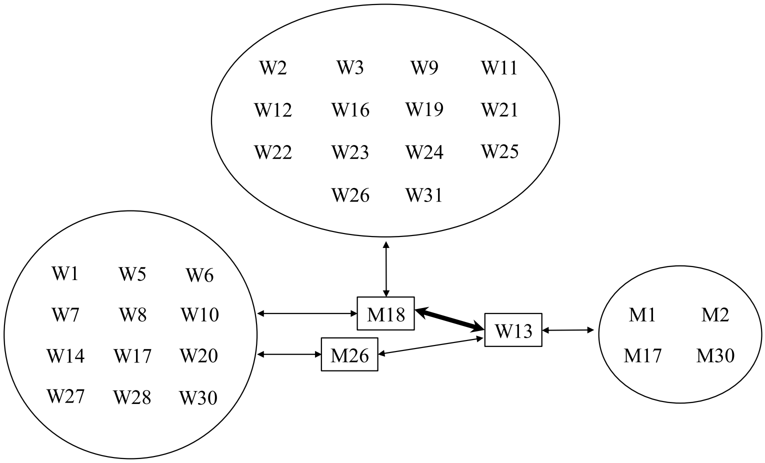

Figure 1. The instability network of one marriage market in our dataset. Arrows indicate profitable remarriage options; the thick arrow denotes the mutually best outside option (i.e., they are each others' BPmax).

Appendix OA3: Instability indices and household and marriage market characteristics

Table I. Shadow wages, land prices, and stability indices for alternative values of the tuning parameter α; correlations between results for $\alpha = 10^{-6}$ and $\alpha = 10^{-5}$, 10^{-4}, 10^{-3}, and 10^{-2}.

	$\alpha = 10^{-5}$	$\alpha = 10^{-4}$	$\alpha = 10^{-3}$	$\alpha = 10^{-2}$
Shadow land price	1	1	1	1
Shadow wage, woman	1	1	1	1
Shadow wage, man	1	1	1	0.993
BPmax, woman	1	1	0.995	0.796
BPmax, man	1	1	0.997	0.926
BPavg, woman	1	1	0.989	0.770
BPavg, man	1	1	0.994	0.831
IR, man	1	1	0.999	0.935

Table II. OLS regressions of instability indices in 2010 on household and marriage market characteristics in 2010.

	BPmax (w) (1)	BPavg (w) (2)	BPmax (m) (3)	BPavg (m) (4)	IR (m) (5)
Primary edu.	−0.470	−0.023	−0.213	−0.118	−0.004
	(0.212)	(0.009)	(0.179)	(0.092)	(0.002)
Secondary edu.	−0.547	−0.020	0.073	0.023	−0.004
	(0.241)	(0.011)	(0.186)	(0.098)	(0.002)
Tertiary edu.	−2.046	−0.063	0.005	0.233	−0.002
	(0.549)	(0.024)	(0.474)	(0.335)	(0.008)
Age (m)	−0.005	−0.000	0.013	0.007	0.000
	(0.010)	(0.000)	(0.007)	(0.004)	(0.000)
Age (w)	0.004	0.000	−0.016	−0.006	−0.000
	(0.011)	(0.000)	(0.008)	(0.005)	(0.000)
N children	−0.099	−0.004	−0.052	−0.020	−0.001
	(0.031)	(0.001)	(0.017)	(0.012)	(0.000)
Same age	0.007	0.006	0.116	0.086	0.003
	(0.205)	(0.010)	(0.131)	(0.075)	(0.002)
Same edu	−0.098	−0.010	−0.131	−0.115	−0.003
	(0.200)	(0.009)	(0.168)	(0.085)	(0.002)
MM dist. to road	−0.051	−0.002	−0.023	−0.005	−0.000
	(0.019)	(0.001)	(0.005)	(0.002)	(0.000)
MM dist. to urban centre	−0.003	−0.000	−0.001	−0.000	−0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)

(*Continues*)

TABLE II. *Continued.*

	BPmax (w) (1)	BPavg (w) (2)	BPmax (m) (3)	BPavg (m) (4)	IR (m) (5)
MM *N* churches	0.023	0.001	0.002	0.002	0.000
	(0.022)	(0.001)	(0.008)	(0.003)	(0.000)
MM *N* households	0.054	−0.001	0.030	−0.000	0.001
	(0.014)	(0.001)	(0.007)	(0.002)	(0.000)
MM Literacy	−0.594	0.036	0.652	0.024	0.010
	(1.361)	(0.059)	(0.476)	(0.212)	(0.008)
MM Female prim. edu.	0.303	−0.056	−0.028	0.126	0.002
	(1.503)	(0.065)	(0.709)	(0.236)	(0.011)
MM Children in school	−0.539	−0.043	−0.478	−0.068	−0.014
	(1.302)	(0.070)	(0.440)	(0.215)	(0.009)
N	5943	5943	5943	5943	5943
R^2	0.093	0.029	0.057	0.005	0.119

Note: Standard errors in parentheses. This table reports OLS regressions.

TABLE III. Partial R^2 for OLS regressions of instability indices in 2010 on household and marriage market characteristics in 2010.

	BPmax (w) (1)	BPavg (w) (2)	BPmax (m) (3)	BPavg (m) (4)	IR (m) (5)
Head education	0.002	0.002	0.001	0.001	0.001
Age of spouses	0.000	0.000	0.001	0.001	0.001
N children	0.002	0.001	0.001	0.001	0.004
Same education and age	0.000	0.000	0.000	0.001	0.001
MM Distances to road and urban	0.022	0.021	0.010	0.001	0.018
MM *N* churches	0.002	0.001	0.000	0.000	0.001
MM *N* households	0.026	0.002	0.021	0.000	0.046
MM Literacy	0.000	0.001	0.001	0.000	0.001
MM Female prim. edu.	0.000	0.001	0.000	0.000	0.000
MM Children in school	0.000	0.001	0.001	0.000	0.003
N	5943	5943	5943	5943	5943

Note: This table reports Partial R^2 values capturing the difference in R^2 when the full model is estimated in Table II and when one variable or set of variables is dropped from the full model. The rows in this table indicate which variables have been dropped: for example, Head Education implies that the dummy variables for the head's education level is not used.

APPENDIX OA4: FURTHER RESULTS

This appendix displays further results tables. Table V shows sensitivity of the main results to the inclusion of successive control variables. Table VI shows logit regressions of divorcing between 2010 and 2013 on instability indices and control variables, parallel to the results in Table 6 in the main text, with marginal effects evaluated at means. The estimated coefficients are very similar to those in the main table, although the coefficient on *BPmax* is not statistically significant. Next, Table VII replaces *BPmax* in Table 6 in the

TABLE IV. Partial R^2 for OLS regressions of instability indices in 2010 on household and marriage market characteristics in 2010 and budget components.

	BPmax (w) (1)	BPavg (w) (2)	BPmax (m) (3)	BPavg (m) (4)	IR (m) (5)
Head education	0.002	0.002	0.001	0.002	0.000
Age of spouses	0.000	0.000	0.001	0.001	0.000
N children	0.001	0.000	0.000	0.001	0.000
Same education and age	0.000	0.000	0.000	0.000	0.001
MM Distances to road and urban	0.014	0.012	0.006	0.000	0.007
MM *N* churches	0.003	0.001	0.000	0.000	0.000
MM *N* households	0.026	0.001	0.027	0.000	0.062
MM Literacy	0.000	0.001	0.002	0.000	0.002
MM Female prim. edu.	0.000	0.001	0.002	0.000	0.002
MM Children in school	0.000	0.000	0.000	0.000	0.002
Spouses' wages	0.019	0.021	0.028	0.031	0.076
Spouses' NLI	0.030	0.037	0.129	0.113	0.250
Spouses' land income	0.001	0.001	0.001	0.002	0.001
Fertilizer	0.004	0.004	0.001	0.000	0.004
N	5943	5943	5943	5943	5943

Note: This table reports Partial R^2 values capturing the difference in R^2 when the full model is estimated in Table II additionally including the components of the structural model, and when one variable or set of variables is dropped from the full model. The rows in this table indicate which variables have been dropped: for example, Head education implies that the dummy variables for the head's education level is not used.

TABLE V. OLS regressions of divorce between 2010–2013 on instability indices in 2010 and other control variables: sensitivity to controls.

	Divorced in 2013					
	(1)	(2)	(3)	(4)	(5)	(6)
BPmax (woman)	0.003	0.016	0.016	0.016	0.014	0.014
	(0.003)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
BPmax (man)	0.001	−0.004	−0.001	−0.004	0.001	0.001
	(0.007)	(0.031)	(0.031)	(0.030)	(0.030)	(0.030)
IR (man)	0.112	1.435	1.143	1.453	0.846	0.864
	(0.378)	(2.556)	(2.529)	(2.475)	(2.504)	(2.491)
Age (m)			−0.002	−0.002	−0.002	−0.001
			(0.002)	(0.002)	(0.002)	(0.002)
Age (w)			−0.001	−0.001	−0.001	−0.001
			(0.002)	(0.002)	(0.002)	(0.002)
Primary edu.				0.077	0.071	0.077
				(0.042)	(0.041)	(0.058)
Secondary edu.				0.018	0.013	0.018
				(0.037)	(0.037)	(0.054)

(*Continues*)

TABLE V. *Continued.*

	Divorced in 2013					
	(1)	(2)	(3)	(4)	(5)	(6)
Tertiary edu.				0.009	−0.003	0.001
				(0.092)	(0.094)	(0.101)
N children					−0.013	−0.013
					(0.006)	(0.006)
Same age						0.013
						(0.039)
Same edu.						0.006
						(0.040)
N	1406	1406	1406	1406	1406	1406
R^2	0.002	0.103	0.118	0.122	0.126	0.126

Note: Standard errors in parentheses. This table reports OLS regressions. Each column adds successive control variables. Columns (2) to (6) include marriage market fixed effects.

TABLE VI. Logit regressions of divorce between 2010–2013 on instability indices in 2010 and other control variables.

	Divorced in 2013	
	(1)	(2)
BPavg (woman)	0.018	
	(0.011)	
BPavg (man)	−0.000	
	(0.044)	
IR (man)	1.090	0.043
	(3.729)	(2.125)
BPmax (woman)		0.683
		(0.253)
BPmax (man)		0.019
		(0.029)
N	1021	1021

Note: Standard errors in parentheses. This table reports Logit regressions. All regressions include marriage market fixed effects, the age of the husband and wife in 2010, fixed effects for the education of the household head, the number of children the household had in 2010, and dummy variables indicating whether the couple are within 2 years of age of each other, and whether they have the same level of education. The sample size is lower than in the equivalent OLS specification (Table X) because some observations are dropped due to no variation in the outcome variable within a marriage market.

main text with the 95th percentile of an individual's distribution of blocking pairs (including zeros). The results are, again, similar. Tables VIII, IX, and X estimate the impact of the instability indices on remarriage versus being single. In particular, Table VIII estimates the effect of the average instability indices in an OLS regression, while Table IX

TABLE VII. OLS regressions of divorce between 2010–2013 on 95th percentile instability indices in 2010 and other control variables.

	Divorced in 2013 (1)
BP95 (woman)	0.046
	(0.015)
BP95 (man)	0.005
	(0.021)
IR (man)	0.441
	(1.683)
N	1406
R^2	0.129

Note: Standard errors in parentheses. This table reports OLS regressions. BP95 is the value of the 95th percentile of the individual's blocking pairs. All regressions include marriage market fixed effects, the age of the husband and wife in 2010, fixed effects for the education of the household head, the number of children the household had in 2010, and dummy variables indicating whether the couple are within 2 years of age of each other, and whether they have the same level of education.

TABLE VIII. OLS regressions of marital status in 2013 on average instability indices in 2010 and other control variables.

	Marital Status of Woman		Marital Status of Man	
	(1) *Remarried*	(2) *Single*	(3) *Remarried*	(4) *Single*
BPavg (woman)	0.275	0.303	0.397	0.092
	(0.079)	(0.130)	(0.128)	(0.062)
BPavg (man)	0.007	0.016	0.006	0.002
	(0.010)	(0.021)	(0.010)	(0.009)
IR (man)	−0.093	0.026	0.071	0.206
	(0.695)	(1.115)	(1.086)	(0.520)
N	1380	1380	1347	1347
R^2	0.114	0.110	0.125	0.094

Note: Standard errors in parentheses. This table reports OLS regressions. All regressions include marriage market fixed effects, the age of the husband and wife in 2010, fixed effects for the education of the household head, the number of children the household had in 2010 and dummy variables indicating whether the couple are within 2 years of age of each other, and whether they have the same level of education.

and X show relative risk ratios in multinomial logit regressions of the maximum and average indices, respectively. The results are generally consistent with the main text, although we find that *BPavg* significantly predicts divorcing and being single, in addition to remarriage, for the man. Further, the impact of the *IR* index, although insignificant, is very large in the logit model, and in the single outcomes in the multinomial models. This is indicative of insufficient variation in the IR index among divorced couples. Finally, Table XI replicates the heterogeneity analysis in Table 9 in the main text, but replacing the maximum indices with average indices. The results are similar, with significantly esti-

TABLE IX. Multinomial logit regressions of marital status in 2013 on instability indices in 2010 and other control variables.

	(1)—Marital Status of Man		(2)—Marital Status of Woman	
	Remarried	*Single*	*Remarried*	*Single*
BPmax (woman)	1.118	1.050	1.062	1.035
	(0.036)	(0.060)	(0.030)	(0.039)
BPmax (man)	1.174	1.010	0.914	1.019
	(0.106)	(0.131)	(0.108)	(0.073)
IR (man)	0.001	76.732	0.317	1047.921
	(0.005)	(785.978)	(1.897)	(5378.269)
N	1347		1380	

Note: Standard errors in parentheses. This table reports the odds ratios from multinomial logit regressions. All regressions include district fixed effects (there was insufficient variation in outcomes within marriage markets), the age of the husband and wife in 2010, fixed effects for the education of the household head, the number of children the household had in 2010 and dummy variables indicating whether the couple are within 2 years of age of each other, and whether they have the same level of education.

TABLE X. Multinomial logit regressions of marital status in 2013 on instability indices in 2010 and other control variables.

	(1)—Marital Status of Man		(2)—Marital Status of Woman	
	Remarried	*Single*	*Remarried*	*Single*
BPavg (woman)	15.361	4.445	2.772	5.455
	(10.551)	(7.004)	(1.543)	(4.976)
BPavg (man)	1.253	1.101	1.058	1.103
	(0.125)	(0.202)	(0.151)	(0.109)
IR (man)	0.002	1.236	0.001	38.309
	(0.014)	(10.895)	(0.007)	(193.271)
N	1347		1380	

Note: Standard errors in parentheses. This table reports the odds ratios from multinomial logit regressions. All regressions include district fixed effects (there was insufficient variation in outcomes within marriage markets), the age of the husband and wife in 2010, fixed effects for the education of the household head, the number of children the household had in 2010 and dummy variables indicating whether the couple are within 2 years of age of each other, and whether they have the same level of education.

mated negative effects of the number of children on the relationship between the *BP* indices and divorce. Put differently, the estimated positive effect of the *BP* indices on divorce is decreasing in the number of children that a household has in 2010.

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 14 January, 2019; final version accepted 20 October, 2020; available online 24 November, 2020.

TABLE XI. OLS regressions of divorce between 2010–2013 on average instability indices in 2010 and their interactions with other variables.

	Divorced in 2013		*Divorced in 2013*		*Divorced in 2013*
	(1)		(2)		(3)
BPavg (woman)	1.061	*BPavg* (woman)	0.509	*BPavg* (female)	−0.193
	(0.256)		(0.163)		(0.920)
Age**BPavg* (woman)	−0.018	Same age**BPavg* (woman)	0.124	Sex ratio**BPavg* (female)	0.801
	(0.005)		(0.252)		(1.019)
BPavg (man)	0.064	*BPavg* (man)	0.020	*BPavg* (man)	−0.009
	(0.030)		(0.026)		(0.057)
Age**BPavg* (man)	−0.001	Same age**BPavg* (man)	−0.015	Sex ratio**BPavg* (man)	0.030
	(0.000)		(0.015)		(0.060)
N	1406		1406		1406

	(4)		(5)
BPavg (woman)	0.567	*BPavg* (woman)	0.561
	(0.170)		(0.178)
Same educ**BPavg* (woman)	−0.062	# Children**BPavg* (woman)	−0.025
	(0.197)		(0.031)
BPavg (man)	0.003	*BPavg* (man)	0.035
	(0.027)		(0.027)
Same educ**BPavg* (man)	0.018	# Children**BPavg* (man)	−0.006
	(0.018)		(0.003)
N	1406		1406

Note: Standard errors in parentheses. This table reports OLS regressions. All regressions include marriage market fixed effects, the age of the husband and wife in 2010, fixed effects for the education of the household head and the number of children the household had in 2010, and dummy variables for whether the couple are within two years of age of each other, and whether they have the same level of education. They also include the variable that is being interacted with the indices (e.g., Sex ratio in regression (3)).