AMERICAN SOCIOLOGICAL REVIEW

OFFICIAL JOURNAL OF THE AMERICAN SOCIOLOGICAL ASSOCIATION

ONLINE SUPPLEMENT
to article in
AMERICAN SOCIOLOGICAL REVIEW, 2021, VOL. 86

Marriage and Masculinity: Male-Breadwinner Culture, Unemployment, and Separation Risk in 29 Countries

Pilar Gonalons-Pons
University of Pennsylvania

Markus Gangl
Goethe University Frankfurt am Main

Part A. DATA PREPARATION AND ANALYSIS SAMPLE

Data Harmonization

Our analyses use harmonized panel data from five major household surveys: the U.S. Survey of Income and Program Participation (SIPP), the European Union Statistics on Income and Living Conditions (EU-SILC), the German Socioeconomic Panel (GSOEP), the British Household Panel Survey (BHPS), and the Understanding Societies Survey (UKHLS).

We selected this set of panel data surveys because they contain the most high-quality longitudinal information on family income dynamics in the United States and Europe. The only similar multi-country longitudinal dataset that exists, the Cross-National Equivalent File (CNEF), includes data from eight countries, which is insufficient to estimate multilevel logistic regression models that require a larger sample size at the higher cluster level (Bryan and Jenkins 2015; Heisig and Schaeffer 2019). Other cross-national datasets, like the Luxembourg Income Study, are not longitudinal.

We use the GSOEP data for Germany instead of the EU-SILC German sample because the EU-SILC German data has not been released to the public (SOEP 2019). We use the BHPS and UKHLS data for the UK instead of the EU-SILC UK sample due to discrepancies in the reference period used to collect income information (Mack and Lange 2015; Eurostat 2019). The EU-SILC UK sample is the only EU-SILC sample that collects income information using the current year as a reference period instead of the previous year. To construct a measure of pervious year income using the EU-SILC UK data, we would need to drop one of the waves and further constrain the already short window of observation in our data.

The five longitudinal surveys we use to construct our analysis sample are remarkably similar. They are all nationally representative probability random samples of households that collect information on households' sociodemographic characteristics, employment, and economic conditions. The basic follow-up rules are also the same across surveys. All panel surveys follow original sample people (OSM) during the life of the panel or rotation sample. OSM are respondents recruited the first year the sample starts. All surveys follow persons moving into households with OSM as long as they live with OSM. Children of OSM become part of the "original sample people" when they become 16. For more information on additional characteristics of each survey, see section "About the Surveys" below.

There are two important differences in the structure of these longitudinal surveys that we harmonized to build our analysis sample: (a) panel length and sample rotation, (b) interview schedule.

The EU-SILC and SIPP have a sample rotation panel structure and each respondent is followed for a maximum of four to six years. The GSOEP, BHPS, and UKHLS have a simple longitudinal structure and each respondent is followed for the entire duration of the survey. The GSOEP is now one of the longest running longitudinal survey in Europe, and the original sample has now been followed for over 30 years.

We harmonized the panel length and sample rotation structure across surveys. We adapted all surveys to follow the EU-SILC four-year rotating panel structure, which is the most constringent structure and thus offers a maximum common denominator template. We did this in two ways. First, for surveys with rotating panels where respondents are eligible to be followed for more than four waves (this applies to some EU-SILC samples and the SIPP), we restricted all respondents to four observations only. Second, for longer surveys without rotating panels, we created a sample rotation structure. This applied to the GSOEP and BHPS; we did not do this for the UKHLS because this survey only included four waves of data at the time of this study. To replicate the EU-SILC overlapping sample rotation, we split the sample into equal rotation groups and assigned them different start dates (four rotation groups for GSOEP and two for BHSP). When the rotation sample ends after four waves, respondents' observations are reused for new rotation samples. For instance, GSOEP rotation group 1 starts in 2004 and is followed until 2007, and rotation group 2 starts in 2005 and is followed until 2008; respondents in rotation group 1 and rotation group 2 can enter new rotation groups after 2007 and 2008, respectively.

The interview schedule also varies across these longitudinal surveys. The SIPP has a quarterly data collection, and the remaining surveys follow an annual interview schedule. We harmonized the SIPP to mirror the other surveys by collapsing the quarterly data into an annual file, utilizing the quarterly data to construct annual measures on employment and income corresponding to the other surveys.

About the Surveys

The **EU-SILC** is a 31-country longitudinal survey including all 28 EU member states plus Norway, Iceland, and Switzerland. The EU-SILC is organized under a "common framework" that aims to maximize comparability of the information produced. The common framework defines: (a) harmonized lists of target variables, (b) recommendations for sample design, (c) common criteria for imputation, weighting, and sampling error calculation, and (d) common concepts and classifications. Using this common framework, the national statistical agencies of each country are responsible for identifying the sample and carrying out the survey. EU-SILC data collected by national statistical agencies comes largely from surveys, but some countries use a combination of survey and data from administrative registers. Register countries are Denmark, Finland, Iceland, Norway, and Sweden (Eurostat 2013, 2017, 2019; Jäntti et al. 2013).

The majority of countries participating in the EU-SILC follow a four-year rotating panel structure recommended by Eurostat. This sample design is based on four subsamples or rotation groups that start in different years and are followed for four waves. The rotation groups overlap and a new sample for each rotation group is obtained every four years. For instance, rotation group 1 starts in 2004 and is followed until 2007, and rotation group 2 starts in 2005 and is followed until 2008. New samples for rotation group 1 and 2 start in 2008 and 2009, respectively. A few countries deviate from this structure and use longer rotation panels. For instance, France and Norway follow rotation groups for eight and nine years, respectively (Eurostat, 2013, 2017, 2019).

The **SIPP** is a longitudinal household survey in the United States composed of a series of continuous and independent multi-year panels that began in 1984. We use the 2004 and 2008 panels comprising two samples of respondents followed for up to four and five years, respectively. The SIPP follows a quarterly interview schedule; every member is interviewed every four months (U.S. Census Bureau 2001, 2015).

The **GSOEP** is a household longitudinal survey in Germany that started in 1984 and follows respondents once per year. It was originally a representative sample of West Germany, and in 1990 it expanded to include the former DGR and become representative of the German residential population. In addition to the original West Germany sample and the former GDR extension, the survey includes one refresher sample that started in 1998 and several samples on specific populations (e.g., immigrants or high-income individuals). All respondents recruited through these additional samples become OSM and are followed using the same criteria.

The **BHPS** is a household longitudinal survey in the United Kingdom that started in 1991 with a sample of 5,000 households, including approximately 10,000 individual interviews. This survey incorporated additional refresher and special samples over time. All respondents recruited through these additional samples became Original Sample Members and are followed using the same criteria. The BHPS study officially ended in 2008, but respondents were invited to continue their participation in Understanding Society (UKHLS), the new longitudinal survey that substituted the BHPS. About 85 percent of BHPS participants accepted and were incorporated into the UKHLS (ISER 2019a; Taylor et al. 2010).

The **UKHLS** is a household longitudinal survey in the United Kingdom that started in 2009 and follows respondents once per year. It consists of a large general population sample, initially including 47,520 addresses; the survey also includes other sample components, such as the Ethnic Minority Boost sample and the Innovation Panel. At Wave 2, the BHPS members were incorporated to the sample (ISER 2019b; Knies 2014).

Analysis Sample

We construct an unbalanced sample of couples at risk of separation during the panel. To identify this analytic sample, we begin with a core sample of women age 15 to 60 at the beginning of the panel and select those who are observed in a union and matched to their partners during the panel. After matching spouses, we obtain the analytic sample of couples at risk of separation. Table S1 shows how we obtain this sample in each survey.

Couples can enter the sample at any point during the panel. For instance, an original sample person who moves in with her partner at wave two is incorporated into our analysis sample at wave 2. We conducted robustness checks using a semi-balanced panel that restricts the sample to respondents who are cohabiting or married at wave 1 and our results do not change.

Couples remain in our analytic sample as long as they stay together, and they exit the sample when they separate, when they attrite out of the survey, or when their panel ends (right-censored), whichever comes first. Table S1 summarizes the percent of separations, attrition, and right-censored couples in our sample.

Table S2 reports additional descriptive statistics for variables used in our analysis; this complements Table 2 in the main text. Our data are unweighted because each survey uses a different approach to calculate weights, and our analyses control for the things typically used to construct those weights (Winship and Radbill 1994).

Table S1. Sample Selection

	(1) Women age 15-60 during panel	(2) Women age 15-60 with completed information on marital and relationship status	(3) Women age 15-60 in union	(4) Women age 15-60 in union and matched to partner	(5) Proportion of initial sample in union and matched with partner	(6) % separation	(7) % attrition	(8) Average observations	(9) % missing at least one covariate	(10) Final analysis sample
EU-SILC	474907	473073	298222	283754	59.7	4.3	15.9	3.2	9.0	283754
	1331031	1323859	840093	801106	60.2				6.5	748639
SIPP	54263	54263	38986	35618	65.6	5.4	15.1	3.3	0	35618
	198745	198745	141462	125030	62.9				0	125030
GSOEP	29799	29748	21137	18653	62.6	6.9	14.6	3.4	10.0	18653
	87985	87830	60232	53514	60.8				11.8	47192
BHPS	7732	7732	5102	4481	58.0	10.2	10.1	3.5	8.5	4481
	33764	33746	21665	18664	55.3				6.7	17417
UKHLS	25353	25349	14275	13391	52.8	7.3	17.0	3.3	1.9	13391
	77871	77821	43718	40864	52.5				2.3	39922
TOTAL	592054	590165	377722	355897	60.1	6.8	14.5	3.3	7.9	355897
	1729396	1722001	1107170	1039178	60.1				5.9	978200

Data sources: EU-SILC, SIPP, GSOEP, BHPS, UKHLS.

Note: **(5)** The proportion of the initial sample in union and matched with partner is obtained by calculating the percent of women in the initial sample who are in unions and matched to partners, or (4) / (1) * 100. **(6)** The % separation is calculated as the number of couples who separate during the survey over all women in unions and matched to partners. **(7)** The % attrition is calculated as the number of respondents who exit the panel before it ends and not due to separation or right-censoring for other reasons (i.e., becoming no longer eligible to be in the analytic sample due to age) over all women in unions and matched to partners. **(8)** The average observations is the average number of years couples are observed in the analysis sample. **(9)** The % missing at least one covariate lists item response missing data in at least one of the covariates used in our models: employment status, age, education, income, children, and household ownership. **(10)** The final couple-level sample is the same as listed in (4), and the final couple-year sample is obtained by subtracting observations with item response missing data, or (4) - (9).

Table S2. Sample Descriptive Statistics for Additional Variables Not Reported in Table 2

	N couples	*N* couple years	His Unemployment		Her Unemployment		Her income	His income	Household tenure	Have children
			Ever	At first interview	Ever	At first interview				
Pooled sample	355897	978200	10.7	5.5	11.8	6.0	23059.5	12277.3	75.7	70.9
AT	9285	24701	6.9	3.6	7.6	3.3	28941.2	13367.6	60.8	65.4
BE	8830	23272	8.0	5.2	10.0	6.5	27384.5	15573.8	73.6	65.7
BG	6123	18548	22.9	14.9	25.1	17.0	1845.0	1225.6	85.9	75.4
CY	6028	18399	13.6	5.3	13.3	6.0	20947.3	10314.8	77.9	77.0
CZ	10416	33657	6.9	3.3	11.6	6.0	8140.6	4207.4	79.0	71.0
DE	18653	47192	10.3	5.8	9.6	4.9	29342.5	11403.4	52.8	85.4
DK	9300	25367	4.7	1.6	7.2	2.9	43235.5	29863.8	81.6	65.1
EE	6899	19902	15.2	7.4	12.5	5.6	5537.5	3091.9	86.1	74.5
EL	9137	26193	12.4	6.8	16.3	8.7	14130.4	6687.3	74.0	75.3
ES	20924	58085	20.3	11.5	25.2	13.4	15494.0	7877.3	80.1	75.5
FI	12755	35472	8.9	4.7	8.6	4.1	28672.0	18912.7	78.6	63.8
FR	11511	45728	11.6	5.5	14.0	6.8	23945.1	14064.8	60.1	65.5
HU	13517	38447	13.3	7.0	14.3	7.7	4023.3	2526.2	88.4	72.4
IE	5326	12202	16.8	11.7	6.3	2.8	35433.7	16471.6	76.9	75.4
IS	5438	14442	5.7	2.5	5.3	2.2	25595.6	13945.5	84.1	77.0
IT	20818	82928	9.2	4.4	13.2	6.3	22947.5	10850.4	68.6	77.5
LT	2946	9099	16.9	8.9	15.4	7.8	3859.8	2793.1	91.5	73.0
LU	7060	24459	8.2	3.4	9.6	4.1	38137.2	18767.4	63.1	70.3
LV	6580	18302	23.4	12.6	19.7	9.9	3856.8	2568.0	82.9	72.2
NL	16877	46649	3.5	1.1	3.1	.5	39827.3	17206.4	84.3	67.5
NO	8681	23367	3.0	1.2	4.6	1.9	47856.5	27291.0	85.9	70.3
PL	17239	53652	11.3	6.5	18.0	11.4	5323.0	3052.2	71.8	77.6
PT	1371	2772	12.7	8.7	14.4	10.5	10330.6	6308.3	75.2	69.2
RO	6978	20228	5.5	3.4	2.9	1.3	1811.1	1033.5	95.8	49.8
SE	9862	25667	5.9	2.7	7.2	3.7	28902.8	19111.1	73.7	66.8
SI	19455	49857	13.1	7.7	17.7	10.8	11587.9	9375.2	83.2	83.9
SK	2573	8002	9.3	5.3	14.5	9.2	5426.6	3250.4	87.8	82.4
UK	17872	57339	11.6	6.7	8.5	3.5	26574.6	14583.5	70.1	52.2
US	35618	125030	11.1	3.8	10.5	3.8	35517.6	17910.1	74.6	67.7

Data sources: EU-SILC, SIPP, GSOEP, BHPS, UKHLS.

Part B. SENSITIVITY ANALYSES

Sensitivity to Different Measures of Male-Breadwinner Values

Our results could be invalidated by measurement error if key findings were sensitive to the way we constructed the male-breadwinner-norms variable. This variable uses items from two different surveys, the European Social Survey (ESS) and the International Social Survey Programme (ISSP). ESS asks respondents whether they agree/disagree that "men should have more right to jobs than women when jobs are scarce." ISSP asks respondents whether they agree/disagree that "men's job is to earn money, women's job to look after home."

Table S3 shows descriptive statistics for the ESS and ISSP male-breadwinner values items, including information on data availability (years of data available for each country from each dataset) and individual-level variability in the variable of interest in each survey.

The ESS-ISSP combination we use to construct our model maximizes country coverage and the number of countries with the same item. The ESS and ISSP items are remarkably similar, but some notable differences exist. In general, there is greater agreement with the ISSP item than with the ESS item, which is not surprising given the difference in statements.

We checked the robustness of the cross-level interaction to different ways of constructing the male-breadwinner-norms variable.

First, we ran analyses using only ESS countries. This reduces the number of countries in our sample from 29 to 27, excluding the United States and Latvia, which have no ESS data. Table S4, Model S1, shows the cross-level interaction finding is robust.

Second, we ran analyses swapping ESS values for ISSP values for countries that are in both surveys. There are 23 countries that have both ISSP and ESS values; the correlation between the two male-breadwinner items is high, .74. Table S2, Model S2, shows the cross-level interaction is robust.

Finally, we ran models adding cross-level interactions for other gender attitudes that are not as directly linked to norms about men's employment, using items from both ISSP and ESS surveys. Table S4, Models 3 to 7, show the cross-level interaction between husbands' unemployment and male-breadwinner values remains robust. These analyses show no statistically significant cross-level interactions between husbands' unemployment and other gender attitudes items. This result reinforces our confidence that the male-breadwinner measure captures attitudes about men's employment and emphasizes the importance of choosing concrete items to measure different dimensions of context-level gender norms (Knight and Brinton 2017).

Sensitivity to Unemployment Transitions

Our measure of unemployment is based on respondents' status at the time of the interview. For a subset of countries, it is possible to use monthly employment calendar information to restrict the measure of unemployment to those who lost jobs and not include those who become unemployed after a period of schooling or inactivity.

We ran analyses on this subset of countries (24 countries, all except Denmark, Finland, Norway, Sweden, and the Netherlands) and confirm that results are robust. Table S5 shows results for the cross-level interaction model.

Sensitivity to Confounders

Table S6 presents analyses that test the sensitivity of our conclusion to more stringent criteria to control for potential sources of unobserved heterogeneity at the individual and country levels.

Model S9 presents results from Mundlak correction, which address the concern that our estimate of the effect of unemployment on divorce might be confounded with individual-level unobserved traits that correlate with unemployment incidence and separation. The Mundlak correction adds individual-level averages for all time-varying variables in an effort to control for time-fixed unobserved heterogeneity. Running this model in our data is not ideal because we only have a few observations per couple, and the average estimate absorbs a lot of the time-variation we observe in the short window of observation. Results show that the cross-level interaction between his time-varying unemployment measure and the context-level male-breadwinner values remains statistically significant. The main effects for men's and women's unemployment are not statistically significant in this specification, which is not surprising given that we only have four years of data per couple.

Models S10 and S11 present results for regressions with country fixed-effects and with country fixed-effects interacted with men's unemployment. These models address the concern that our measure of male-breadwinner norms is picking up unobserved variation in other country-level characteristics, such as institutions (Giesselmann and Schmidt-Catran 2019). Model S10 shows the moderating effect of male-breadwinner norms remains robust when only time-varying within-country variation is leveraged. Model S11 shows the moderating effect of male-breadwinner norms is not driven by country-specific effects for his unemployment. The results reinforce the robustness of our findings.

Model S12 presents results with year fixed-effects, which address the concern that changes in male-breadwinner values could be capturing a larger time trend in correlated unobserved processes shared by all countries included in the analysis. Our results are robust to this specification too.

Finally, Models S13, S14, and S15 exclude three outliers, Greece (EL), Poland (PL), and Netherlands (NL). The outliers are selected based on patterns in Figure 1. All analyses show that our key findings remain robust.

Sensitivity to Lacking Data on Union Duration

The EUSILC does not include information on marital parity or duration, and we cannot incorporate these control variables in our analyses. Our models include women's age as a proxy for couples' duration; but is age a good enough proxy for marital duration? The results would be severely biased if marital duration mediated or moderated the association between men's unemployment and divorce, potentially shifting the interaction with male-breadwinner norms. We assessed this source of potential bias in two ways.

First, we ran analyses for a subsample of young couples (women under age 46) and confirmed that results are robust. Table S7, Model S16, shows results for the cross-level interaction with male-breadwinner norms.

Second, we used marital duration information in the SIPP sample and tested whether the effect of men's unemployment systematically varied across marital duration. Table S7, Model S17, shows no evidence of an interaction between men's unemployment and marital duration. Table S7, Models S18 to S20, show the coefficient for men's unemployment does not vary substantially across models that control for women's age only, marital duration only, or both.

Third, we tested the interaction between his age (as a proxy for marital duration) and unemployment to address the possibility that union duration moderates the relationship between unemployment and separation and may confound our key estimates of interest. We fully recognize that using age as a proxy for union duration is far from perfect, but for the purpose at hand, one should not fail to notice that respondent age and union duration are strongly correlated (r = .67 in the SIPP), and thus age can serve as a useful control in our study, in terms of capturing true age (or birth-cohort effects) and as a proxy control to indirectly capture some of the true duration or (marriage/union cohort) effects and to net out the respective correlations with unemployment incidence. Table S7 reports results showing that our findings remain robust. Unlike the results with SIPP data reported above, the interaction between men's unemployment and age is statistically significant and the effect is positive (which implies potential for downward bias, based on the simulation results presented below). In any case, the interaction between men's age and unemployment does not alter any of the key coefficients of interest, namely the relationship between (men's) unemployment and separation and the interaction between (men's) unemployment and male-breadwinner norms.

Sensitivity to the Structure of the Model

Our three-level models include the random slope parameters for country-level controls and cross-level interactions at the country level, even though our country-level measures vary within countries across years. This modeling specification responds to substantive and computation reasons. We model the interaction at the country level because most of the variation in male-breadwinner norms is between rather than within countries. This specification is also computationally more efficient than similar specifications including random slopes, at the country and country-year levels. Furthermore, to keep the model setup reasonably parsimonious and computationally feasible, our preferred specification includes random slope parameters for all variables involved in a cross-level interaction, as well as for covariates where previous studies have established an association with relationship outcomes that varies across countries, namely cohabitation and women's education.

To demonstrate that our results are not sensitive to this particular model structure or list of random slope parameters, we present robustness tests from analyses that change the structure of the model and that include a more expanded set of random slope parameters. Table S8 presents various versions of our preferred model (Model 6 from Table 3) and shows that all key patterns are robust to changes in the structure of the model and in the list of random slope parameters. Model S25 shows our results are robust to including random slopes at the country-year and country levels, Model S26 shows our results are robust to including additional random slope parameters, and Model S27 shows our results are robust to estimating a two-level model, instead of a three-level, model.

Table S3. Male-Breadwinner Macro-Level Indicators, Comparing ESS and ISSP Data

Country	BWN			ESS BWN			ISSP BWN						ISSP-ESS
	% agree men's primary role is breadwinning (ESS data for all countries except US and LT using ISSP data)			% who agree with the statement: "Men should have more right to jobs than women when jobs are scarce"			% who agree with the statement: "Men’s job is to earn money, women’s job to look after home"						
	%	Min	Max	Years	SD in individual data (1)	Education gap (2)	%	Min	Max	Years	SD in individual data (1)	Education gap (2)	Difference
AT	18.5	16.0	23.6	04;08	1.2	9.9	26.0	25.0	28.7	02;05;12	1.0	18.7	7.4
BE	24.0	20.0	33.2	04;08	1.2	19.0	21.9	16.5	24.0	02;05;12	1.0	18.7	–2.1
BG	33.5	33.5	33.5	08	1.6	16.0	36.0	36.0	36.0	02;05;12	1.0	9.8	2.6
CY	39.9	39.9	39.9	08	1.3	21.2	30.1	30.1	30.1	02;05	.8	25.4	–9.8
CZ	28.2	25.6	34.8	04;08	1.5	11.5	40.4	37.4	41.4	02;05;12	1.1	12.6	12.2
DE	19.5	16.4	24.1	04;08	1.2	13.7	19.3	17.8	19.8	02;05;12	1.0	12.6	–.2
DK	5.3	3.3	9.7	04;08	.9	4.7	12.1	6.8	13.2	02;05;12	.8	7.7	6.8
EE	22.4	17.5	34.9	04;08	1.3	10.8							
EL	47.3	47.2	47.3	04;08	1.3	20.6							
ES	20.1	17.4	28.5	04;08	1.3	18.2	22.3	17.9	24.3	02;05;12	.9	21.6	2.3
FI	7.1	5.7	11.8	04;08	1.0	5.4	10.5	9.0	11.4	02;05;12	.7	5.7	3.4
FR	18.4	14.8	29.2	04;08	1.3	17.5	18.1	12.2	21.6	02;05;12	.9	10.5	–.3
HU	40.9	35.6	58.8	04;08	1.4	14.5	40.9	39.2	44.7	02;05;12	1.1	19.7	.1
IE	14.0	10.2	21.6	04;08	1.0	11.1	15.6	13.2	16.8	02;05;12	.8	10.1	1.6
IS	13.8	13.8	13.8	04	1.3	13.0	6.7	6.7	6.7	12	.6	5.4	–7.1
IT	26.2	26.2	26.2	04	3.3	22.5							
LT	33.0	33.0	33.0				33.0	33.0	33.0	12	.9	5.8	
LU	26.9	26.9	26.9	04	1.3	19.0							
LV	20.1	20.1	20.1	08	1.4	8.5	37.1	36.0	39.6	02;05;12	1.1	5.1	17.0
NL	13.4	10.4	20.1	04;08	1.0	10.2	12.3	12.2	12.6	02;05;12	.8	10.1	–1.1
NO	5.4	4.5	9.1	04;08	.8	4.0	7.4	5.1	9.3	02;05;12	.7	6.7	2.0
PL	30.9	27.5	40.3	04;08	1.4	23.5	42.7	42.5	43.3	02;05;12	1.1	28.4	11.9
PT	27.1	22.4	37.8	04;08	1.3	14.7	29.8	24.6	33.2	02;05;12	1.0	28.8	2.7
RO	33.3	33.3	33.3	08	1.7	12.9							
SE	4.3	2.8	7.5	04;08	.8	3.4	7.0	5.7	7.5	02;05;12	0.6	5.2	2.8
SI	18.7	16.9	23.5	04;08	1.2	17.1	26.6	20.6	28.8	02;05;12	1.0	22.8	7.8
SK	30.4	29.9	31.9	04;08	1.3	14.1	42.4	42.2	44.4	02;05;12	1.1	6.0	12.0
UK	16.6	12.5	25.1	04;08	1.1	12.3	17.9	12.6	19.6	02;05;12	.9	14.1	1.3
US	23.0	21.6	23.3				23.0	21.6	23.3	02;05;12	1.0	11.3	

Data sources: ESS and ISSP.
Note: SD = Standard Deviation. (1) SD in individual data is calculated using the original measure used in the survey and including all respondents. In ESS, the original measure ranges from 1 to 5 (5 indicates more egalitarian). In the ISSP, the original measure ranges from 1 to 4 (4 means more egalitarian). (2) Education gap is calculated as the difference between the percent of support for the male-breadwinner statement among respondents without a college degree and among respondents with a college degree. A value of 9, for instance, means the percentage of support for male-breadwinner among those without college is 9 percentage points higher than among those with a college degree.

Table S4. Sensitivity to Male-Breadwinner Measures: Multi-Level Logistic Regression on the Annual Probability of Separation

VARIABLES	Model S1 ESS countries only	Model S2 swapping ESS values for ISSP	Model S3 ISSP item 1: warm	Model S4 ISSP item 2: suffer	Model S5 ISSP item 3: contribute	Model S6 ESS item 1: men home	Model S7 ESS item 2: women cut
Women's unemployment	.124***	.144***	.133***	.143***	.143***	.102*	.103*
	(.037)	(.038)	(.038)	(.038)	(.037)	(.041)	(.041)
Men's unemployment	.323***	.297***	.295***	.298***	.294***	.327***	.337***
	(.036)	(.034)	(.034)	(.035)	(.034)	(.041)	(.040)
Male-breadwinner values	.007	.002	.012*	.011*	.011*	.012*	.013
	(.005)	(.005)	(.005)	(.006)	(.005)	(.006)	(.007)
# W Unemp	-.000	-.001	.000	.003	.001	-.002	-.001
	(.004)	(.003)	(.004)	(.005)	(.004)	(.004)	(.005)
# M Unemp	.009**	.006**	.007*	.008**	.008**	.010**	.010**
	(.004)	(.002)	(.003)	(.003)	(.003)	(.003)	(.003)
ISSP warm			-.004				
			(.007)				
# M Unemp			.007				
			(.005)				
ISSP suffer				.001			
				(.00383)			
# M Unemp				-.001			
				(.003)			
ISSP contribute					.002		
					(.004)		
# M Unemp					-.001		
					(.003)		
ESS men home						.009	
						(.008)	
# M Unemp						.008	
						(.006)	
ESS women cut							-.005
							(.005)
# M Unemp							-.002
							(.003)
Wives' earnings	.001	.003	.007	.001	.001	.001	.000
	(.001)	(.008)	(.007)	(.001)	(.001)	(.001)	(.001)
Husbands' earnings	-.003***	-.032***	-.034***	-.003***	-.003***	-.002***	-.003***
	(.001)	(.006)	(.006)	(.001)	(.001)	(.001)	(.001)
Wives' education							
secondary	-.027	.001	.011	.011	.011	-.047	-.030
	(.028)	(.025)	(.025)	(.025)	(.025)	(.029)	(.030)
college	-.155***	-.179***	-.164***	-.165***	-.165***	-.183***	-.153***
	(.034)	(.032)	(.032)	(.032)	(.032)	(.036)	(.036)
Husbands' education							
secondary	.056*	-.004	.005	.005	.005	.053	.047
	(.027)	(.025)	(.025)	(.025)	(.025)	(.029)	(.029)
college	-.048	-.136***	-.120***	-.121***	-.121***	-.053	-.050
	(.034)	(.032)	(.032)	(.032)	(.032)	(.036)	(.036)
Cohabitation	1.515***	1.601***	1.636***	1.636***	1.636***	1.510***	1.481***
	(.023)	(.022)	(.022)	(.022)	(.022)	(.025)	(.025)

Household tenure	-.410***	-.407***	-.419***	-.419***	-.418***	-.420***	-.447***
	(.023)	(.022)	(.022)	(.022)	(.022)	(.025)	(.025)
Constant	.0751	.118	.163	.163	.157	.498	-.422
	(.444)	(.438)	(.419)	(.419)	(.419)	(.468)	(.476)
Random intercepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random slopes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	798,636	842,175	767,605	767,605	767,605	798,636	798,636
Number of groups	272	282	244	244	244	272	272

Note: All models also control for women's age (quadratic), women's and men's inactivity, incomes (logged), and educational levels, parental status, cohabitation, household tenure, GDP, UR, UGEN, GWG, WLFP. Models also include interactions with each macro-level item and women's unemployment, we omit these coefficients due to space and because they are not statistically significant. ISSP item 1 warm = "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work"; ISSP item 2 suffer = "A preschool child is likely to suffer if his or her mother works"; ISSP item 3 contribute = "Both the man and woman should contribute to the household income"; ESS item 1 = "Men should take as much responsibility as women for home and children"; ESS item 2 = "Women should be prepared to cut down on paid work for sake of family."

Standard errors in parentheses. *** p<.001, ** p<.01, * p<.05

VARIABLES	Model S1 ESS countries only	Model S2 swapping ESS values for ISSP	Model S3 ISSP item 1: warm	Model S4 ISSP item 2: suffer	Model S5 ISSP item 3: contribute	Model S6 ESS item 1: men home	Model S7 ESS item 2: women cut
Women's unemployment	.124***	.144***	.133***	.143***	.143***	.102*	.103*
	(.037)	(.038)	(.038)	(.038)	(.037)	(.041)	(.041)
Men's unemployment	.323***	.297***	.295***	.298***	.294***	.327***	.337***
	(.036)	(.034)	(.034)	(.035)	(.034)	(.041)	(.040)
Male-breadwinner values	.007	.002	.012*	.011*	.011*	.012*	.013
	(.005)	(.005)	(.005)	(.006)	(.005)	(.006)	(.007)
# W Unemp	-.000	-.001	.000	.003	.001	-.002	-.001
	(.004)	(.003)	(.004)	(.005)	(.004)	(.004)	(.005)
# M Unemp	.009**	.006**	.007*	.008**	.008**	.010**	.010**
	(.004)	(.002)	(.003)	(.003)	(.003)	(.003)	(.003)
ISSP warm			-.004				
			(.007)				
# M Unemp			.007				
			(.005)				
ISSP suffer				.001			
				(.00383)			
# M Unemp				-.001			
				(.003)			
ISSP contribute					.002		
					(.004)		
# M Unemp					-.001		
					(.003)		
ESS men home						.009	
						(.008)	
# M Unemp						.008	
						(.006)	
ESS women cut							-.005
							(.005)
# M Unemp							-.002

							(.003)
Wives' earnings	.001	.003	.007	.001	.001	.001	.000
	(.001)	(.008)	(.007)	(.001)	(.001)	(.001)	(.001)
Husbands' earnings	-.003***	-.032***	-.034***	-.003***	-.003***	-.002***	-.003***
	(.001)	(.006)	(.006)	(.001)	(.001)	(.001)	(.001)
Wives' education							
secondary	-.027	.001	.011	.011	.011	-.047	-.030
	(.028)	(.025)	(.025)	(.025)	(.025)	(.029)	(.030)
college	-.155***	-.179***	-.164***	-.165***	-.165***	-.183***	-.153***
	(.034)	(.032)	(.032)	(.032)	(.032)	(.036)	(.036)
Husbands' education							
secondary	.056*	-.004	.005	.005	.005	.053	.047
	(.027)	(.025)	(.025)	(.025)	(.025)	(.029)	(.029)
college	-.048	-.136***	-.120***	-.121***	-.121***	-.053	-.050
	(.034)	(.032)	(.032)	(.032)	(.032)	(.036)	(.036)
Cohabitation	1.515***	1.601***	1.636***	1.636***	1.636***	1.510***	1.481***
	(.023)	(.022)	(.022)	(.022)	(.022)	(.025)	(.025)
Household tenure	-.410***	-.407***	-.419***	-.419***	-.418***	-.420***	-.447***
	(.023)	(.022)	(.022)	(.022)	(.022)	(.025)	(.025)
Constant	.0751	.118	.163	.163	.157	.498	-.422
	(.444)	(.438)	(.419)	(.419)	(.419)	(.468)	(.476)
Random intercepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random slopes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	798,636	842,175	767,605	767,605	767,605	798,636	798,636
Number of groups	272	282	244	244	244	272	272

Notes: All models also control for women's age (quadratic), women's and men's inactivity, incomes (logged), and educational levels, parental status, cohabitation, household tenure, GDP, UR, UGEN, GWG, WLFP. Models also include interactions with each macro-level item and women's unemployment, we omit these coefficients due to space and because they are not statistically significant. ISSP item 1 warm = "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work"; ISSP item 2 suffer = "A preschool child is likely to suffer if his or her mother works"; ISSP item 3 contribute = "Both the man and woman should contribute to the household income"; ESS item 1 = "Men should take as much responsibility as women for home and children"; ESS item 2 = "Women should be prepared to cut down on paid work for sake of family."

Standard errors in parentheses. *** p<.001, ** p<.01, * p<.05

Table S5. Sensitivity to Unemployment Definition: Multi-Level Logistic Regression on the Annual Probability of Separation

VARIABLES	Model S8 Using employment calendar data
Women's unemployment	.140***
	(.040)
Men's unemployment	.274***
	(.036)
Male-breadwinner values	.009
	(.005)
# W Unemp	.001
	(.005)
# M Unemp	.012**

	(.004)
Women's earnings (logged)	.008
	(.007)
Men's earnings (logged)	-.034***
	(.006)
Women's education	
secondary	.005
	(.026)
college	-.194***
	(.033)
Men's education	
secondary	.043
	(.025)
college	-.092**
	(.033)
Cohabitation	1.620***
	(.023)
Constant	-1.079**
	(.443)
Random intercepts	Yes
Random slopes	Yes
Observations	762,790
Number of groups	241

Notes: Models also control for women's age (quadratic), parental status, women's and men's inactivity, household tenure, GDP, UR, UGEN, GWG, WLFP.

Standard errors in parentheses. *** p<.001, ** p<.01, * p<.05

Table S6. Sensitivity to Other Specifications; Multi-Level Logistic Regression on the Annual Probability of Separation: Mundlak Correction, Country Fixed Effects, Year Fixed Effects, Omitting Recession Years, and Omitting Outliers

VARIABLES	Model S9 Mundlak correction	Model S10 country fixed effects	Model S11 country fixed effects + MU interacted with country dummies	Model S12 year fixed effects	Model S13 w/o recession years	Model S14 excluding EL	Model S145 excluding PL	Model S16 excluding NL
Women's unemployment	-.069	.207***	.209***	.145***	.220***	.138***	.137***	.132***
	(.060)	(.037)	(.037)	(.037)	(.056)	(.036)	(.036)	(.036)
Men's unemployment	-.022	.274***	.115	.325***	.326***	.302***	.290***	.305***
	(.059)	(.034)	(.205)	(.045)	(.054)	(.032)	(.032)	(.032)
Male-breadwinner values	.021**	.045***	.044***	.029***	.017	.011*	.009*	.005
	(.008)	(.009)	(.009)	(.004)	(.010)	(.005)	(.005)	(.005)
# W Unemp	-.001	.001	.001	.001	.010	.002	.001	.001
	(.004)	(.004)	(.004)	(.004)	(.006)	(.004)	(.004)	(.004)
# M Unemp	.009**	.008**	.018*	.012**	.013**	.009**	.008**	.011**
	(.003)	(.003)	(.009)	(.005)	(.005)	(.003)	(.003)	(.004)
Cohabitation	1.653***	1.609***	1.610***	1.636***	1.678***	1.646***	1.645***	1.640***
	(.022)	(.021)	(.021)	(.021)	(.030)	(.021)	(.021)	(.021)
Household tenure	-.416***	-.409***	-.408***	-.403***	-.380***	-.408***	-.400***	-.409***
	(.021)	(.021)	(.021)	(.021)	(.029)	(.021)	(.021)	(.021)
Mean W Unemp	.472***							
	(.0747)							
Mean H Unemp	.444***							
	(.0735)							
Constant	.226	-.442	-.383	.309	-1.479***	-.167	-.0935	-.207
	(.420)	(.456)	(.457)	(.423)	(.574)	(.402)	(.404)	(.403)
Random intercepts	Yes	Yes (year-level only)	Yes (year-level only)	Yes (country-level only)	Yes	Yes	Yes	Yes
Random slopes	Yes	Yes (year-level only)	Yes (year-level only)	Yes (country-level only)	Yes	Yes	Yes	Yes
Observations	978200	978200	978200	978200	488748	952007	924548	961323
Number of groups	292	292	292	292	258	276	276	276

Note: All models also control for women's age (quadratic), parental status, women's and men's inactivity and education levels and incomes (logged), household tenure, GDP, UR, UGEN, GWG, WLFP. Model S9 includes time-varying and time-fixed variables for his and her income. In Models S10 and S11 country dummies are omitted, and Model S11 also omits interactions between men's unemployment and country dummies. Random intercepts at the country and year levels. Random slopes for men's and women's unemployment, college education, and cohabitation. Coefficients for the random components are omitted in the interest of space.
Standard errors are clustered at the country and country-year levels. *** p<.001, ** p<.01, * p<.05

Table S7. Sensitivity to Omission of Marital Duration: Multi-Level Logistic Regression on the Annual Probability of Separation

VARIABLES	Model S17 young couples	Model S18 SIPP interaction mardur*M Unemp	Model S19 SIPP age only	Model S20 SIPP mardur only	Model S21 SIPP age and mardur	Model S22 M age#M Unemp	Model S23 M age#M Unemp M edu#M Unemp
Women's unemployment	.083	.175	.235**	.176	.174	.198***	.199***
	(.044)	(.105)	(.084)	(.104)	(.105)	(.037)	(.037)
Men's unemployment	.289***	.293**	.315***	.286**	.292**	.346***	.322***
	(.036)	(.095)	(.077)	(.095)	(.095)	(.041)	(.043)
Male-breadwinner values	.008					.016*	.016*
	(.005)					(.007)	(.007)
# W Unemp	.001					.004	.004
	(.004)					(.004)	(.004)
# M Unemp	.010**					.008**	.009**
	(.003)					(.003)	(.003)
Wives' age	-.586***	-.179	-.200*		-.186	-.0171***	-.0171***
	(.084)	(.117)	(.079)		(.117)	(.00197)	(.00197)
Wives' age^2	.016***	.004	.005*		.005		
	(.003)	(.003)	(.002)		(.003)		
Marital duration		-.037***		-.118***	-.039***		
		(.004)		(.023)	(.004)		
# H Unemp		-.010					
		(.008)					
Marital duration^2				.004**			
				(.001)			
Men's age						-.008***	-.008***
						(.002)	(.002)
# H Unemp						.006*	.006*
						(.003)	(.003)
Husbands' education							
secondary	.013	-.226**	-.0137	-.227**	-.226**	-.001	-.002
	(.027)	(.0939)	(.081)	(.094)	(.094)	(.023)	(.023)
college	-.120***	-.564***	-.290***	-.569***	-.563***	-.215***	-.228***
	(.035)	(.111)	(.096)	(.111)	(.111)	(.031)	(.031)

# H Unemp							.212*
							(.097)
Cohabitation	1.474***	2.630***	2.275***	2.659***	2.629***	1.720***	1.720***
	(.024)	(.062)	(.048)	(.063)	(.062)	(.022)	(.022)
Household tenure	-.387***	-.333***	-.397***	-.344***	-.335***	-.414***	-.414***
	(.023)	(.058)	(.050)	(.057)	(.058)	(.021)	(.021)
Constant	2.801***	-.736	-1.016	-3.228***	-.639	-5.036***	-5.030***
	(.857)	(1.501)	(.961)	(.116)	(1.498)	(.083)	(.083)
Random intercepts	Yes	No	No	No	No	Yes	Yes
Random slopes	Yes	No	No	No	No	Yes	Yes
Observations	536,544	83,713	88,417	83,713	83,713	978,200	978,200
Number of groups	282	N/A	N/A	N/A	N/A	292	292

Notes: All models include controls for women's and men's education levels, incomes (logged), household tenure, and parenthood status. Model S16 also includes controls for macro-level variables GDP, UR, UGEN, GWG, and WLFP. Random intercepts at the country and year levels. Random slopes for men's and women's unemployment, college education, and cohabitation. Coefficients for the random components are omitted in the interest of space.
Standard errors are clustered at the country and country-year levels. *** $p<.001$, ** $p<.01$, * $p<.05$

Table S8. Sensitivity to the Structure of the Model

VARIABLES	Model S24 Model 5 from Table 3 (preferred model)	Model S25 Model with random slopes at both levels	Model S26 Model with additional random slopes	Model S27 Two-level model with additional random slopes
Women's unemployment	.224***	.225***	.227***	.225***
	(.038)	(.038)	(.038)	(.038)
Men's unemployment	.332***	.353***	.330***	.329***
	(.036)	(.048)	(.048)	(.038)
Male-breadwinner values	.033**	.029**	.030**	.014*
	(.010)	(.010)	(.010)	(.007)
# W Unemp	.004	.005	.004	.004
	(.004)	(.004)	(.004)	(.004)
# M Unemp	.012**	.009*	.008*	.007*
	(.004)	(.004)	(.004)	(.003)
Women's earnings (logged)	.023***	.019***	.022***	.020***
	(.004)	(.006)	(.004)	(.004)
Men's earnings (logged)	-.020***	-.026***	-.020***	-.025***
	(.004)	(.006)	(.004)	(.005)
Women's education				
secondary	.011	.009	.009	.004
	(.025)	(.025)	(.025)	(.024)
college	-.139***	-.138***	-.141***	-.166***
	(.040)	(.040)	(.041)	(.034)
Men's education				
secondary	.008	.006	.006	.006
	(.024)	(.024)	(.024)	(.024)
college	-.200***	-.194***	-.203***	-.202***
	(.031)	(.038)	(.031)	(.033)
Women's inactivity	.036	.038	.033	.031
	(.033)	(.034)	(.033)	(.034)
Men's inactivity	.184***	.184***	.189***	.181***
	(.041)	(.042)	(.042)	(.042)
Cohabitation	1.789***	1.787***	1.739***	1.587***
	(.137)	(.136)	(.134)	(.052)
Household tenure	-.398***	-.395***	-.397***	-.403***
	(.021)	(.021)	(.021)	(.021)
Have young child	-.127***	-.125***	-.125***	-.128***
	(.021)	(.021)	(.021)	(.021)
Women's age	-.255***	-.261***	-.259***	-.251***
	(.035)	(.035)	(.035)	(.035)
# Women's age	.007***	.007***	.007***	.006***
	(.001)	(.001)	(.001)	(.001)
GDP	-.005	-.005	-.006	-.001
	(.005)	(.005)	(.005)	(.003)
GWG	.058***	.057***	.057***	.058***
	(.015)	(.015)	(.015)	(.010)

WLFP	-.945	-.637	-.912	-.977
	(1.021)	(1.016)	(.997)	(.757)
UGEN	.010*	.012**	.010*	.009***
	(.004)	(.004)	(.004)	(.003)
UR	.057*	.056*	.050	.042
	(.027)	(.027)	(.026)	(.031)
Constant	-1.434***	-1.433***	-1.425***	-1.295***
	(.431)	(.432)	(.438)	(.423)
Random part				
Level 2 (country-years)				
Constant	.476	.412	.481	.580
	(.063)	(.059)	(.063)	(.088)
Women's education college		.003		.016
		(.010)		(.012)
Women's unemployment		.000		.000
		(.000)		(.000)
Men's unemployment		.009		.008
		(.013)		(.013)
Cohabitation		.128		.461
		(.026)		(.060)
Men's education college				.011
				(.010)
Women's age				.000
				(.000)
Women's inactivity				.000
				(.000)
Men's inactivity				.031
				(.027)
Women's earnings (logged)				.000
				(.000)
Men's earnings (logged)				.000
				(.000)
Level 3 (countries)				
Constant	.311	.327	.277	
	(.107)	(.109)	(.118)	
Women's education college	.010	.010	.008	
	(.006)	(.007)	(.005)	
Women's unemployment	.000	.000	.000	
	(.000)	(.000)	(.000)	
Men's unemployment	.018	.016	.014	
	(.018)	(.018)	(.018)	
Cohabitation	.495	.455	.503	
	(.143)	(.138)	(.147)	
Men's education college			.006	
			(.005)	
Women's age			.000	
			(.000)	
Women's inactivity			.036	
			(.019)	

Men's inactivity			.038	
			(.025)	
Women's earnings (logged)			.000	
			(.000)	
Men's earnings (logged)			.000	
			(.000)	
Respondents	978200	978200	978200	978200
Number of level-2 groups (country-years)	292	292	292	292
Number of level-3 groups (countries)	29	29	29	N/A

Note: Standard errors are clustered at the country and country-year levels. *** p<.001, ** p<.01, * p<.05

Part C. SUPPLEMENTARY TESTS FOR DIFFERENCES IN KEY COEFFICIENTS OF INTEREST

Figures S1 to S6 plot Wald tests of statistical significance for differences in estimates of interest. Figure S1, for instance, shows tests for differences in men's unemployment coefficients across all levels of male-breadwinner norms; with Panel A testing average marginal effects (AMEs) and Panel B testing logistic marginal coefficients. We show results from key regression models presented in the main text. Tests are performed at different levels of male-breadwinner norms as the baseline, as indicated in the legend (at 0%, 5%, 10%, 15%, and 20% levels of support for male-breadwinner norms). For example, black dots correspond to tests against BWN = 0, or no support for male-breadwinner norms, meaning the black dot at 45 on the *x*-axis tests the difference in men's unemployment AME between a context with no support for male-breadwinner norms and a context where 45% of the population supports male-breadwinner values. All estimates are shown with 95% confidence intervals.

Figure S1 shows differences between men's unemployment AMEs across levels of male-breadwinner norms are statistically significant, especially when support for male-breadwinner norms is high. Results are generally consistent across models and across metrics (AME versus logistic coefficients).

Figure S2 tests for differences between men's and women's unemployment effects across male-breadwinner contexts. It shows men's unemployment effects are generally higher than women's, and those differences are usually, although not always, statistically significant. Differences between men's and women's unemployment effects are larger in contexts with high support for male-breadwinner norms.

Figure S3 tests for differences between men's unemployment estimates across male-breadwinner norms using the quadratic specification. It shows differences are accentuated, particularly in contexts with high levels of support for the male-breadwinner norm.

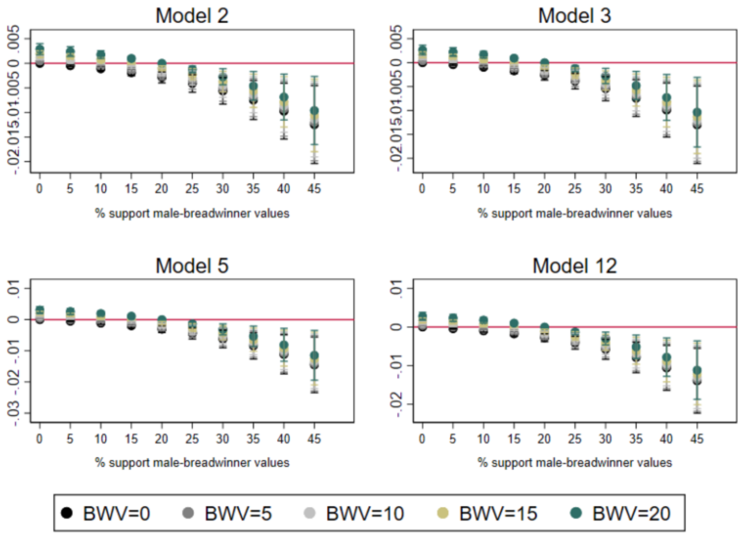
Figure S4 tests for differences between men's and women's unemployment effects across male-breadwinner contexts using the quadratic specification.

Figure S5 shows tests for differences between men's unemployment effects by union type and across male-breadwinner contexts. It shows variation in men's unemployment effects across male-breadwinner values is only statistically significant among married couples, not among cohabiting couples.

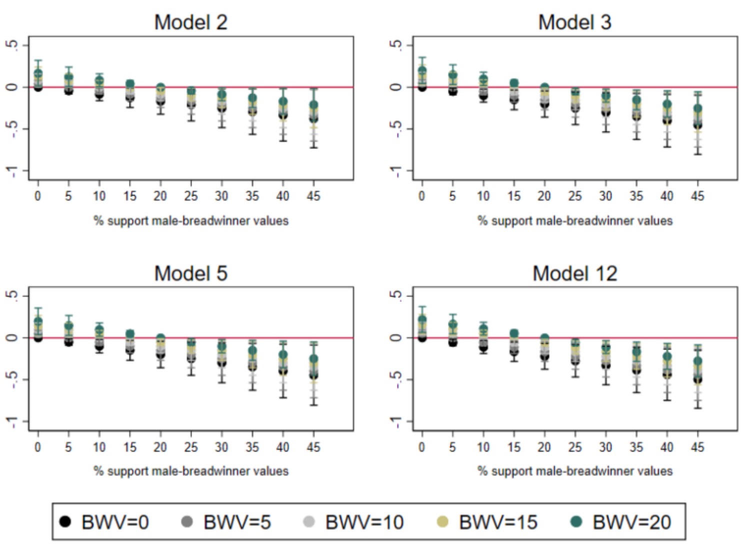
Figure S6 shows tests for differences between men's and women's unemployment effects by union type across male-breadwinner contexts. It shows the difference between men's and women's unemployment effects is only statistically significant among married couples, not among cohabiting couples.

Figure S1. Wald Tests for Differences in Men's Unemployment Estimates across Male-Breadwinner Norms Contexts, Selected Models

Panel A. Average Marginal Effects



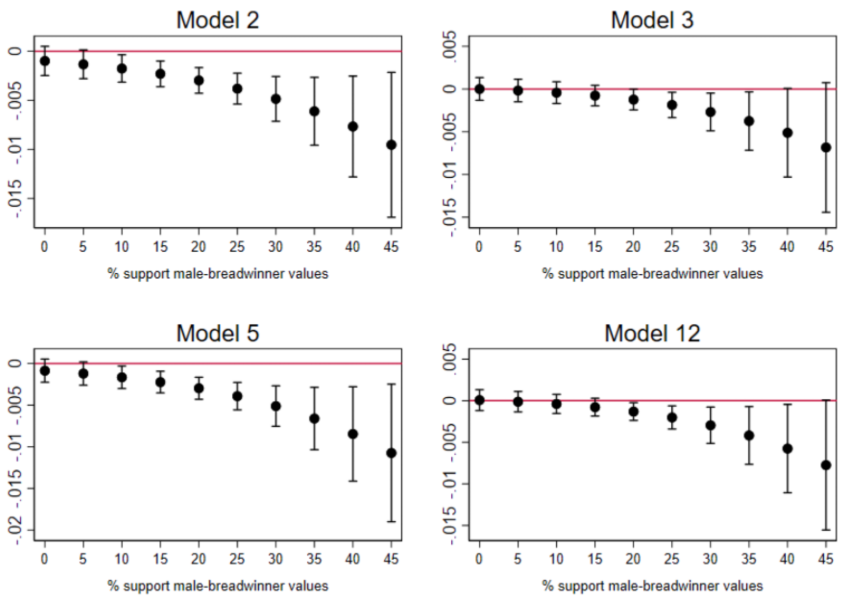
Panel B. Logistic Marginal Coefficients



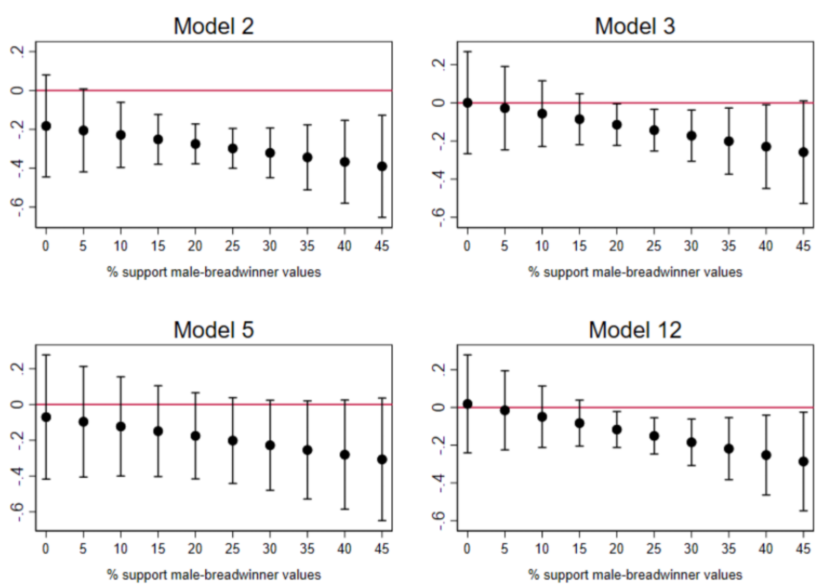
Note: Figures summarize two-sided Wald tests for differences in men's unemployment estimates across all levels of male-breadwinner norms; with Panel A presenting average marginal effects (AMEs) and Panel B presenting logistic coefficients. Each panel shows tests using different levels of male-breadwinner norms as the baseline. For instance, BWN = 0 tests differences between men's unemployment estimate in a context where nobody supports the male-breadwinner model and men's unemployment estimate across all other levels of male-breadwinner norms on the *x*-axis. The BWV = 0 point at value 45 on the *x*-axis = men's unemployment at 0 male-breadwinner norms minus men's unemployment at 45% male-breadwinner norms. Tests plot 95% confidence intervals.
Data sources: EU-SILC, SIPP, GSOEP, BHPS, UKHLS.

Figure S2. Wald Tests for Differences between Men's and Women's Unemployment Effects across Male-Breadwinner Contexts, Selected Models

Panel A. Average Marginal Effects



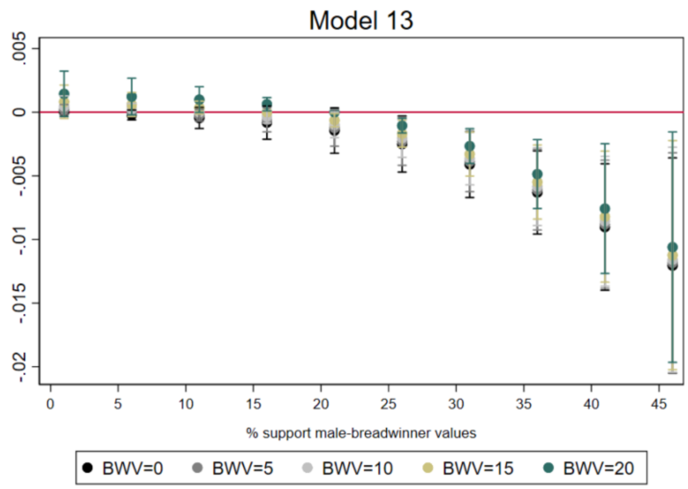
Panel B. Logistic Marginal Effects



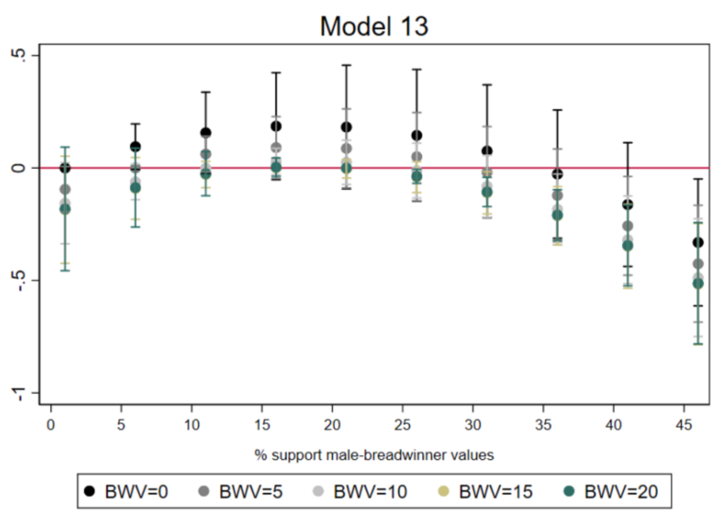
Note: Figures summarize two-sided Wald tests for differences in men's unemployment and women's unemployment estimates across all levels of male-breadwinner norms; with Panel A presenting average marginal effects (AMEs) and Panel B presenting logistic coefficients. Each panel shows tests using different levels of male-breadwinner norms as the baseline. For instance, point estimates at value 20 on the *x*-axis tests differences between men's and women's unemployment estimate in a context where 20% of the population supports the male-breadwinner model. Tests plot 95% confidence intervals.
Data sources: EU-SILC, SIPP, GSOEP, BHPS, UKHLS.

Figure S3. Wald Tests for Differences in Men's Unemployment Effects across Male-Breadwinner Contexts, Quadratic Specification

Panel A. Average Marginal Effects



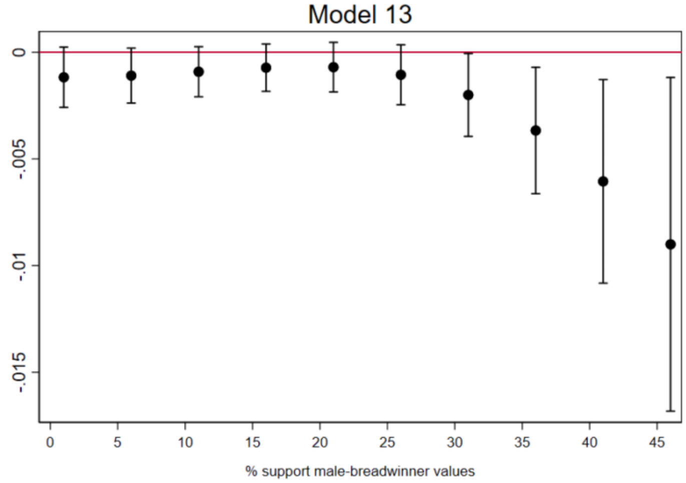
Panel B. Logistic Marginal Effects



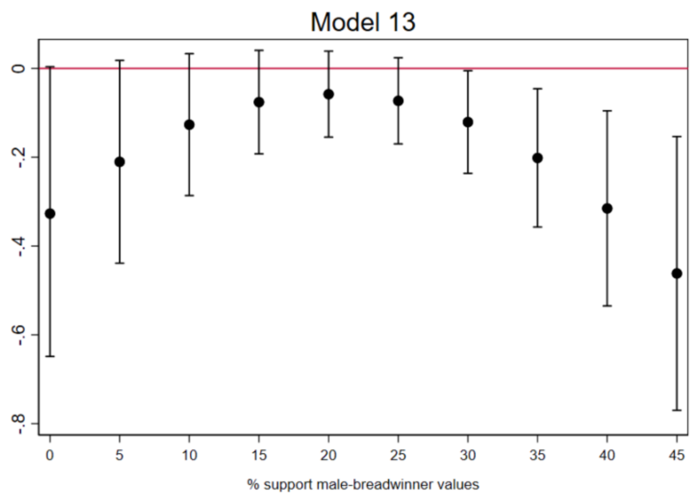
Note: Figures summarize two-sided Wald tests for differences in men's unemployment estimates across all levels of male-breadwinner norms; with Panel A presenting average marginal effects (AMEs) and Panel B presenting logistic coefficients. Each panel shows tests using different levels of male-breadwinner norms as the baseline. For instance, BWN = 0 tests differences between men's unemployment estimate in a context where nobody supports the male-breadwinner model and men's unemployment estimate across all other levels of male-breadwinner norms on the *x*-axis. The BWW = 0 point at value 45 on the *x*-axis = men's unemployment at 0 male-breadwinner norms minus men's unemployment at 45% male-breadwinner norms. Tests plot 95% confidence intervals.
Data sources: EU-SILC, SIPP, GSOEP, BHPS, UKHLS.

Figure S4. Wald Tests for Differences between Men's and Women's Unemployment Effects across Male-Breadwinner Contexts, Quadratic Specification

Panel A. Average Marginal Effects



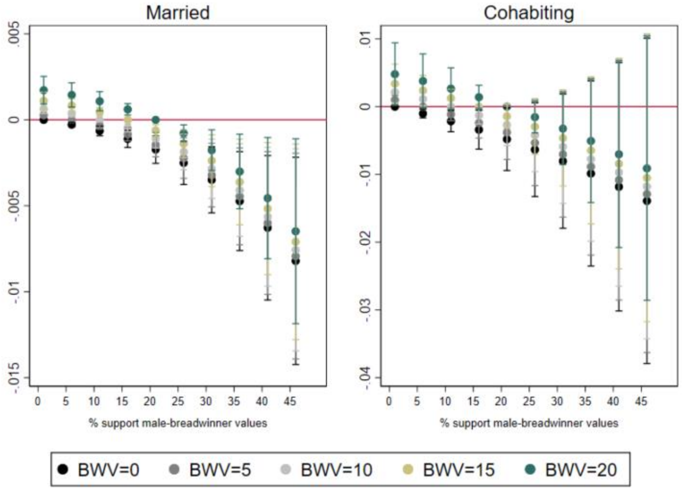
Panel B. Logistic Marginal Effects



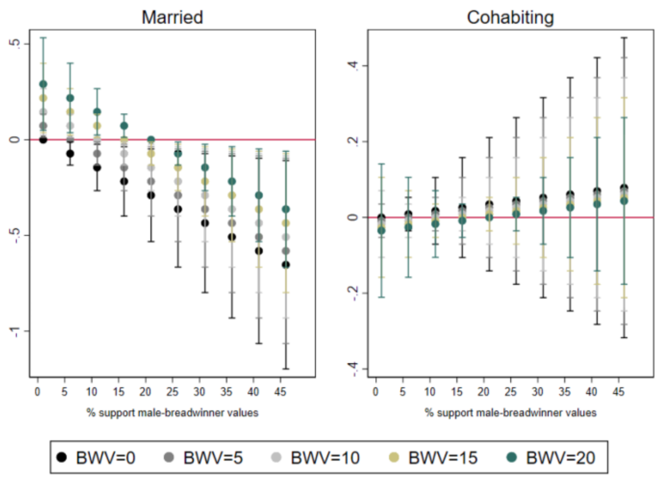
Note: Figures summarize two-sided Wald tests for differences in men's unemployment and women's unemployment estimates across all levels of male-breadwinner norms; with Panel A presenting average marginal effects (AMEs) and Panel B presenting logistic coefficients. Each panel shows tests using different levels of male-breadwinner norms as the baseline. For instance, point estimates at value 20 on the *x*-axis tests differences between men's and women's unemployment estimate in a context where 20% of the population supports the male-breadwinner model. Tests plot 95% confidence intervals.
Data sources: EU-SILC, SIPP, GSOEP, BHPS, UKHLS.

Figure S5. Wald Tests for Differences in Men's Unemployment Effects across Male-Breadwinner Contexts, by Type of Union

Panel A. Average Marginal Effects



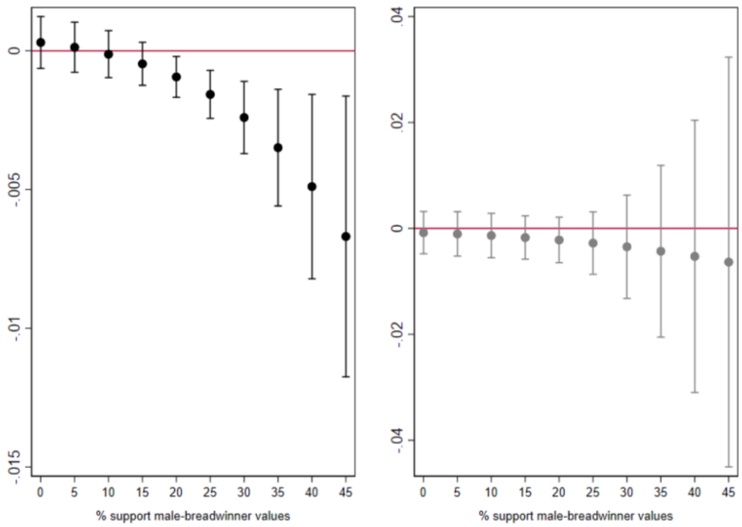
Panel B. Logistic Marginal Effects



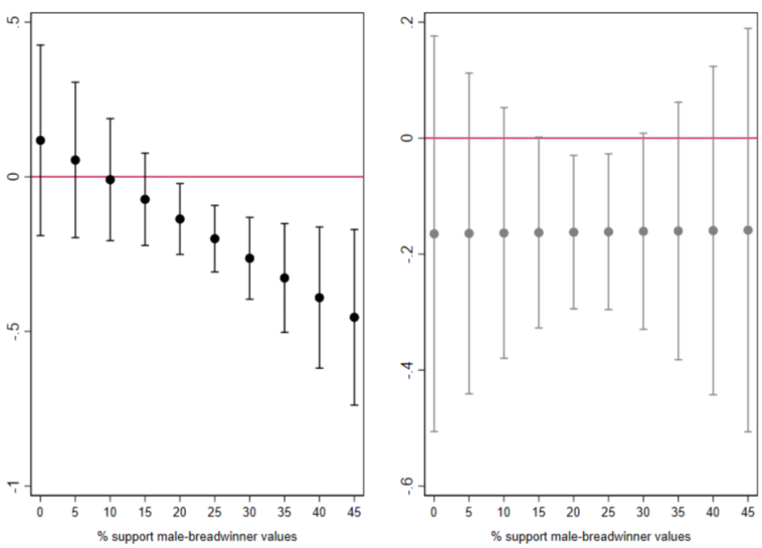
Note: Figures summarize two-sided Wald tests for differences in men's unemployment estimates across all levels of male-breadwinner norms; with Panel A presenting average marginal effects (AMEs) and Panel B presenting logistic coefficients. Each panel shows tests using different levels of male-breadwinner norms as the baseline. For instance, BWN = 0 tests differences between men's unemployment estimate in a context where nobody supports the male-breadwinner model and men's unemployment estimate across all other levels of male-breadwinner norms on the *x*-axis. The BWW = 0 point at value 45 on the *x*-axis = men's unemployment at 0 male-breadwinner norms minus men's unemployment at 45% male-breadwinner norms. Tests plot 95% confidence intervals.
Data sources: EU-SILC, SIPP, GSOEP, BHPS, UKHLS.

Figure S6. Wald Tests for Differences between Men's and Women's Unemployment Effects across Male-Breadwinner Contexts, by Type of Union

Panel A. Average Marginal Effects



Panel B. Logistic Marginal Effects



Note: Figures summarize two-sided Wald tests for differences in men's and women's unemployment estimates across all levels of male-breadwinner norms; with Panel A presenting average marginal effects (AMEs) and Panel B presenting logistic coefficients. Each panel shows tests using different levels of male-breadwinner norms as the baseline. For instance, point estimates at value 20 on the *x*-axis tests differences between men's and women's unemployment estimate in a context where 20% of the population supports the male-breadwinner model. Tests plot 95% confidence intervals.
Data sources: EU-SILC, SIPP, GSOEP, BHPS, UKHLS.

Part D. SIMULATION STUDY TO EVALUATE POTENTIAL OMITTED VARIABLE BIAS FROM UNMEASURED DURATION DEPENDENCE

Our dataset does not have information on the start date of marital and cohabiting unions, a shortcoming known in the event history analysis literature as "left-truncation with unknown start dates" (Guo 1993:219ff.). As is usual with many household panel surveys, the constituent surveys used to construct the cross-nationally comparative dataset are based on representative samples of respondents drawn at a particular point in time (or indeed at several points in time in case of the EU-SILC rotating panel survey architecture), who are then followed up in (typically) annual interviews over at least several years. From the perspective of analyzing separation processes in marriages and consensual unions, the resulting sample is a stock sample where all existing unions enter the survey observation window at the point of the initial sampling, irrespective of the elapsed union duration prior to joining the survey. Relative to an inflow sample or cohort design that would permit observation of marriage (or union formation) cohorts from their common starting point, unions are only observed conditional on being sufficiently long-lasting to become included in the stock sample, and then also contribute data selectively for only those periods of the consensual union that happen to coincide with the survey observation window and that may represent earlier or later stages of the relationship, depending on union duration at the point of entry to the survey. The corresponding left-truncation of the sample, occasionally also described as "late entry to the risk set" (Singer and Willett 2003:595ff.), creates an "initial conditions" problem (Heckman and Singer 1986) that is conventionally resolved by conditioning the analysis on process time at entry to the survey observation period (cf. Guo 1993:228ff.).

Owing to the fact that the EU-SILC lacks questions from which marriage or union start dates could be recovered, our analysis is faced with the additional problem of "left-censoring" (in the terminology of Singer and Willett 2003:319; Tuma and Hannan 1984; Yamaguchi 1991) or "unknown start dates" (in the terminology of Guo 1993). When start dates are unknown, it is of course impossible to condition the analysis on union duration at entry to the survey, and the only remaining option is to estimate the empirical model under the assumption of a time-constant baseline hazard function (cf. Guo 1993:222ff.).[1] In the present context, this requirement is conceptually unattractive because all available empirical evidence confirms that the separation hazard is not time-constant but exhibits negative duration dependence or some bell-, hump-, or sickle-shaped (inverted-U) pattern over process time. The assumption of a time-constant baseline hazard that is implied by the incomplete data available to the present analysis is therefore patently false in a descriptive sense. However, given that the inferential goal of our article is neither in recovering the baseline hazard itself, nor in providing a complete model of separation processes, but in isolating the effect of (men's and women's) unemployment experiences on separation risk, the main question is whether the lack of information on union start dates is likely to create

[1] As an alternative, Guo recommends restricting the sample to spells where the starting date is actually observed because the start date occurs after the initial stock sampling. This general alternative is conceptually unattractive in our specific study given the unfavorable ratio between the very short four-year observation window in our comparative dataset and the ideal length of the observation window in a (marriage) cohort study that aims for a near complete observation of the union separation process. Given that the four-year window would translate into a maximum length of observation of about 2.5 years for any newly formed union, restricting the sample to these couples would drastically redefine the substantive hypothesis being tested, because we would only be estimating the effect of (men's or women's) unemployment among recently formed couple relationships in their first two years of being together (the maximum identifiable case being a respondent who is without a partner or with an earlier partner at survey wave 1, who is observed as newly partnered by survey wave 2, and who is then being re-interviewed in two further follow-ups at survey waves 3 and 4). However, as the parameter that is identified clearly differs between the two analyses, conducting Guo's alternative test is non-informative in the present study because a difference in estimates could always be due to either our main analysis being non-robust or to there being significant treatment effect heterogeneity in separation risks and the effect of either partner's unemployment or cross-national differences therein being very different among recently formed couples relative to their older peers (e.g., due to genuine cohort or life-cycle effects).

a particular type of omitted variable bias in our estimates. This is a legitimate concern because union duration is well-known to be among the important determinants of separation risk, and because it is straightforward to presume our estimates must be affected by some bias due to unobserved "initial conditions" (and union duration obviously prime among these) for the marriages and consensual unions in our sample.

In the main text and in several supplementary analyses we offered a range of empirical reasons and robustness checks that suggest our results are not likely to be critically affected by omitted variable bias despite the non-ideal database we are drawing on. However, to also make a more principled argument why left-truncation without known start dates may have less severe consequences for our estimates than standard intuition about omitted variable bias might suggest, we conducted a small-scale simulation analysis on the joint impact of left-truncation (a.k.a. late entry to the risk set) and left-censoring (i.e., unknown start dates). We sought to mimic some broad features of our actual data in the simulation, for example, in terms of baseline separation risk, likely patterns of duration dependence in the baseline hazard, or the magnitude of the treatment effect of men's unemployment on separation risk. We deliberately cast the simulation in the form of an underlying separation process driven by an event history model in continuous time, where the resulting data are incompletely observed and where the main workhorse in the data analysis is a logistic model for the transition probability (rather than the hazard rate in continuous time). We explore the role of different shapes of the baseline hazard and different degrees of duration dependence in the process, and we address the difference between the case where the generative process conforms to a proportional hazards (PH) assumption and the case of time-varying effects of unemployment that result in a non-PH process.

In line with standard intuition, our simulation demonstrates that left-censoring (i.e., the inability to allow for duration dependence and the necessity to work with a model that implies a time-constant baseline hazard) creates bias in parameter estimates. Our simulation evidence also underscores two very important qualifications that need to be added to this general intuition, and that imply the (perhaps counterintuitive) result that the estimates we report in the main text are likely to be conservative, that is, are likely to exhibit mild downward bias, if anything. The first of these qualifications relates to the fact that while left-censoring implies a potential for bias, the direction of bias depends on the type of covariate concerned. For time-constant covariates, left-censoring implies upward bias in the estimates, conforming to standard intuition. For time-varying covariates, however, left-censoring creates downward bias in parameter estimates, and this downward bias increases in the degree of positive correlation between the covariate (i.e., men's unemployment) and process time (i.e., union duration). Thus, unlike in the standard case of a cross-sectional regression model, the specific type of omitted variable bias that results from left-censored data may go either way in an event history setting. And because our key covariate of interest (men's unemployment) is a time-varying covariate, it follows that our reported estimates are likely to exhibit some degree of downward bias, if anything; that is, they are likely to be conservative relative to the "true" causal effect in question. The main intuition for this conclusion is that when there is negative duration dependence in the hazard (as is the well-known case in the separation hazard), time-varying covariates tend to index variation that occurs at some later point in process time, that is, something that happens when the hazard already is below average. Therefore, the bias that results from left-censoring is downward bias because it results from the misattribution of the low hazard to the covariate of interest rather than to unobserved process time.

Second, although some downward bias may be present in our reported estimates, it is likely to be small. This follows from the fact that left-truncation (i.e., late entry to the risk set in the sense of incompletely observed spells) is mitigating and counteracts any bias created by left-censoring, so the overall bias in stock-sample estimates is the sum of two sources of bias that work in opposite directions and therefore tend to largely cancel each other out. The intuition for this result is also rather straightforward. The bias from left-censoring occurs with completely observed spell data, where exposure to the event in question (i.e., the separation risk set) is well defined and fully observed, but where the estimate for the hazard (or, equivalently, the transition probability in discrete-time) then cannot be conditioned on time-specific survival without information on starting times. Left-truncation implicitly corrects for this problem because while each individual spell is incompletely observed (e.g., only in four-year snippets of full relationship histories in our dataset), truncation is equivalent to an implicit conditioning on initial survival up to the point of the first survey wave, and the resulting left-truncated data therefore no longer contain the heaping of exposure time units that is the source of bias in completely observed but left-censored spell data in the first place.

As a result, the simulation exercise may shed some conceptual and fundamental light on why the estimates we obtain from imperfectly observed data nevertheless succeed in arriving at parameter estimates that broadly fall in line with those from other studies based on superior (single-country) data sources, and why they support our conviction that even when based on imperfect data, our reported estimates are unlikely to be so severely biased that our substantive inferences would be called into question. Against this general conclusion, we report the simulation results in greater detail in the following.

References

Bryan, Mark L., and Stephen P. Jenkins. 2016. "Multilevel Modelling of Country Effects: A Cautionary Tale." *European Sociological Review* 32(1):3–22.

Eurostat. 2013. "2010 EU Comparative Final Quality Report." European Commission.

Eurostat. 2017. "Methodological Guidelines and Description of EU-SILC Target Variables." European Commission.

Eurostat. 2019. "EU Statistics on Income and Living Conditions (EU-SILC) Methodology." European Commission.

Giesselmann, Marco, and Alexander W. Schmidt-Catran. 2019. "Getting the Within Estimator of Cross-Level Interactions in Multilevel Models with Pooled Cross-Sections: Why Country Dummies (Sometimes) Do Not Do the Job." *Sociological Methodology* 49(1): 190–219 (https://doi.org/10.1177/0081175018809150).

Guo, Guang. 1993. "Event-History Analysis for Left-Truncated Data." *Sociological Methodology* 23:217–43.

Heckman, James, and Burton Singer. 1986. "Econometric Analysis of Longitudinal Data." Pp. 1689–1763 in *Handbook of Econometrics*, Vol. 3, edited by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland.

Heisig, Jan P., and Merlin Schaeffer. 2019. "Why You Should Always Include a Random Slope for the Lower-Level Variable Involved in a Cross-Level Interaction." *European Sociological Review* 35(2):258–79 (https://doi.org/10.1093/esr/jcy053).

ISER. 2019a. "British Household Panel Survey." https://www.iser.essex.ac.uk/bhps.

ISER. 2019b. "Understanding Society. The UK Household Longitudinal Study." https://www.understandingsociety.ac.uk/about/about-the-study.

Jäntti, Markus, Veli-Matti Törmälehto, and Eric Marlier. 2013. *The Use of Registers in the Context of EU-SILC Challenges and Opportunities*. Eurostat Statistical Working Papers Luxembourg: European Commission (https://doi.org/10.2785/45854).

Knies, Gundi. 2014. "Understanding Society: The UK Household Longitudinal Study Waves 1-4 User Manual." Colchester: University of Essex.

Knight, Carly R., and Mary C. Brinton. 2017. "One Egalitarianism or Several? Two Decades of Gender-Role Attitude Change in Europe." *American Journal of Sociology* 112(5): 1485–1532 (https://doi.org/10.1086/689814).

Mack, Alexander, and Barbara Lange. 2015. "Harmonization of Income Data in EU-SILC." *GESIS Papers*, no. 18: 5–17.

Singer, Judith D., and John B. Willett. 2003. *Applied Longitudinal Data Analysis*. Oxford: Oxford University Press.

SOEP. 2019. "European Union Statistics on Income and Living Conditions (EU-SILC)-like Panel for Germany Based on the Socio-Economic Panel (SOEP)." DIW German Institute for Economic Research (https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.592700.de/eu-silc_codebook_de_version_june_18.pdf).

Taylor, Marcia, John Brice, Nick Buck, and Elaine Prentice-Lane. 2010. "British Household Panel Survey User Manual. Volume A. Introduction, Technical Report, and Appendices." Colchester: University of Essex.

Tuma, Nancy B., and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods*. Orlando: Academic Press.

U.S. Census Bureau. 2001. "Survey of Income and Program Participation Users' Guide. Third Edition." (https://www2.census.gov/programs-surveys/sipp/guidance/SIPP_USERS_Guide_Third_Edition_2001.pdf).

U.S. Census Bureau. 2015. *Survey of Income and Program Participation*. Washington DC: U S. Census Bureau. (https://www.census.gov/programs-surveys/sipp/data/datasets.html).

Winship, Christopher, and Larry Radbill. 1994. “Sampling Weights and Regression Analysis.” *Sociological Methods & Research* 23(3): 230–57 (https://doi.org/10.1177/0049124194023002004).

Yamaguchi, Kazuo. 1991. *Event History Analysis*. Newbury Park, CA: Sage.

Simulation results for omitted variable bias due to left-censoring, by type of negative duration dependence, type of covariate of interest, and left truncation of sample

To systematically understand and document the potential for omitted variable bias that is created by the fact that our study relies on left-truncated data with unknown start times of couple relationships, we programmed a simulation study on the behavior of the parameter of interest under alternative scenarios for the unobserved baseline hazard in the separation process. We base our simulation on a sample of N = 100,000 spells for which we first create observable (union) duration data according to a prespecified transition model, and then, to mimic the constraint inherent in the left-censoring of the data (i.e., the unknown start dates), ascertain whether we are able to recover the parameter of interest from estimating a model that assumes a time-constant baseline hazard (i.e., that does not control for an effect of process time on transition outcomes) in a second step. We conduct all simulations for two types of baseline hazard functions, one exhibiting monotonous negative duration dependence and the other a bell-shaped (a.k.a. hump-shaped or sickle-shaped) pattern of duration dependence that is regularly confirmed in union separation data, and we report simulation results from a range of transition processes that differ in the degree of duration dependence and the degree of the bell-shape in the hazard, respectively. Intending to broadly align the simulation to empirical features of our data, we constrain all models to be consistent with the total separation risk implied by the survivor function $G(t) = e^{-rt}$ for a baseline time constant hazard rate $r = .005$ after simulating the process for $t = 100$ time units; in this, the baseline hazard r was chosen to roughly reflect the average annual separation probability among couples not experiencing unemployment in our actual data.

In each simulation analysis we conduct, we focus on the parameter estimate for the multiplicative effect of a single observed covariate U on the transition rate $r(t)$. To again align the simulation to the estimated average effect of male partner's unemployment on separation risk in the actual study, we set the true value of the model coefficient to $b = ln(2)$, corresponding to a hazard ratio of $e^{b} = 2$ throughout. We randomly assign 10 percent of all spells to the treatment condition of experiencing U (i.e., set $U = 1$) and the remaining spells to the control group. To focus the simulation study on essentials, we abstract from the presence of any additional covariates in the model, that is, we assume any confounding other than from unobserved process time (i.e., from left-censoring) is absent or has been perfectly controlled for through appropriate controls in the actual research. Instead, we systematically vary the type of observed covariate in our simulation, and we present three case studies for each simulated model, where the covariate of interest is either a time-constant factor, a time-varying covariate that occurs randomly across process time, or a time-varying covariate that exhibits a mildly positive correlation with process time. As it turns out, the type of covariate studied will critically affect the direction of omitted variable bias due to the (negative, null, or positive) correlation with process time that is implied.

In Part I of the simulation study, we study omitted variable bias due to left-censoring and left-truncation in the context of proportional-hazard (PH) transition models. We simulate a class of models that exhibit monotonous duration dependence in the baseline hazard from

(D-1)	$r(t) = \frac{f(t)}{G(t)} \approx Pr(T = t \mid T \geq t) = e^{(c+at+bU)}$.

The simulated model is a continuous-time exponential model with linear duration dependence and a (standard) multiplicative effect of U on $r(t)$ or its discrete-time analogue $Pr(T = t \mid T \geq t)$ (cf. Allison 1982:72, equation 10), respectively. Even though our actual research is estimating logit models for the (annual) separation probability in discrete-time, we liberally use a basic model in continuous time as the workhorse for our simulation because time aggregation bias is negligible at the very low baseline hazard of $r = .005$ observed in our actual data and used in our derived simulation exercise (also see Petersen 1991:273, Table 1 in particular). Our simulation data show that although strictly speaking, as a proportional odds model, the logit model of course violates the PH assumption of the exponential model, the difference in parameter estimates is again minuscule in practice, given the low level of the baseline hazard in our actual data and in the simulation study.

We present simulation evidence for a range of transition models that vary in the parameter a to express and simulate different degrees of duration dependence, but where choosing $a < 0$ guarantees to create negative duration dependence in principle (see Figure D1 for a graphical display of the specific baseline hazards assumed in our simulation study). We simulate the effect of left-censoring (i.e., unknown start times) by estimating a discrete-time logit model on the complete data but without any control for process time (i.e., assuming a time-constant transition probability), and by conducting a Wald test against H_0: $b = ln(2)$ to evaluate any resulting bias in the parameter estimate of interest. To check whether alternative model specifications would be more successful in identifying the "true" coefficient, we compare the parameter estimates from the logit model with time-constant baseline hazard to a logit model that allows for duration dependence as well as to a range of model specifications in continuous time (i.e., exponential, Weibull, and Cox regression models). Among these, the comparison between the logit and the exponential model with time-constant baseline hazard (i.e., without controls for process time) demonstrates that the subtle mathematical difference between proportional hazards and proportional odds specifications always results in quantitatively negligible differences in the practical context of our study. In a second step, we repeat the simulation exercise but randomly left-truncate the spell sample and observe each spell over a fixed observation window afterward, that is, we simulate random "late entry to the observed risk set" and a subsequent fixed panel observation window to mimic the respective features of our stock sample data in the actual study.

To allow for slightly more complicated bell-shaped patterns in the baseline hazard, we then repeat the exercise for the slightly more complex class of transition models given by

(D-2)	$r(t) = e^{(c+a_1t+a_2t^2+bU)}$,

where a bell-shaped pattern of duration dependence results from appropriately fixing $a_1 > 0$ and $a_2 < 0$ in a polynomial function of process time; for the reader's benefit, Figure D2 depicts the four different bell-shaped baseline hazards utilized in our concrete simulations. As before, we set the true value of $b = ln(2)$ and we conduct the respective hypothesis tests in left-censored but otherwise completely observed data and in a left-truncated sample that also exhibits left-censored duration data (i.e., unknown starting times coupled with incompletely observed exposure). We summarize the main results from our simulations in Table D1, and report more detailed case-by-case evidence in Tables D3 to D14. The bottom-line result is that left-censoring and left-truncation create bias in opposite directions. As a consequence, the bias that remains in scenarios that most closely resemble our actual data is mild and downward, not upward. If anything, the left-truncated data with unknown start times that we have available are likely to yield slightly conservative inferences.

Table D1.

Part I: Summary of simulation results for PH models

	Type of selectivity of job loss by relationship duration		
Scenario	Case 1: U time-constant factor	Case 2: U time-varying and randomly distributed over T	Case 3: U time-varying and positively selected on process time
Monotonous hazard			
A: negative duration dependence	upward bias in parameter of interest; bias increases with the degree of negative duration dependence	mild downward bias in parameter of interest; bias increases with the degree of negative duration dependence	downward bias in parameter of interest; bias increases with the degree of negative duration dependence
B: A + left-truncation of sample	mild upward bias in parameter of interest, but only with strong negative duration dependence	no bias in parameter of interest at all; opposing effects cancel out	very mild downward bias in parameter of interest with strong duration dependence, no bias otherwise
Non-monotonous hazard			
C: bell-shaped (hump-shaped, sickle-shaped) hazard function	upward bias in parameter of interest (stronger than in A); bias increases with the degree of hump-shape	no bias in parameter of interest with mild-moderate hump-shape; downward bias in parameter of interest with pronounced hump-shaped hazard	downward bias in parameter of interest; bias increases with more pronounced hump-shape of the hazard
D: C + left-truncation of sample	upward bias in parameter of interest (slightly attenuated relative to C, stronger than in A)	no bias in parameter of interest	no bias in parameter of interest except with very pronounced hump-shape of the hazard; downward bias in that case

Note: Statements of bias refer to the difference between parameter estimate from a discrete-time logit model with time-constant baseline hazard and the true parameter of $b = ln(2)$. Our actual study concerns the effect of a time-varying covariate (men's unemployment) estimated from stock-sample data, that is, left-truncated data with unknown start times. The simulation results most relevant (i.e., that most closely resemble the actual features of our dataset) are simulations B-2/B-3 and D-2/D-3, depending on what unobserved pattern of duration dependence and what degree of positive selection on process time (i.e., union duration) one wishes to assume when judging the interpretations advanced in the main text.

In Part II of the simulation exercise, we add a further layer of complexity by examining the bias from left-censoring and left-truncation in the context of transition models that lack the PH property, that is, that permit the effect of the main covariate of interest U to vary with process time. For this simulation we expand the earlier class of models exhibiting monotonous negative duration dependence by adding an interaction term $(t \times U)$ between U and process time t to yield

(D-3)	$r(t) = e^{(c+at+bU+d(t^*\times U))}$.

To fix the specification, we set $d = 1.01$ to simulate a time-varying coefficient of U that increases with process time, that is, where the total effect of U on $r(t)$ is given by $b + dt^*$. To simplify the presentation of results as well as the actual hypothesis testing, we construct the interaction term with the demeaned process time $t^* = t - 50.5$ to keep the scaling of b in line with Part I of the simulation study. In Part II, in other words, we keep to testing against H_0: $b = ln(2)$ with the understanding this represents the average effect of U on $r(t)$ over the observed process time t with $1 \leq t \leq 100$. We keep the setup of the simulation exactly parallel to Part I otherwise, and we incorporate a simulation study on omitted variable bias in a class of non-PH transition models exhibiting a bell-shaped baseline hazard given by

(D-4)	$r(t) = e^{\left(c+a_1t+a_2t^2+bU+d(t^*\times U)\right)}$.

In line with our earlier practice, we summarize the main results from these simulations in Table D2, we graphically display the simulated (non-PH) baseline hazards in the treatment and control group in Figures D3 and D4, and we report more detailed case-by-case evidence in Tables D15 to D26. In case of non-PH specifications, the bottom-line result is that the effects of left-censoring and left-truncation are less benign, but there is considerable potential for bias, and bias will be in an upward direction when it is realistic to assume the covariate of interest U (i.e., men's unemployment) may have stronger effects on separation risk among young couples, but downward (i.e., our reported estimates may be conservative) if stronger effects apply among older couples. At the same time, our robustness checks provide no indication of the PH assumption being violated (at least in U.S. data), and we also note that what we report as "bias" in the non-PH case could equally well be interpreted as a specific type of treatment effect heterogeneity (cf. Xie et al. 2012) that we are unable to capture in our present study, but that may represent at best a sideline interest in a study to compare the average effects of men's unemployment on separation risk in Western countries.

Table D2.

Part II: Summary of simulation results for non-PH models

	Type of selectivity of job loss by relationship duration		
Scenario	Case 1: U time-constant factor	Case 2: U time-varying and randomly distributed over T	Case 3: U time-varying and positively selected on process time
Monotonous hazard			
E: negative duration dependence, time-varying coefficient (= non-PH model)	downward bias in parameter estimate, increasing with more pronounced duration dependence	downward bias in parameter estimate, bias more pronounced than in case E-1	downward bias in parameter estimate, bias slightly more pronounced than in case E-2
F: E + left-truncation of sample	mild downward bias in parameter estimate, but only with strong duration dependence	mild downward bias in parameter estimate, but only with moderate-strong duration dependence	mild downward bias in parameter estimate, but only with moderate-strong duration dependence
Non-monotonous hazard			
G: bell-shaped (hump-shaped,) hazard, time-varying coefficient (= non-PH model)	clear downward bias in parameter estimate, more pronounced bias pattern than in E	clear downward bias in parameter estimate, slightly more pronounced patterns than in G-1	clear downward bias in parameter estimate, slightly more pronounced patterns than in G-2
H: G + left-truncation of sample	clear downward bias in parameter estimate, somewhat attenuated relative to G-1	clear downward bias in parameter estimate, somewhat attenuated relative to G-2	clear downward bias in parameter estimate, somewhat attenuated relative to G-3

Note: Statements of bias refer to the difference between parameter estimate from a discrete-time logit model with time-constant baseline hazard and the true parameter of $b = ln(2)$. Our actual study concerns the effect of a time-varying covariate (men's unemployment) estimated from stock-sample data, that is, left-truncated data with unknown start times. The simulation results most relevant (i.e., that most closely resemble the actual features of our dataset) are simulations F-2/F-3 and H-2/H-3, depending on what unobserved pattern of duration dependence and what degree of positive selection on process time (i.e., union duration) one wishes to assume when judging the interpretations advanced in the main text.

References

Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories." *Sociological Methodology* 13:61–98.

Petersen, Trond. 1991. "Time-Aggregation Bias in Continuous-Time Hazard-Rate Models." *Sociological Methodology* 21:263–90.

Xie, Yu, Jennie E. Brand, and Ben Jann. 2012. "Estimating Heterogeneous Treatment Effects with Observational Data." *Sociological Methodology* 42:314–47.

Detailed simulation results

Part I: PH transition models
Scenarios A and B: models exhibiting monotonous negative duration dependence

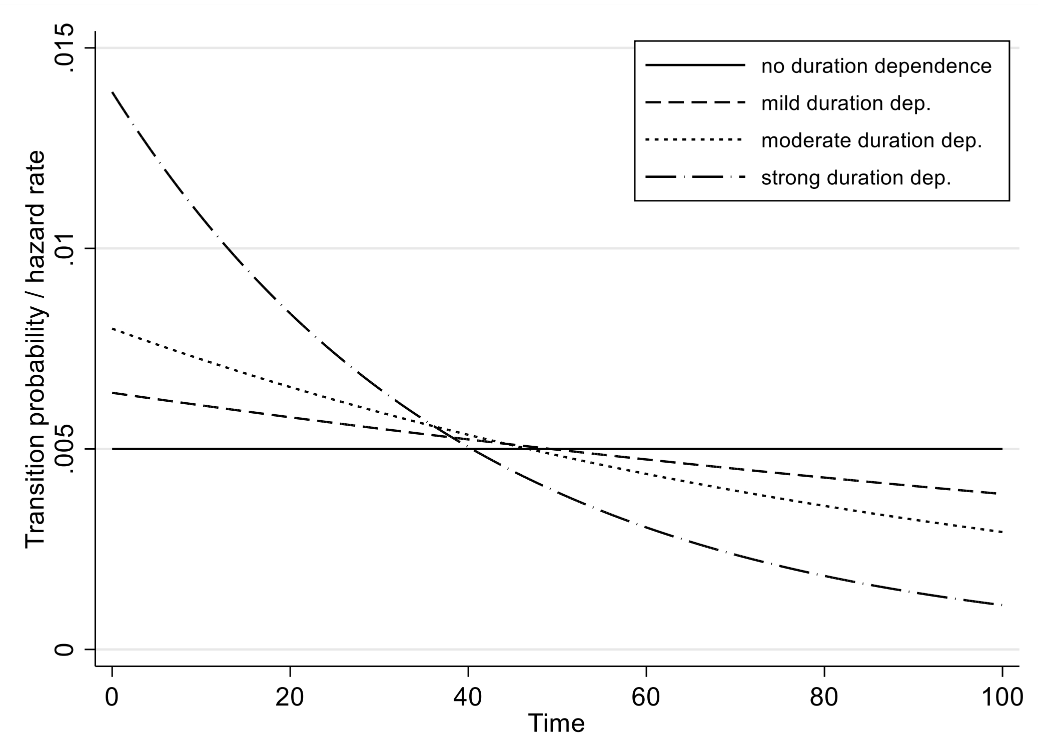


Figure D1.
Monotonically declining baseline hazards utilized in the simulation study

Table D3.

Simulation A1: Monotonous negative duration dependence, *U* time-constant covariate

Estimator	Simulation 1: time-constant baseline hazard (e^a = 1)	Simulation 2: mild duration dependence (e^a = .995)	Simulation 3: moderate dur. dependence (e^a = .99)	Simulation 4: strong duration dependence (e^a = .975)
Discrete-time				
Logit	.699 (.014)	**.725 (.014)**	**.747 (.014)**	**.812 (.014)**
- Wald test, b=.69	.673	**.020***	**.000*****	**.000*****
Logit w t linear		.704 (.014)	.705 (.014)	.713 (.014)
- Wald test, b=.69		.439	.368	.139
Continuous-time				
Exponential	.694 (.014)	*.719 (.014)*	**.741 (.014)**	**.805 (.014)**
- Wald test, b=.69	.961	*.053*	**.000*****	**.000*****
Exp. w t linear		.698 (.014)	.699 (.014)	.705 (.014)
- Wald test, b=.69		.708	.645	.400
Weibull	.698 (.014)	.709 (.014)	.715 (.014)	**.736 (.014)**
- Wald test, b=.69	.712	.258	.102	**.002****
Cox regression	.694 (.014)	.698 (.014)	.699 (.014)	.704 (.014)
- Wald test, b=.69	.977	.711	.650	.411

Note: *N* = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D4.

Simulation A2: Monotonous negative duration dependence, *U* time-varying covariate but uncorrelated with process time

Estimator	Simulation 1: time-constant baseline hazard (e^a = 1)	Simulation 2: mild duration dependence (e^a = .995)	Simulation 3: moderate dur. dependence (e^a = .99)	Simulation 4: strong duration dependence (e^a = .975)
Discrete-time				
Logit	.700 (.014)	.682 (.015)	**.660 (.015)**	**.551 (.016)**
- Wald test, b=.69	.654	.448	**.024***	**.000*****
Logit w t linear		.703 (.015)	.708 (.015)	.713 (.016)
- Wald test, b=.69		.484	.325	.218
Continuous-time				
Exponential	.695 (.014)	.677 (.015)	**.655 (.015)**	**.546 (.016)**
- Wald test, b=.69	.925	.266	**.009****	**.000*****
Exp. w t linear		.698 (.015)	.702 (.015)	.706 (.016)
- Wald test, b=.69		.733	.538	.411
Weibull	.686 (.014)	.699 (.015)	.711 (.015)	.713 (.016)
- Wald test, b=.69	.628	.670	.233	.217
Cox regression	.696 (.014)	.699 (.015)	.703 (.015)	.708 (.016)
- Wald test, b=.69	.844	.672	.510	.340

Note: N = 100.000, true b = .693 (hazard ratio of e^b = 2); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at p < .05)

Table D5.

Simulation A3: Monotonous negative duration dependence, *U* time-varying covariate and positively correlated with process time

Estimator	Simulation 1: time-constant baseline hazard ($e^a = 1$)	Simulation 2: mild duration dependence ($e^a = .995$)	Simulation 3: moderate dur. dependence ($e^a = .99$)	Simulation 4: strong duration dependence ($e^a = .975$)
Discrete-time				
Logit	.699 (.015)	*.666 (.015)*	**.634 (.015)**	**.487 (.016)**
- Wald test, b=.69	.688	*.074*	**.000*** **	**.000*** **
Logit w t linear		.698 (.015)	.702 (.015)	.705 (.016)
- Wald test, b=.69		.754	.542	.487
Continuous-time				
Exponential	.694 (.015)	**.662 (.015)**	**.629 (.015)**	**.484 (.016)**
- Wald test, b=.69	.957	**.033* **	**.000*** **	**.000*** **
Exp. w t linear		.693 (.015)	.697 (.015)	.698 (.016)
- Wald test, b=.69		.975	.795	.748
Weibull	.684 (.015)	.688 (.015)	.694 (.015)	.671 (.016)
- Wald test, b=.69	.523	.724	.948	.176
Cox regression	.696 (.015)	.694 (.015)	.698 (.015)	.701 (.014)
- Wald test, b=.69	.865	.950	.736	.643

Note: *N* = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at $p < .05$)

Table D6.
Simulation B1: Monotonous negative duration dependence, *U* time-constant covariate, left-truncated sample

Estimator	Simulation 1: time-constant baseline hazard ($e^a = 1$)	Simulation 2: mild duration dependence ($e^a = .995$)	Simulation 3: moderate dur. dependence ($e^a = .99$)	Simulation 4: strong duration dependence ($e^a = .975$)
Discrete-time				
Logit	.691 (.020)	.689 (.020)	.726 (.020)	**.759 (.022)**
- Wald test, b=.69	.906	.826	.104	**.003****
Logit w t linear		.679 (.020)	.706 (.020)	.696 (.022)
- Wald test, b=.69		.491	.522	.915
Continuous-time				
Exponential	.686 (.020)	.684 (.020)	.721 (.020)	**.754 (.022)**
- Wald test, b=.69	.709	.638	.173	**.006****
Exp. w t linear		.674 (.020)	.701 (.020)	.689 (.022)
- Wald test, b=.69		.347	.712	.856
Weibull	.690 (.020)	.684 (.020)	.717 (.020)	**.739 (.022)**
- Wald test, b=.69	.874	.636	.241	**.035***
Cox regression	.686 (.020)	.678 (.020)	.709 (.020)	.728 (.022)
- Wald test, b=.69	.715	.463	.422	.110

Note: *N* = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D7.

Simulation B2: Monotonous negative duration dependence, *U* time-varying covariate but uncorrelated with process time, left-truncated sample

Estimator	Simulation 1: time-constant baseline hazard (e^a = 1)	Simulation 2: mild duration dependence (e^a = .995)	Simulation 3: moderate dur. dependence (e^a = .99)	Simulation 4: strong duration dependence (e^a = .975)
Discrete-time				
Logit	.701 (.020)	.689 (.020)	.709 (.020)	.688 (.022)
- Wald test, b=.69	.683	.853	.445	.818
Logit w t linear		.688 (.020)	.714 (.020)	.713 (.022)
- Wald test, b=.69		.794	.295	.371
Continuous-time				
Exponential	.696 (.019)	.684 (.020)	.703 (.020)	.683 (.022)
- Wald test, b=.69	.882	.658	.613	.654
Exp. w t linear		.683 (.020)	.709 (.020)	.707 (.022)
- Wald test, b=.69		.603	.436	.528
Weibull	.695 (.019)	.684 (.020)	.705 (.020)	.692 (.022)
- Wald test, b=.69	.941	.657	.556	.960
Cox regression	.696 (.019)	.685 (.020)	.704 (.020)	.690 (.022)
- Wald test, b=.69	.888	.673	.585	.894

Note: N = 100.000, true b = .693 (hazard ratio of e^b = 2); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D8.

Simulation B3: Monotonous negative duration dependence, *U* time-varying covariate and positively correlated with process time, left-truncated sample

Estimator	Simulation 1: time-constant baseline hazard ($e^a = 1$)	Simulation 2: mild duration dependence ($e^a = .995$)	Simulation 3: moderate dur. dependence ($e^a = .99$)	Simulation 4: strong duration dependence ($e^a = .975$)
Discrete-time				
Logit	.692 (.020)	.691 (.020)	.690 (.020)	**.641 (.022)**
- Wald test, b=.69	.938	.904	.868	**.020***
Logit w t linear		.697 (.020)	.708 (.020)	.702 (.022)
- Wald test, b=.69		.839	.457	.707
Continuous-time				
Exponential	.687 (.020)	.686 (.020)	.685 (.020)	**.637 (.022)**
- Wald test, b=.69	.737	.708	.679	**.011***
Exp. w t linear		.692 (.020)	.703 (.020)	.696 (.022)
- Wald test, b=.69		.959	.625	.893
Weibull	.683 (.020)	.685 (.020)	.688 (.020)	*.652 (.022)*
- Wald test, b=.69	.622	.690	.792	*.065*
Cox regression	.687 (.020)	.688 (.020)	.691 (.020)	*.656 (.022)*
- Wald test, b=.69	.767	.810	.926	*.095*

Note: N = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at $p < .05$)

Part I: PH transition models
Scenarios C and D: models featuring bell-shaped baseline hazard function

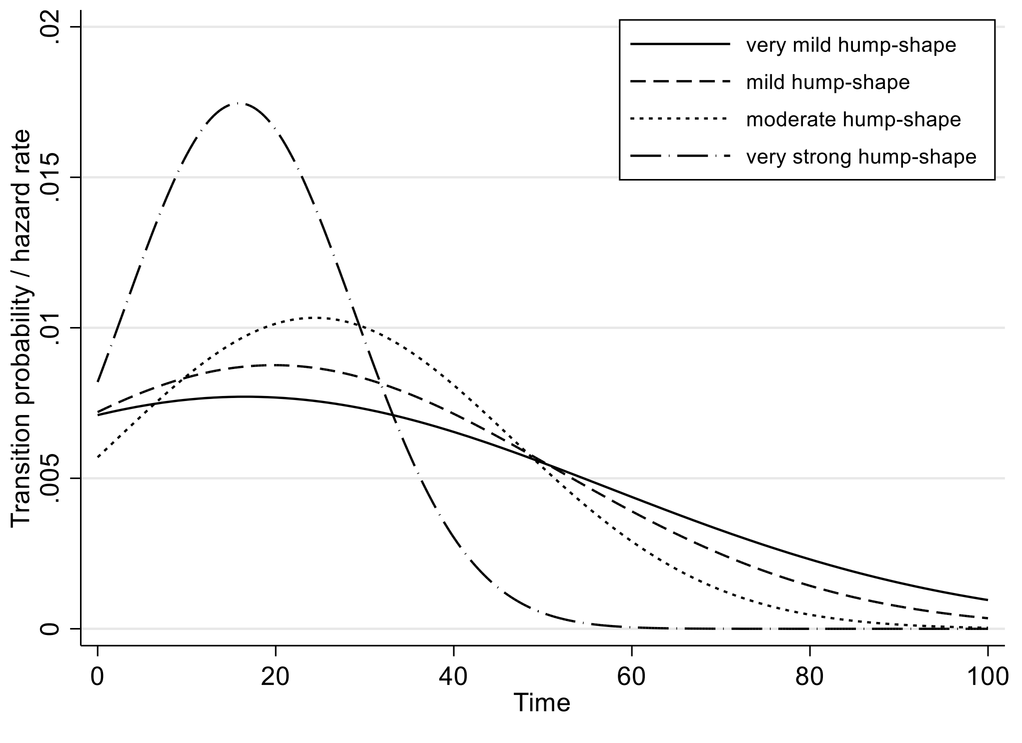


Figure D2.
Bell-shaped baseline hazards utilized in the simulation study

Table D9.

Simulation C1: Bell-shaped baseline hazard, *U* time-constant covariate

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	**.770 (.014)**	**.796 (.014)**	**.805 (.014)**	**.864 (.014)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Logit w t linear	.702 (.014)	.705 (.014)	.694 (.014)	.680 (.014)
- Wald test, b=.69	.538	.397	.945	.342
Logit w t quad.	.708 (.014)	*.716 (.014)*	.715 (.014)	.713 (.014)
- Wald test, b=.69	.277	*.089*	.115	.147
Continuous-time				
Exponential	**.763 (.014)**	**.789 (.014)**	**.798 (.014)**	**.856 (.014)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t linear	.695 (.014)	.697 (.014)	.686 (.014)	**.666 (.014)**
- Wald test, b=.69	.905	.775	.588	**.047** **
Exp. w t quad.	.701 (.014)	.709 (.014)	.706 (.014)	.699 (.014)
- Wald test, b=.69	.552	.252	.343	.678
Weibull	**.728 (.014)**	**.742 (.014)**	**.745 (.014)**	**.751 (.014)**
- Wald test, b=.69	**.011* **	**.000*** **	**.0001*** **	**.000*** **
Cox regression	.701 (.014)	.709 (.014)	.706 (.014)	.699 (.014)
- Wald test, b=.69	.553	.252	.343	.682

Note: *N* = 100.000, true b = .693 (hazard ratio of e^b = 2); Wald tests of b = ln(2)

* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D10.

Simulation C2: Bell-shaped baseline hazard, *U* time-varying covariate but uncorrelated with process time

Estimator	Simulation 1: mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	.686 (.015)	.690 (.015)	*.719 (.015)*	**.545 (.016)**
- Wald test, b=.69	.619	.839	*.080*	**.000*** **
Logit w t linear	**.766 (.015)**	**.803 (.015)**	**.863 (.015)**	**.942 (.016)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Logit w t quad.	.709 (.015)	.714 (.015)	.713 (.015)	.711 (.016)
- Wald test, b=.69	.287	.164	.186	.285
Continuous-time				
Exponential	.680 (.015)	.685 (.015)	.713 (.015)	**.540 (.016)**
- Wald test, b=.69	.388	.566	.173	**.000*** **
Exp. w t linear	**.759 (.015)**	**.796 (.015)**	**.854 (.015)**	**.929 (.016)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t quad.	.702 (.015)	.707 (.015)	.704 (.015)	.696 (.016)
- Wald test, b=.69	.530	.369	.465	.851
Weibull	**.754 (.015)**	**.781 (.015)**	**.818 (.015)**	**.783 (.016)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Cox regression	.702 (.015)	.707 (.015)	.705 (.015)	.696 (.016)
- Wald test, b=.69	.530	.371	.446	.859

Note: N = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at $p < .05$)

Table D11.

Simulation C3: Bell-shaped baseline hazard, *U* time-varying covariate and positively correlated with process time

Estimator	Simulation 1: mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	**.647 (.015)**	**.633 (.015)**	**.650 (.016)**	**.401 (.017)**
- Wald test, b=.69	**.003****	**.000*****	**.006****	**.000*****
Logit w t linear	**.762 (.015)**	**.794 (.016)**	**.853 (.016)**	**.913 (.018)**
- Wald test, b=.69	**.000*****	**.000*****	**.000*****	**.000*****
Logit w t quad.	.703 (.015)	.702 (.016)	.702 (.016)	.710 (.018)
- Wald test, b=.69	.533	.569	.574	.346
Continuous-time				
Exponential	**.642 (.015)**	**.628 (.015)**	**.645 (.015)**	**.398 (.017)**
- Wald test, b=.69	**.0009*****	**.000*****	**.002*****	**.000*****
Exp. w t linear	**.756 (.015)**	**.787 (.016)**	**.845 (.016)**	**.900 (.017)**
- Wald test, b=.69	**.000*****	**.000*****	**.000*****	**.000*****
Exp. w t quad.	.696 (.015)	.695 (.016)	.693 (.016)	.696 (.017)
- Wald test, b=.69	.832	.912	.989	.885
Weibull	**.728 (.015)**	**.740 (.016)**	**.767 (.016)**	.669 (.017)
- Wald test, b=.69	**.022***	**.003****	**.000*****	.160
Cox regression	.696 (.015)	.695 (.016)	.694 (.016)	.696 (.017)
- Wald test, b=.69	.829	.913	.968	.887

Note: N = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at $p < .05$)

Table D12.

Simulation D1: Bell-shaped baseline hazard, *U* time-constant covariate, left-truncated sample

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	**.735 (.020)**	**.778 (.020)**	**.786 (.020)**	**.825 (.024)**
- Wald test, b=.69	**.034***	**.000*****	**.000*****	**.000*****
Logit w t linear	.697 (.020)	.718 (.020)	.712 (.020)	.672 (.025)
- Wald test, b=.69	.852	.204	.336	.383
Logit w t quad.	.700 (.020)	*.727 (.020)*	.725 (.020)	.707 (.025)
- Wald test, b=.69	.746	*.091*	.115	.581
Continuous-time				
Exponential	*.729 (.020)*	**.772 (.020)**	**.779 (.020)**	**.820 (.024)**
- Wald test, b=.69	*.068*	**.0001*****	**.000*****	**.000*****
Exp. w t linear	.690 (.020)	.711 (.020)	.704 (.020)	.657 (.024)
- Wald test, b=.69	.893	.366	.591	.135
Exp. w t quad.	.693 (.020)	.719 (.020)	.716 (.020)	.693 (.024)
- Wald test, b=.69	.996	.184	.248	.989
Weibull	.721 (.020)	**.758 (.020)**	**.759 (.020)**	**.781 (.024)**
- Wald test, b=.69	.166	**.001****	**.001*****	**.000*****
Cox regression	.709 (.020)	**.744 (.020)**	**.740 (.020)**	**.758 (.024)**
- Wald test, b=.69	.422	**.010***	**.017***	**.008****

Note: N = 100.000, true b = .693 (hazard ratio of e^b = 2); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at p < .05)

Table D13.

Simulation D2: Bell-shaped baseline hazard, *U* time-varying covariate but uncorrelated with process time, left-truncated sample

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	.693 (.020)	.722 (.020)	**.737 (.020)**	.653 (.025)
- Wald test, b=.69	.997	.156	**.031***	.105
Logit w t linear	.700 (.020)	*.729 (.020)*	**.766 (.020)**	**.850 (.025)**
- Wald test, b=.69	.732	*.070*	**.000*****	**.000*****
Logit w t quad.	.684 (.020)	.702 (.020)	.713 (.020)	.721 (.025)
- Wald test, b=.69	.635	.660	.325	.267
Continuous-time				
Exponential	.688 (.020)	.716 (.020)	*.730 (.020)*	*.649 (.025)*
- Wald test, b=.69	.777	.259	*.063*	*.074*
Exp. w t linear	.694 (.020)	.722 (.020)	**.758 (.020)**	**.837 (.025)**
- Wald test, b=.69	.967	.141	**.001****	**.000*****
Exp. w t quad.	.678 (.020)	.695 (.020)	.705 (.020)	.708 (.025)
- Wald test, b=.69	.434	.928	.568	.557
Weibull	.690 (.020)	.720 (.020)	**.736 (.020)**	.680 (.025)
- Wald test, b=.69	.880	.185	**.034****	.607
Cox regression	.685 (.020)	.712 (.020)	.724 (.020)	.678 (.025)
- Wald test, b=.69	.699	.346	.117	.553

Note: *N* = 100.000, true b = .693 (hazard ratio of $e^{b} = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

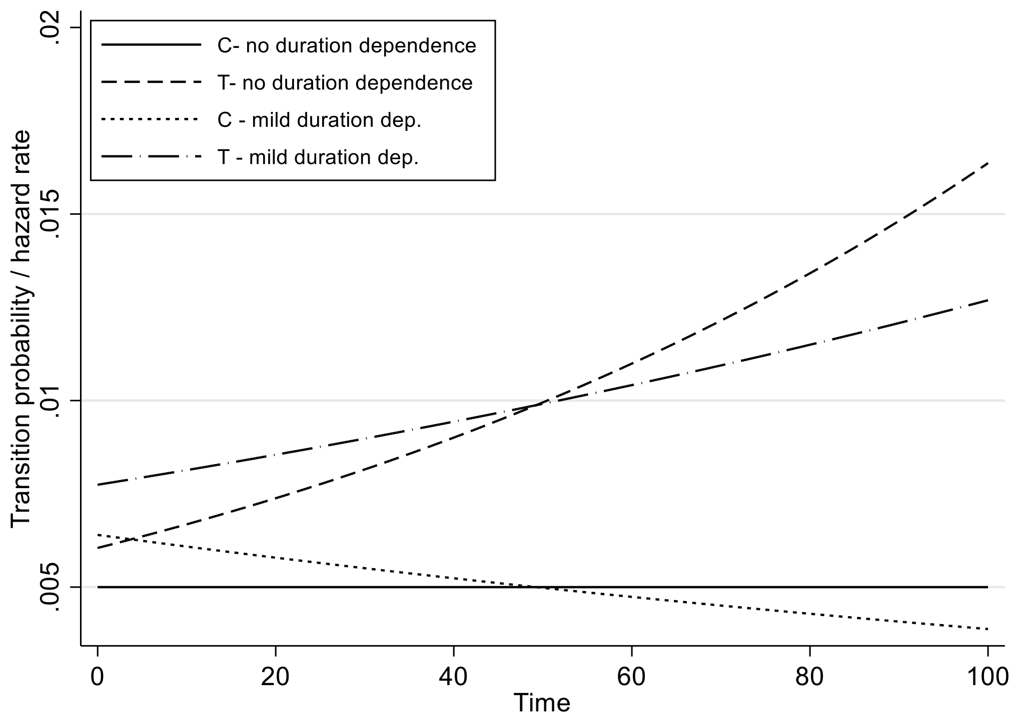
Table D14.

Simulation D3: Bell-shaped baseline hazard, *U* time-varying covariate and positively correlated with process time, left-truncated sample

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	.705 (.020)	.686 (.020)	.682 (.021)	**.513 (.026)**
- Wald test, b=.69	.565	.712	.603	**.000*** **
Logit w t linear	*.731 (.020)*	.725 (.020)	**.743 (.021)**	**.806 (.027)**
- Wald test, b=.69	*.057*	.119	**.016* **	**.000*** **
Logit w t quad.	.712 (.020)	.692 (.020)	.682 (.021)	.693 (.027)
- Wald test, b=.69	.358	.961	.577	.998
Continuous-time				
Exponential	.699 (.020)	.680 (.020)	.677 (.021)	**.510 (.026)**
- Wald test, b=.69	.772	.517	.424	**.000*** **
Exp. w t linear	.725 (.020)	.718 (.020)	**.736 (.021)**	**.795 (.027)**
- Wald test, b=.69	.111	.215	**.039* **	**.000*** **
Exp. w t quad.	.705 (.020)	.685 (.020)	.674 (.021)	.681 (.026)
- Wald test, b=.69	.546	.703	.348	.649
Weibull	.706 (.020)	.691 (.020)	.692 (.021)	**.567 (.026)**
- Wald test, b=.69	.503	.900	.960	**.000*** **
Cox regression	.708 (.020)	.690 (.020)	.692 (.021)	**.589 (.026)**
- Wald test, b=.69	.456	.890	.944	**.000*** **

Note: *N* = 100.000, true b = .693 (hazard ratio of e^b = 2); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Part II: Non-PH transition models
Scenarios E and F: models exhibiting monotonous negative duration dependence



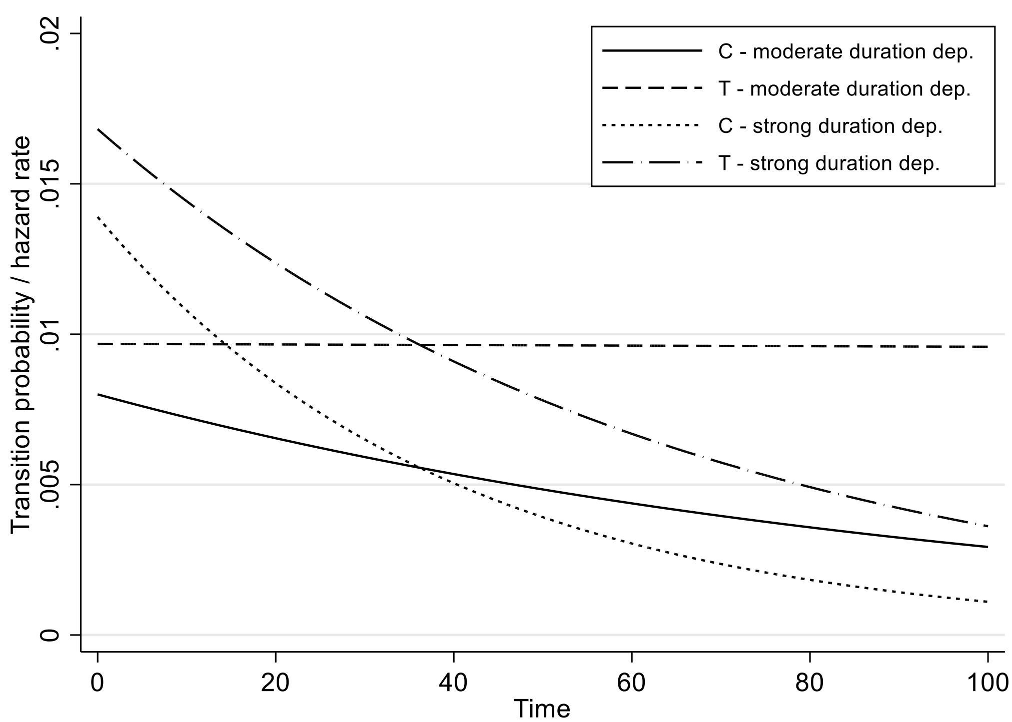


Figure D3.
Non-PH specifications combined with monotonically declining baseline hazards
(C – control group, T – treatment group)

Table D15.

Simulation E1: Non-PH specification, monotonous negative duration dependence, *U* time-constant covariate

Estimator	Simulation 1: time-constant baseline hazard (e^a = 1)	Simulation 2: mild duration dependence (e^a = .995)	Simulation 3: moderate dur. dependence (e^a = .99)	Simulation 4: strong duration dependence (e^a = .975)
Discrete-time				
Logit	**.651 (.014)**	**.627 (.014)**	**.606 (.014)**	**.565 (.015)**
- Wald test, b=.69	**.002****	**.000*****	**.000*****	**.000*****
Logit w t linear		**.613 (.014)**	**.576 (.014)**	**.498 (.014)**
- Wald test, b=.69		**.000*****	**.000*****	**.000*****
Logit w t tvcoeff		.688 (.014)	.686 (.015)	.700 (.018)
- Wald test, b=.69		.688	.627	.698
Continuous-time				
Exponential	**.647 (.013)**	**.622 (.014)**	**.601 (.014)**	**.560 (.014)**
- Wald test, b=.69	**.001*****	**.000*****	**.000*****	**.000*****
Exp. w t linear		**.609 (.014)**	**.571 (.014)**	**.493 (.014)**
- Wald test, b=.69		**.000*****	**.000*****	**.000*****
Exp. w t tvcoeff		.683 (.014)	.681 (.014)	.696 (.018)
- Wald test, b=.69		.446	.420	.863
Weibull	**.655 (.013)**	**.617 (.014)**	**.585 (.014)**	**.519 (.014)**
- Wald test, b=.69	**.004****	**.000*****	**.000*****	**.000*****
Weibull w anc p	**.652 (.014)**	**.608 (.014)**	**.571 (.014)**	**.494 (.014)**
- Wald test, b=.69	**.002****	**.000*****	**.000*****	**.000*****
Cox regression	**.653 (.014)**	**.609 (.014)**	**.572 (.014)**	**.493 (.014)**
- Wald test, b=.69	**.003****	**.000*****	**.000*****	**.000*****
Cox w tvcoeff	.689 (.014)	.683 (.014)	.681 (.014)	.695 (.018)
- Wald test, b=.69	.771	.444	.415	.900

Note: *N* = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D16.

Simulation E2: Non-PH specification, monotonous negative duration dependence, *U* time-varying covariate but uncorrelated with process time

Estimator	Simulation 1: time-constant baseline hazard ($e^a = 1$)	Simulation 2: mild duration dependence ($e^a = .995$)	Simulation 3: moderate dur. dependence ($e^a = .99$)	Simulation 4: strong duration dependence ($e^a = .975$)
Discrete-time				
Logit	.714 (.014)	**.656 (.014)**	**.601 (.015)**	**.421 (.016)**
- Wald test, b=.69	.133	**.020***	**.000*****	**.000*****
Logit w t linear		.675 (.014)	**.644 (.015)**	**.583 (.016)**
- Wald test, b=.69		.216	**.001*****	**.000*****
Logit w t tvcoeff		.694 (.015)	.691 (.015)	.706 (.018)
- Wald test, b=.69		.951	.883	.468
Continuous-time				
Exponential	.709 (.014)	**.655 (.014)**	**.597 (.015)**	**.418 (.016)**
- Wald test, b=.69	.257	**.007****	**.000*****	**.000*****
Exp. w t linear		.670 (.014)	**.639 (.015)**	**.578 (.016)**
- Wald test, b=.69		.112	**.000*****	**.000*****
Exp. w t tvcoeff		.689 (.014)	.686 (.015)	.702 (.018)
- Wald test, b=.69		.776	.646	.609
Weibull	.696 (.014)	.672 (.014)	**.648 (.015)**	**.583 (.016)**
- Wald test, b=.69	.841	.137	**.003****	**.000*****
Weibull w anc p	.691 (.014)	**.652 (.014)**	**.617 (.015)**	**.544 (.016)**
- Wald test, b=.69	.896	**.004****	**.000*****	**.000*****
Cox regression	.712 (.014)	.677 (.014)	**.646 (.015)**	**.586 (.016)**
- Wald test, b=.69	.178	.259	**.001****	**.000*****
Cox w tvcoeff	.697 (.014)	.690 (.015)	.687 (.015)	.702 (.018)
- Wald test, b=.69	.804	.840	.688	.610

Note: *N* = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)

* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at $p < .05$)

Table D17.

Simulation E3: Non-PH specification, monotonous negative duration dependence, *U* time-varying covariate and positively correlated with process time

Estimator	Simulation 1: time-constant baseline hazard ($e^a = 1$)	Simulation 2: mild duration dependence ($e^a = .995$)	Simulation 3: moderate dur. dependence ($e^a = .99$)	Simulation 4: strong duration dependence ($e^a = .975$)
Discrete-time				
Logit	**.731 (.014)**	*.668 (.015)*	**.600 (.015)**	**.381 (.016)**
- Wald test, b=.69	**.009****	*.084*	**.000*****	**.000*****
Logit w t linear		.691 (.015)	**.660 (.015)**	**.593 (.017)**
- Wald test, b=.69		.904	**.030***	**.000*****
Logit w t tvcoeff		.696 (.015)	.691 (.015)	.703 (.018)
- Wald test, b=.69		.861	.900	.590
Continuous-time				
Exponential	**.725 (.014)**	**.662 (.015)**	**.595 (.015)**	**.379 (.016)**
- Wald test, b=.69	**.024***	**.038***	**.000*****	**.000*****
Exp. w t linear		.686 (.015)	**.655 (.015)**	**.588 (.017)**
- Wald test, b=.69		.639	**.012***	**.000*****
Exp. w t tvcoeff		.691 (.015)	.687 (.015)	.699 (.018)
- Wald test, b=.69		.869	.666	.743
Weibull	.710 (.014)	.683 (.015)	**.655 (.015)**	**.563 (.017)**
- Wald test, b=.69	.238	.504	**.012***	**.000*****
Weibull w anc p	.706 (.014)	*.668 (.015)*	**.633 (.015)**	**.557 (.017)**
- Wald test, b=.69	.369	*.088*	**.000*****	**.000*****
Cox regression	**.727 (.014)**	.693 (.015)	**.662 (.015)**	**.596 (.017)**
- Wald test, b=.69	**.019***	.975	**.040***	**.000*****
Cox w tvcoeff	.698 (.015)	.692 (.015)	.688 (.015)	.699 (.018)
- Wald test, b=.69	.722	.941	.727	.740

Note: *N* = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D18.

Simulation F1: Non-PH specification, monotonous negative duration dependence, *U* time-constant covariate, left-truncated sample

Estimator	Simulation 1: time-constant baseline hazard ($e^a = 1$)	Simulation 2: mild duration dependence ($e^a = .995$)	Simulation 3: moderate dur. dependence ($e^a = .99$)	Simulation 4: strong duration dependence ($e^a = .975$)
Discrete-time				
Logit	.694 (.019)	.669 (.019)	.662 (.020)	**.616 (.022)**
- Wald test, b=.69	.976	.214	.114	**.000*** **
Logit w t linear		.664 (.019)	**.647 (.020)**	**.578 (.022)**
- Wald test, b=.69		.133	**.019***	**.000*** **
Logit w t tvcoeff		.699 (.020)	.703 (.020)	.693 (.025)
- Wald test, b=.69		.761	.618	.995
Continuous-time				
Exponential	.689 (.019)	.664 (.019)	*.657 (.020)*	**.612 (.022)**
- Wald test, b=.69	.811	.134	*.067*	**.000*** **
Exp. w t linear		*.659 (.019)*	**.642 (.020)**	**.573 (.022)**
- Wald test, b=.69		*.078*	**.009** **	**.000*** **
Exp. w t tvcoeff		.694 (.019)	.698 (.020)	.689 (.025)
- Wald test, b=.69		.963	.793	.870
Weibull	.694 (.019)	.666 (.019)	*.655 (.020)*	**.602 (.022)**
- Wald test, b=.69	.971	.154	*.051*	**.000*** **
Weibull w anc p	.692 (.019)	*.660 (.019)*	**.643 (.020)**	**.575 (.022)**
- Wald test, b=.69	.964	*.088*	**.011***	**.000*** **
Cox regression	.690 (.019)	*.661 (.019)*	**.649 (.020)**	**.592 (.022)**
- Wald test, b=.69	.881	*.096*	**.024***	**.000*** **
Cox w tvcoeff	.714 (.019)	.682 (.019)	.669 (.020)	**.552 (.024)**
- Wald test, b=.69	.267	.571	.231	**.000*** **

Note: N = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at p < .05)

Table D19.

Simulation F2: Non-PH specification, monotonous negative duration dependence, *U* time-varying covariate but uncorrelated with process time, left-truncated sample

Estimator	Simulation 1: time-constant baseline hazard ($e^a = 1$)	Simulation 2: mild duration dependence ($e^a = .995$)	Simulation 3: moderate dur. dependence ($e^a = .99$)	Simulation 4: strong duration dependence ($e^a = .975$)
Discrete-time				
Logit	.675 (.019)	*.658 (.019)*	**.644 (.020)**	**.586 (.022)**
- Wald test, b=.69	.335	*.072*	**.015***	**.000*****
Logit w t linear		*.658 (.019)*	**.651 (.020)**	**.614 (.022)**
- Wald test, b=.69		*.073*	**.036***	**.000*****
Logit w t tvcoeff		.687 (.020)	.691 (.020)	.697 (.024)
- Wald test, b=.69		.751	.912	.867
Continuous-time				
Exponential	.670 (.019)	**.653 (.019)**	**.640 (.020)**	**.582 (.022)**
- Wald test, b=.69	.221	**.040***	**.007****	**.000*****
Exp. w t linear		**.653 (.019)**	**.646 (.020)**	**.609 (.022)**
- Wald test, b=.69		**.040***	**.019****	**.0001*****
Exp. w t tvcoeff		.682 (.019)	.686 (.020)	.693 (.024)
- Wald test, b=.69		.565	.732	.993
Weibull	.668 (.019)	**.653 (.019)**	**.641 (.020)**	**.591 (.022)**
- Wald test, b=.69	.189	**.037***	**.009****	**.000*****
Weibull w anc p	.667 (.019)	**.647 (.019)**	**.637 (.020)**	**.588 (.022)**
- Wald test, b=.69	.164	**.017****	**.005****	**.000*****
Cox regression	.670 (.019)	**.655 (.019)**	**.640 (.020)**	**.591 (.022)**
- Wald test, b=.69	.225	**.047***	**.008****	**.000*****
Cox w tvcoeff	.680 (.019)	.671 (.019)	**.651 (.020)**	**.551 (.024)**
- Wald test, b=.69	.502	.253	**.034***	**.000*****

Note: N = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* $p<.05$, ** $p<.01$, *** $p<.001$; biased parameter estimates in boldface (significance level at least at $p < .05$)

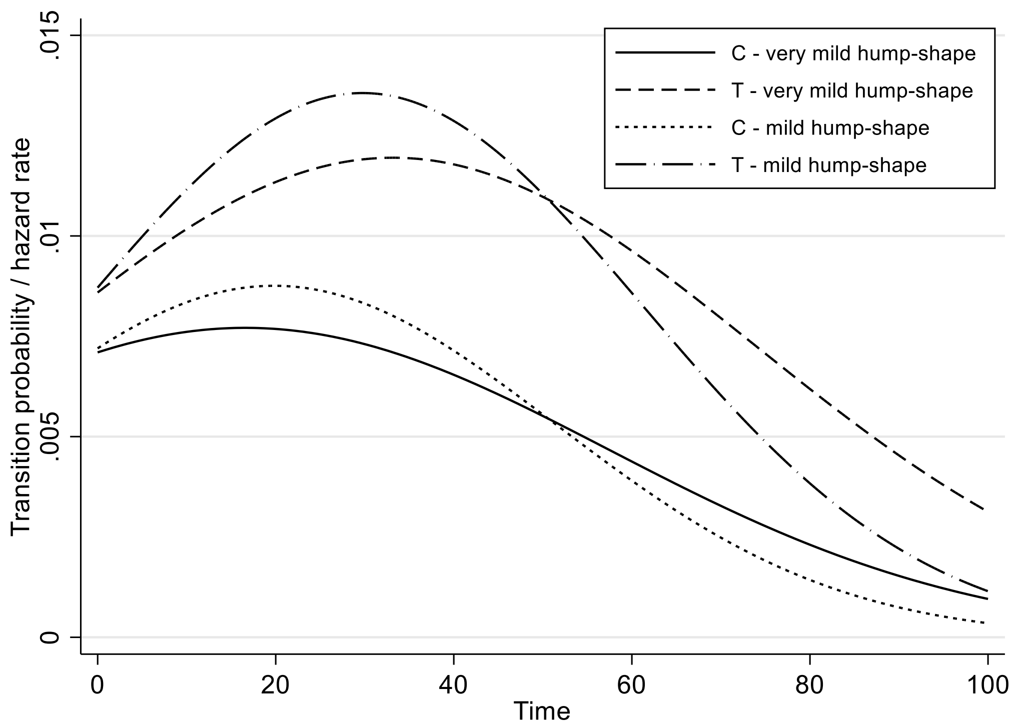
Table D20.

Simulation F3: Non-PH specification, monotonous negative duration dependence, *U* time-varying covariate and positively correlated with process time, left-truncated sample

Estimator	Simulation 1: time-constant baseline hazard (e^a = 1)	Simulation 2: mild duration dependence (e^a = .995)	Simulation 3: moderate dur. dependence (e^a = .99)	Simulation 4: strong duration dependence (e^a = .975)
Discrete-time				
Logit	.685 (.019)	*.655 (.020)*	**.635 (.020)**	**.574 (.022)**
- Wald test, b=.69	.663	*.051*	**.004****	**.000*****
Logit w t linear		*.659 (.020)*	**.648 (.020)**	**.636 (.022)**
- Wald test, b=.69		*.087*	**.027***	**.010****
Logit w t tvcoeff		.679 (.020)	.682 (.021)	.714 (.024)
- Wald test, b=.69		.485	.597	.378
Continuous-time				
Exponential	.680 (.019)	**.650 (.020)**	**.631 (.020)**	**.571 (.022)**
- Wald test, b=.69	.488	**.028***	**.002****	**.000*****
Exp. w t linear		**.654 (.020)**	**.643 (.020)**	**.631 (.022)**
- Wald test, b=.69		**.050***	**.014****	**.005****
Exp. w t tvcoeff		.674 (.020)	.678 (.020)	.710 (.024)
- Wald test, b=.69		.343	.447	.472
Weibull	.675 (.019)	**.649 (.020)**	**.633 (.020)**	**.587 (.022)**
- Wald test, b=.69	.352	**.025***	**.003****	**.000*****
Weibull w anc p	.674 (.019)	**.647 (.020)**	**.633 (.020)**	**.610 (.022)**
- Wald test, b=.69	.334	**.020***	**.003****	**.000*****
Cox regression	.679 (.019)	**.652 (.020)**	**.637 (.020)**	**.592 (.022)**
- Wald test, b=.69	.472	**.039***	**.005****	**.000*****
Cox w tvcoeff	.678 (.019)	.663 (.020)	**.648 (.020)**	**.568 (.024)**
- Wald test, b=.69	.442	.125	**.025***	**.000*****

Note: *N* = 100.000, true b = .693 (hazard ratio of e^b=2); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Scenarios G and H: models featuring bell-shaped baseline hazard function



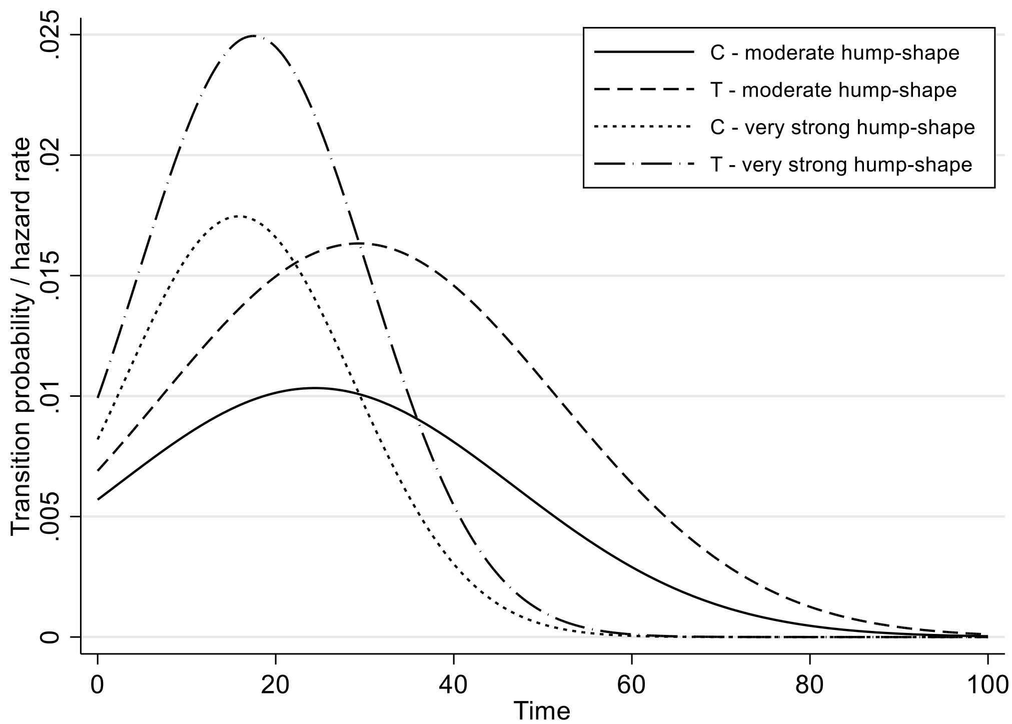


Figure D4.
Non-PH specifications combined with bell-shaped baseline hazards
(C – control group, T – treatment group)

Table D21.

Simulation G1: Non-PH specification, bell-shaped baseline hazard, *U* time-constant covariate

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	**.592 (.014)**	**.573 (.014)**	**.544 (.014)**	**.462 (.015)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Logit w t quad.	**.541 (.014)**	**.511 (.014)**	**.475 (.014)**	**.377 (.015)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Logit w t tvcoeff	.699 (.016)	.698 (.018)	.682 (.023)	.708 (.049)
- Wald test, b=.69	.713	.778	.634	.763
Continuous-time				
Exponential	**.587 (.014)**	**.569 (.014)**	**.540 (.014)**	**.459 (.015)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t quad.	**.536 (.014)**	**.506 (.014)**	**.470 (.014)**	**.371 (.015)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t tvcoeff	.695 (.016)	.694 (.018)	.677 (.023)	.701 (.048)
- Wald test, b=.69	.930	.974	.476	.878
Weibull	**.564 (.014)**	**.539 (.014)**	**.508 (.014)**	**.409 (.015)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Weibull w anc p	**.536 (.014)**	**.505 (.014)**	**.468 (.014)**	**.370 (.015)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Cox regression	**.536 (.014)**	**.506 (.014)**	**.470 (.014)**	**.371 (.015)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Cox w tvcoeff	.695 (.016)	.694 (.018)	.677 (.023)	.700 (.048)
- Wald test, b=.69	.928	.973	.482	.880

Note: *N* = 100.000, true b = .693 (hazard ratio of e^b=2); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001, biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D22.

Simulation G2: Non-PH specification, bell-shaped baseline hazard, *U* time-varying covariate but uncorrelated with process time

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	**.590 (.015)**	**.548 (.015)**	**.524 (.016)**	**.222 (.017)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Logit w t quad.	**.611 (.015)**	**.574 (.015)**	**.530 (.016)**	**.428 (.018)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Logit w t tvcoeff	.706 (.016)	.701 (.018)	.675 (.023)	.694 (.056)
- Wald test, b=.69	.434	.681	.414	.989
Continuous-time				
Exponential	**.585 (.015)**	**.544 (.015)**	**.520 (.015)**	**.220 (.017)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t quad.	**.606 (.015)**	**.569 (.015)**	**.524 (.016)**	**.420 (.018)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t tvcoeff	.701 (.016)	.696 (.018)	.670 (.023)	.691 (.055)
- Wald test, b=.69	.622	.875	.300	.968
Weibull	**.656 (.015)**	**.639 (.015)**	**.628 (.016)**	**.472 (.018)**
- Wald test, b=.69	**.014***	**.0005*** **	**.000*** **	**.000*** **
Weibull w anc p	**.588 (.015)**	**.553 (.015)**	**.513 (.016)**	**.426 (.017)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Cox regression	**.607 (.015)**	**.570 (.015)**	**.525 (.016)**	**.420 (.018)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Cox w tvcoeff	.701 (.016)	.696 (.018)	.670 (.023)	.695 (.055)
- Wald test, b=.69	.610	.862	.310	.975

Note: *N* = 100.000, true b = .693 (hazard ratio of e^b=2); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D23.

Simulation G3: Non-PH specification, bell-shaped baseline hazard, *U* time-varying covariate and positively correlated with process time

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	**.574 (.015)**	**.516 (.016)**	**.477 (.016)**	**.097 (.019)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Logit w t quad.	**.624 (.016)**	**.582 (.016)**	**.534 (.016)**	**.431 (.019)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Logit w t tvcoeff	.705 (.016)	.697 (.018)	.671 (.023)	.703 (.057)
- Wald test, b=.69	.475	.831	.334	.861
Continuous-time				
Exponential	**.570 (.015)**	**.512 (.016)**	**.473 (.016)**	**.096 (.019)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t quad.	**.618 (.015)**	**.577 (.016)**	**.528 (.016)**	**.424 (.019)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t tvcoeff	.700 (.016)	.692 (.018)	.666 (.023)	.700 (.057)
- Wald test, b=.69	.669	.969	.238	.902
Weibull	**.651 (.015)**	**.622 (.016)**	**.596 (.016)**	**.372 (.019)**
- Wald test, b=.69	**.007** **	**.000*** **	**.000*** **	**.000*** **
Weibull w anc p	**.606 (.015)**	**.570 (.016)**	**.530 (.016)**	**.432 (.019)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Cox regression	**.619 (.015)**	**.578 (.016)**	**.529 (.016)**	**.424 (.019)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Cox w tvcoeff	.700 (.016)	.693 (.018)	.666 (.023)	.703 (.057)
- Wald test, b=.69	.660	.977	.240	.860

Note: N = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at $p < .05$)

Table D24.
Simulation H1: Non-PH specification, bell-shaped baseline hazard, *U* time-constant covariate, left-truncated sample

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	**.648 (.019)**	**.627 (.020)**	**.614 (.020)**	**.518 (.025)**
- Wald test, b=.69	**.020***	**.001*****	**.000*****	**.000*****
Logit w t quad.	**.619 (.019)**	**.588 (.020)**	**.564 (.020)**	**.448 (.025)**
- Wald test, b=.69	**.000*****	**.000*****	**.000*****	**.000*****
Logit w t tvcoeff	.700 (.021)	.703 (.023)	.699 (.027)	.720 (.067)
- Wald test, b=.69	.733	.655	.831	.694
Continuous-time				
Exponential	**.643 (.019)**	**.622 (.020)**	**.610 (.020)**	**.515 (.025)**
- Wald test, b=.69	**.009****	**.000*****	**.000*****	**.000*****
Exp. w t quad.	**.614 (.019)**	**.582 (.020)**	**.558 (.020)**	**.441 (.025)**
- Wald test, b=.69	**.000*****	**.000*****	**.000*****	**.000*****
Exp. w t tvcoeff	.695 (.021)	.699 (.023)	.694 (.027)	.716 (.067)
- Wald test, b=.69	.918	.815	.987	.734
Weibull	**.638 (.019)**	**.612 (.020)**	**.597 (.020)**	**.494 (.025)**
- Wald test, b=.69	**.004****	**.000*****	**.000*****	**.000*****
Weibull w anc p	**.615 (.019)**	**.582 (.020)**	**.556 (.020)**	**.431 (.025)**
- Wald test, b=.69	**.000*****	**.000*****	**.000*****	**.000*****
Cox regression	**.628 (.019)**	**.601 (.020)**	**.583 (.020)**	**.481 (.025)**
- Wald test, b=.69	**.001*****	**.000*****	**.000*****	**.000*****
Cox w tvcoeff	**.624 (.021)**	**.572 (.022)**	**.474 (.024)**	**-.349 (.048)**
- Wald test, b=.69	**.001*****	**.000*****	**.000*****	**.000*****

Note: *N* = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D25.

Simulation H2: Non-PH specification, bell-shaped baseline hazard, *U* time-varying covariate but uncorrelated with process time, left-truncated sample

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	**.628 (.020)**	**.627 (.020)**	**.597 (.020)**	**.402 (.026)**
- Wald test, b=.69	**.001*** **	**.001*** **	**.000*** **	**.000*** **
Logit w t quad.	**.616 (.020)**	**.606 (.020)**	**.570 (.020)**	**.500 (.026)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Logit w t tvcoeff	.686 (.021)	.714 (.023)	.683 (.027)	**.837 (.070)**
- Wald test, b=.69	.722	.370	.715	**.041* **
Continuous-time				
Exponential	**.623 (.020)**	**.622 (.020)**	**.593 (.020)**	**.400 (.026)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t quad.	**.611 (.020)**	**.600 (.020)**	**.564 (.020)**	**.492 (.026)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Exp. w t tvcoeff	.681 (.021)	.709 (.023)	.678 (.027)	**.834 (.070)**
- Wald test, b=.69	.557	.498	.580	**.043* **
Weibull	**.626 (.020)**	**.627 (.020)**	**.601 (.020)**	**.437 (.026)**
- Wald test, b=.69	**.001*** **	**.001** **	**.000*** **	**.000*** **
Weibull w anc p	**.612 (.020)**	**.608 (.020)**	**.592 (.020)**	**.545 (.026)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Cox regression	**.623 (.020)**	**.621 (.020)**	**.592 (.020)**	**.441 (.026)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **
Cox w tvcoeff	**.611 (.021)**	**.585 (.022)**	**.454 (.024)**	**-.484 (.051)**
- Wald test, b=.69	**.000*** **	**.000*** **	**.000*** **	**.000*** **

Note: *N* = 100.000, true b = .693 (hazard ratio of e^{b} = 2); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at *p* < .05)

Table D26.

Simulation H3: Non-PH specification, bell-shaped baseline hazard, *U* time-varying covariate and positively correlated with process time, left-truncated sample

Estimator	Simulation 1: very mild hump-shape (e^{a1} = 1.01, e^{a2} = .9997)	Simulation 2: mild hump-shape (e^{a1} = 1.02, e^{a2} = .9995)	Simulation 3: moderate hump-shape (e^{a1} = 1.05, e^{a2} = .999)	Simulation 4: very strong hump-shape (e^{a1} = 1.10, e^{a2} = .997)
Discrete-time				
Logit	*.660 (.020)*	**.606 (.020)**	**.551 (.021)**	**.275 (.028)**
- Wald test, b=.69	*.099*	**.000*****	**.000*****	**.000*****
Logit w t quad.	.665 (.020)	**.616 (.020)**	**.555 (.021)**	**.453 (.028)**
- Wald test, b=.69	.154	**.000*****	**.000*****	**.000*****
Logit w t tvcoeff	.726 (.021)	.691 (.023)	.669 (.027)	.795 (.072)
- Wald test, b=.69	.110	.924	.359	.156
Continuous-time				
Exponential	*.655 (.020)*	**.602 (.020)**	**.547 (.021)**	**.273 (.028)**
- Wald test, b=.69	*.055*	**.000*****	**.000*****	**.000*****
Exp. w t quad.	*.659 (.020)*	**.610 (.020)**	**.549 (.021)**	**.446 (.028)**
- Wald test, b=.69	*.086*	**.000*****	**.000*****	**.000*****
Exp. w t tvcoeff	.721 (.021)	.686 (.023)	.664 (.027)	.793 (.072)
- Wald test, b=.69	.177	.756	.271	.164
Weibull	.661 (.020)	**.612 (.020)**	**.564 (.021)**	**.332 (.028)**
- Wald test, b=.69	.102	**.000*****	**.000*****	**.000*****
Weibull w anc p	.663 (.020)	**.626 (.020)**	**.583 (.021)**	**.500 (.028)**
- Wald test, b=.69	.123	**.001*****	**.000*****	**.000*****
Cox regression	.662 (.020)	**.612 (.020)**	**.565 (.021)**	**.360 (.028)**
- Wald test, b=.69	.114	**.000*****	**.000*****	**.000*****
Cox w tvcoeff	*.656 (.020)*	**.563 (.022)**	**.444 (.024)**	–**.528 (.053)**
- Wald test, b=.69	*.069*	**.000*****	**.000*****	**.000*****

Note: N = 100.000, true b = .693 (hazard ratio of $e^b = 2$); Wald tests of b = ln(2)
* p<.05, ** p<.01, *** p<.001; biased parameter estimates in boldface (significance level at least at $p < .05$)