Online appendix to

"Intertemporal Substitution in Labor Supply: A Meta-Analysis"*

Ali Elminejad[a], Tomas Havranek[a,b,c], Roman Horvath[a], and Zuzana Irsova[a]

[a]Charles University, Prague

[b]Centre for Economic Policy Research, London

[c]Meta-Research Innovation Center at Stanford

October 4, 2023

Abstract

The online appendix provides a meta-analysis of Frisch elasticities at the intensive margins, details on data collection, and robustness checks.

Keywords: Frisch elasticity, labor supply, meta-analysis, publication bias, Bayesian model averaging

JEL Codes: C83, E24, J21

*Corresponding author: Zuzana Irsova, zuzana.irsova@ies-prague.org.

A Intensive Margin Elasticities

This appendix summarizes the meta-analysis of intensive margin elasticities. Our approach here is analogous to the meta-analysis of extensive margin elasticities presented in the main body of the paper, so we only briefly describe the results. All the concepts and techniques are explained in detail in the main body of the paper; the reader should inspect these sections before turning to this appendix. Again we use Google Scholar to search for the estimates of Frisch elasticities at the intensive margin, and the details of the search strategy are described in Figure B2. We find 40 studies, listed in Table A1, which together provide 709 estimates of the intensive margin elasticity; details on the extraction of estimates from individual studies are available in Table B1. For comparison, on the extensive margin elasticity we found 38 studies with 762 estimates, so the size of the dataset is almost the same. But for the intensive margin we only have 8 quasi-experimental studies, compared to 14 for the extensive margin. The relative scarcity of quasi-experimental evidence for the intensive margin elasticity compared to the extensive margin elasticity was noted by Chetty *et al.* (2013) and persists to this day.

As shown in Figure A1, the reported intensive margin elasticities are most commonly between 0 and 0.7, and their density is relatively flat in this interval. The mean is about 0.5 and the median 0.4. Estimated elasticities below -0.1 and above 1 are quite rare in the literature. Note the jump in the distribution at 0, which is consistent with bias against negative estimates of the elasticity; we observed a similar pattern for the extensive margin. Figure A2 shows some stylized facts in the data. Similarly to the extensive margin, estimates corresponding to workers near retirement are larger than estimates corresponding to prime-age workers. Estimates are larger for women than men and for macro data than micro data. In contrast to the extensive margin, however, for the intensive margin quasi-experimental estimates tend to be substantially larger than the rest of the micro estimates. For the intensive margin, quasi-experimental evidence does not contradict macro evidence, which was also noted by Chetty *et al.* (2013). We confirm that this finding holds with more recent data, and additionally the mean of quasi-experimental estimates (0.6) is similar to that reported by Chetty *et al.* (2013, 0.54).

But the mean of reported estimates is a misleading statistic affected in many fields (including the extensive margin Frisch elasticity, as we showed in the main body of the paper) by publication selection bias. Once again we find evidence of this bias, as apparent from Figure A3 and Table A2. The funnel plot is clearly asymmetrical, though perhaps less so than in the case of the extensive margin. All statistical tests find evidence of publication bias, and the mean elasticities corrected for this bias range between 0.2 and 0.4, with a median of 0.3. This finding implies a slightly weaker publication bias for the intensive margin compared to elasticities at the extensive margin: for both margins, the mean reported (uncorrected) elasticity is around 0.5. After correction for the bias (and ignoring for a while methodology and demographics considerations that also affect the estimates), the mean estimate is a bit smaller for the extensive margin (about 0.25) than for the intensive margin (about 0.3). One potential explanation is that with a larger underlying effect (intensive margin elasticity), less p-hacking is needed to produce statistically significant estimates.

In Table A3, we repeat the analysis of publication bias previously reported in Table A2 for two subsamples: quasi-experimental estimates and IV estimates with first-stage robust F-statistics above 10. Many authors would consider those two groups of studies as especially relevant for a proper identification of the underlying intensive margin elasticity. In addition, Keane & Neal (2023) show that for instrumental variables, estimates and standard errors are correlated by construction when instruments are weak. So we need to check whether the correlation persists even for strong instruments. (They recommend a much larger cut-off for first-stage F-statistic than the commonly used 10, but that would leave only a handful of papers in the subsample.) Even with a much reduced sample, almost all specifications in Table A2 find evidence of publication selection bias. For quasi-experimental estimates, the corrected mean effect ranges between 0 and 0.25, with a median of 0.1. For IV estimates with relatively strong instruments (first-stage F-statistics above 10), the corrected mean ranges between 0.2 and 0.6, with a median of 0.3. We conclude that evidence for publication bias is solid in the case of intensive margin elasticities, and values between 0.1 and 0.3 can be quite easily defended for the calibration of representative agent models.

Next, we focus on heterogeneity in the estimated elasticities. Table A4 summarizes the variables that reflect the context in which intensive margin elasticities are estimated; the variables are the same as in the case of the extensive margin with the exception of a few that had to be omitted (*Ratio, Indivisible, Probit*) due to their limited variation in the intensive elasticity dataset, lack of relevance, or high correlation with other variables. The relatively modest correlations of the remaining variables are shown in Figure A4. Table A5 and Figure A5 report the results of Bayesian model averaging. BMA corroborates publication bias among intensive margin elasticities. Similarly to the extensive margin, for the intensive margin macro estimates tend to be larger than micro estimates, prime-age workers display smaller elasticities than workers near retirement, and women display larger elasticities than men. In contrast to the extensive margin, for the intensive margin data frequency can be important, recent studies tend to report estimates larger than those in older studies, estimates for the US are larger than for other countries, and quasi-experimental estimates are larger than other micro estimates. The results hold across several robustness checks, Bayesian or frequentist, reported in Table A7.

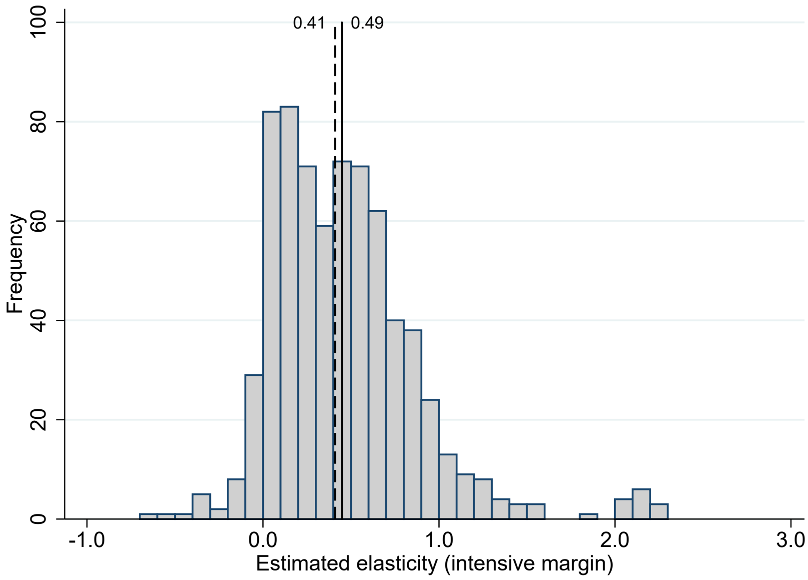
The bottom line of the meta-analysis of intensive margin elasticities is reported, together with the corresponding evidence for the extensive margin, in the main body of the paper. The table presents implied elasticities in various contexts: that is, mean elasticities corrected for publication bias and conditional on a definition of best practice methodology. The definition is then plugged into the results of the model averaging exercise, from which fitted values for the estimated elasticities are computed. The overall mean implied elasticity at the intensive margin is 0.24 when using our subjective definition of best practice and 0.18 when defining best practice according to Martinez *et al.* (2021), a large recent quasi-experimental study published in the American Economic Review.

To avoid spurious precision, we recommend 0.2 for the calibration of the intensive margin elasticity in representative agent models. As we have noted earlier, this value is also in the middle of the interval consistent with bias-corrected means for quasi-experimental estimates and structural estimates with strong instruments. The intensive margin elasticity is larger for women and workers near retirement. Single workers seem to have smaller intensive margin elasticities, but this result should be interpreted with caution because the corresponding variable in BMA has a posterior inclusion probability smaller than 0.75, and only a small fraction of studies focus on single workers in the context of the intensive margin elasticity of intertemporal substitution in labor supply.

Table A1: Studies included in the meta-analysis of intensive margin elasticities

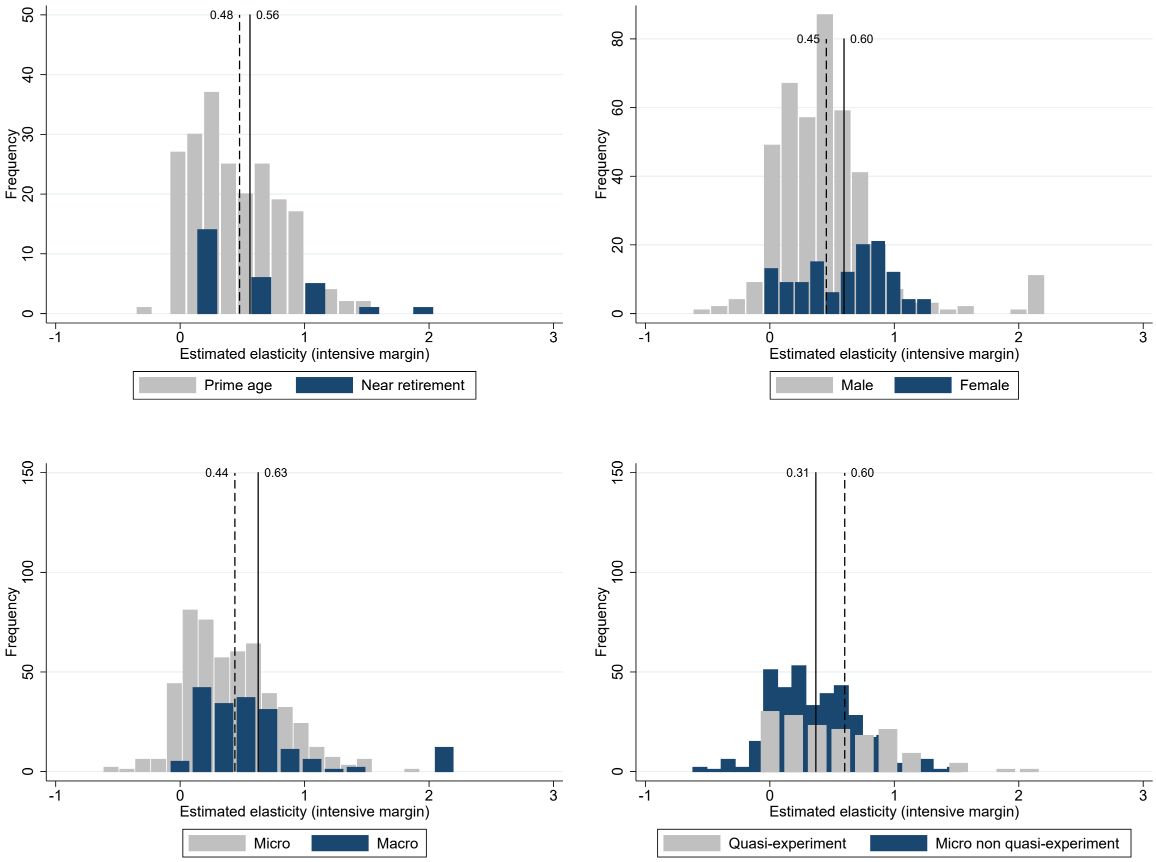
Aaronson & French (2009)	Ham & Reilly (2002)
Altonji (1986)	Inoue (2015)
Angrist (1991)	Karabarbounis (2016)
Angrist *et al.* (2021)	Keane & Wasi (2016)
Attanasio *et al.* (2018)	Kimmel & Kniesner (1998)
Battisti *et al.* (2023)	Kneip *et al.* (2019)
Beffy *et al.* (2019)	Kuroda & Yamamoto (2008)
Blundell *et al.* (2016a)	Lee (2001)
Blundell *et al.* (2016b)	Looney & Singhal (2006)
Borella *et al.* (2023)	MaCurdy (1981)
Bredemeier *et al.* (2019)	Martinez *et al.* (2021)
Caldwell & Oehlsen (2022)	Ong (2019)
Chang *et al.* (2011)	Peterman (2016)
Domeij & Floden (2006)	Pistaferri (2003)
Erosa *et al.* (2016)	Saez (2003)
Farber (2015)	Sigurdsson (2023)
Fiorito & Zanella (2012)	Stafford (2015)
French (2005)	Theloudis (2021)
French & Stafford (2017)	Wallenius (2011)
Haan & Uhlendorff (2013)	Ziliak & Kniesner (2005)

Figure A1: Estimates between 0 and 0.7 are almost equally common



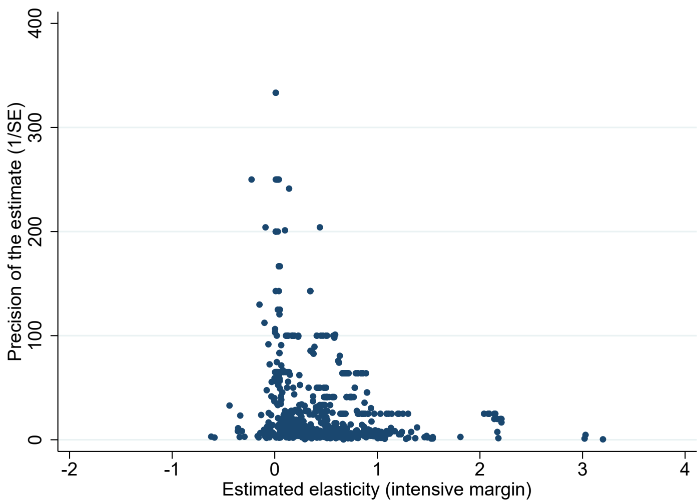
Notes: The solid line denotes the sample mean (0.49); the dashed line denotes the sample median (0.41). Note the jump at 0. Estimates smaller than -1 and larger than 3 are excluded from the figure for ease of exposition but included in all tests.

Figure A2: Stylized facts in the data



Notes: The dashed line denotes the mean elasticity for the subset mentioned first in the legend (depicted in light gray); the solid line denotes the mean for the second subset (dark). Estimates smaller than −1 and larger than 3 are excluded from the figure for ease of exposition but included in all tests.

Figure A3: The funnel plot suggests publication bias



Notes: In the absence of bias the plot should form a symmetrical funnel. Extreme values are excluded from the figure for ease of exposition but included in all tests.

Table A2: Linear and nonlinear tests document publication bias

Panel A: Linear tests					
	OLS	FE	Precision	Study	MAIVE
Publication bias	0.590**	0.928***	1.179***	0.780***	5.163**
(*Standard error*)	(0.266)	(0.110)	(0.440)	(0.257)	(2.159)
	[-0.01, 1.22]	-	[0.23, 2.17]	[0.23, 1.38]	{0.73, 3.72}
Effect beyond bias	0.373***	0.329***	0.297***	0.331***	0.279***
(*Constant*)	(0.0567)	(0.0170)	(0.0666)	(0.0467)	(0.0505)
	[0.24, 0.49]	-	[0.14, 0.50]	[0.23, 0.43]	{0.04, 0.20}
First stage F-stat					9.9
Observations	709	709	709	709	663
Studies	40	40	40	40	39

Panel B: Nonlinear tests					
	Ioannidis *et al.* (2017)	Andrews & Kasy (2019)	Bom & Rachinger (2019)	Furukawa (2021)	van Aert & van Assen (2023)
Effect beyond bias	0.199***	0.295***	0.213***	0.343***	0.387***
	(0.045)	(0.003)	(0.014)	(0.126)	(0.065)
Observations	709	709	709	709	709
Studies	40	40	40	40	40

Notes: Panel A presents the results of regression $\hat{\eta}_{ij} = \eta_0 + \delta \cdot SE(\hat{\eta}_{ij}) + e_{ij}$, where $\hat{\eta}_{ij}$ and $SE(\hat{\eta}_{ij})$ are the i-th estimated Frisch intensive margin elasticity and its standard error reported in the j-th study. OLS = ordinary least squares. FE = study fixed effects. Precision = estimates are weighted by the inverse of their variance. Study = estimates are weighted by the inverse of the number of estimates reported per study. MAIVE = meta-analysis instrumental variable estimator (Irsova *et al.*, 2023); the inverse of the square root of the number of observations is used as an instrument for the standard error (the number of observations is not available for all studies). We cluster standard errors at the study level; if applicable, we also report 95% confidence intervals from wild bootstrap clustering in square brackets. For MAIVE, in curly brackets we show the weak-instrument-robust Anderson-Rubin 95% confidence intervals. Panel B presents the mean elasticity corrected for publication bias using nonlinear techniques. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Publication bias in subsamples of the literature

Part 1: Quasi-experimental estimates

Panel A: Linear tests

	OLS	FE	Precision	Study	MAIVE
Publication bias	1.744***	2.036***	2.803***	1.703***	8.698*
(*Standard error*)	(0.630)	(0.251)	(0.624)	(0.521)	(4.616)
	[-0.32, 3.55]	-	[1.15, 3.92]	[0.63, 3.43]	{-0.35, 17.75}
Effect beyond bias	0.224**	0.176***	0.0513	0.224***	0.121
(*Constant*)	(0.111)	(0.0451)	(0.0470)	(0.0610)	(0.0837)
	[0.04, 0.59]	-	[-0.78, 0.51]	[0.08, 0.52]	{-0.004, 0.25}
First stage F-stat					3.1
Observations	162	162	162	162	132
Studies	8	8	8	8	8

Panel B: Nonlinear tests

	Ioannidis *et al.* (2017)	Andrews & Kasy (2019)	Bom & Rachinger (2019)	Furukawa (2021)	van Aert & van Assen (2023)
Effect beyond bias	0.028	-0.027	-0.002	0.234*	0.155
	(NA)	(0.02)	(0.008)	(0.122)	(0.504)
Observations	162	162	162	162	162
Studies	8	8	8	8	8

Part 2: IV estimates with first-stage F-statistics > 10

Panel A: Linear tests

	OLS	FE	Precision	Study	MAIVE
Publication bias	0.523**	0.327	0.728**	0.692**	-3.393*
(*Standard error*)	(0.239)	(0.335)	(0.309)	(0.293)	(1.821)
	[-0.10, 0.93]	-	[-0.20, 0.96]	[-0.13, 1.20]	{-8.98, -0.69}
Effect beyond bias	0.285***	0.327***	0.246***	0.262***	0.587***
(*Constant*)	(0.0586)	(0.0724)	(0.0714)	(0.0620)	(0.110)
	[0.21, 0.45]	-	[0.19, 0.48]	[0.21, 0.45]	{0.12, 1.55}
First stage F-stat					19.2
Observations	92	92	92	92	92
Studies	6	6	6	6	6

Panel B: Nonlinear tests

	Ioannidis *et al.* (2017)	Andrews & Kasy (2019)	Bom & Rachinger (2019)	Furukawa (2021)	van Aert & van Assen (2023)
Effect beyond bias	0.247**	0.421***	0.204***	0.277**	0.375***
	(0.112)	(0.06)	(0.055)	(0.121)	(0.145)
Observations	92	92	92	92	92
Studies	6	6	6	6	6

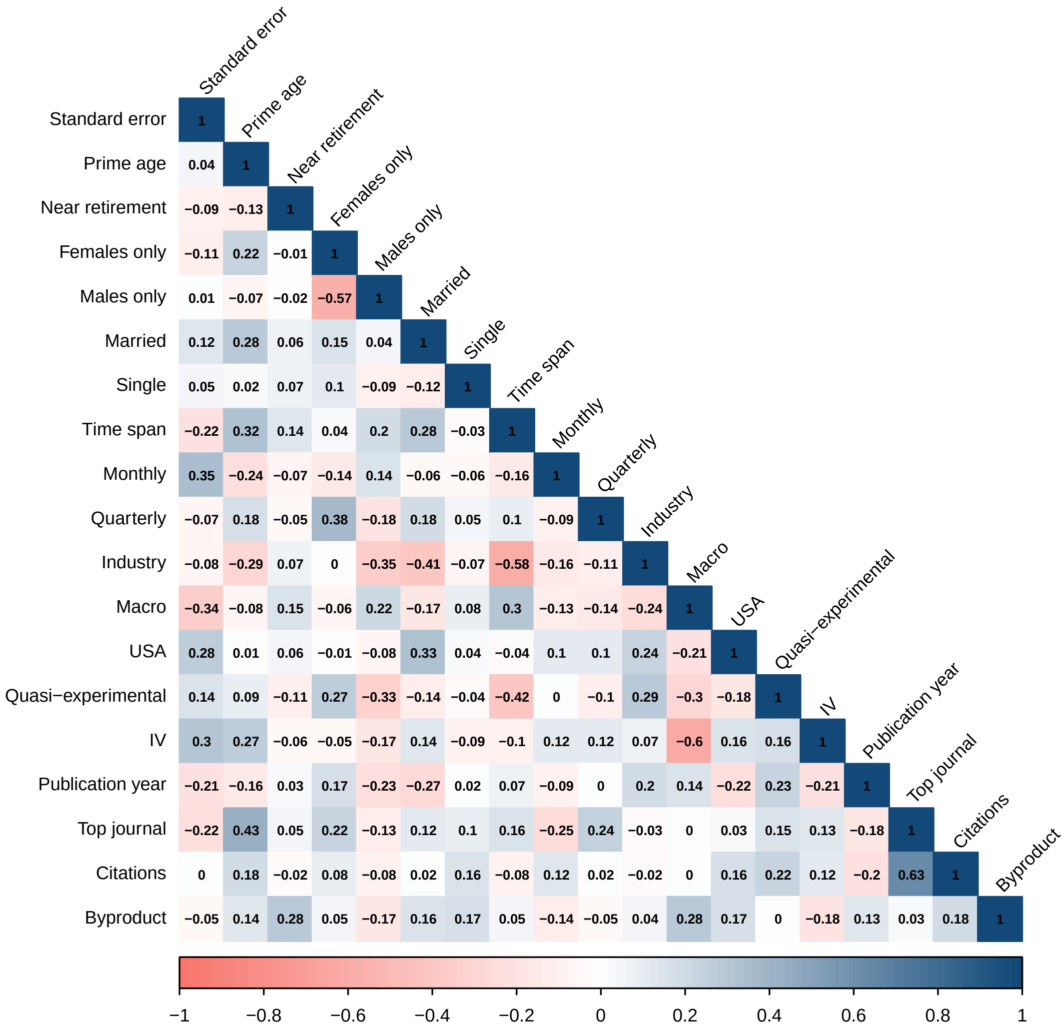
Notes: Panel A presents the results of regression $\hat{\eta}_{ij} = \eta_0 + \delta \cdot SE(\hat{\eta}_{ij}) + e_{ij}$, where $\hat{\eta}_{ij}$ and $SE(\hat{\eta}_{ij})$ are the i-th estimated Frisch intensive margin elasticity and its standard error reported in the j-th study. OLS = ordinary least squares. FE = study fixed effects. Precision = estimates are weighted by the inverse of their variance. Study = estimates are weighted by the inverse of the number of estimates reported per study. MAIVE = meta-analysis instrumental variable estimator (Irsova *et al.*, 2023); the inverse of the square root of the number of observations is used as an instrument for the standard error (the number of observations is not available for all studies). In square brackets we report 95% confidence intervals from wild bootstrap clustering. In curly brackets we show the Anderson-Rubin 95% confidence intervals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Definition and summary statistics of regression variables

Variable	Description	Mean	SD
Frisch elasticity	Estimate of the intensive margin Frisch elasticity (response variable).	0.49	0.55
Standard error (SE)	Standard error of the estimate (the variable is important for gauging publication bias).	0.15	0.22
Demographic characteristics			
Prime age	= 1 if the sample only consists of people between 25 and 55 years of age.	0.30	0.46
Near retirement	= 1 if the sample only consists of people older than 55.	0.04	0.19
Females only	= 1 if the sample consists of females only.	0.18	0.38
Males only	= 1 if the sample consists of males only.	0.60	0.49
Married	= 1 if the sample consists of married people only.	0.47	0.50
Single	= 1 if the sample consists of single people only.	0.02	0.15
Data characteristics			
Time span	The logarithm of the data time span used to estimate the elasticity.	2.55	0.84
Monthly	= 1 if the data frequency is monthly (reference category: annual).	0.12	0.32
Quarterly	= 1 if the data frequency is quarterly (reference category: annual).	0.06	0.23
Industry	= 1 if the sample consists of workers in a specific industry (reference category: whole economy data).	0.16	0.37
Macro	= 1 if the estimate uses aggregated data (reference category: micro).	0.26	0.44
USA	= 1 if the estimate uses data for the US.	0.77	0.42
Specification characteristics			
Quasi-experimental	= 1 if the estimation framework uses quasi-experimental identification.	0.23	0.42
IV	= 1 if instrumental variable methods are used for the estimate (reference category: OLS).	0.56	0.50
Publication characteristics			
Publication year	The logarithm of the year the study was published.	3.42	0.53
Top journal	= 1 if the estimate is published in a top five journal in economics.	0.32	0.47
Citations	The logarithm of the number of per-year citations of the study in Google Scholar.	2.05	1.42
Byproduct	= 1 if the information reported in the study allows for the computation of the elasticity but the elasticity is not interpreted in the paper.	0.13	0.33

Notes: SD = standard deviation. The table excludes the definition and summary statistics of the reference categories, which are omitted from the regressions.

Figure A4: Correlations among explanatory variables



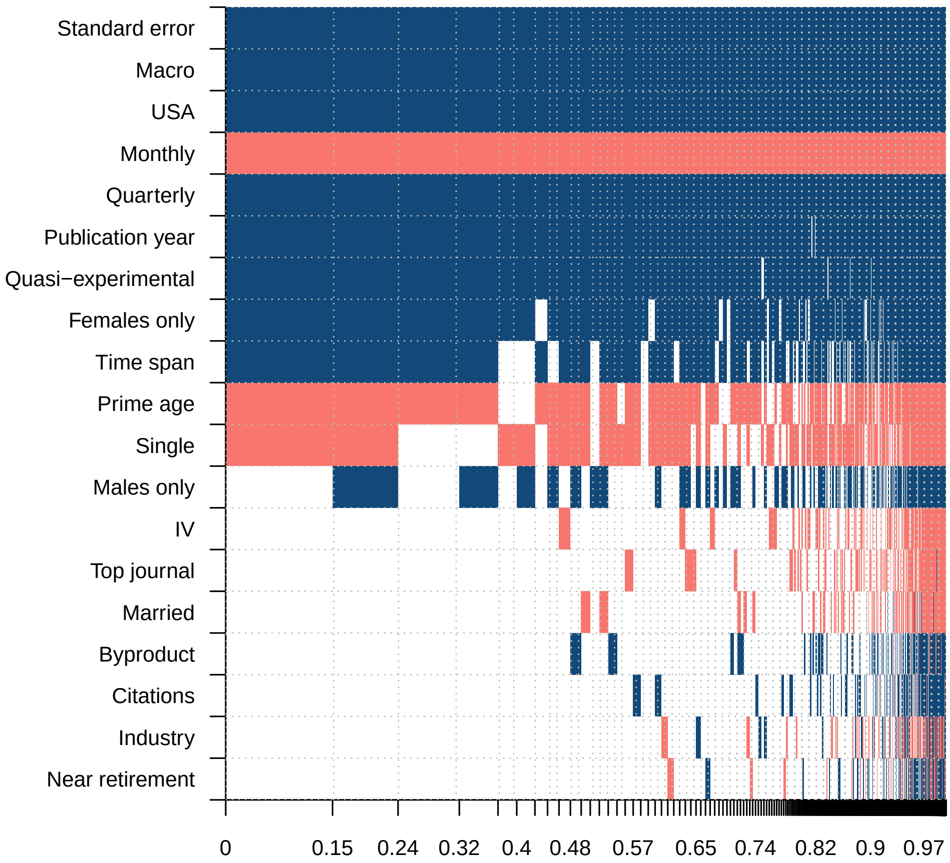
Notes: The figure shows Pearson correlation coefficients for the variables described in Table A4; only intensive margin elasticities are used for the computation.

Table A5: Why do estimates of the elasticity vary?

Response variable: Frisch elasticity (intensive margin)	Bayesian model averaging (baseline model)			Ordinary least squares (frequentist check)		
	P. mean	P. SD	PIP	Mean	SE	p-value
Intercept	-0.405	NA	1.000	-0.391	0.190	0.046
Standard error	1.025	0.104	1.000	1.022	0.222	0.000
Demographic characteristics						
Prime age	-0.073	0.047	0.787	-0.098	0.062	0.122
Near retirement	0.001	0.016	0.060			
Females only	0.122	0.055	0.924	0.106	0.067	0.122
Males only	0.028	0.039	0.408			
Married	-0.002	0.012	0.089			
Single	-0.137	0.115	0.665			
Data characteristics						
Time span	0.044	0.028	0.799	0.062	0.038	0.112
Monthly	-0.190	0.040	1.000	-0.185	0.071	0.013
Quarterly	0.261	0.058	0.999	0.260	0.193	0.186
Industry	-0.001	0.017	0.075			
Macro	0.252	0.032	1.000	0.251	0.066	0.001
USA	0.208	0.030	1.000	0.203	0.070	0.006
Specification characteristics						
Quasi-experimental	0.157	0.045	0.988	0.171	0.080	0.039
IV	-0.003	0.014	0.101			
Publication characteristics						
Publication year	0.101	0.029	0.991	0.090	0.046	0.060
Top journal	-0.003	0.015	0.096			
Citations	0.001	0.004	0.076			
Byproduct	0.003	0.016	0.089			
Observations	709			709		
Studies	40			40		

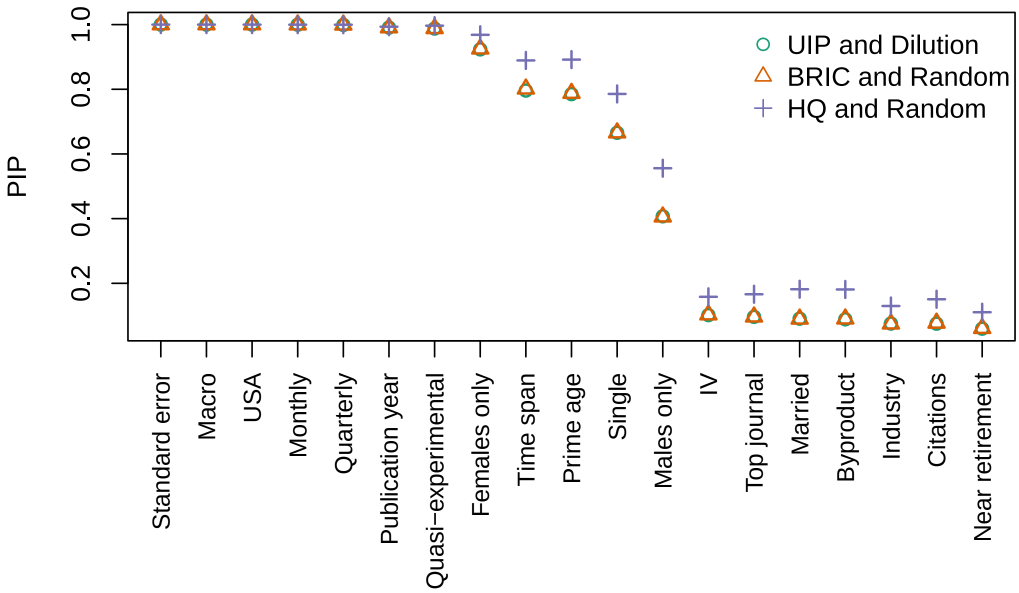
Notes: The response variable is the Frisch elasticity of labor supply at the intensive margin. P. mean = posterior mean, P. SD = posterior standard deviation, PIP = Posterior inclusion probability, SE = standard error. The left-hand panel applies BMA based on the UIP g-prior and the dilution prior (Eicher *et al.* 2011; George 2010). The right-hand panel reports a frequentist check using OLS, which includes variables with PIPs higher than 0.75 in BMA. Standard errors in the frequentist check are clustered at the study level. Table A4 presents a detailed description of all the variables.

Figure A5: Model inclusion in Bayesian model averaging (UIP and dilution prior)



Notes: The response variable is the reported estimate of the Frisch elasticity of labor supply at the intensive margin. The columns denote individual models; variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes the cumulative posterior model probabilities. The estimation is based on the unit information prior (UIP) recommended by Eicher *et al.* (2011) and the dilution prior suggested by George (2010), which takes collinearity into account. Blue color (darker in grayscale) = the variable has a positive estimated sign. Red color (lighter in grayscale) = the variable has a negative estimated sign. No color = the variable is excluded from the given model. Table A4 presents a detailed description of all variables. The numerical results are reported in Table A7.

Figure A6: Posterior inclusion probabilities hold across different priors



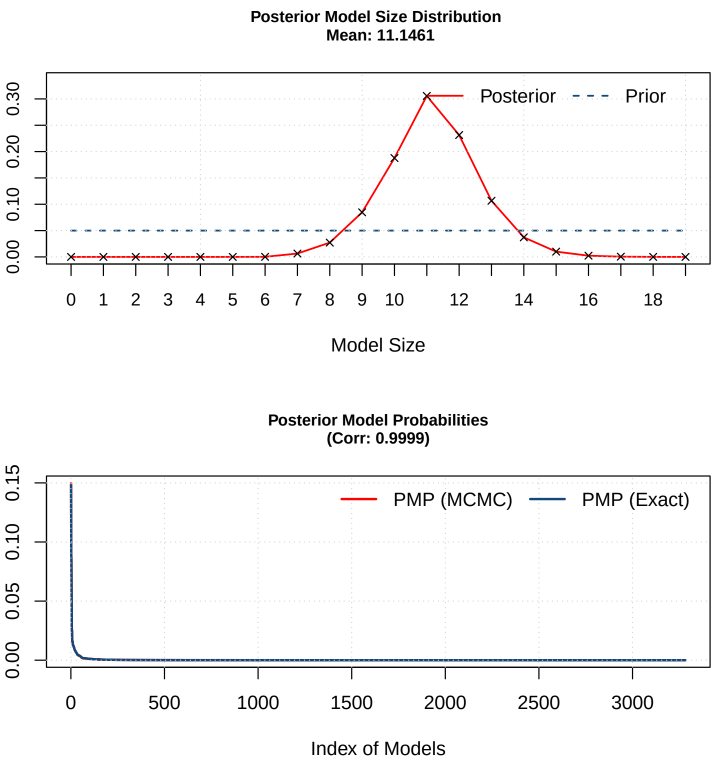
Notes: UIP and Dilution = priors according to Eicher *et al.* (2011) and George (2010). BRIC and Random = the benchmark g-prior for parameters with the beta-binomial model prior (each model size has equal prior probability). The HQ prior asymptotically mimics the Hannan-Quinn criterion. PIP = posterior inclusion probability.

Table A6: Summary of the BMA estimation (UIP and dilution prior)

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
11.1461	$3 \cdot 10^6$	$1 \cdot 10^6$	12.08 mins	688,859
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
$5.24 \cdot 10^5$	131.0%	100%	0.9999	709
Model prior	*g-prior*	*Shrinkage-stats*		
Random/9.5	UIP	Av = 0.9986		

Notes: The results of this BMA specification are reported in Table A5. Based on Eicher *et al.* (2011) we employ unit information prior and, as suggested by George (2010), the dilution prior that takes into account potential collinearity.

Figure A7: Model size and convergence in the BMA model (UIP and dilution prior)



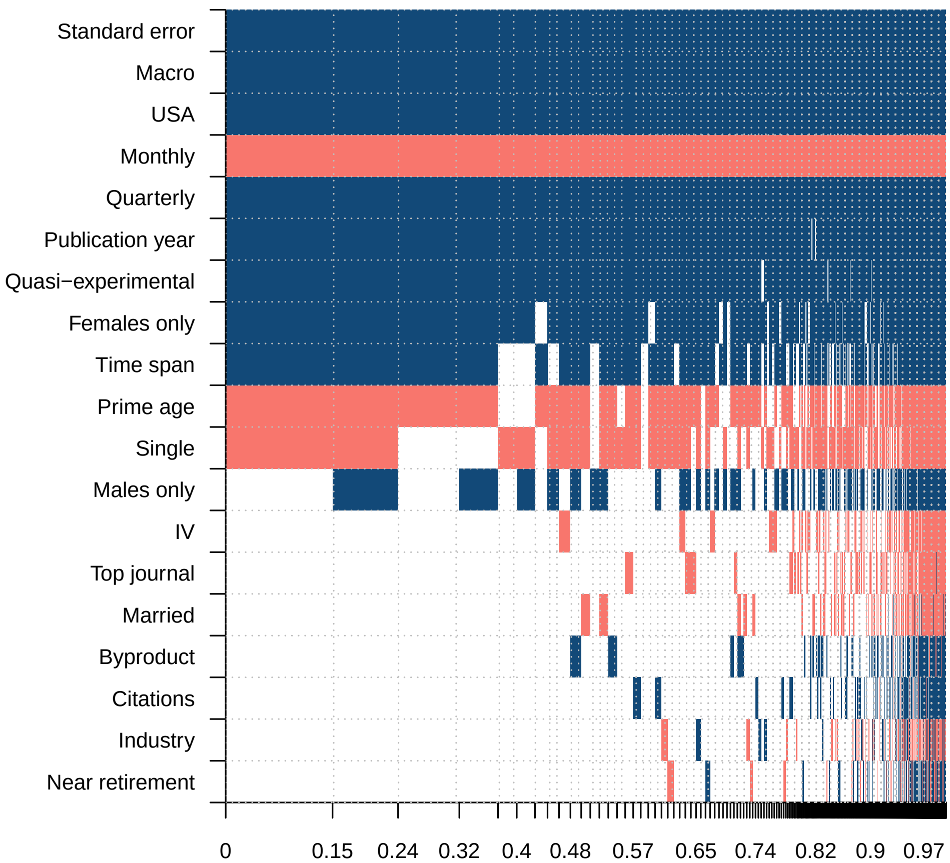
Notes: The figure depicts the posterior model size distribution and the posterior model probabilities of the BMA exercise reported in Table A5.

Table A7: Results of BMA with alternative priors and results of FMA

Response variable: Frisch elasticity (intensive margin)	Bayesian model averaging (BRIC g-prior)			Bayesian model averaging (HQ g-prior)			Frequentist model averaging		
	P. mean	P. SD	PIP	P. mean	P. SD	PIP	Coef.	SE	p-value
Intercept	-0.405	NA	1.000	-0.413	NA	1.000	-0.417	0.114	0.000
Standard error	1.025	0.104	1.000	1.032	0.104	1.000	1.042	0.111	0.000
Demographic characteristics									
Prime age	-0.073	0.047	0.788	-0.083	0.041	0.892	-0.089	0.034	0.008
Near retirement	0.001	0.016	0.061	0.001	0.021	0.111	0.010	0.064	0.876
Females only	0.122	0.055	0.925	0.135	0.050	0.968	0.169	0.043	0.000
Males only	0.028	0.039	0.406	0.038	0.041	0.556	0.071	0.032	0.029
Married	-0.002	0.012	0.090	-0.006	0.018	0.182	-0.043	0.033	0.188
Single	-0.137	0.115	0.666	-0.162	0.108	0.786	-0.238	0.082	0.004
Data characteristics									
Time span	0.045	0.028	0.801	0.049	0.025	0.889	0.061	0.022	0.005
Monthly	-0.190	0.040	1.000	-0.195	0.040	1.000	-0.221	0.044	0.000
Quarterly	0.261	0.058	0.999	0.261	0.057	1.000	0.279	0.058	0.000
Industry	-0.001	0.017	0.075	-0.001	0.022	0.130	-0.009	0.060	0.876
Macro	0.252	0.032	1.000	0.246	0.033	1.000	0.211	0.039	0.000
USA	0.208	0.030	1.000	0.208	0.031	1.000	0.211	0.040	0.000
Specification characteristics									
Quasi-experimental	0.157	0.045	0.988	0.161	0.041	0.997	0.164	0.038	0.000
IV	-0.003	0.015	0.103	-0.004	0.016	0.158	-0.018	0.031	0.552
Publication characteristics									
Publication year	0.101	0.029	0.991	0.098	0.028	0.994	0.089	0.028	0.001
Top journal	-0.003	0.015	0.096	-0.006	0.020	0.166	-0.044	0.041	0.280
Citations	0.001	0.004	0.077	0.001	0.006	0.151	0.014	0.013	0.273
Byproduct	0.003	0.016	0.090	0.008	0.024	0.181	0.046	0.044	0.287
Observations	709			709			709		
Studies	40			40			40		

Notes: The response variable is the Frisch elasticity of labor supply at the intensive margin. P. mean = posterior mean, P. SD = posterior standard deviation, PIP = Posterior inclusion probability, SE = standard error. In the left-hand panel we apply BMA based on BRIC g-prior (the benchmark g-prior for parameters with the beta-binomial model prior). The middle panel reports the results of BMA based on HQ g-prior, which asymptotically mimics the Hannan-Quinn criterion. Table A4 presents a detailed description of all variables. In the right-hand panel we use Mallow's weights Hansen (2007) and the orthogonalization of the covariate space suggested by Amini & Parmeter (2012) to conduct the frequentist model averaging exercise.

Figure A8: Model inclusion in Bayesian model averaging (Random and BRIC)



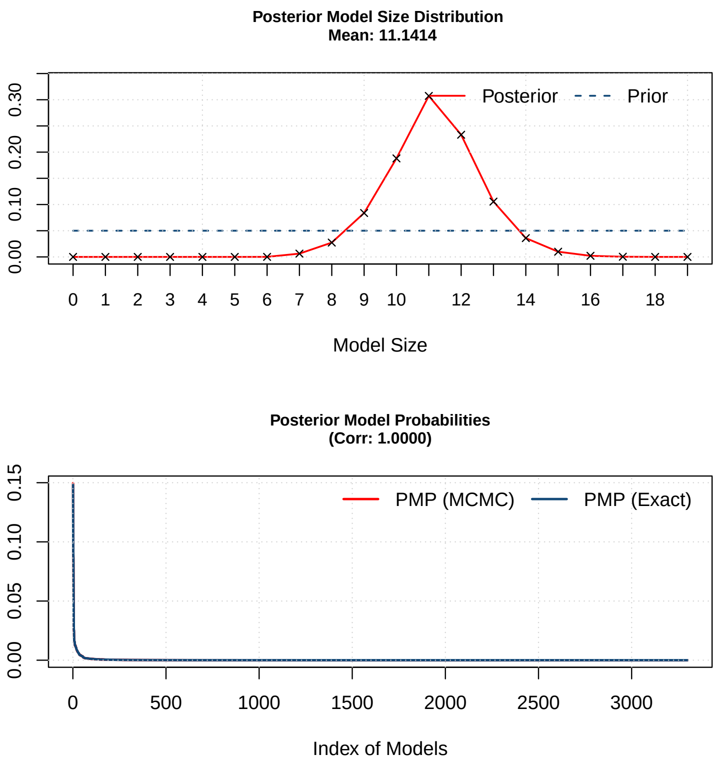
Notes: The response variable is the estimate of the Frisch elasticity of labor supply at the intensive margin. The columns denote individual models; variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes the cumulative posterior model probabilities. The estimation is based on BRIC g-prior (the benchmark g-prior for parameters with the beta-binomial model prior) and random model prior. Blue color (darker in grayscale) = the variable has a positive estimated sign. Red color (lighter in grayscale) = the variable has a negative estimated sign. No color = the variable is excluded from the given model. The numerical results are reported in Table A7.

Table A8: Summary of the BMA (Random and BRIC)

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
11.1414	$3 \cdot 10^6$	$1 \cdot 10^6$	12.05 mins	684,908
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
$5.24 \cdot 10^5$	131.0%	100%	1.0000	709
Model prior	*g-prior*	*Shrinkage-stats*		
Random/9.5	BRIC	Av = 0.9986		

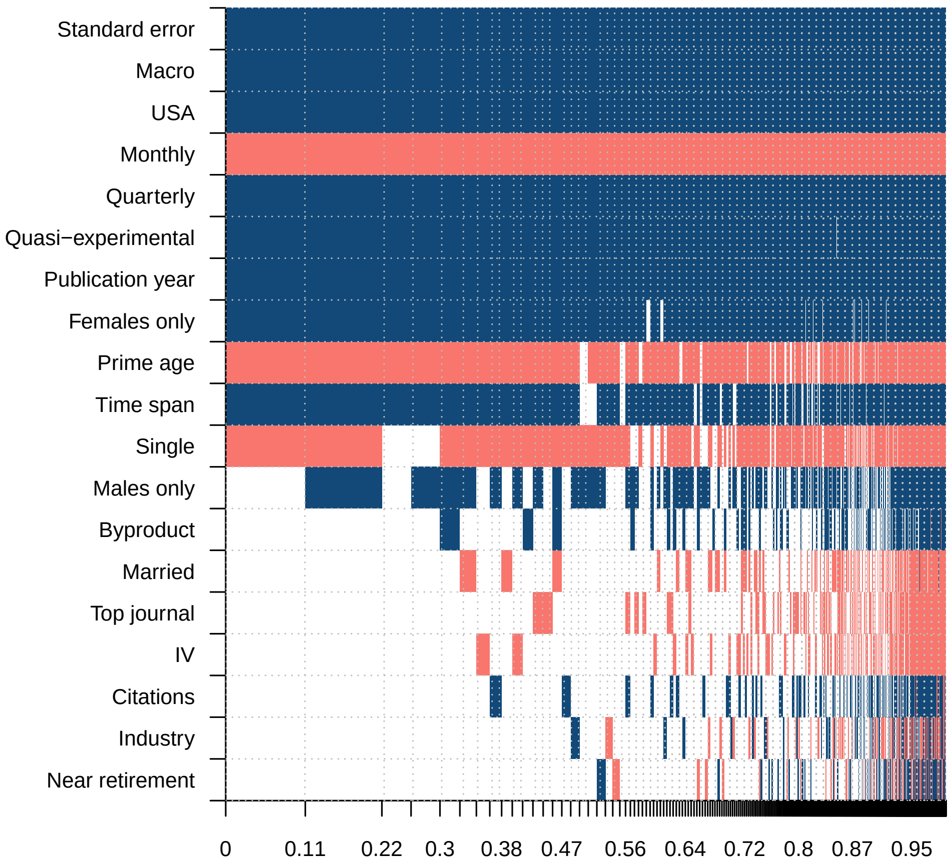
Notes: The results of this BMA specification are reported in Table A7. The estimation is based on BRIC g-prior suggested by Fernandez *et al.* (2001) and the beta-binomial model prior according to Ley & Steel (2009).

Figure A9: Model size and convergence in the BMA (Random and BRIC)



Notes: The figure depicts the posterior model size distribution and the posterior model probabilities of the BMA (random and BRIC prior) exercise reported in Table A7.

Figure A10: Model inclusion in BMA (Random and HQ g-prior)



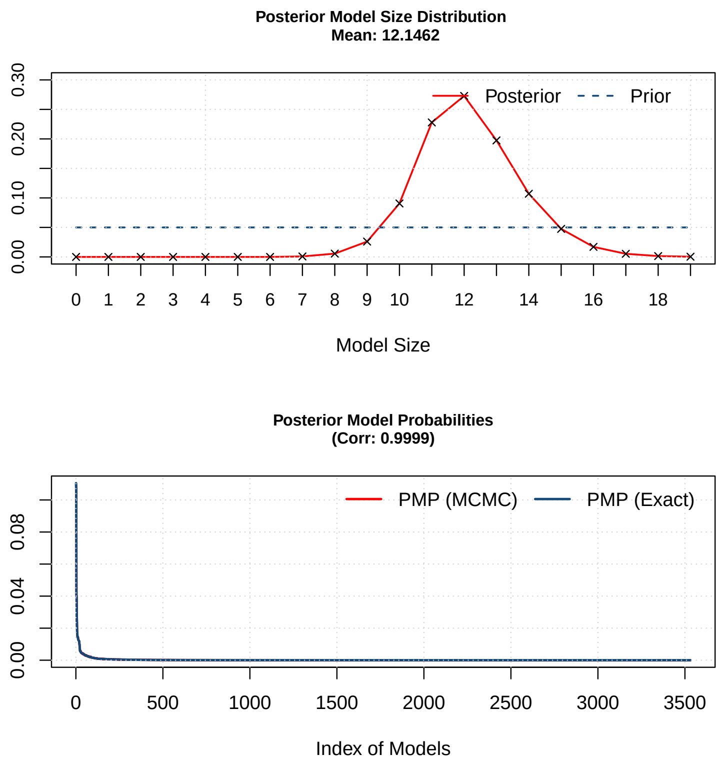
Notes: The response variable is the estimate of the Frisch intensive elasticity reported in a primary study. The columns denote individual models; variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes the cumulative posterior model probabilities. The estimation is based on HQ g-prior that asymptotically mimics the Hannan-Quinn criterion and random model prior. Blue color (darker in grayscale) = the variable has a positive estimated sign. Red color (lighter in grayscale) = the variable has a negative estimated sign. No color = the variable is excluded from the given model. The numerical results are reported in Table A7.

Table A9: Summary of the BMA (Random and HQ g-prior)

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
12.1462	$3 \cdot 10^6$	$1 \cdot 10^6$	13.61 mins	801,966
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
$5.24 \cdot 10^5$	153.0%	100%	1.0000	709
Model prior	*g-prior*	*Shrinkage-stats*		
Random/9.5	Hannan-Quinn	Av = 0.9965		

Notes: The results of this BMA specification are reported in Table A7. The estimation is based on HQ g-prior that asymptotically mimics the Hannan-Quinn criterion and random model prior as suggested by Fernandez *et al.* (2001).

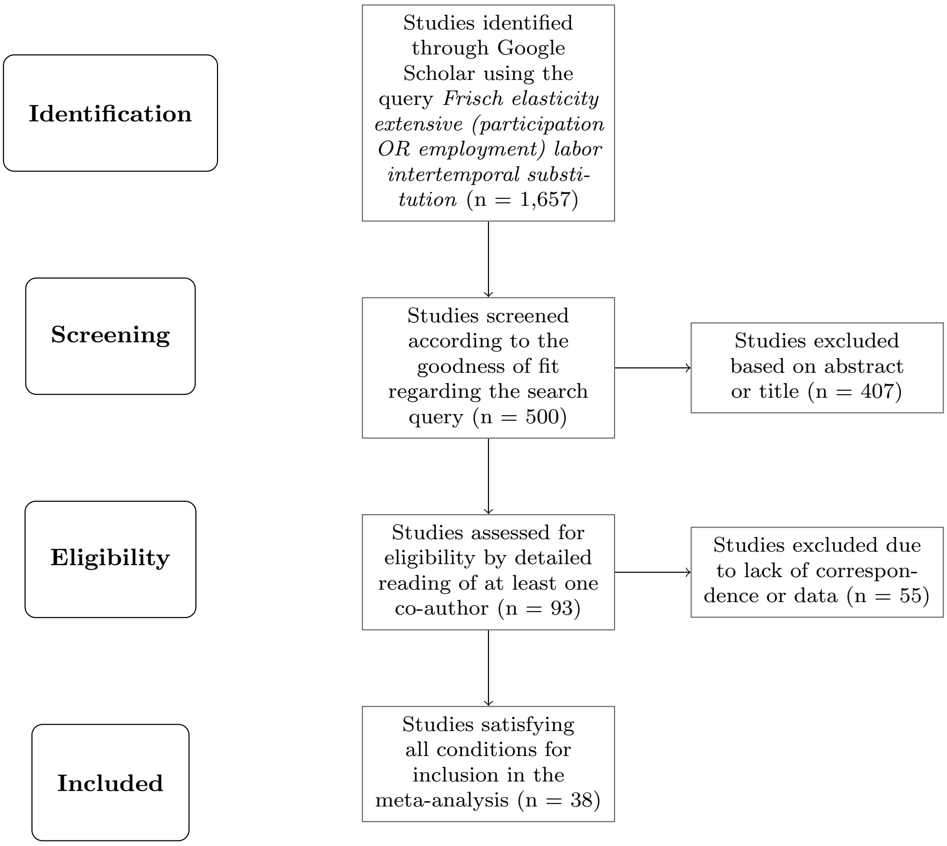
Figure A11: Model size and convergence in the BMA (Random and HQ g-prior)



Notes: The figure depicts the posterior model size distribution and the posterior model probabilities of the BMA (random and HQ g-prior) exercise reported in Table A7.

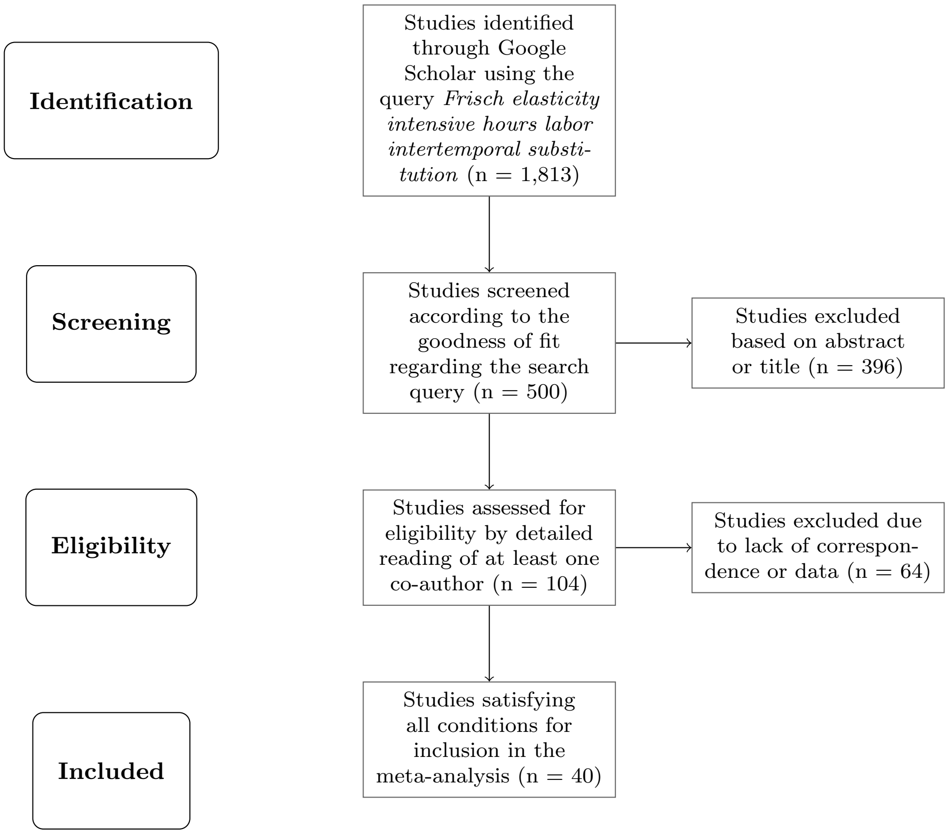
B Details on Literature Search and Data Collection

Figure B1: The PRISMA flow diagram (extensive margin elasticities)



Notes: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is an evidence-based set of items for reporting in systematic reviews and meta-analyses. More details on PRISMA and reporting standard of meta-analysis in general are provided by Havranek *et al.* (2020).

Figure B2: The PRISMA flow diagram (intensive margin elasticities)



Notes: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is an evidence-based set of items for reporting in systematic reviews and meta-analyses. More details on PRISMA and reporting standard of meta-analysis in general are provided by Havranek *et al.* (2020).

Table B1: Sources for estimates collected from individual papers

Extensive margin	Source	Intensive margin	Source
Attanasio *et al.* (2018)	Tables VIII-X[1]	Aaronson & French (2009)	Tables 2-3
Beffy *et al.* (2019)	Table 11	Altonji (1986)	Tables 1-2, 4
Bianchi *et al.* (2001)	Tables 4-6, 8	Angrist (1991)	Tables 2, 4
Blundell *et al.* (2016a)	Table XIV	Angrist *et al.* (2021)	Table 5
Blundell *et al.* (2016b)	Table 7	Attanasio *et al.* (2018)	Table VIII-X
Borella *et al.* (2023)	Table 4	Battisti *et al.* (2023)	Table 5
Brown (2013)	Via Chetty *et al.* (2013)[2]	Beffy *et al.* (2019)	Table 11
Caldwell (2019)	Table 3.7	Blundell *et al.* (2016a)	Table XIV
Card & Hyslop (2005)	Via Chetty *et al.* (2013)[3]	Blundell *et al.* (2016b)	Tables 4-6
Carrington (1996)	Table 2	Borella *et al.* (2023)	Table 4
Chang & Kim (2006)	Table 8	Bredemeier *et al.* (2019)	Tables 1-5, B2-F4
Chang *et al.* (2019)	Table 7	Caldwell & Oehlsen (2022)	Tables 4, A6-7
Erosa *et al.* (2016)	Tables 4-5	Chang *et al.* (2011)	Table 1
Espino *et al.* (2017)	Table 4	Domeij & Floden (2006)	Tables 2, 4-7
Fiorito & Zanella (2012)	Table 3, 6[4]	Erosa *et al.* (2016)	Table 4
French & Stafford (2017)	Tables 2-3	Farber (2015)	Tables IV-VI
Gine *et al.* (2017)	Table 6	Fiorito & Zanella (2012)	Table 6
Gourio & Noual (2009)	Abstract and Table 7	French (2005)	Tables 2, 5
Gruber & Wise (1999)	Via Chetty *et al.* (2013)[5]	French & Stafford (2017)	Tables 2-3
Haan & Uhlendorff (2013)	Table 6	Haan & Uhlendorff (2013)	Table 6
Inoue (2015)	Tables 3-6	Ham & Reilly (2002)	Table 1
Karabarbounis (2016)	Table 3	Inoue (2015)	Tables 3-6
Keane & Wasi (2016)	Figure 19[6]	Karabarbounis (2016)	Table 3
Kimmel & Kniesner (1998)	Table 1	Keane & Wasi (2016)	Figure 20[7]
Kneip *et al.* (2019)	Tables 3, E.2, F.1-3	Kimmel & Kniesner (1998)	Table 1
Kuroda & Yamamoto (2008)	Tables 2-5[8]	Kneip *et al.* (2019)	Tables 3, D.2, E.2, F.1-3
Looney & Singhal (2006)	Table 36	Kuroda & Yamamoto (2008)	Tables 3, 5
Manoli & Weber (2011)	Tables 3-4, 5A-B	Lee (2001)	Tables 1-2
Manoli & Weber (2016)	Table 3	Looney & Singhal (2006)	Tables 5, 8
Martinez *et al.* (2021)	Tables 3-4	MaCurdy (1981)	Table 1
Mustre-del Rio (2011)	Table 5	Martinez *et al.* (2021)	Tables 2-5
Mustre-del Rio (2015)	Table 8	Ong (2019)	Tables 2, A2
Oettinger (1999)	Table 5	Peterman (2016)	Tables 2-4, 9
Ong (2019)	Tables 2-3, A3	Pistaferri (2003)	Tables 2-3
Park (2020)	Tables 1, 8	Saez (2003)	Tables 5-6
Peterman (2016)	Table 5	Sigurdsson (2023)	Tables 1, A.1
Sigurdsson (2023)	Tables 2, A.10, A.28	Stafford (2015)	Tables 2, 4
Stafford (2015)	Tables 2, 4	Theloudis (2021)	Table 4
		Wallenius (2011)	Tables 1-3
		Ziliak & Kniesner (2005)	Tables 2-3

[1] The difference between reported total hours elasticities and median intensive elasticities.
[2] Computed based on the approach described in Chetty *et al.* (2013).
[3] Computed based on the approach described in Chetty *et al.* (2013).
[4] The difference between total hours elasticities in Table 3 and pure intensive elasticities in Table 6.
[5] Computed based on the approach described in Chetty *et al.* (2013).
[6] Elasticity of employment for ages 25, 40, and 55 with a college education.
[7] Elasticity of employment for ages 25, 40, and 55 with a college education.
[8] The difference between total hours and intensive elasticities in Tables 2-3 and 4-5.

C Estimating the Elasticities

In this section we provide a brief introduction to the Frisch elasticity and its estimation. For details on the theoretical background and empirical approaches, see Chang & Kim (2006), Keane (2011), and Attanasio *et al.* (2018). Put simply, the Frisch elasticity measures how much more people want to work when their net wage increases temporarily. So the Frisch elasticity corresponds to the elasticity of substitution of labor supply. The total effect can be disentangled into two margins: extensive (a decision whether to work at all) and intensive (a decision on how many hours to work given that one is already employed). The modern quasi-experimental literature has focused primarily on the extensive margin, and this is also the focus of our meta-analysis. In practice, the extensive margin elasticity is often computed simply as the change in the logarithm of employment rates divided by the change in the logarithm of net wages, and the latter is often instrumented. For more context, let us start with the definition of the total hours Frisch elasticity:

$$\eta = \frac{\partial h_t}{\partial w_t} \frac{w_t}{h_t} ||_\lambda, \tag{1}$$

where h and w denote hours of work and wage, respectively. The elasticity measures the marginal change in hours worked due to the marginal change in wages while the marginal utility of lifetime wealth (λ) is held constant. Following MaCurdy (1981), in a dynamic setting without uncertainty where a temporally separable utility function (with the discount factor β), represents the household's preferences over a life cycle, the equation for estimating the elasticity can be written as:

$$\ln h_t = \alpha_i + \rho + \theta x_t + \eta \ln w_t + \varepsilon_t, \tag{2}$$

where $\alpha_i = \eta \ln \lambda$, $\rho = -\eta \ln (\beta R)$, R is the interest rate, x is a vector of characteristics affecting the household's taste for work, and ε_t is an error term.

The estimated elasticity based on this equation is usually interpreted as the total hours response of labor supply, including both extensive and intensive margins. Assuming labor indivisibility, we can abstract from the intensive margin to address only the participation decision that operates at the extensive margin. Then the dependent variable takes a binary value, and the elasticity can be estimated by using a probit model for the participation decision. The optimal participation (employment) decision can be written as

$$h_t = \begin{cases} \bar{h}, & \text{if } w_t \geq w_t^R \\ 0, & \text{if } w_t \leq w_t^R. \end{cases} \tag{3}$$

The worker participates in the labor market and works $\bar{h}$ hours if the offered wage w_t is equal or larger than the reservation wage, w_t^R. Hence, the distribution of reservation wages plays a crucial role in determining the aggregate elasticity's magnitude at the extensive margin.

Alternatively, one can disentangle the total hours elasticity into the intensive and extensive margins using macro data. As in Fiorito & Zanella (2012), the variance of the log of aggregate labor can be decomposed as:

$$\operatorname{var}(\ln H_t) = \operatorname{var}(\ln n_t) + \operatorname{var}(\ln \bar{h}_t) + 2\operatorname{cov}(\ln n_t, \ln \bar{h}_t), \tag{4}$$

where n_t is the number of employed individuals, $\bar{h}_t$ is the average number of hours worked, and aggregate labor is $H_t = n_t \bar{h}_t$. Using (4), the decomposition of total hours Frisch elasticity can be written as

$$\eta = \frac{\operatorname{cov}(\Delta \ln H, \Delta \ln W)}{\operatorname{var}(\Delta \ln W)} = \frac{\operatorname{cov}(\Delta \ln \bar{h}, \Delta \ln W)}{\operatorname{var}(\Delta \ln W)} + \frac{\operatorname{cov}(\Delta \ln n, \Delta \ln W)}{\operatorname{var}(\Delta \ln W)}, \tag{5}$$

where Δ is the first-difference operator and W denotes the aggregate wage rate. The first term on the right-hand side is the intensive margin, and the second term corresponds to the extensive margin. In the extreme case where there is no heterogeneity among workers and employment is constant over the population, the extensive margin is eliminated as $\operatorname{cov}(\Delta \ln n, \Delta \ln W) = 0$.

Apart from conventional estimation methods, some studies use nonparametric or simulation-based methods to estimate the Frisch elasticity (Erosa *et al.* 2016; Kneip *et al.* 2019). When these estimates directly capture the response of labor supply at the extensive margin, we include them as well together with controls that capture the context in which the estimates were obtained. We discuss these aspects in detail in the main text.

D Diagnostics and Robustness Checks of the Meta-Analysis of Extensive Margin Elasticities

Table D1: Publication bias tests in a subsample of quasi-experimental estimates

Panel A: Linear tests					
	OLS	FE	Precision	Study	MAIVE
Publication bias	0.992^{**}	0.0415	1.479^{**}	1.498^{**}	0.643
(*Standard error*)	(0.488)	(0.283)	(0.720)	(0.683)	(0.460)
	[-0.20, 2.92]	-	[-3.12, 7.74]	[0.23, 3.13]	{-0.04, 2.33}
Effect beyond bias	0.153^{***}	0.211^{***}	0.123^{***}	0.170^{***}	0.188^{***}
(*Constant*)	(0.0469)	(0.0213)	(0.0467)	(0.0479)	(0.0393)
	[-0.01, 0.28]	-	[-0.01, 0.22]	[0.05, 0.29]	{-0.01, 0.68}
First stage F-stat					10.3
Observations	202	202	202	202	179
Studies	14	14	14	14	13
Panel B: Nonlinear tests					
	Ioannidis *et al.* (2017)	Andrews & Kasy (2019)	Bom & Rachinger (2019)	Furukawa (2021)	van Aert & van Assen (2023)
Effect beyond bias	0.112^{**}	0.211^{***}	0.083^{***}	0.095	0.217^{***}
	(0.049)	(0.048)	(0.015)	(0.082)	(0.057)
Observations	202	202	202	202	202
Studies	14	14	14	14	14

Notes: Panel A presents the results of regression $\hat{\eta}_{ij} = \eta_0 + \delta \cdot SE(\hat{\eta}_{ij}) + e_{ij}$, where $\hat{\eta}_{ij}$ and $SE(\hat{\eta}_{ij})$ are the i-th estimated Frisch extensive margin elasticity and its standard error reported in the j-th study. OLS = ordinary least squares. FE = study fixed effects. Precision = estimates are weighted by the inverse of their variance. Study = estimates are weighted by the inverse of the number of estimates reported per study. MAIVE = meta-analysis instrumental variable estimator (Irsova *et al.*, 2023); the inverse of the square root of the number of observations is used as an instrument for the standard error (the number of observations is not available for all studies). We cluster standard errors at the study level; if applicable, we also report 95% confidence intervals from wild bootstrap clustering in square brackets. In curly brackets we show the Anderson-Rubin 95% confidence interval. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D2: Correlation between elasticities and standard errors is weaker for stronger instruments

	OLS
Standard error (SE)	1.876^{***}
	(0.518)
SE * First-stage F-stat	-0.0110^{**}
	(0.00430)
Constant	0.133^{*}
	(0.0725)
Observations	22
Studies	4

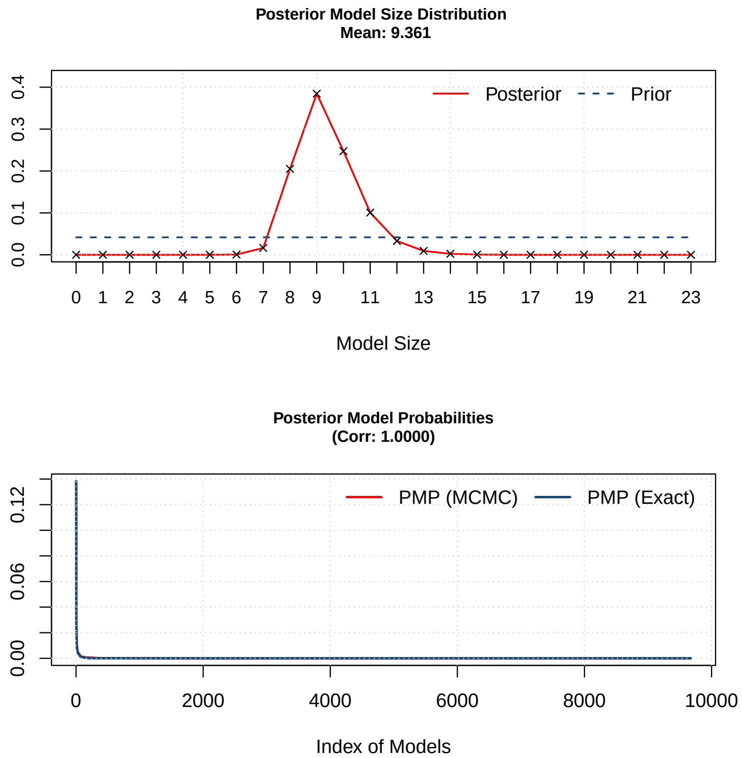
Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D3: Summary of the benchmark BMA estimation

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
9.361	$3 \cdot 10^6$	$1 \cdot 10^6$	12.89 mins	546,667
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
$8.39 \cdot 10^6$	6.5%	100%	1.0000	762
Model prior	*g-prior*	*Shrinkage-stats*		
Random/11.5	UIP	Av = 0.9987		

Notes: Based on Eicher *et al.* (2011) we employ unit information prior and, as suggested by George (2010), the dilution prior that takes into account potential collinearity.

Figure D1: Model size and convergence in the benchmark BMA model



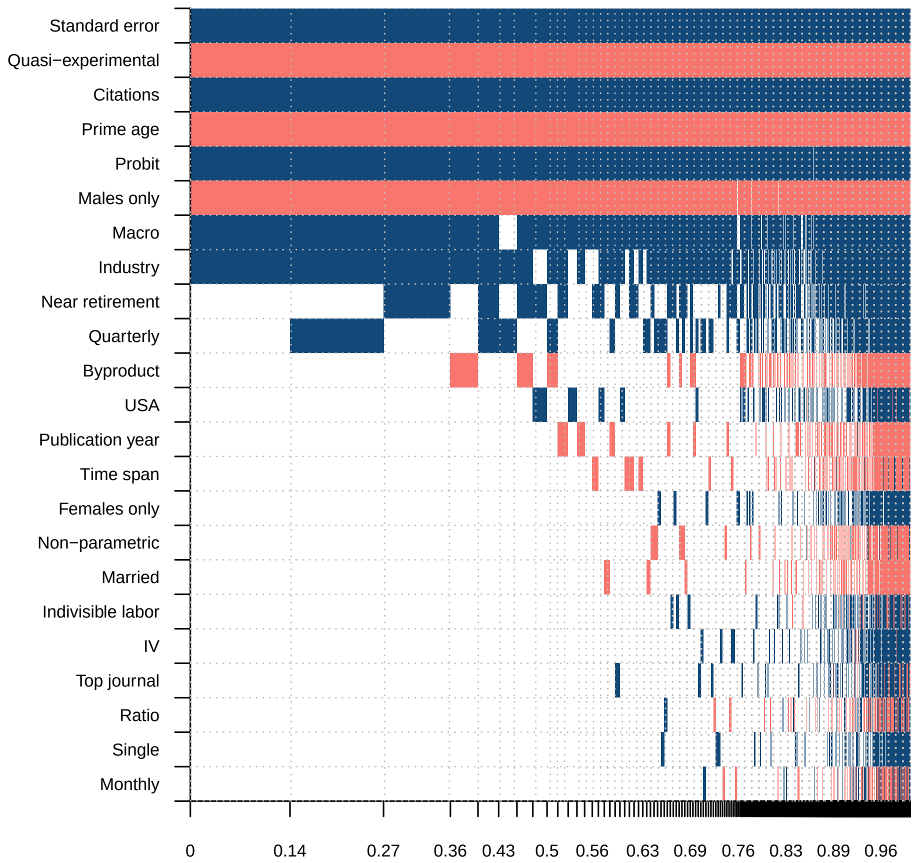
Notes: The figure depicts the posterior model size distribution and the posterior model probabilities of the BMA exercise reported in the main text.

Table D4: Results of BMA with alternative priors and results of FMA

Response variable: Frisch elasticity (extensive margin)	Bayesian model averaging (BRIC g-prior)			Bayesian model averaging (HQ g-prior)			Frequentist model averaging		
	P. mean	P. SD	PIP	P. mean	P. SD	PIP	Coef.	SE	p-value
Intercept	0.326	NA	1.000	0.353	NA	1.000	0.876	0.310	0.005
Standard error	1.381	0.120	1.000	1.371	0.124	1.000	1.254	0.173	0.000
Demographic characteristics									
Prime age	-0.150	0.030	1.000	-0.146	0.031	1.000	-0.127	0.033	0.000
Near retirement	0.034	0.047	0.389	0.047	0.051	0.535	0.112	0.038	0.003
Females only	0.003	0.014	0.057	0.005	0.020	0.109	0.089	0.038	0.017
Males only	-0.113	0.032	0.980	-0.113	0.033	0.976	-0.057	0.038	0.130
Married	-0.002	0.015	0.047	-0.004	0.018	0.079	-0.019	0.048	0.697
Single	0.001	0.012	0.035	0.003	0.017	0.068	0.072	0.054	0.183
Data characteristics									
Time span	-0.002	0.010	0.074	-0.002	0.010	0.098	0.032	0.028	0.239
Monthly	0.000	0.015	0.029	0.000	0.020	0.054	0.004	0.083	0.963
Quarterly	0.030	0.045	0.363	0.032	0.044	0.411	0.103	0.048	0.030
Ratio	0.000	0.008	0.037	0.000	0.010	0.063	0.052	0.041	0.200
Industry	0.129	0.066	0.859	0.134	0.064	0.886	0.297	0.088	0.001
Macro	0.134	0.051	0.942	0.140	0.049	0.964	0.217	0.051	0.000
USA	0.007	0.023	0.111	0.007	0.024	0.137	-0.014	0.044	0.757
Specification characteristics									
Indivisible labor	0.002	0.013	0.045	0.004	0.021	0.088	0.109	0.058	0.062
Quasi-experimental	-0.285	0.042	1.000	-0.287	0.042	1.000	-0.277	0.058	0.000
Probit	0.232	0.057	0.995	0.229	0.057	0.996	0.178	0.065	0.006
Non-parametric	-0.002	0.014	0.056	-0.006	0.022	0.118	-0.062	0.052	0.239
IV	0.001	0.012	0.042	0.003	0.017	0.080	0.034	0.057	0.559
Publication characteristics									
Publication year	-0.010	0.039	0.089	-0.018	0.052	0.158	-0.232	0.098	0.018
Top journal	0.001	0.010	0.040	0.002	0.013	0.071	-0.014	0.045	0.754
Citations	0.067	0.013	1.000	0.067	0.013	1.000	0.070	0.016	0.000
Byproduct	-0.016	0.042	0.164	-0.026	0.051	0.266	-0.127	0.055	0.022
Observations	762			762			762		
Studies	38			38			38		

Notes: The response variable is the Frisch elasticity of labor supply at the extensive margin. P. mean = posterior mean, P. SD = posterior standard deviation, PIP = Posterior inclusion probability, SE = standard error. In the left-hand panel we apply BMA based on BRIC g-prior (the benchmark g-prior for parameters with the beta-binomial model prior). The middle panel reports the results of BMA based on HQ g-prior, which asymptotically mimics the Hannan-Quinn criterion. In the right-hand panel we use Mallow's weights Hansen (2007) and the orthogonalization of the covariate space suggested by Amini & Parmeter (2012) to conduct the frequentist model averaging exercise.

Figure D2: Model inclusion in BMA (BRIC g-prior)



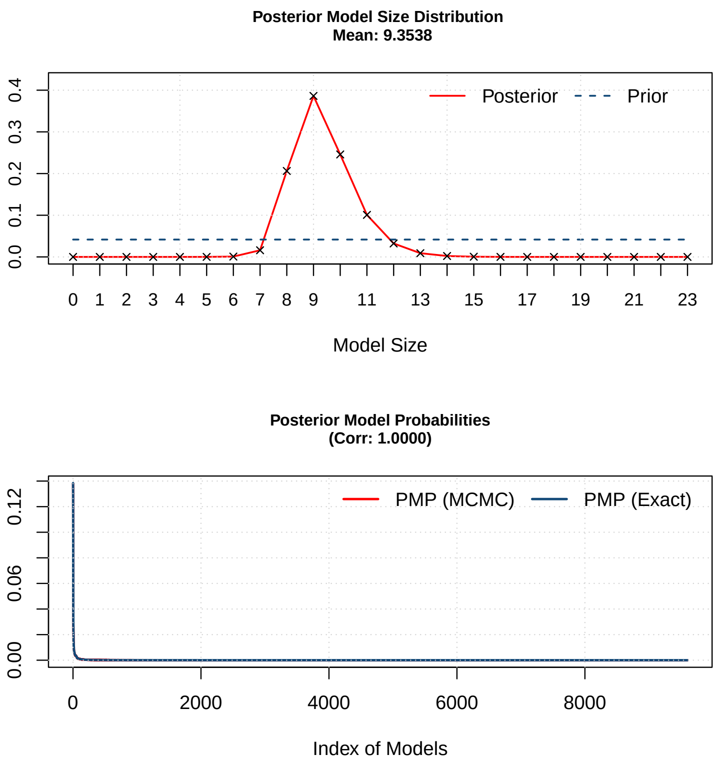
Notes: The response variable is the estimate of the Frisch extensive elasticity reported in a primary study. The columns denote individual models; variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes the cumulative posterior model probabilities. The estimation is based on BRIC g-prior (the benchmark g-prior for parameters with the beta-binomial model prior) and random model prior. Blue color (darker in grayscale) = the variable has a positive estimated sign. Red color (lighter in grayscale) = the variable has a negative estimated sign. No color = the variable is excluded from the given model. The numerical results are reported in Table D4.

Table D5: Summary of the BMA (BRIC g-prior)

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
9.3538	$3 \cdot 10^6$	$1 \cdot 10^6$	13.07 mins	544,779
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
$8.39 \cdot 10^6$	6.5%	100%	1.0000	762
Model prior	*g-prior*	*Shrinkage-stats*		
Random/11.5	BRIC	Av = 0.9987		

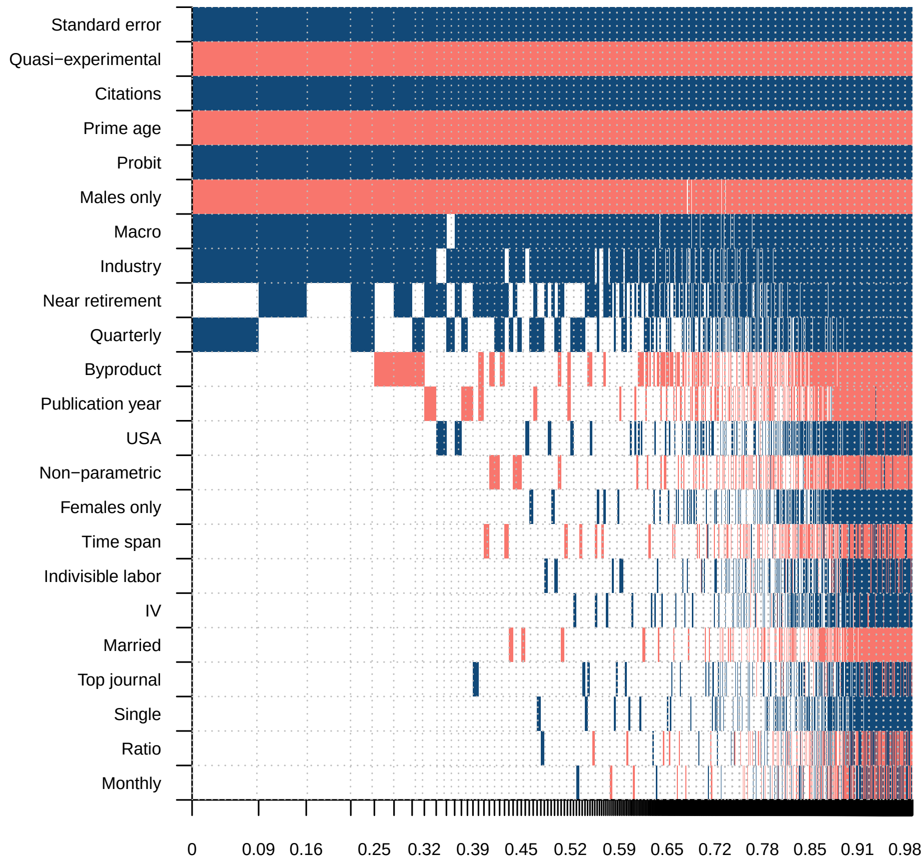
Notes: The results of this BMA specification are reported in Table D4. The estimation is based on BRIC g-prior suggested by Fernandez *et al.* (2001) and the beta-binomial model prior according to Ley & Steel (2009).

Figure D3: Model size and convergence in the BMA (BRIC g-prior)



Notes: The figure depicts the posterior model size distribution and the posterior model probabilities of the BMA (random and BRIC prior) exercise reported in Table D4.

Figure D4: Model inclusion in BMA (Random and HQ g-prior)



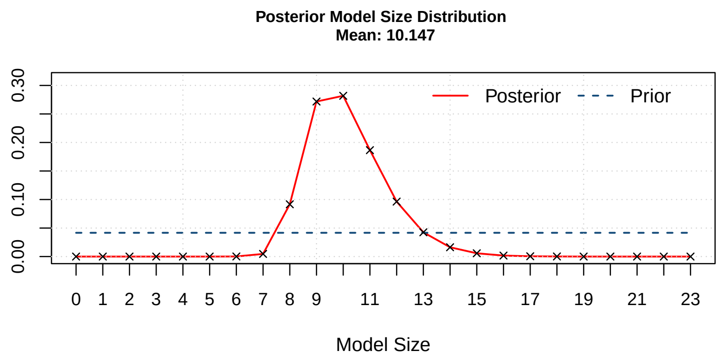
Notes: The response variable is the estimate of the Frisch extensive elasticity reported in a primary study. The columns denote individual models; variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes the cumulative posterior model probabilities. The estimation is based on HQ g-prior that asymptotically mimics the Hannan-Quinn criterion and random model prior. Blue color (darker in grayscale) = the variable has a positive estimated sign. Red color (lighter in grayscale) = the variable has a negative estimated sign. No color = the variable is excluded from the given model. The numerical results are reported in Table D4.

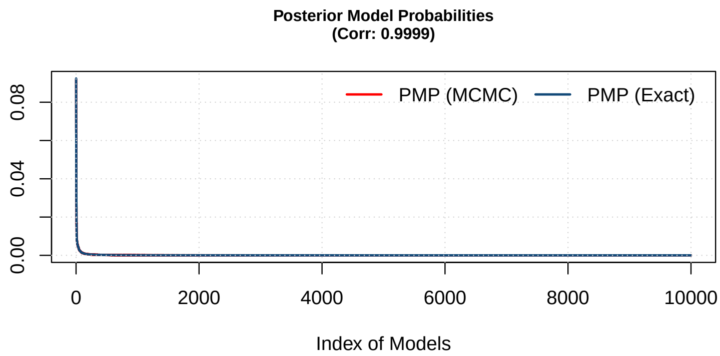
Table D6: Summary of the BMA (Random and HQ g-prior)

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
10.147	$3 \cdot 10^6$	$1 \cdot 10^6$	16.38 mins	718,854
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
$8.39 \cdot 10^6$	8.6%	99%	0.9999	762
Model prior	*g-prior*	*Shrinkage-stats*		
Random/11.5	Hannan-Quinn	Av = 0.9966		

Notes: The results of this BMA specification are reported in Table D4. The estimation is based on HQ g-prior that asymptotically mimics the Hannan-Quinn criterion and random model prior as suggested by Fernandez *et al.* (2001).

Figure D5: Model size and convergence in the BMA (Random and HQ g-prior)





Notes: The figure depicts the posterior model size distribution and the posterior model probabilities of the BMA (random and HQ g-prior) exercise reported in Table D4.

References

Aaronson, D. & E. French (2009): "The effects of progressive taxation on labor supply when hours and wages are jointly determined." *Journal of Human Resources* **44(2)**: pp. 386–408.

van Aert, R. C. & M. van Assen (2023): "Correcting for publication bias in a meta-analysis with the p-uniform* method." *Working paper*, Tilburg University & Utrecht University.

Altonji, J. G. (1986): "Intertemporal substitution in labor supply: Evidence from micro data." *Journal of Political Economy* **94(3, Part 2)**: pp. S176–S215.

Amini, S. M. & C. F. Parmeter (2012): "Comparison of model averaging techniques: Assessing growth determinants." *Journal of Applied Econometrics* **27(5)**: pp. 870–876.

Andrews, I. & M. Kasy (2019): "Identification of and correction for publication bias." *American Economic Review* **109(8)**: pp. 2766–2794.

Angrist, J. D. (1991): "Grouped-data estimation and testing in simple labor-supply models." *Journal of Econometrics* **47(2-3)**: pp. 243–266.

Angrist, J. D., S. Caldwell, & J. V. Hall (2021): "Uber versus taxi: A driver's eye view." *American Economic Journal: Applied Economics* **13(3)**: pp. 272–308.

Attanasio, O., P. Levell, H. Low, & V. Sánchez-Marcos (2018): "Aggregating elasticities: Intensive and extensive margins of women's labor supply." *Econometrica* **86(6)**: pp. 2049–2082.

Battisti, M., R. Michaels, & C. Park (2023): "Labor supply within the firm." *Journal of Labor Economics* **(forthcoming)**.

Beffy, M., R. Blundell, A. Bozio, G. Laroque, & M. To (2019): "Labour supply and taxation with restricted choices." *Journal of Econometrics* **211(1)**: pp. 16–46.

Bianchi, M., B. R. Gudmundsson, & G. Zoega (2001): "Iceland's natural experiment in supply-side economics." *American Economic Review* **91(5)**: pp. 1564–1579.

Blundell, R., M. Costa Dias, C. Meghir, & J. Shaw (2016a): "Female labor supply, human capital, and welfare reform." *Econometrica* **84(5)**: pp. 1705–1753.

Blundell, R., L. Pistaferri, & I. Saporta-Eksten (2016b): "Consumption inequality and family labor supply." *American Economic Review* **106(2)**: pp. 387–435.

Bom, P. R. & H. Rachinger (2019): "A kinked meta-regression model for publication bias correction." *Research Synthesis Methods* **10(4)**: pp. 497–514.

Borella, M., M. De Nardi, & F. Yang (2023): "Are marriage-related taxes and social security benefits holding back female labour supply?" *The Review of Economic Studies* **90(1)**: pp. 102–131.

Bredemeier, C., J. Gravert, & F. Juessen (2019): "Estimating labor supply elasticities with joint borrowing constraints of couples." *Journal of Labor Economics* **37(4)**: pp. 1215–1265.

Brown, K. M. (2013): "The link between pensions and retirement timing: Lessons from California teachers." *Journal of Public Economics* **98(C)**: pp. 1–14.

Caldwell, S. & E. Oehlsen (2022): "Gender differences in labor supply: Experimental evidence from the gig economy." *Working paper*, University of California, Berkeley Working Paper.

Caldwell, S. C. (2019): *Essays on imperfect competition in the labor market*. Ph.D. thesis, Massachusetts Institute of Technology.

Card, D. & D. R. Hyslop (2005): "Estimating the effects of a time-limited earnings subsidy for welfare-leavers." *Econometrica* **73(6)**: pp. 1723–1770.

Carrington, W. J. (1996): "The Alaskan labor market during the pipeline era." *Journal of Political Economy* **104(1)**: pp. 186–218.

Chang, Y. & S.-B. Kim (2006): "From individual to aggregate labor supply: A quantitative analysis based on a heterogeneous agent macroeconomy." *International Economic Review* **47(1)**: pp. 1–27.

Chang, Y., S.-B. Kim, K. Kwon, & R. Rogerson (2011): "Interpreting labor supply regressions in a model of full-and part-time work." *American Economic Review* **101(3)**: pp. 476–481.

Chang, Y., S.-B. Kim, K. Kwon, & R. Rogerson (2019): "2018 Klein lecture: individual and aggregate labor supply in heterogeneous agent economies with intensive and extensive margins." *International Economic Review* **60(1)**: pp. 3–24.

Chetty, R., A. Guren, D. Manoli, & A. Weber (2013): "Does indivisible labor explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities." *NBER Macroeconomics Annual* **27(1)**: pp. 1–56.

Domeij, D. & M. Floden (2006): "The labor-supply elasticity and borrowing constraints: Why estimates are biased." *Review of Economic Dynamics* **9(2)**: pp. 242–262.

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Erosa, A., L. Fuster, & G. Kambourov (2016): "Towards a micro-founded theory of aggregate labour supply." *The Review of Economic Studies* **83(3)**: pp. 1001–1039.

Espino, A., F. Isabella, M. Leites, & A. Machado (2017): "Do women have different labor supply behaviors? Evidence based on educational groups in Uruguay." *Feminist Economics* **23(4)**: pp. 143–169.

Farber, H. S. (2015): "Why you can't find a taxi in the rain and other labor supply lessons from cab drivers." *The Quarterly Journal of Economics* **130(4)**: pp. 1975–2026.

Fernandez, C., E. Ley, & M. F. Steel (2001): "Benchmark priors for Bayesian model averaging." *Journal of Econometrics* **100(2)**: pp. 381–427.

Fiorito, R. & G. Zanella (2012): "The anatomy of the aggregate labor supply elasticity." *Review of Economic Dynamics* **15(2)**: pp. 171–187.

French, E. (2005): "The effects of health, wealth, and wages on labour supply and retirement behaviour." *The Review of Economic Studies* **72(2)**: pp. 395–427.

French, S. & T. Stafford (2017): "Returns to experience and the elasticity of labor supply." *Working paper 2017-15*, UNSW Business School.

Furukawa, C. (2021): "Publication bias under aggregation frictions: From communication model to new correction method." *Working paper*, MIT, mimeo.

George, E. I. (2010): "Dilution priors: Compensating for model space redundancy." In "Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown," pp. 158–165. Institute of Mathematical Statistics.

Gine, X., M. Martinez-Bravo, & M. Vidal-Fernandez (2017): "Are labor supply decisions consistent with neoclassical preferences? Evidence from Indian boat owners." *Journal of Economic Behavior & Organization* **142(C)**: pp. 331–347.

Gourio, F. & P.-A. Noual (2009): "The marginal worker and the aggregate elasticity of labor supply." *Working Papers Series 2006-009*, Boston University Dept. of Economics.

Gruber, J. & D. A. Wise (1999): *Social security and retirement around the world*. University of Chicago Press.

Haan, P. & A. Uhlendorff (2013): "Intertemporal labor supply and involuntary unemployment." *Empirical Economics* **44(2)**: pp. 661–683.

Ham, J. C. & K. T. Reilly (2002): "Testing intertemporal substitution, implicit contracts, and hours restriction models of the labor market using micro data." *American Economic Review* **92(4)**: pp. 905–927.

Hansen, B. E. (2007): "Least squares model averaging." *Econometrica* **75(4)**: pp. 1175–1189.

Havranek, T., T. D. Stanley, H. Doucouliagos, P. Bom, J. Geyer-Klingeberg, I. Iwasaki, W. R. Reed, K. Rost, & R. C. M. van Aert (2020): "Reporting guidelines for meta-analysis in economics." *Journal of Economic Surveys* **34(3)**: pp. 469–475.

Inoue, Y. (2015): "Intensive and extensive margins of Japanese male and female workers: Evidence from the tax policy reform in Japan." *Working paper*, Panel Data Research Center at Keio University.

Ioannidis, J. P., T. D. Stanley, & H. Doucouliagos (2017): "The power of bias in economics research." *The Economic Journal* **127(605)**: pp. F236–F265.

Irsova, Z., P. R. D. Bom, T. Havranek, & H. Rachinger (2023): "Spurious Precision in Meta-Analysis." *CEPR Discussion Papers 17927*, Centre for Economic Policy Research.

Karabarbounis, M. (2016): "A road map for efficiently taxing heterogeneous agents." *American Economic Journal: Macroeconomics* **8(2)**: pp. 182–214.

Keane, M. & T. Neal (2023): "Instrument strength in IV estimation and inference: A guide to theory and practice." *Journal of Econometrics* **235(2)**: pp. 1625–1653.

Keane, M. P. (2011): "Labor supply and taxes: A survey." *Journal of Economic Literature* **49(4)**: pp. 961–1075.

Keane, M. P. & N. Wasi (2016): "Labour supply: The roles of human capital and the extensive margin." *The Economic Journal* **126(592)**: pp. 578–617.

Kimmel, J. & T. J. Kniesner (1998): "New evidence on labor supply: Employment versus hours elasticities by sex and marital status." *Journal of Monetary Economics* **42(2)**: pp. 289–301.

Kneip, A., M. Merz, & L. Storjohann (2019): "Aggregation and labor supply elasticities." *Journal of the European Economic Association* **18(5)**: pp. 2315–2358.

Kuroda, S. & I. Yamamoto (2008): "Estimating Frisch labor supply elasticity in Japan." *Journal of the Japanese and International Economies* **22(4)**: pp. 566–585.

Lee, C.-I. (2001): "Finite sample bias in IV estimation of intertemporal labor supply models: Is the intertemporal substitution elasticity really small?" *The Review of Economics and Statistics* **83(4)**: pp. 638–646.

Ley, E. & M. F. Steel (2009): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regression." *Journal of Applied Econometrics* **24(4)**: pp. 651–674.

Looney, A. & M. Singhal (2006): "The effect of anticipated tax changes on intertemporal labor supply and the realization of taxable income." *Working paper 12417*, National Bureau of Economic Research.

MaCurdy, T. E. (1981): "An empirical model of labor supply in a life-cycle setting." *Journal of Political Economy* **89(6)**: pp. 1059–1085.

Manoli, D. & A. Weber (2011): "Nonparametric evidence on the effects of retirement benefits on labor force participation decisions." *Working Papers 2011-24*, Center for Retirement Research, Boston College.

Manoli, D. & A. Weber (2016): "Nonparametric evidence on the effects of financial incentives on retirement decisions." *American Economic Journal: Economic Policy* **8(4)**: pp. 160–182.

Martinez, I. Z., E. Saez, & M. Siegenthaler (2021): "Intertemporal labor supply substitution? Evidence from the Swiss income tax holidays." *American Economic Review* **111(2)**: pp. 506–546.

Oettinger, G. S. (1999): "An empirical analysis of the daily labor supply of stadium venors." *Journal of Political Economy* **107(2)**: pp. 360–392.

Ong, P. (2019): "The effect of child support on labor supply: An estimate of the Frisch elasticity." *Working paper*, Department of Economics, Northwestern University.

Park, C. (2020): "Consumption, reservation wages, and aggregate labor supply." *Review of Economic Dynamics* **37(1)**: pp. 54–80.

Peterman, W. B. (2016): "Reconciling micro and macro estimates of the Frisch labor supply elasticity." *Economic Inquiry* **54(1)**: pp. 100–120.

Pistaferri, L. (2003): "Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply." *Journal of Labor Economics* **21(3)**: pp. 729–754.

Mustre-del Rio, J. (2011): "The aggregate implications of individual labor supply heterogeneity." *Working paper*, Federal Research Bank of Kansas City, Research Division.

Mustre-del Rio, J. (2015): "Wealth and labor supply heterogeneity." *Review of Economic Dynamics* **18(3)**: pp. 619–634.

Saez, E. (2003): "The effect of marginal tax rates on income: A panel study of 'bracket creep'." *Journal of Public Economics* **87(5-6)**: pp. 1231–1258.

Sigurdsson, J. (2023): "Labor supply responses and adjustment frictions: A tax-free year in Iceland." *American Economic Journal: Economic Policy* **(forthcoming)**.

Stafford, T. M. (2015): "What do fishermen tell us that taxi drivers do not? An empirical investigation of labor supply." *Journal of Labor Economics* **33(3)**: pp. 683–710.

Theloudis, A. (2021): "Consumption inequality across heterogeneous families." *European Economic Review* **136(C)**: p. 103765.

Wallenius, J. (2011): "Human capital accumulation and the intertemporal elasticity of substitution of labor: How large is the bias?" *Review of Economic Dynamics* **14(4)**: pp. 577–591.

Ziliak, J. P. & T. J. Kniesner (2005): "The effect of income taxation on consumption and labor supply." *Journal of Labor Economics* **23(4)**: pp. 769–796.