# Additional Results

## to “Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting”*

Tomas Havranek

Czech National Bank and Charles University, Prague

tomas.havranek@ies-prague.org

September 15, 2014

**Abstract**

This appendix provides three sets of results additional to the main body of the paper. The first set presents Hedges’ model of selective reporting. The second set shows robustness checks and statistics for my results concerning selective reporting. The third set presents Bayesian model averaging, used here to describe the heterogeneity in the estimates of the elasticity of intertemporal substitution.

## 1 Hedges’ Model of Selective Reporting

Hedges (1992) introduces a model of selective reporting which assumes that the probability of reporting of estimates is determined by their statistical significance. The probability of reporting only changes when a psychologically important p-value is reached: in economics these threshold values are 0.01, 0.05, and 0.1. When no reporting bias is present, all estimates, significant and insignificant at the conventional levels, should have the same probability of being published. I estimate both the original model of Hedges (1992) and the augmented model developed by

*An online appendix providing the paper, data, and code is available at meta-analysis.cz/eis.

Ashenfelter *et al.* (1999), which allows for heterogeneity in the estimates of the underlying effect. The augmented log-likelihood function is (Ashenfelter *et al.*, 1999, p. 468)

$$L = c + \sum_{i=1}^{n} \log w_i(X_i, \omega) - \frac{1}{2} \sum_{i=1}^{n} \left( \frac{X_i - \mathbf{Z}_i \Delta}{\eta_i} \right)^2 - \sum_{i=1}^{n} \log(\eta_i) - \sum_{i=1}^{n} \log \left[ \sum_{j=1}^{4} \omega_j B_{ij}(\mathbf{Z}_i \Delta, \sigma) \right], \quad (1)$$

where $X_i \sim N(\Delta, \eta_i)$ are the estimates of the elasticity of intertemporal substitution. The parameter $\Delta$ is the average underlying elasticity, and $\eta_i = \sigma_i^2 + \sigma^2$, where $\sigma_i$ are the reported standard errors of the estimates and $\sigma$ measures heterogeneity in the estimates. The probability of reporting is determined by the weight function $w(X_i)$. In this model $w(X_i)$ is a step function associated with the p-values of the estimates. I choose four steps reflecting different levels of conventional statistical significance of the estimates: p-value $< 0.01$, $0.01 <$ p-value $< 0.05$, $0.05 <$ p-value $< 0.1$, and p-value $> 0.1$. $B_{ij}(\Delta, \sigma)$ represents the probability that an estimate $X_i$ will be assigned weight $\omega_i$. For the first step, p-value $< 0.01$, I normalize $\omega$ to 1 and evaluate whether the remaining three weights differ from this value. $Z_i$ is a vector of the characteristics of estimate $X_i$.

Table 1: Hedges' model of selective reporting

| | Unrestricted model | | Restricted ($\omega_j = 1$) | |
|---|---|---|---|---|
| | Coefficient | Standard error | Coefficient | Standard error |
| $\omega_2$ | -1.568 | 0.205 | | |
| $\omega_3$ | -1.091 | 0.210 | | |
| $\omega_4$ | -0.236 | 0.107 | | |
| Constant | 0.264 | 0.015 | 0.247 | 0.008 |
| $\sigma$ | 0.348 | 0.008 | 0.327 | 0.007 |
| Log likelihood | -558.2 | | -466.5 | |
| Observations | 2,735 | | 2,735 | |
| $\chi^2$ ($H_0$: all estimates have the same probability of reporting): 183.4, $p$-value $< 0.001$. | | | | |

*Notes:* In the absence of reporting bias estimates with different statistical significance should have the same probability of being reported. $\omega_1$, the weight associated with the probability of reporting for estimates significant at the 1% level, is set to 1. $\omega_2$, $\omega_3$, and $\omega_4$ show the relative probabilities for estimates only significant at the 5% level, estimates only significant at the 10% level, and insignificant estimates. $\sigma$ is the estimated measure of heterogeneity (standard deviation) of the estimates of the EIS.

Table 1 shows the estimation results of the model where $Z$ only includes a constant (that is, no heterogeneity in the estimates of the elasticity of intertemporal substitution is assumed). The table includes two models, an unrestricted model and a restricted model with restriction $\omega_2 = \omega_3 = \omega_4 = 1$. The unrestricted model assumes reporting bias, while the restricted model assumes no bias (all coefficients have the same probability of being published, their different

Table 2: Hedges' model of selective reporting with heterogeneity

| | Unrestricted model | | Restricted ($\omega_j = 1$) | |
|---|---|---|---|---|
| | Coefficient | Standard error | Coefficient | Standard error |
| $\omega_2$ | -1.105 | 0.170 | | |
| $\omega_3$ | -0.711 | 0.173 | | |
| $\omega_4$ | -0.070 | 0.093 | | |
| *Utility* | | | | |
| Epstein-Zin | 0.007 | 0.044 | 0.003 | 0.042 |
| Habits | 0.090 | 0.041 | 0.093 | 0.039 |
| Nonsep. durables | 0.142 | 0.037 | 0.138 | 0.035 |
| Nonsep. public | 0.140 | 0.039 | 0.138 | 0.037 |
| Nonsep. tradables | 0.359 | 0.039 | 0.334 | 0.038 |
| *Data* | | | | |
| No. of households | -0.016 | 0.010 | -0.018 | 0.009 |
| No. of years | -0.069 | 0.022 | -0.072 | 0.034 |
| Average year | 2.638 | 1.921 | 2.645 | 3.907 |
| Micro data | 0.117 | 0.073 | 0.132 | 0.071 |
| Annual data | 0.012 | 0.023 | 0.016 | 0.024 |
| Monthly data | 0.130 | 0.033 | 0.127 | 0.033 |
| *Design* | | | | |
| Quasipanel | 0.030 | 0.052 | 0.003 | 0.050 |
| Inverse estimation | 0.119 | 0.021 | 0.124 | 0.020 |
| Asset holders | 0.391 | 0.046 | 0.391 | 0.043 |
| First lag instrument | 0.087 | 0.020 | 0.083 | 0.021 |
| No year dummies | -0.200 | 0.097 | -0.232 | 0.092 |
| Income | -0.056 | 0.019 | -0.056 | 0.019 |
| Taste shifters | 0.042 | 0.025 | 0.039 | 0.024 |
| *Variable definition* | | | | |
| Total consumption | 0.043 | 0.025 | 0.035 | 0.023 |
| Food | 0.183 | 0.087 | 0.201 | 0.082 |
| Stock return | -0.108 | 0.020 | -0.117 | 0.019 |
| Capital return | -0.099 | 0.027 | -0.103 | 0.025 |
| *Estimation* | | | | |
| Exact Euler | 0.107 | 0.025 | 0.096 | 0.024 |
| ML | -0.015 | 0.037 | -0.018 | 0.035 |
| TSLS | 0.021 | 0.022 | 0.024 | 0.021 |
| OLS | 0.127 | 0.034 | 0.122 | 0.032 |
| *Publication* | | | | |
| Publication year | -0.784 | 2.414 | -0.719 | 4.876 |
| Citations | 0.011 | 0.008 | 0.013 | 0.008 |
| Top journal | 0.148 | 0.029 | 0.147 | 0.029 |
| Impact | -0.016 | 0.006 | -0.015 | 0.005 |
| Constant | -13.75 | 16.96 | -14.29 | 20.51 |
| $\sigma$ | 0.277 | 0.007 | 0.265 | 0.006 |
| Log likelihood | -223.8 | | -158.9 | |
| Observations | 2,735 | | 2,735 | |
| $\chi^2$ ($H_0$: all estimates have the same probability of publication): 129.8, $p$-value $< 0.001$. | | | | |

*Notes:* In the absence of reporting bias estimates with different statistical significance should have the same probability of being reported. $\omega_1$, the weight associated with the probability of reporting for estimates significant at the 1% level, is set to 1. $\omega_2$, $\omega_3$, and $\omega_4$ show the relative probabilities for estimates only significant at the 5% level, estimates only significant at the 10% level, and insignificant estimates. $\sigma$ is the estimated measure of heterogeneity (standard deviation) of the estimates of the EIS.

statistical significance notwithstanding). The restriction is rejected, which suggests reporting bias: estimates significant at the 1% level are much more likely to get published than all other estimates. The results are similar when I allow for heterogeneity in the estimates of the elasticity of intertemporal substitution (Table 2).

## 2 Robustness Checks and Tests

Table 3 provides several robustness checks of the results concerning selective reporting shown in the main body of the paper. The first column provides the baseline results using published studies. In the second column I use estimates from 50 unpublished papers, the latest available version of which was written in 2007 or before. I expect that these old working papers are unlikely to be published, so I can test whether there is a difference in selective reporting between published and unpublished studies. The results show a slightly smaller reporting bias for unpublished studies, but the difference is not statistically significant (column 3). My results suggest that selective reporting arises because of the authors' priors concerning the plausible values of the EIS, not because of a tendency of editors and referees to select some particular results for publication.

Table 3: The reported estimates are correlated with their standard errors

| | Pub | Unpub | All | Finance | Asym | LLR | Boot |
|---|---|---|---|---|---|---|---|
| SE | 2.115$^{***}$ | 1.554$^{***}$ | 2.407$^{***}$ | 1.889$^{***}$ | 2.108$^{***}$ | 2.111$^{***}$ | 2.065$^{***}$ |
| | (0.205) | (0.0735) | (0.398) | (0.0249) | (0.207) | (0.207) | (0.117) |
| SE*Unpub | | | -1.006 | | | | |
| | | | (0.864) | | | | |
| Constant | 0.0145 | -0.000730 | 0.00195 | 0.00438$^{***}$ | 0.0145 | 0.0145 | 0.0102 |
| | (0.00881) | (0.00128) | (0.00218) | (0.00120) | (0.00881) | (0.00881) | (0.00651) |
| Observations | 2735 | 945 | 3680 | 548 | 2702 | 2716 | 2671 |
| Studies | 169 | 50 | 219 | 31 | 163 | 166 | 164 |

*Notes:* The table presents the results of regression $EIS_{ij} = EIS_0 + \beta \cdot SE(EIS_{ij}) + u_{ij}$. $EIS_{ij}$ and $SE(EIS_{ij})$ are the $i$-th estimates of the elasticity of intertemporal substitution and their standard errors reported in the $j$-th studies. Estimated by weighted least squares with the inverse of the reported estimate's standard error taken as the weight. Standard errors of regression parameters are robust, clustered at the study level, and shown in parentheses. Pub = only published studies (the baseline result reported in the main body of the paper). Unpub = only unpublished studies. All = published and unpublished studies. Finance = only studies from finance journals (listed in the category "Business, Finance" in Journal Citation Reports by Thomson Reuters). Asym = estimates with asymmetric confidence intervals are excluded. LLR = studies using the long-run risks model are excluded. Boot = estimates with bootstrapped standard errors are excluded. $^{***}$, $^{**}$, and $^{*}$ denote statistical significance at the 1%, 5%, and 10% level.

In the next column I only include estimates from studies published in finance journals to see whether my analysis is affected by the use of more economics than finance estimates of the EIS. To define finance journals I use the category "Business, finance" in Journal Citation Reports published by Thomson Reuters. The results are similar to the baseline case concerning both reporting bias and the underlying mean EIS. Next, some reported estimates of the EIS have asymmetric confidence intervals. In the main body of the paper I construct an approximate standard error based on the average distance between the lower and upper bounds of the confidence intervals, but this shortcut may affect my analysis. Column "Asym" in Table 3 shows results computed when the estimates with asymmetric confidence intervals are omitted; the results are very similar to the baseline case. In the next columns I exclude estimates based on the long-run risks model and bootstrapped standard errors, respectively, but again the results are little affected by this restriction.

Bansal *et al.* (2012) stress the difference between consumers' decision frequency and the econometrician's sampling frequency of the data. They estimate the decision frequency to be approximately monthly and show that estimating the EIS with a lower sampling frequency without taking into account temporal aggregation leads to a substantial downward bias. Table 4 shows mean reported estimates of the EIS when different sampling frequencies are used. In line with the argument of Bansal *et al.* (2012), it seems that studies using higher frequencies report larger estimates of the EIS, but the differences are small (0.1 both between monthly and quarterly frequency and between quarterly and annual frequency). In Section 4 of the main body of the paper I construct dummy variables for annual and monthly sampling frequencies with the most common quarterly frequency as the base. My findings suggest that different sampling frequencies do not lead to systematically different results when the other method choices are controlled for.

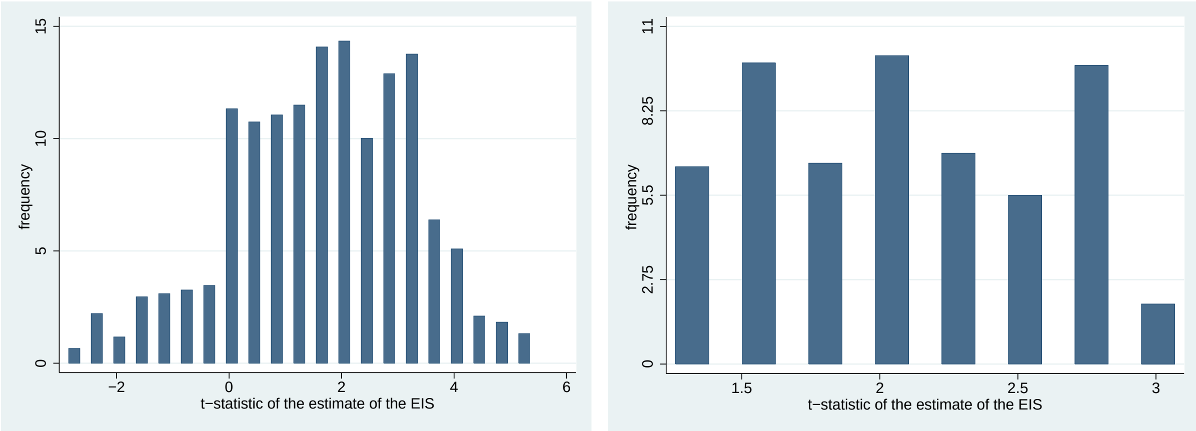
Table 4: Sampling frequency and the EIS

| Sampling frequency | monthly | | quarterly | | annual | |
|---|---|---|---|---|---|---|
| Mean EIS | 0.87 | (0.21) | 0.79 | (0.13) | 0.67 | (0.08) |
| Studies | 17 | | 93 | | 74 | |

*Notes:* Standard errors, clustered at the study level, in parentheses. Weighted by the inverse of the number of estimates reported per study. Some studies use more sampling frequencies.

Figure 1: Negative and marginally insignificant estimates are underreported

(a) Histogram of $t$-statistics (b) $t$-statistics around $t = 2$




*Notes:* In the absence of selective reporting the distribution of the $t$-statistics should be approximately normal. Weighted by the inverse of the number of estimates reported per study. I exclude estimates with large $t$-statistics from the figure but include all in the regressions.

Figure 1 shows histograms of the reported $t$-statistics. In the main body of the paper I use the median $t$-statistics reported in individual studies, because some studies report many estimates with similar $t$-statistics, which would distort the histogram. In this appendix I show the histogram of all $t$-statistics, weighted by the inverse of the number of estimates reported per study. In the left-hand panel we can see how negative estimates are less likely to be reported and how the threshold values for statistical significance are associated with changes in the frequency of the reported estimates. As in the main body of the paper, here I also report a histogram of $t$-statistics around two (the right-hand panel of the figure), which shows the importance of each conventional threshold for statistical significance.

Statistical tests reject the hypothesis that the distribution of the reported $t$-statistics corresponding to the estimates of the EIS is close to normal. First, I consider only the median $t$-statistics reported in the studies, as depicted in the main body of the paper. The $z$-statistic of the Shapiro-Wilk test of normality is 9.8, rejecting the null hypothesis of normality at the 1% level. The $z$-statistic of the Shapiro-Francia test is 9.0, again rejecting the null hypothesis at the 1% level. When extremely large $t$-statistics (above 5 in absolute value) are excluded, normality is still rejected, but only at the 5% level of significance. The reported $t$-statistics deviate from the normal distribution especially in terms of skewness, but not in terms of kurtosis (I cannot reject the hypothesis of normal kurtosis), which again highlights the importance of selective reporting. When all the estimates, not only the study medians, are included, the results are similar: the hypothesis of normal distribution is rejected in all cases at the 1% level. Nevertheless, I cannot reject the hypothesis of normal kurtosis.

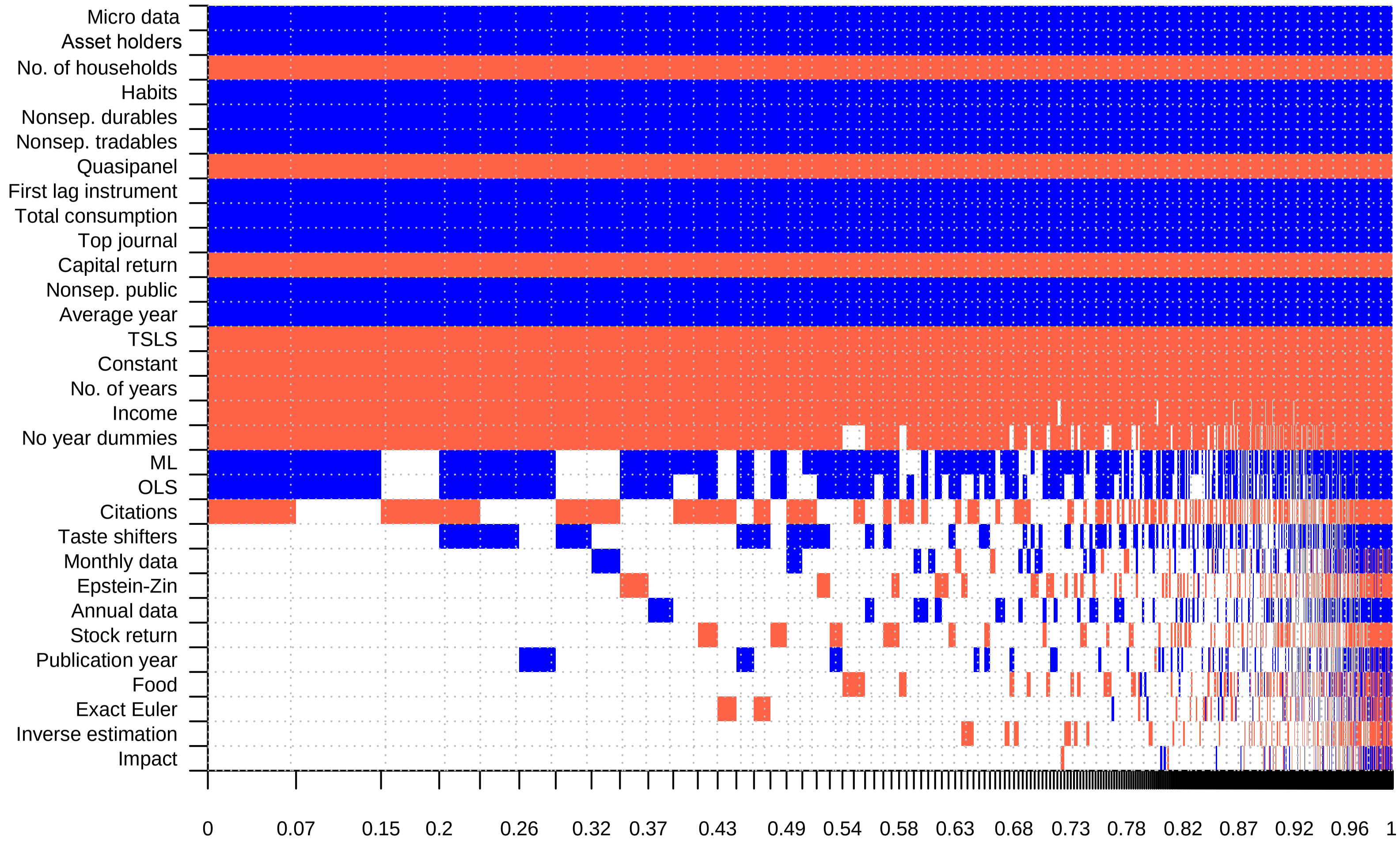
## 3 Bayesian Model Averaging

Many potential variables can be included in the model to explain the heterogeneity in the estimates of the elasticity of intertemporal substitution, but I am not sure which variables are important. The theory does not say much about why, for instance, a particular characteristic of the data should lead to systematically different estimates of the elasticity. If I included all potential variables at the same time, my regression would probably be misspecified. One solution to this model uncertainty is Bayesian model averaging.

In Bayesian model averaging I estimate many regressions with the possible subsets of all potential explanatory variables on the right-hand side and construct a weighted average over these regressions. The weights used in the Bayesian model averaging are the so-called posterior model probabilities. Posterior model probabilities can be thought of as a measure of the fit of the model, analogous to information criteria. For each explanatory variable I then compute the posterior inclusion probability, which represents the sum of the posterior model probabilities of all models that contain this particular variable. The posterior inclusion probability expresses how likely it is that the particular variable should be included in the "true" regression. For the estimation of Bayesian model averaging I use the `bms` package available in R (developed by Feldkircher & Zeugner, 2009, who also provide a detailed explanation of Bayesian model averaging). More details on the Bayesian model averaging procedure employed in this appendix are available in Table 6 and Figure 3.

The results of Bayesian model averaging are reported in Figure 2 and Table 5. The figure shows the individual regression models, which are displayed in the columns. The explanatory variables are sorted by posterior inclusion probability in descending order. Blue color (darker in grayscale) means that the variable is included and the estimated sign of its regression parameter is positive. Red color (lighter in grayscale) means that the variable is included and the estimated sign is negative. The horizontal axis measures the cumulative posterior model probabilities; that is, the best models are displayed on the left. The table shows the weighted average over all

Figure 2: Bayesian model averaging, model inclusion




*Notes:* Response variable: the estimate of the elasticity of intertemporal substitution reported in studies. Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative. No color = the variable is not included in the model. The horizontal axis measures cumulative posterior model probabilities.

Table 5: Explaining the differences in the reported estimates of the EIS

| Variable | PIP | Posterior mean | Posterior std. dev. | Standardized coef. |
|---|---|---|---|---|
| *Utility* | | | | |
| Epstein-Zin | 0.141 | -0.002 | 0.005 | -0.004 |
| Habits | 1.000 | 0.291 | 0.016 | 0.285 |
| Nonsep. durables | 1.000 | 0.043 | 0.006 | 0.500 |
| Nonsep. public | 1.000 | 0.176 | 0.033 | 0.061 |
| Nonsep. tradables | 1.000 | 0.160 | 0.019 | 0.104 |
| *Data* | | | | |
| No. of households | 1.000 | -0.082 | 0.004 | -1.000 |
| No. of years | 0.999 | -0.019 | 0.004 | -1.105 |
| Average year | 1.000 | 3.684 | 0.829 | 442.2 |
| Micro data | 1.000 | 0.507 | 0.022 | 1.134 |
| Annual data | 0.135 | 0.002 | 0.005 | 0.005 |
| Monthly data | 0.142 | 0.001 | 0.007 | 0.007 |
| *Design* | | | | |
| Quasipanel | 1.000 | -0.120 | 0.015 | -0.199 |
| Inverse estimation | 0.061 | -0.001 | 0.004 | -0.008 |
| Asset holders | 1.000 | 0.281 | 0.030 | 0.112 |
| First lag instrument | 1.000 | 0.046 | 0.009 | 0.422 |
| No year dummies | 0.905 | -0.225 | 0.087 | -0.123 |
| Income | 0.986 | -0.022 | 0.006 | -0.068 |
| Taste shifters | 0.321 | 0.014 | 0.022 | 0.014 |
| *Variable definition* | | | | |
| Total consumption | 1.000 | 0.079 | 0.007 | 0.159 |
| Food | 0.117 | -0.019 | 0.074 | -0.011 |
| Stock return | 0.130 | -0.001 | 0.002 | -0.004 |
| Capital return | 1.000 | -0.021 | 0.004 | -0.213 |
| *Estimation* | | | | |
| Exact Euler | 0.073 | -0.001 | 0.003 | -0.008 |
| ML | 0.715 | 0.020 | 0.016 | 0.102 |
| TSLS | 1.000 | -0.040 | 0.008 | -0.082 |
| OLS | 0.613 | 0.015 | 0.014 | 0.029 |
| *Publication* | | | | |
| Publication year | 0.122 | 0.240 | 0.768 | 28.84 |
| Citations | 0.497 | -0.003 | 0.003 | -0.066 |
| Top journal | 1.000 | 0.242 | 0.013 | 0.433 |
| Impact | 0.027 | 0.000 | 0.001 | 0.000 |
| Constant | 0.999 | -29.71 | 8.20 | -470.0 |

*Notes:* Estimated by Bayesian model averaging. Response variable: the estimate of the elasticity of intertemporal substitution reported in the studies. PIP = posterior inclusion probability. The posterior mean is analogous to the estimate of the regression coefficient in a standard regression; the posterior standard deviation is analogous to the standard error of the regression coefficient in a standard regression.

models with weights equal to the posterior model probabilities. The results of Bayesian model averaging are similar to the main results reported in the paper: the estimate of the mean elasticity of intertemporal substitution related to asset holders conditional on my definition of best practice is 0.42.

Table 6: Summary of BMA estimation

| | | | |
|---|---|---|---|
| *Mean no. regressors* | *Draws* | *Burn-ins* | *Time* |
| 20.9816 | $2 \cdot 10^8$ | $1 \cdot 10^8$ | 17.39666 hours |
| *No. models visited* | *Modelspace* | *Visited* | *Topmodels* |
| $31,863,505$ | $2.1 \cdot 10^9$ | 1.5% | 100% |
| *Corr PMP* | *No. obs.* | *Model prior* | *g-Prior* |
| 1.0000 | 2,735 | uniform / 15.5 | UIP |
| *Shrinkage-Stats* | | | |
| Av= 0.9996 | | | |

*Notes:* UIP = unit information prior, PMP = posterior model probability. I follow the advice of Eicher *et al.* (2011), who suggest using the uniform model prior and the unit information prior because these priors perform well in forecasting exercises.

## References


Ashenfelter, O., C. Harmon, & H. Oosterbeek (1999): "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." *Labour Economics* **6(4)**: pp. 453–470.

Bansal, R., D. Kiku, & A. Yaron (2012): "Risks For the Long Run: Estimation with Time Aggregation." *NBER Working Papers 18305*, National Bureau of Economic Research, Inc.

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.

Feldkircher, M. & S. Zeugner (2009): "Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging." *IMF Working Papers 09/202*, International Monetary Fund.

Hedges, L. V. (1992): "Modeling Publication Selection Effects in Meta-Analysis." *Statistical Science* **7(2)**: pp. 246–255.

Figure 3: Model size and convergence