

Habit Formation in Consumption: A Meta-Analysis*

Tomas Havranek^{a,b}, Marek Rusnak^{a,b}, and Anna Sokolova^{c,d}

^aCzech National Bank

^bCharles University, Prague

^cNRU Higher School of Economics, Moscow

^dUniversity of Nevada, Reno

March 16, 2017

Abstract

We examine 597 estimates of habit formation reported in 81 published studies. The mean reported strength of habit formation equals 0.4, but the estimates vary widely both within and across studies. We use Bayesian and frequentist model averaging to assign a pattern to this variance while taking into account model uncertainty. Studies employing macro data report consistently larger estimates than micro studies: 0.6 vs. 0.1 on average. The difference remains 0.5 when we control for 30 factors that reflect the context in which researchers obtain their estimates, such as data frequency, geographical coverage, variable definition, estimation approach, and publication characteristics. We also find that evidence for habits strengthens when researchers use lower data frequencies, employ log-linear approximation of the Euler equation, and utilize open-economy DSGE models. Moreover, estimates of habits differ systematically across countries.

Keywords: Habit formation, consumption, meta-analysis, Bayesian model averaging, frequentist model averaging

JEL Codes: C83, D12, E21

*An online appendix with data, code, and additional results is available at meta-analysis.cz/habits. The corresponding author is Tomas Havranek, e-mail: tomas.havranek@ies-prague.org. We gratefully acknowledge helpful suggestions by the editor Eric Leeper, an anonymous associate editor, and two anonymous referees. We thank Jan Babecky, Alexei Boulatov, Kamil Galuscak, Zuzana Irsova, Simona Malovana, Udara Peiris, Sergey Pekarski, Tom Stanley, Federica Teppa, Dimitrios Tsomocos, and Stefan Zimmermann for their comments on earlier versions of the manuscript. We acknowledge support from the Czech Science Foundation (grant #15-02411S). Anna Sokolova acknowledges support from the Basic Research Program at the National Research University Higher School of Economics. The views expressed here are ours and not necessarily those of the Czech National Bank.

1 Introduction

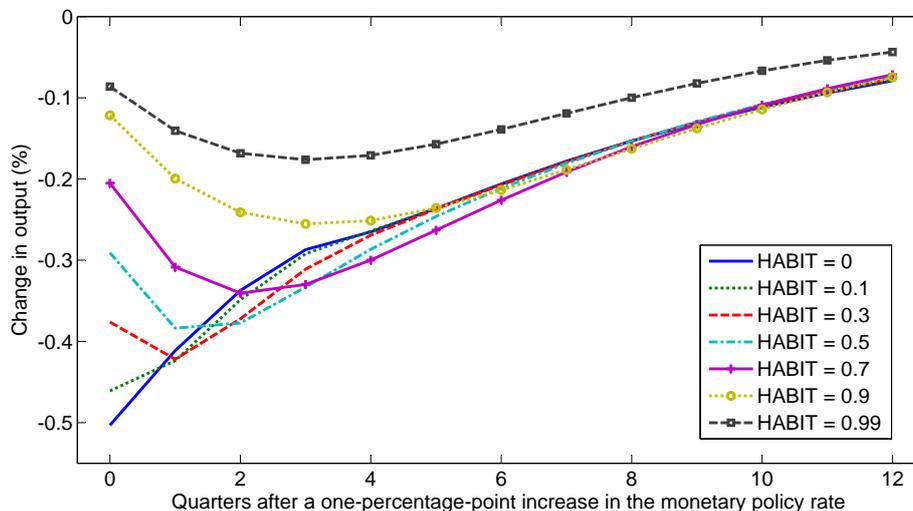
The concept of habit formation in consumption is crucial for the explanation of various stylized facts in macroeconomics and finance. For example, in the asset pricing literature consumption habit helps reconcile the theory with the observed moments of asset returns. Mehra & Prescott (1985) show that the standard Lucas (1978) tree model fails to replicate the high equity premium and low risk-free rate at reasonable model parameters. Constantinides (1990) argues that habit formation solves this problem, as it can generate large variability in the marginal rate of substitution in consumption alongside smooth consumption growth—a feature that allows one to replicate high risk premium without having to rely on large risk aversion. Further refinements of the model suggested by Campbell & Cochrane (1999), Abel (1999), and Allais (2004) make it possible also to generate plausible variability in equity returns and the risk-free rate, while adding habits to a real business cycle framework helps explain the joint behavior of asset prices and consumption (Boldrin *et al.*, 2001).

The presence of habit formation implies that past consumption choices affect current preferences. This notion violates the independence axiom used by Koopmans (1960) to derive the classic discounted utility model. Due to the growing popularity of models with habits, researchers have made efforts to develop theoretical underpinning for utility that is non-separable over time and features habit formation. Rozen (2010) lays out axiomatic foundation for a utility function displaying linear internal habits, describing a decision maker whose preferences depend on the history of past consumption choices. Rustichini & Siconolfi (2014) present a general axiomatic approach that allows for time-separable and non-separable utility as special cases, while He *et al.* (2013) put forward a model that incorporates habits as well as satiation in utility.

Studies that feature general equilibrium models have come to rely on consumption habit as a means of replicating a delayed hump-shaped response of macro variables to policy shocks (Fuhrer, 2000; Del Negro *et al.*, 2007). This is because habit formation makes abrupt changes in consumption costly, thereby inducing smoothness in consumption dynamics (for a detailed discussion see Kano & Nason, 2014). But the quantitative predictions of such models largely depend on the size of the parameter specifying the strength of habit formation. Figure 1 shows how the impulse response of output to a nominal interest rate shock changes in the popular model by Smets & Wouters (2007) when we assume different values of habit formation: the

modeled behavior of the economy within one year after the shock depends heavily on the assumed strength of habits.¹

Figure 1: The importance of habit formation for DSGE models



Notes: The figure shows simulated impulse responses of GDP to a one-percentage-point increase in the monetary policy rate. We use a calibrated version of the model developed by Smets & Wouters (2007) and vary the value of the habit formation parameter while leaving all other parameters calibrated at the posterior values from Smets & Wouters (2007). For the simulations we use Matlab code from the Macroeconomic Model Data Base (Wieland *et al.*, 2012).

Dozens of papers have estimated the habit formation parameter, but their results vary widely. The variance can be partially attributed to differences in the data used in the estimation: some studies analyze Euler equations for aggregate consumption (Fuhrer, 2000; Carroll *et al.*, 2011; Everaert & Pozzi, 2014), some employ micro panel data sets (Dynan, 2000; Collado & Browning, 2007; Alessie & Teppa, 2010), and others use DSGE models (Christiano *et al.*, 2005; Smets & Wouters, 2007), often employing prior values for the habit parameter. A brief look at the results of the seminal studies in each category suggests that the estimates are all over the place: Fuhrer (2000) shows that habit formation is crucial for his model to fit the data and obtains estimates that lie within the range 0.8–0.9. In contrast, Dynan (2000) uses panel

¹Figure 1 closely resembles Kano & Nason (2014, Figure 1) depicting the impulse-response function of consumption growth rate to a real interest rate shock for different values of the habits parameter. Remarkably, the result of Kano & Nason (2014) does not rely on the full New Keynesian DSGE model: they derive it using a log-linear approximation of the Euler equation and an AR(1) process for the real interest rate. Absent habit formation, the Euler equation sets consumption growth equal to the real interest rate, an AR(1); therefore, the impulse-response function peaks at 0 and decays afterward. With habit formation, current changes in consumption are associated with a utility loss in the future—in consequence, changes in the interest rate lead to a gradual adjustment of consumption growth.

household data and finds no evidence of habit formation. Christiano *et al.* (2005) estimate the same parameter using a DSGE model and obtain a value in the range 0.5–0.7.²

In this paper we investigate whether this diversity in the estimates of the habit parameter can be explained through differences in study designs used by researchers. We present what to our knowledge is the first quantitative synthesis—or a meta-analysis—of the evidence from the literature estimating habit formation. Meta-analyses attempt to trace variation in the estimates reported in the literature to differences in how the studies are conducted; it is the quantitative method of research review frequently used in medical research, which has recently become used by economists as well (Stanley, 2001). In economics the method has been applied to a wide range of topics: the effect of the minimum wage on unemployment (Card & Krueger, 1995), returns from education (Ashenfelter *et al.*, 1999), the effect of distance on trade (Disdier & Head, 2008), the intertemporal elasticity of substitution in labor supply (Chetty *et al.*, 2011), and the impact of FDI on domestic firms’ productivity (Havranek & Irsova, 2011), among others.

We gather 81 published studies presenting estimates of habit formation and collect 31 aspects related to study design, such as the estimation techniques used, variable definition, data characteristics, geographical coverage, and model specification. We attempt to establish whether these aspects systematically affect the reported estimates of the habit parameter.

We cannot claim that our method allows us to explain variation in the true degree of habit formation; instead, we attempt to explain differences in its *estimates* reported in previous studies—a task that meta-analysis can accomplish. One obstacle that we face is the uncertainty over which of the 31 study characteristics should be included in the model approximating the process that generates habit estimates. To address this problem we employ Bayesian model averaging (BMA; Raftery *et al.*, 1997; Moral-Benito, 2015)—a method that estimates many regressions consisting of subsets of the potential explanatory variables and weights them by model fit and model complexity. As a robustness check we use frequentist model averaging, which does not rely on Bayesian methods.

Our results show that the difference between micro estimates (think Dynan, 2000) and macro estimates (think Fuhrer, 2000) remains large even after controlling for other aspects of study design. This finding resonates with Chetty *et al.* (2011), who report similar divergence

²Sampling uncertainty would seem to suggest that habit estimates of 0.5 and 0.9, for example, are not that far apart, but Figure 1 shows that the economic implications of consumption habit differ greatly across this range of estimates.

between micro and macro estimates in the literature estimating the intertemporal elasticity of labor supply. Furthermore, the frequency of the data used in the estimation matters: estimates from studies employing monthly data tend to be substantially smaller than those obtained with lower frequencies, with the largest estimates being associated with the use of annual data. We also find that the use of second-order approximation of the Euler equation yields smaller estimates, which indicates that it is important to account for the precautionary saving motive when evaluating habit formation. Estimates obtained using US data tend to be larger than those reported for Japan, Europe, and other regions. Additionally, our results suggest that among the DSGE studies the ones that rely on the open-economy framework typically require higher degrees of habit formation to match the dynamics of the observables.

By contrast, we find that studies using the moments of asset returns do not report estimates that differ systematically from those obtained without the use of stock market data. In a similar vein, given the features of the data employed by the particular study, the use of the DSGE methodology itself does not result in estimates that are systematically different from those obtained by other methods. This finding suggests that reproducing empirical moments of the data within structural models requires roughly the same degree of habit formation as what would typically arise from reduced-form estimation with similar data sets.

We do not find evidence of systematic differences between the estimated magnitude of external habits (“keeping up with the Joneses”) and internal habits (past own consumption decreases present utility) when other data and method characteristics are controlled for. The result is in line with Dennis (2009), who shows that the distinction between internal and external habits has a limited effect on the business cycle characteristics of New Keynesian models, and Kano & Nason (2014), who show in their online appendix that for log-linear approximation of the Euler equation under additive habits there is observational equivalence between external and internal specifications.³ We also find that estimates of habits formed at the level of individual goods do not systematically differ from those of habits formed over the whole consumption bundle. Furthermore, studies using total non-durable consumption, food expenditures, or measures that include durable consumption come up with estimates that are roughly the same. However, we find that the use of simple panel techniques that do not rely on instrumental variables system-

³The online appendix to Kano & Nason (2014) is available at <http://hermes-ir.lib.hit-u.ac.jp/rs/bitstream/10086/23297/1/070econDP12-08.pdf>.

atically affects the results. We also observe a correlation between the reported estimates and the characteristics of the journal where the study is published.

The remainder of the paper is structured as follows. Section 2 describes the approach we use to collect estimates of habit formation and presents the summary statistics for our data set. Section 3 investigates the sources of heterogeneity in the estimated habit formation parameters. Section 4 concludes. Appendix A provides the correlation matrix of the variables used, shows diagnostics of the Bayesian model averaging exercise, tests for publication bias in the literature, and provides a robustness check using an alternative set of priors. Appendix B discusses issues related to model uncertainty. Appendix C shows the list of studies included in the data set. An online appendix with data, code, and additional results is available at meta-analysis.cz/habits.

2 The Data Set of Habit Formation Estimates

2.1 Estimating the Degree of Habit Formation

Modeling habit formation usually involves the following utility function:

$$\sum_t \beta^t u(c_{i,t} - \gamma h_{i,t}), \quad (1)$$

where β is a discount factor, $u(\cdot)$ denotes the instantaneous utility function, $c_{i,t}$ is the consumption of individual i in period t , $h_{i,t}$ is the reference habit stock, and $\gamma \in [0, 1)$ captures the strength of habit formation ($\gamma = 0$ gives the standard time-separable utility function). Papers that explore internal habits assume $h_{i,t} = c_{i,t-1}$: lagged own consumption decreases current utility. Under internal habits, therefore, utility is determined by consumption growth, not just the level of current consumption. Papers studying external habits (“catching up with the Joneses,” Abel, 1990) assume that utility is determined by the difference between the current consumption of an individual and the consumption of the corresponding reference group (for instance, the city where the consumer lives). External habits can be modeled by defining $h_{i,t} = \tilde{c}_{t-1}$, where \tilde{c}_{t-1} denotes aggregate consumption in the preceding period. Several studies investigate “deep” habits formed at the level of individual goods rather than the whole consumption bundle (e.g., Ravn *et al.*, 2006; Lubik & Teo, 2014). Additionally, instead of using consumption directly, some papers use the variable “habit stock” defined by an autoregressive process (for example,

Fuhrer, 2000). Finally, a few studies model habits using a multiplicative rather than an additive specification; for example, Andrés *et al.* (2009) and Bjornland *et al.* (2011).

A common approach to estimating γ is to evaluate an approximation of the consumption Euler equation that incorporates habits. For example, with internal habit formation instantaneous utility depends on the household's past consumption; therefore, from the households' perspective, an increase in consumption today affects not only utility of the current period, but also future utility—by affecting future habit stock. The marginal effect on welfare of an increase in current consumption c_t is then given by

$$\lambda_t = u'(c_t - \gamma c_{t-1}) - \gamma \beta E_t u'(c_{t+1} - \gamma c_t). \quad (2)$$

Habit-forming households will internalize this effect when making consumption decisions; this is reflected in the first-order condition of the households' problem:

$$\lambda_t = \beta E_t \left\{ \frac{\lambda_{t+1} R_{t+1}}{1 + \pi_{t+1}} \right\}, \quad (3)$$

which relates expected marginal effects of changes in c_t and c_{t+1} to the nominal interest rate R_{t+1} and inflation π_{t+1} .

As shown in Kano & Nason (2014), assuming that utility is logarithmic and total factor productivity is driven by a random walk, Euler equation (3) can be approximated by

$$\Delta c_t = \phi_1 \Delta c_{t-1} + \frac{(\alpha^* - \beta\gamma)(\alpha^* - \gamma)}{\alpha^{*2} \phi_2} \sum_{j=0}^{\infty} \phi_2^{-j} E_t q_{t+j}, \quad (4)$$

where q_t is the demeaned real interest rate, α^* is a steady-state growth, $\phi_1 = \gamma \alpha^{*-1}$ and $\phi_2 = \alpha^* (\beta\gamma)^{-1}$ (see Kano & Nason 2014, Equation 1). Therefore, Euler equation (3) implies a relationship between current and past consumption growth, and a forward-looking component related to the expected discounted sum of future interest rates. Several studies derive their estimates of the habit parameter from a simplified version of approximation (4) that assumes a constant interest rate (e.g., Dynan, 2000; Carroll *et al.*, 2011; Sommer, 2007). Furthermore, some studies employ higher-order approximations and account for the precautionary saving motive by including a measure of consumption risk in the estimated specification (e.g., Guariglia, 2002).

Approximation (4) is derived from the problem faced by an individual household, and a number of studies obtain estimates of habit formation by using individual household data (for example, Dynan, 2000; Guariglia, 2002; Alessie & Teppa, 2010). But micro studies often have

data covering only short periods of time, and only on a fraction of consumption (such as food expenditures), and micro data are also often noisy, yielding imprecise estimates. Therefore, in practice similar specifications are often estimated on aggregate data. However, such treatment of the Euler equation may bias estimation results.

Attanasio & Weber (1993) point out that a correct aggregation of log-linear representation of the Euler equation necessitates examining a sum of the logarithms of the individual expenditures. Nevertheless, aggregate national accounts data only provide information on the sum of expenditures; researchers relying on aggregate data then take logarithms of the sum. Attanasio & Weber (1993) argue that if the cross-sectional distribution of expenditures varies over time, then the difference between the two measures will not be constant, and in fact is likely to be serially correlated. This induces serial correlation in the error term of log-specifications estimated on the aggregate data, a problem that could potentially lead to spurious results when estimating the habit parameter.

Another potential problem associated with the use of aggregate data stems from the fact that macro studies typically cannot account for households' taste shifters such as age, number of children, and employment status, all of which are likely to influence consumption decisions. Attanasio & Weber (1993) point out that when changes in these taste shifters do not cancel out after aggregating across the population, then the Euler equation cannot be consistently estimated on aggregate data. For example, it is well known that the consumption profile varies within the life cycle and depends on the individual's age (e.g., Attanasio & Weber 1995). If population composition with respect to young and old changes over time, then the aggregate specification that does not account for this effect will be prone to omitted variable bias. Attanasio & Weber (1993) also argue that aggregation bias may arise if households do not have full information about aggregate events: that would make aggregate instruments invalid for identifying household preference parameters. All these issues may bias estimates of the habit parameter obtained using aggregate data.

The voluminous macro literature that estimates habits is diverse, employing various data sets and approaches to estimation, as we discuss below. These papers estimate consumption habit while studying issues like sticky consumption growth (Carroll *et al.*, 2011), habit persistence in current account data (Gruber, 2004; Kano, 2009), predictability of aggregate consumption

growth (Everaert & Pozzi, 2014), inflation dynamics (Fuhrer, 2000), and moments of asset returns (Heaton, 1995). Many estimates of the habit formation parameter come from dynamic stochastic general equilibrium models. Those estimates can be obtained by minimizing the distance between the model predictions and the empirical impulse response function (Christiano *et al.*, 2005), by maximizing the likelihood of the state space representation of the model (Bouakez *et al.*, 2005), or by using Bayesian methods (Smets & Wouters, 2007).

2.2 Collecting Estimates of γ

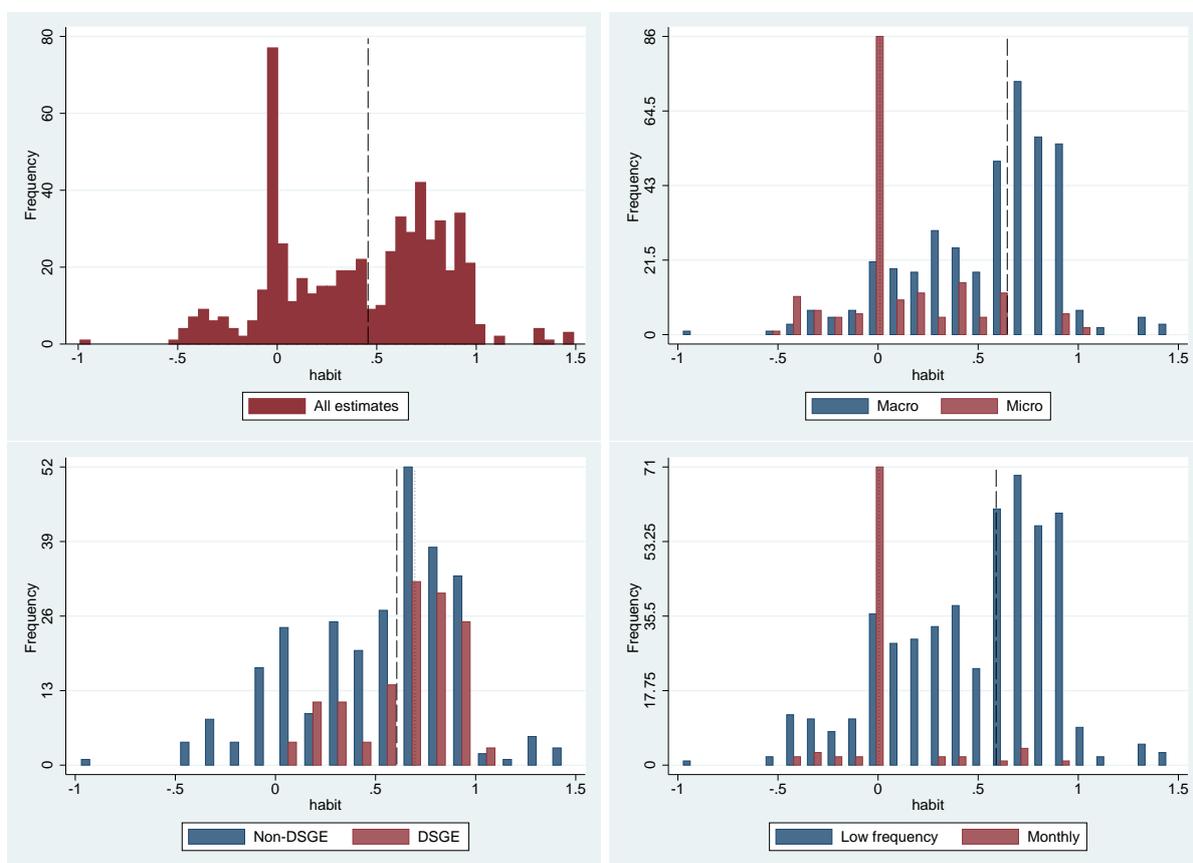
The first step of any meta-analysis is to gather the empirical studies on the topic, usually referred to as “primary studies.” To collect primary studies, meta-analyses in economics often employ the RePEc or EconLit databases. We use Google Scholar because it provides powerful full-text search, whereas RePEc and EconLit only allow searching through abstracts and keywords related to the studies, thereby making it harder to devise an exhaustive search query. We first collect papers that contain the exact phrases “habit formation” or “habit persistence” and, at the same time, feature occurrences or synonyms of the following words: consumption, estimate, regression, and empirical. After reading the abstracts of the studies returned by our search query we download those that show any promise of containing empirical estimates of the habit formation parameter. In the next step we extend our search to the references of these studies and add the last study on March 1, 2016.

We apply the following three inclusion criteria. First, the study must provide an empirical estimate of the habit formation parameter. Second, the study must include an estimate of the standard error (or a statistic from which the standard error can be computed). Finally, the third inclusion criterion is that the study must be published in a peer-reviewed journal. Meta-analyses differ in their treatment of unpublished results—sometimes they include unpublished papers as well, especially when the resulting data set would otherwise be small. Since there are many published studies estimating the habit formation parameter, we prefer to focus on studies that have been subjected to a peer-review process. We find 81 studies that comply with our selection criteria, and we list them in Appendix C.

Each primary study typically reports several estimates, and the median number of estimates per study is four. It is hard to pin down each study’s representative estimate, because the au-

thors themselves rarely say explicitly which one they prefer. Therefore, we collect all estimates reported in each study. This approach results in an unbalanced data set, as some studies report many more estimates than others—nevertheless, it allows us to exploit the differences in data and method choices both within and across individual studies. Wherever possible, we include study fixed effects to filter out the effects of study-level characteristics that are otherwise unobservable. All studies combined provide us with 597 estimates of the habit formation parameter, and for each of them we collect 31 variables reflecting the context in which researchers obtain the estimates.

Figure 2: Estimates of habit formation vary widely



Notes: The figure shows histograms of the estimates of the habit formation parameter reported in individual studies. The top-left panel shows all estimates, the top-right panel splits the sample based on the level of aggregation in the data, the bottom-left panel examines a subsample of macro studies, splitting it based on whether the DSGE methodology is used, and the bottom-right panel splits the overall sample based on the frequency of the data. Dotted and dashed lines correspond to the respective medians.

The top-left panel of Figure 2 present a histogram of the estimated parameters, providing additional insights. First, the distribution of the estimates is far from normal, and both the lower and upper boundaries of the range 0–1, consistent with habit formation, seem to affect the

probability of an estimate being reported. This result, however, may also reflect the constraints that researchers use in the process of estimation. Second, while not normal, the distribution of estimates is relatively symmetric, as both the lower and the upper tails are cut off, and the mean estimate virtually equals the median. Third, the histogram has multiple peaks, suggesting heterogeneity generated by different estimation methods.

To shed some light on the sources of heterogeneity, we split the sample of all estimates into subsamples, depending on whether the study uses household-level or aggregate data (the top-right panel, Figure 2). The histogram of micro estimates peaks at a much lower level of the habit parameter than that of macro estimates. Furthermore, neither distribution seems to be symmetrical: for micro studies, the right tail is heavier, while for macro studies the opposite holds. We further split the sample of macro studies, distinguishing between studies that estimate the habit parameter within DSGE models and those that use other techniques. The shapes of the histograms displayed on the bottom-left panel of Figure 2 are similar, suggesting that two groups of estimates may come from the same distribution. Finally, we investigate the role played by the frequency of the data by splitting the overall sample into estimates obtained using monthly data versus data using lower frequencies (the bottom-right panel). Data frequency seems to affect the estimates' distribution, and the use of monthly data is likely to result in lower estimates of the habit parameter.

Table 1: Habit formation estimates for different data and methods

	Unweighted				Weighted				No. of est.
	Mean	Median	5%	95%	Mean	Median	5%	95%	
All estimates	0.43	0.47	-0.32	0.97	0.55	0.62	-0.21	0.98	597
Micro studies	0.10	0.00	-0.39	0.62	0.12	0.08	-0.41	0.62	183
Macro studies	0.57	0.66	-0.11	0.98	0.62	0.69	0.00	0.99	414
Internal	0.28	0.15	-0.38	0.95	0.41	0.44	-0.34	0.96	369
External	0.66	0.67	0.16	1.00	0.72	0.71	0.16	1.48	228
Asset returns	0.43	0.62	-0.44	0.96	0.47	0.57	-0.29	0.96	87
Micro - internal	0.03	0.00	-0.40	0.60	0.09	0.01	-0.41	0.62	147
Micro - external	0.40	0.37	0.06	0.96	0.40	0.37	0.06	0.96	36
Macro - internal	0.45	0.61	-0.33	0.97	0.51	0.63	-0.22	0.98	222
Macro - external	0.70	0.71	0.21	1.00	0.73	0.73	0.19	1.48	192
Macro - non DSGE	0.52	0.62	-0.28	1.10	0.52	0.55	-0.28	1.10	279
Macro - DSGE	0.67	0.71	0.16	0.97	0.68	0.71	0.18	0.98	135

Notes: 5% and 95% denote the corresponding percentiles. Weighted = summary statistics based on the observations weighted by the inverse of the number of estimates reported per individual study. In such case each study receives the same weight in the computation of the summary statistics.

We compute average and median values for different groups of estimates and display them in Table 1. The overall mean of the reported estimates is approximately 0.4. Studies using micro data deliver much smaller estimates on average—about 0.1. By contrast, macro studies tend to generate larger estimates: around 0.6. Among the macro approaches to assessing habit formation, DSGE studies tend to yield slightly larger estimates. The nature of the habit formation process seems to matter, too. Estimates of internal habit formation average 0.3, while estimates of external habits tend to be more than twice as large at around 0.7. The difference between estimates of external and internal habits remains substantial, albeit smaller, even when we calculate the means separately for macro and micro studies. For macro data, estimates of external habits are still larger—this finding seems to contradict the argument of Carroll *et al.* (1997), who suggest that estimates of external and internal habits are empirically indistinguishable when using macro data. We will revisit the difference between internal and external habits in Section 3, where we will control for other aspects of study design. The conclusions outlined above remain intact even when we weight the estimates by the inverse of the number of estimates reported in each study, thereby giving each study the same weight regardless of the number of estimates the study produces.

Table 2: Habit formation differs across countries

	Unweighted				Weighted				No. of est.
	Mean	Median	5%	95%	Mean	Median	5%	95%	
<i>All estimates</i>									
US	0.42	0.40	-0.08	0.96	0.60	0.67	-0.04	1.00	377
EU countries	0.51	0.63	-0.27	1.00	0.48	0.61	-0.27	0.91	151
Japan	0.07	-0.23	-0.46	0.94	0.32	0.39	-0.41	0.96	27
Other countries	0.34	0.30	-0.03	0.78	0.36	0.31	-0.03	0.98	42
<i>Micro estimates</i>									
US	0.13	0.00	-0.09	0.63	0.15	0.08	-0.06	0.49	126
EU countries	0.10	0.07	-0.46	0.99	0.08	0.03	-0.46	0.62	36
Japan	-0.37	-0.39	-0.50	-0.23	-0.37	-0.39	-0.50	-0.23	14
Other countries	0.59	0.58	0.56	0.62	0.59	0.58	0.56	0.62	7
<i>Macro estimates</i>									
US	0.58	0.67	-0.26	0.98	0.66	0.72	0.00	1.00	251
EU countries	0.64	0.70	-0.08	1.12	0.60	0.69	0.07	0.91	115
Japan	0.55	0.64	0.02	0.96	0.50	0.39	0.09	0.96	13
Other countries	0.29	0.24	-0.04	0.93	0.30	0.21	-0.04	0.98	35

Notes: 5% and 95% denote the corresponding percentiles. Weighted = summary statistics based on the observations weighted by the inverse of the number of estimates reported per individual study. In such case each study receives the same weight in the computation of the summary statistics.

Most estimates in our data set are obtained using US data (63%). All studies combined provide results for 17 countries, arguably contributing to the heterogeneity we observe, but the number of countries is not large enough to connect the differences in estimates to the structural differences among the economies. Nevertheless, in Table 2 we compare group averages for the US, Japan, countries belonging to the EU, and the rest of the countries (other OECD economies, such as Australia, Canada, New Zealand, and Korea) and notice several regularities. The estimates of habit formation for the US and EU tend to be larger on average than the estimates for Japan and other countries. At the same time, for macro studies the difference between Japan, the US, and the EU is smaller, while for micro data the highest estimates correspond to the group “other countries,” which, however, only includes seven observations. It is not clear how to interpret these differences, as seeming cross-country diversity may be driven by differences among other features of the data sets, such as their length or frequency. Cross-country papers focusing on habit formation are rare, and the prominent study of this category, Carroll *et al.* (2011), finds homogeneous coefficients for a number of countries in our sample. Thus, we refrain from making any conclusions at this point, but will return to this issue in the more detailed analysis in the next section.

3 Why Do Estimates of Habit Formation Vary?

3.1 Explanatory Variables

We have noted that the estimates of habit formation differ substantially both within and between studies. In this section we attempt to relate the differences in the estimates to differences in the design of primary studies. To this end we collect 31 variables that reflect each study’s data characteristics, geographical coverage, variable definitions, estimation technique, specification features (for studies estimating DSGE models), and publication characteristics (for example the number of citations). We cannot hope that these 31 variables will explain all differences across estimates—the set of potential explanatory variables is close to unlimited—but we believe that our selection reflects the most common choices faced by researchers who estimate habit formation.

Data characteristics For each study we collect the number of observations and the average year of the data used. We specify whether the study employs micro or aggregate data, as the discussion in Subsection 2.1 and the statistics in Subsection 2.2 suggest that this dimension may have crucial effect on the estimates. We also control for the frequency of the data: Bansal *et al.* (2012) argue that studies estimating consumption Euler equations should account for the difference between the econometrician’s sampling frequency and consumers’ decision frequency; the authors estimate the latter to be approximately monthly. Habit formation estimates are likely to be affected by the data frequency because at sufficiently high frequencies every consumption good displays durability, rendering the habit formation parameter negative: a full meal makes people saturated for the next few hours. Most studies employ quarterly data; for those using monthly and annual data we include controls.

Countries examined Although habit formation is supposed to be a so-called deep parameter, differences in structural characteristics of economies (such as culture) might cause the parameter to vary across countries. Havranek *et al.* (2015a) find substantial cross-country heterogeneity in the elasticity of intertemporal substitution in consumption associated with cross-country differences in income and stock market participation. Since the number of countries investigated by the studies in our sample is small, we only use regional dummy variables instead of the underlying characteristics of the countries. We include dummies for US data, data on Japan, and data from countries that are members of the European Union. The remaining studies estimate the habit formation parameter for other non-European OECD countries.

Variable definitions In Section 2 we show that the mean reported estimates of internal and external habit formation differ. To see whether the difference holds after we control for other aspects of data and methodology, we create a dummy variable attributed to the type of habits under investigation. We also create a control for studies that investigate “deep” habits, for which the habit is formed over individual goods rather than the whole consumption basket. Such formulation has an important effect on the dynamics of DSGE models: it implies that habits affect not only the demand side of the economy, but also the supply side, as demand for individual goods becomes dependent on current sales, altering the optimal pricing behavior of firms and yielding countercyclical mark-ups of prices over marginal costs.

Estimates may also differ depending on the consumption good used in the estimation. Studies that include durable goods should obtain lower estimates of the habit formation parameter, while estimates based on food consumption may be biased if food is non-separable from other consumption goods (Attanasio & Weber, 1995). We distinguish three categories of consumption proxies: food consumption, total non-durable consumption, and measures that include durable consumption; non-durable consumption represents our reference category. Finally, a prominent group of studies obtain habit formation parameters by scrutinizing asset pricing moments—we create a control signifying whether the study uses financial data other than capturing returns on government bonds or the economy-wide risk-free interest rate.

Estimation approach A common wisdom in empirical economics is that different estimation approaches often yield different results. We want to find out whether the use of a particular method is associated with systematic differences in the reported habit formation parameter. Most studies estimate habit formation by using reduced-form regression models. For such studies the most common method choice is GMM, although some assume homoskedasticity and employ TSLS. A few panel studies use fixed effects estimation that does not account for the Nickell (1981) bias, or random effects estimation, the assumptions of which are unlikely to hold in consumption Euler equations. Finally, a small fraction of studies estimate habit formation with OLS—we use this estimation approach as the reference group (using, for example, Bayesian estimation as the baseline does not affect the results).

Most of the regression-based estimates are obtained by first-order approximation of the Euler equation. This approach is criticized by Carroll (2001), who argues that terms of higher order are correlated with structural parameters and thus cannot be ignored. Some studies in our sample use second-order approximation, which allows the researchers to account for the precautionary saving motive, relating consumption growth to the degree of income uncertainty. Finally, some studies obtain the habit formation parameter by estimating dynamic stochastic general equilibrium models. These studies use maximum-likelihood-based methods, minimum distance estimators, or Bayesian techniques.

DSGE specification While most DSGE studies in our sample model closed economies, some extend their analysis to the open-economy framework, introducing exchange rate and current

account fluctuations. All such models in our sample can be classified as small-open-economy models in the sense that they do not allow for feedback from domestic variables to foreign output, inflation, and the interest rate. An exogenous world interest rate in particular may play an important role in how the model identifies consumption habit because it pins down domestic interest rate movements. (Unfortunately it is not feasible to control for the different version of open-economy DSGE models; additional dummy variables focusing on these aspects would have very limited variation.) Another feature of the DSGE approach we control for is the set of observables used for estimation. Guerron-Quintana (2010) argues that using too few observables may lead to identification problems and biased impulse responses. Specifically, excluding consumption or the real wage from the set of observables may cause bimodality in the model’s posterior and strongly affect the estimate of the habit parameter. The author compares model forecasting properties and impulse responses for different sets of observables and finds evidence in favor of the set that includes the seven observables used by Levin *et al.* (2005) and Smets & Wouters (2007).⁴ Moreover, Adolfson *et al.* (2008) suggest that models estimated to match impulse responses to monetary policy shocks (e.g., Christiano *et al.*, 2005) tend to deliver lower real friction parameters than models estimated to match all variation in the observables. We introduce a corresponding dummy variable to account for these specification characteristics.

Publication characteristics Finally, we control for the publication characteristics of individual studies. We include the year of publication to capture methodological advances that are otherwise hard to codify or that have not been employed by a sufficient number of studies yet. To account for approximate study quality beyond the observed differences in data and methodology, we include the number of citations, the recursive impact factor of the journal that published the study, and a dummy variable for studies published in top journals. We collect the data on the impact factor from RePEc: unlike other databases, RePEc covers virtually all economics journals and provides a discounted recursive impact factor well-suited for comparison of outlets in economics.

Table 3 describes the 31 explanatory variables mentioned above, listing their means, standard deviations, and means weighted by the inverse of the number of estimates reported in individual studies. The correlation matrix of all the collected explanatory variables is presented

⁴This set includes output, consumption, investment, real wages, total labor, interest rates, and inflation.

in Figure A1 in Appendix A; it shows that the variables reflect different aspects of the studies. A large correlation appears between micro data and the number of observations: micro-level studies tend to have more observations available than macro studies. Bayesian techniques are often employed within the framework of DSGE models, which renders the variable *Bayes* correlated with controls describing the set of observables. Furthermore, the DSGE models that do not include consumption in the set of observables tend to also omit wages, while studies that replicate responses to monetary policy shocks typically rely on minimum distance estimators. Finally, the positive correlation we observe between the year of publication of the study and the average year of data used in the study is intuitive.

Table 3: Description and summary statistics of regression variables

Variable	Description	Mean	Std. dev.	WM
Habit	The estimate of the habit formation parameter (response variable).	0.43	0.45	0.55
SE	The standard error of the estimate of the habit formation parameter.	0.16	0.26	0.15
<i>Data characteristics</i>				
No. of obs.	The logarithm of the number of observations.	6.14	1.81	5.48
Average year	The midpoint of the sample used for the estimation of habit formation (the base is the sample minimum: 1932).	53.30	11.73	52.52
Micro	= 1 if micro data are used for the estimation.	0.31	0.46	0.15
Monthly	= 1 if the frequency of the data used for the estimation is monthly.	0.15	0.36	0.06
Annual	= 1 if the frequency of the data used for the estimation is annual.	0.32	0.47	0.20
<i>Countries examined</i>				
US	= 1 if habit formation is estimated for the US.	0.63	0.48	0.68
EU	= 1 if habit formation is estimated for a country belonging to the EU.	0.25	0.44	0.21
Japan	= 1 if habit formation is estimated for Japan.	0.05	0.21	0.06
<i>Variable definition</i>				
External	= 1 if external habit formation is estimated.	0.38	0.49	0.45
Deep	= 1 if habits apply to individual goods.	0.05	0.21	0.04
Durable	= 1 if durable consumption goods are included in the measure of consumption.	0.74	0.44	0.76
Food	= 1 if food expenditures are used as a proxy for consumption.	0.12	0.32	0.07
Asset returns	= 1 if data on risky financial assets (e.g., stocks, house prices) are used.	0.15	0.35	0.16
<i>Estimation approach</i>				
GMM	= 1 if the general method of moments is employed for the estimation.	0.46	0.50	0.27
TOLS	= 1 if the two-step-least-squares method is employed for the estimation.	0.14	0.35	0.06
Panel	= 1 if a panel technique (fixed effects, random effects) is employed for the estimation.	0.05	0.23	0.02

Continued on next page

Table 3: Description and summary statistics of regression variables (continued)

Variable	Description	Mean	Std. dev.	WM
Second-order approx.	= 1 if second-order approximation is employed.	0.05	0.21	0.08
DSGE	= 1 if the estimation uses a dynamic stochastic general equilibrium model.	0.23	0.42	0.53
Bayes	= 1 if the estimation uses Bayesian inference.	0.20	0.40	0.42
Minimum distance	= 1 if the minimum distance method is employed for the estimation.	0.06	0.24	0.10
ML	= 1 if the maximum likelihood method is employed.	0.03	0.16	0.09
<i>DSGE specification</i>				
Open-economy DSGE	= 1 if the open-economy DSGE framework is employed.	0.02	0.13	0.05
Matching IR to mon. policy	= 1 if the study matches theoretical and empirical impulse responses to monetary policy shocks.	0.05	0.21	0.09
No. of observables in DSGE	The number of observables the study matches.	1.47	2.89	3.69
Seven observables from SW	= 1 if the list of observables includes proxies for output, consumption, investment, the wage, labor, the interest rate, and inflation.	0.06	0.23	0.16
No consumption	= 1 if the list of observables does not include consumption.	0.13	0.34	0.25
No wage	= 1 if the list of observables does not include real wages.	0.15	0.36	0.31
<i>Publication characteristics</i>				
Publication year	The year in which the study was published (base = 1991).	14.50	6.76	14.74
Citations	The logarithm of the mean number of Google Scholar citations received per year since the study was published (collected in May 2016).	0.54	0.33	0.62
Top journal	= 1 if the study was published in one of the top five journals in economics.	0.08	0.26	0.12
Impact	The recursive discounted RePEc impact factor of the outlet (collected in May 2016).	0.88	0.80	1.00

Notes: The variables are collected from published studies estimating the habit formation parameter. The following journals are considered top journals in economics: American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, and Review of Economic Studies. WM = mean weighted by the inverse of the number of estimates reported in a study.

3.2 Estimation and Results

Estimates of the habit parameter may vary both because of variation in the underlying degree of habit formation for different data sets (e.g., due to cultural differences across countries) and because of differences in estimation methods (e.g., due to differences in how the study approximates the Euler equation). In the previous subsection we pointed out 31 factors that in our view can contribute to explaining the heterogeneity among the estimates. A number of studies in our sample already explore the effects of some of these elements by conducting a series of experiments with data sets and methodologies. Researcher wishing to study variation across data sets can estimate the habit parameter on different data and compare results (e.g., Carroll *et al.* 2011 who make cross-country comparisons, or Ferson & Constantinides 1991 who

compare results for data of different frequencies and definitions of consumption). To examine the consequences of using certain methodology, studies can compare results obtained by applying different methods to the same data (e.g., Gueron-Quintana 2010 who studies the effects of varying the set of observables in DSGE, or Korniotis 2010 who uses a log-linear approximation of the Euler equation that incorporates both internal and external habits to compare the two specifications; he also adds a term capturing consumption risk to the log-linear specification to check whether accounting for the precautionary saving motive alters habit estimates).

While the methodology outlined above could potentially shed light on some of the sources of heterogeneity, it also has major disadvantages. First, with this strategy it is impossible to address all 31 aspects of study design within the same framework. Habit parameter estimates can be obtained using data and methods that differ along many dimensions, some of which impact the estimates' distribution (e.g., see Figure 2, in which we compare the distribution of micro and macro estimates). This means that we would not be able to draw meaningful quantitative comparisons of the associated effects, unless we explicitly assumed that some of the 31 factors could be excluded from consideration without loss of generality. Second, this method would not address the variation *observed* in the literature, as doing so requires factoring in both quantitative effects associated with each aspect of study design as well as data describing the literature itself.

For example, studies that apply DSGE methodology seem to come up with estimates that are larger than the average. This may be because matching dynamics of the observables in DSGE models requires a degree of habit formation that is higher than that estimated from individual Euler equations—an observation that would make the two methods seem inconsistent with each other. An alternative explanation, however, would say that this is because studies that use macro data suffer from aggregation bias, and that includes studies estimating DSGE models. We cannot construct a DSGE study that would use micro data and account for households' taste shifters. We could potentially compare estimates obtained on the same macro data set by using DSGE and non-DSGE approaches, but it is not straightforward which DSGE specification should be used (e.g., open or closed economy), which observables should be matched, etc. Furthermore, it is not clear whether quantitative comparisons drawn from such exercise can be used to infer something about similar differences for different data.

In this paper we would like to make quantitative comparisons of the effects of all 31 explanatory variables on the estimates of the habit parameter, which is why instead of taking the approach outlined above we perform a meta-analysis. Rather than evaluating the degree of habit formation from consumption data while trying to fit all the different approaches and methodologies into a unified framework—a task that we deem impossible to accomplish—we focus on the *estimates* that were previously obtained within the literature and investigate their variation. We consider the following regression:

$$\hat{\gamma}_{ij} = \alpha_0 + \sum_{k=1}^{31} \theta_k Z_{k,ij} + \varepsilon_{ij}, \quad (5)$$

where $\hat{\gamma}_{ij}$ is an i -th estimate from a j -th study, and $Z_{k,ij}$ is a corresponding value of the k -th explanatory variable (introduced in the previous subsection). Model (5) is meant to approximate the process generating estimates of the habit parameter. Estimating (5) would not allow us to comment on the sources of variation in the habit parameter itself, but it would capture some of the variation in the habit parameter *estimates*, and allow for meaningful quantitative comparisons of the effects of choosing different study designs.

We believe that each variable in our set can contribute to explaining the heterogeneity among the estimates. But including all 31 variables in the regression would inflate the standard errors and yield inefficient estimates, because some of the variables are likely to prove redundant. The theory does not give us enough guidance to determine the exact subset of the 31 variables that should be included in the final regression. Sequential t -testing (sometimes called the “general-to-specific approach”), which is often used to decide which variables belong to the underlying model, is not statistically valid and gives rise to the possibility of excluding relevant variables. The large number of potential variables thus brings about problems related to model uncertainty that could result in severely erroneous inference. To address these issues, we employ the Bayesian model averaging technique (BMA)—a method that does not require selecting one individual specification.

Inference in BMA is based on a weighted average of individual regressions that include different combinations of explanatory variables; the weights reflect the posterior model probabilities (PMPs) of the corresponding individual specifications. PMPs can be thought of as a Bayesian analogy of information criteria used in frequentist econometrics (at least under certain

assumptions, such as that model shocks are Gaussian). Researchers typically want to check the robustness of their results by estimating several regressions that include different combinations of explanatory variables; BMA generalizes this approach. Our intention here is to explain the basics of the BMA method and the terms needed for inference, not to give an exhaustive introduction to the BMA procedure; readers interested in such information should consult Koop (2003) for an introduction and Moral-Benito (2015) for a survey of BMA applications in economics. BMA have been used in meta-analysis, among others, by Havranek & Irsova (2017) and Irsova & Havranek (2013).

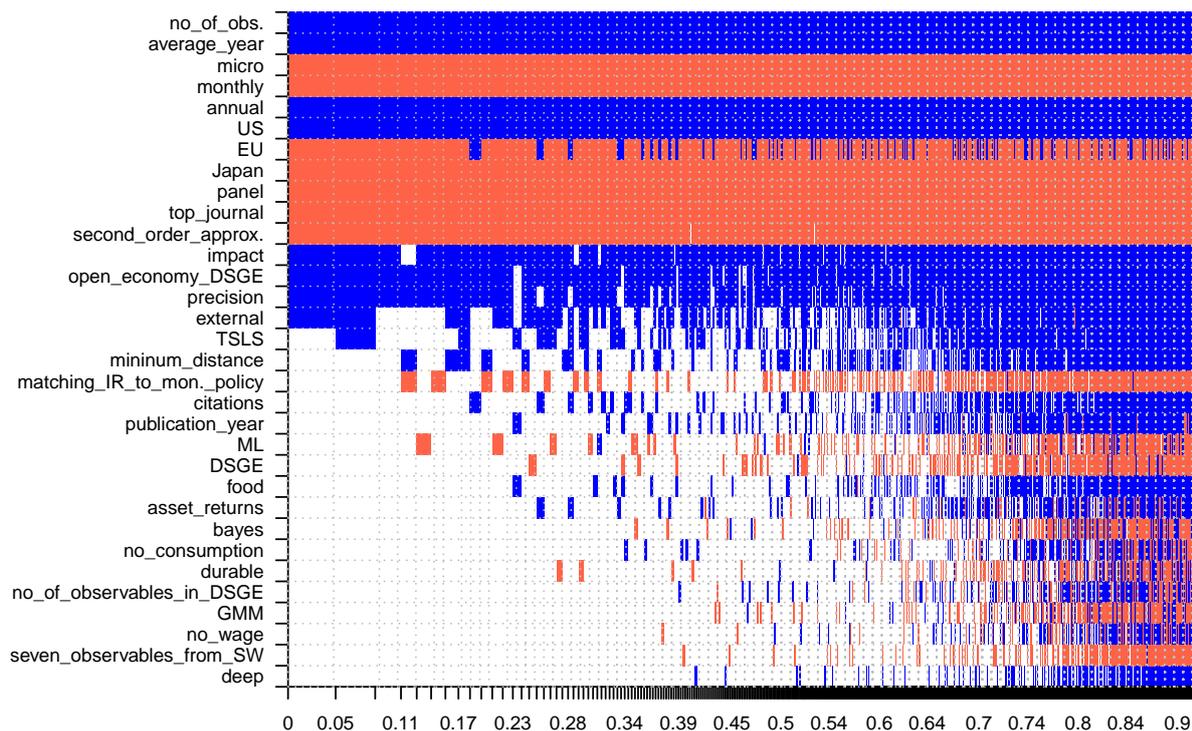
Estimating regression (5) means treating the estimates of habit as if they were observed data points. Nevertheless, each estimate is specific to the data set used in the estimation process and has a degree of uncertainty attached to it. Because of this feature our application of BMA departs from the standard approach: we explore uncertainty over which of the 31 elements should belong in auxiliary regression (5) describing the estimates, while leaving out uncertainty related to the structural models (e.g., the specification of log-linear Euler equation 4) that our primary studies choose to estimate (see Appendix B for further discussion).

We partially address some of the problems arising from treating estimates as data with the following strategy. First, we fix a subset of eight variables pertaining to data characteristics and geographical coverage, so that all eight variables appear in every regression estimated in the BMA exercise. In other words, we condition the estimates on the use of data of similar extent, age, aggregation, frequency, and regional coverage—any regression that fails to control for these factors is likely to suffer from omitted variable bias. Second, following the literature on estimated dependent variable models, we weight each observation by the precision of the estimates $[1/SE(\hat{\gamma}_{ij})]$, effectively giving more weight to estimates that are more precise.

All of the computations are performed using the R package `BMS`. Estimating all 2^{31-8} possible specifications is computationally too demanding—therefore, we approximate the whole model space by using the Model Composition Markov Chain Monte Carlo algorithm (Madigan & York, 1995), which only traverses the most important part of the model space: that is, the models with high posterior model probabilities. Such a simplification is commonly applied in applications of BMA (see, for example, Feldkircher & Zeugner, 2009). For the BMA estimation we also have to choose priors for the parameters and model space. We follow Eicher *et al.* (2011), who

recommend using the unit information prior for the parameters and the uniform model prior for the model space because these priors perform well in predictive exercises. Our prior setting can be interpreted as follows: the unit information prior provides the same amount of information as one observation of data, while the uniform model prior means that each model has the same prior probability (thereby giving higher prior probabilities to medium model *sizes*).

Figure 3: Model inclusion in Bayesian model averaging



Notes: Response variable: the estimate of the habit formation parameter. Columns denote individual models; variables are sorted by posterior inclusion probability in descending order. Blue color (darker in greyscale) = the variable is included and the estimated sign is positive. Red color (lighter in greyscale) = the variable is included and the estimated sign is negative. No color = the variable is not included in the model. The horizontal axis measures the cumulative posterior model probabilities; only the 5,000 models with the highest posterior probabilities are shown. Numerical results of the BMA exercise are reported in Table 4. A detailed description of all variables is available in Table 3.

Figure 3 presents the results of the BMA exercise. The variables are sorted from top to bottom by posterior inclusion probability (which can be thought of as a Bayesian analogy of statistical significance); the columns denote individual models. The color of the cell reflects the sign of the corresponding regression coefficient: negative signs are depicted in red (lighter in greyscale) and positive ones in blue (darker in greyscale); a white cell means that the variable is not included in the given model. The width of the columns is proportional to the posterior model probability (that is, how well the model fits the data relative to its size). Apart from the

eight variables we fix, the most important variables in explaining the heterogeneity among the estimates are *Panel*, *Top*, *Second-order approx.*, *Impact*, *Open*, and *External*. The regression signs for all of these variables are stable regardless of whether or not other control variables are included.

Table 4 presents the numerical results of Bayesian model averaging. In BMA the key statistic is the posterior inclusion probability (PIP), which reflects the importance of each variable. For a given variable the PIP is calculated by summing the posterior model probabilities of all models in which the variable is included. According to the rule of thumb proposed by Jeffreys (1961) and refined by Kass & Raftery (1995), the significance of each regressor is weak, positive, strong, or decisive if the PIP lies between 0.5–0.75, 0.75–0.95, 0.95–0.99, or 0.99–1, respectively. Additionally, we plot the posterior distribution of the estimated parameters corresponding to the first eight variables we fix, because for these variables the PIP is not informative: in the BMA exercise we force it to equal 1 by design (see Figure 4). We can see that the posterior means for the parameter estimates for all the variables except *Japan* and *EU* are more than two posterior standard deviations away from zero.

The level of data aggregation seems to be crucial for explaining the differences among estimates: micro data dramatically reduce estimates of habits (by more than 0.5), which corroborates the conclusion drawn from the histograms and summary statistics in Section 2. This resonates with the findings of Attanasio & Weber (1993) who argue that substituting national accounts data into log-linear Euler equations means incorrectly aggregating Euler equations of individual households, and that not accounting for taste shifters of individual households or household cohorts may make Euler equation estimation inconsistent (see discussion in Section 2).

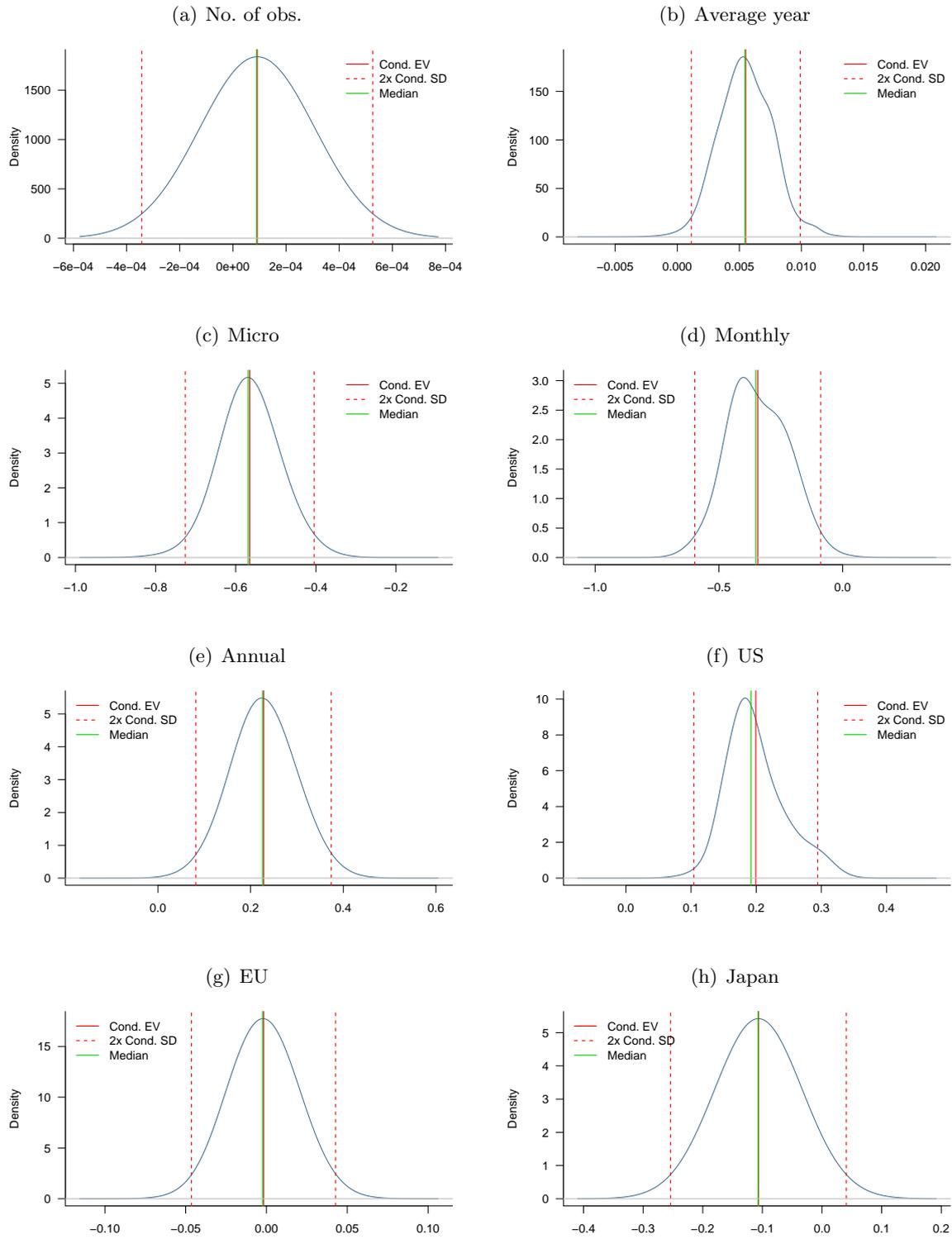
Attanasio & Weber (1993) show that specifications that do not factor in these effects fail to pass the excess sensitivity test, delivering significant correlation between changes in consumption and predictable changes in income. Our results point toward a similar problem. If changes in demographic and labor market characteristics do not even out across the population (e.g., due to population aging, or because labor market participation follows the business cycle), then past changes in consumption may partially proxy for these omitted effects, resulting in a biased estimate of the habits parameter.

Table 4: Explaining the differences in the estimates of habit formation

Response variable: Estimate of habit formation	Bayesian model averaging		
	Post. mean	Post. std. dev.	PIP
Precision	0.258	0.155	0.836
<i>Data characteristics</i>			
No. of obs.	0.000092	0.0002	1.000
Average year	0.006	0.002	1.000
Micro	-0.565	0.080	1.000
Monthly	-0.343	0.127	1.000
Annual	0.228	0.073	1.000
<i>Countries examined</i>			
US	0.199	0.048	1.000
EU	-0.002	0.022	1.000
Japan	-0.107	0.074	1.000
<i>Variable definition</i>			
External	0.093	0.101	0.538
Deep	0.001	0.007	0.041
Durable	-0.002	0.016	0.066
Food	0.016	0.053	0.120
Asset returns	0.013	0.054	0.119
<i>Estimation approach</i>			
GMM	-0.003	0.024	0.058
TSLs	0.080	0.124	0.353
Panel	-0.525	0.069	1.000
Second-order approx.	-0.385	0.118	0.982
DSGE	-0.018	0.060	0.136
Bayes	-0.004	0.025	0.076
Minimum distance	0.085	0.157	0.286
ML	-0.005	0.040	0.161
<i>DSGE specification</i>			
Open	0.226	0.110	0.893
Matching IR to mon. policy	-0.024	0.057	0.249
No. of observables	0.001	0.006	0.062
Seven observables from SW	-0.003	0.026	0.042
No consumption	0.002	0.017	0.067
No wage	0.001	0.013	0.044
<i>Publication characteristics</i>			
Publication year	0.002	0.005	0.164
Citations	0.018	0.047	0.174
Top journal	-0.571	0.119	1.000
Impact	0.125	0.061	0.923
Constant	-2.802	NA	1.000
Studies	81		
Observations	597		

Notes: PIP = posterior inclusion probability. More details on the BMA estimation are available in Table A1 and Figure A2.

Figure 4: Posterior density of parameter estimates for fixed variables



Notes: The figure depicts the densities of the regression parameters corresponding to the set of data characteristics that we include in every evaluated model. The posterior means for the parameter estimates for all the variables except *Japan* and *EU* are more than two posterior standard deviations away from zero, which can be interpreted as analogous to statistical significance at the 5% level.

Another key factor is data frequency: the higher the frequency, the lower the estimate of habit formation, with the lowest estimates corresponding to monthly data. At high frequency substitution effects in consumption get more important, as some consumer goods become durable. For example, clothing expenditure will probably show durability at monthly frequency, but not at annual frequency. This notion is in line with the findings of Eichenbaum & Hansen (1990) and Dunn & Singleton (1986), who report evidence of such substitution at monthly frequencies, and Heaton (1995) and Allais (2004), who show that adding consumption substitution at nearby dates to an asset pricing model featuring habit formation improves model fit for moments of asset returns. Alternatively, the higher estimates reported for low frequencies may result from a bias introduced by time aggregation: Heaton (1995) points out that if the decision frequency is higher than that of the data, time aggregation can induce positive autocorrelation in changes in consumption.

We find some evidence of country heterogeneity in the estimates of habit formation. The parameters estimated for the US tend to be 0.2 larger than those reported for other countries (and Japan in particular). To our knowledge, the only study that discusses cross-country differences in habit formation is Carroll *et al.* (2011), who find little heterogeneity across countries, but do not consider Japan. The cross-country differences in habit formation might reflect cultural differences—nevertheless, the specifics of the data may play a role, too. For instance, Carroll *et al.* (2011) mention several problems with Japanese data on consumption related to adjustments in the Japanese national accounts methodology.

Furthermore, we find that some estimation techniques deliver results systematically different from those obtained via other methods. The use of simple panel data techniques such as fixed effect method results in estimates that are substantially smaller. On the one hand, such methods can take into account heterogeneity between individuals beyond that captured by observed taste shifters. On the other hand, they may be prone to Nickell (1981) bias resulting from not taking into account the endogeneity created by including a lagged value of the dependent variable among the explanatory variables. Our result corroborates observations made by Naik & Moore (1996) who document that the use of fixed effects reduces estimates of the habits parameter. As noted before, studies that employ first-order approximation of the Euler equation cannot account for the precautionary saving motive, in the presence of which growth in consumption

depends positively on the degree of consumption risk, as households postpone consumption when faced with uncertainty. This feature may be important for estimating habits: if consumption uncertainty is correlated with lagged consumption growth, then first-order approximation will bias the estimate of the habit formation parameter because lagged consumption growth would partially proxy for precautionary saving. We find support for this conjecture, as the use of second-order approximation tends to reduce the estimate of habit formation by about 0.4.

The BMA exercise suggests that the specification of habit as external slightly increases the estimated parameter (by about 0.1), even though the PIP is weak according to the classification by Kass & Raftery (1995). This contradicts our observation based on Table 1 that estimates of external habits are 0.4 higher on average. Nevertheless, the contradiction can be explained by three observations. First, micro studies use internal habits about four times more often than macro studies, as shown in Table 1. This feature is likely to increase the average difference between external and internal specifications, as micro studies deliver lower estimates regardless of the method used. Second, all 100 estimates obtained from monthly data pertain to internal habits, which also plays a role, as high-frequency data deliver lower estimates. Third, 26 out of the 28 estimates obtained via second-order approximation employ internal habits, which has a similar downward effect on the average internal habits parameter.

It is well known that to replicate certain empirical facts (i.e., the response of consumption and output to a monetary policy shock) DSGE models require a high degree of habit persistence. Therefore, it is reasonable to expect the DSGE methodology to deliver higher estimates of the habit formation parameter. At the same time, DSGE studies use exclusively macro data, which are prone to aggregation bias. Furthermore, none of the DSGE studies in our sample employ data at monthly frequency. We find that the reason for the higher average DSGE estimates is most likely the fact that DSGE studies use aggregate data of low frequencies, not the DSGE methodology itself. This result is supported by histograms in Figure 2 indicating that estimates obtained within DSGE models seem to belong to the same distribution as other macro estimates. Our finding echoes that of Kano & Nason (2014), who point out the resemblance between the impulse response functions obtained within DSGE models that include consumption habits and those generated using the log-linear approximation of the Euler equation (4) on its own. At the same time, among the DSGE models, those featuring open economies tend to deliver estimates

of habit formation that are about 0.2 higher. This result corroborates the observation made by Adolfson *et al.* (2008), who compare open- and closed-economy estimates of habit formation and find that habits tend to show stronger in open-economy models. Moreover, our results suggest that when other aspects of the data are controlled for, studies scrutinizing moments of asset returns report estimates that are close to those found in the rest of the literature.

We perform a robustness check using an alternative prior setup, employing the benchmark g -priors for the parameters suggested by Fernandez *et al.* (2001) along with the beta-binomial model prior for the model space, which gives each model size equal prior probability (Ley & Steel, 2009). The results, reported in Table A2, are very similar to the baseline specification, with one notable exception: the posterior inclusion probability pertaining to *External* drops below 0.5, rendering this variable ineffective in explaining any variation among the reported estimates of habit formation.

3.3 Frequentist Model Averaging

We have stressed earlier that our dependent variable (habit parameters reported in previous studies) is estimated, which gives rise to conceptual problems for the BMA technique most commonly used in model averaging exercises. We have tried to address this issue in three ways: by including the data characteristics to all models estimated by BMA, by using precision of the estimates as weights, and by discussing the potential implications of this problem for our results (see Appendix B). An alternative approach is to employ a frequentist method of model averaging and for individual regressions utilize the standard technique of the literature on estimated dependent variable models.

The intuition of frequentist model averaging is analogous to that of BMA discussed in detail earlier: many models featuring different combinations of explanatory variables are estimated and weighted according to their goodness of fit and parsimony. The dominance of BMA in model averaging applications is given by the computational ease of Bayesian relative to frequentist methods in this field. For example, we are not aware of any previous meta-analysis that would employ frequentist model averaging. Many studies, especially in the literature on growth determinants, use combinations of Bayesian and frequentist approaches (for example, Sala-I-Martin *et al.*, 2004). The few studies that rely on purely frequentist techniques typically

Table 5: Explaining the differences in the estimates of habit formation (frequentist approach)

Response variable:	Frequentist model averaging		
	Coef.	Std. er.	p-value
Estimate of habit formation			
Precision (1/SE)	0.563	0.457	0.218
<i>Data characteristics</i>			
No. of obs.	0.000	0.000	1.000
Average year	0.000	0.006	1.000
Micro	-0.836	0.365	0.022
Monthly	-0.000	0.276	1.000
Annual	0.356	0.136	0.009
<i>Countries examined</i>			
US	0.264	0.070	0.000
EU	-0.000	0.002	1.000
Japan	-0.000	0.076	1.000
<i>Variable definition</i>			
External	0.098	0.197	0.619
Deep	0.000	0.025	1.000
Durable	-0.000	0.007	1.000
Food	0.000	0.068	1.000
Asset returns	0.000	0.226	1.000
<i>Estimation approach</i>			
GMM	-0.000	0.051	1.000
TSLS	0.000	0.286	1.000
Panel	-0.405	0.120	0.001
Second-order approx.	-0.000	0.367	1.000
DSGE	-0.000	0.025	1.000
Bayes	-0.000	0.056	1.000
Minimum distance	0.000	0.725	1.000
ML	0.000	0.263	1.000
<i>DSGE specification</i>			
Open	0.342	0.056	0.000
Matching IR to mon. policy	-0.000	0.323	1.000
No. of observables	-0.000	0.022	1.000
Seven observables from SW	0.000	0.006	1.000
No consumption	0.000	0.051	1.000
No wage	0.000	0.019	1.000
<i>Publication characteristics</i>			
Publication year	0.000	0.002	1.000
Citations	0.000	0.108	1.000
Top journal	-0.414	0.185	0.025
Impact	0.062	0.105	0.552
Constant	-3.011	0.803	0.000
Studies	81		
Observations	597		

Notes: Frequentist model averaging requires full enumeration of models, which are weighted by information criteria. We employ Mallows's criterion to select the weights since it delivers weights that are asymptotically optimal. Because our model consists of 32 potential explanatory variables, the model space is huge, 2^{32} , and full enumeration would take a prohibitive amount of time. We therefore follow the approach suggested by Amini & Parmeter (2012), who build on Magnus *et al.* (2010), and use orthogonalization of the covariate space, thus reducing the number of models that need to be estimated from 2^{32} to 32.

use information criteria as weights. Nevertheless, Hansen (2007) shows that weights selected by minimizing the Mallows criterion (an estimate of the average squared error from the model average fit) are asymptotically optimal. Another problem is how to simplify the model space: it would take us several months to estimate all the 2^{32} models, and we cannot use the Model Composition Markov Chain Monte Carlo algorithm that helped us in the case of BMA. We therefore follow the approach suggested by Amini & Parmeter (2012), who build on the pioneering insight of Magnus *et al.* (2010) and use orthogonalization of the covariate space, thus reducing the number of models that need to be estimated from 2^{32} to 32. In individual regressions we use inverse-variance weights to account for the estimated dependent variable issue.

The results of frequentist model averaging are shown in Table 5 and are broadly similar to that of BMA. It is worth noting at this point that the standard errors displayed in the table are approximate and probably conservative (Amini & Parmeter, 2012), since a formal asymptotic theory for Mallows model averaging is still to be developed. We can see from the table that, even using the frequentist approach, micro estimates are found to be substantially smaller than macro estimates on average (the difference is even larger than what BMA suggests). Next, the frequency of the data matters, as studies with annual data tend to find substantially more evidence for consumption habit. Habit formation is stronger for the US than for other countries, which is also consistent with the BMA evidence. Once again we find no significant difference between the estimates of internal and external habit once other aspects of data and methodology are controlled for. Simple panel data techniques bring systematically smaller estimates of consumption habit, which might be caused by the Nickell (1981) bias. Open-economy DSGE models are associated with larger habit estimates, and the top journals in economics tend to report, *ceteris paribus*, weaker evidence for habits compared to other outlets.

4 Concluding Remarks

In this paper we collect and examine estimates of the habit formation parameter previously reported in the literature. We document that the mean value of the parameter is 0.4 overall, but that it differs greatly between micro studies (0.1) and macro studies (0.6). None of these values is large enough to explain some of the best-known empirical puzzles in macroeconomics and finance: for example, Constantinides (1990) shows that to account for the equity premium puzzle

the habit formation parameter must exceed 0.8, while Fuhrer (2000) reproduces a humped-shaped response of consumption to various shocks with values of habit formation in the range 0.8–0.9. We find that the mean habit formation parameter produced by studies that estimate DSGE models is close to 0.7, which seems to corroborate the notion that structural estimation requires a high degree of habit formation. Nevertheless, when we turn to a more detailed investigation and control for the context in which researchers obtain their estimates, we get alternative explanations for the large habit formation reported by DSGE studies.

We show that the specifics of the data have a crucial impact on the estimated consumption habit. The difference between the results of micro and macro studies remains large when 30 other aspects of study design are controlled for. The distinction arises because micro and macro studies focus on different sources of variation in consumption: micro studies exploit variation at the level of individual households, but often lack information on consumption patterns over longer time horizons (and typically only use a fraction of consumption, such as food expenditures). By contrast, macro studies make use of consumption variation over time, while neglecting demographic characteristics and taste shifters. Our results also suggest that the frequency of the data matters—estimates obtained employing monthly frequency tend to be substantially smaller than when quarterly and annual frequencies are used. This finding may be due to the fact that at higher frequencies more consumption goods are likely to display durability, or may arise because of the time aggregation problem widely recognized in the asset pricing literature (e.g., Heaton, 1995).

We also find evidence indicating the importance of the order of approximation of the Euler equation: the use of second-order approximation tends to reduce the estimate of consumption habit. This result may signify that the precautionary saving motive plays an important role in the behavior of consumers. By contrast, we find that the use of the DSGE methodology *per se* (when other aspects of study design are controlled for) does not necessarily yield higher estimates of habits. Thus, a part of the explanation of why many estimated structural models require high degrees of habit formation may lie in their use of aggregate and low-frequency data. Additionally, because such studies typically rely on log-linearized specifications, they might be subject to the omitted variable bias, as high estimates of habits may partially capture the precautionary saving motive we have mentioned. Similarly, we find that, everything else

being held equal, studies focusing on moments of asset returns deliver habit parameters that are roughly the same as those reported by other studies. We also show that estimates reported in DSGE studies are affected by model specification: in line with Adolfson *et al.* (2008), our results indicate that open-economy models tend to report higher estimates of habit formation than closed-economy models. Finally, unlike Carroll *et al.* (2011), we find cross-country heterogeneity in habit formation, with the US displaying stronger habit formation than other countries.

References

- ABEL, A. B. (1990): “Asset Prices under Habit Formation and Catching Up with the Joneses.” *American Economic Review* **80**(2): pp. 38–42.
- ABEL, A. B. (1999): “Risk premia and term premia in general equilibrium.” *Journal of Monetary Economics* **43**(1): pp. 3–33.
- ADOLFSON, M., S. LASÉEN, J. LINDÉ, & M. VILLANI (2008): “Empirical Properties Of Closed- And Open-Economy Dsge Models Of The Euro Area.” *Macroeconomic Dynamics* **12**(S1): pp. 2–19.
- ALESSIE, R. & F. TEPPA (2010): “Saving and habit formation: evidence from Dutch panel data.” *Empirical Economics* **38**(2): pp. 385–407.
- ALLAIS, O. (2004): “Local Substitution and Habit Persistence: Matching the Moments of the Equity Premium and the Risk-Free Rate.” *Review of Economic Dynamics* **7**(2): pp. 265–296.
- AMINI, S. M. & C. F. PARMETER (2012): “Comparison Of Model Averaging Techniques: Assessing Growth Determinants.” *Journal of Applied Econometrics* **27**(5): pp. 870–876.
- ANDRÉS, J., J. DAVID LÓPEZ-SALIDO, & E. NELSON (2009): “Money and the natural rate of interest: Structural estimates for the United States and the euro area.” *Journal of Economic Dynamics and Control* **33**(3): pp. 758–776.
- 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.
- ATTANASIO, O. P. & G. WEBER (1993): “Consumption Growth, the Interest Rate and Aggregation.” *Review of Economic Studies* **60**(3): pp. 631–49.
- ATTANASIO, O. P. & G. WEBER (1995): “Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey.” *Journal of Political Economy* **103**(6): pp. 1121–57.
- BANSAL, R., D. KIKU, & A. YARON (2012): “Risks For the Long Run: Estimation with Time Aggregation.” *NBER Working Papers* 18305.
- BJORNLAND, H., K. LEITEMO, & J. MAIH (2011): “Estimating the natural rates in a simple New Keynesian framework.” *Empirical Economics* **40**(3): pp. 755–777.
- BOLDRIN, M., L. J. CHRISTIANO, & J. D. M. FISHER (2001): “Habit Persistence, Asset Returns, and the Business Cycle.” *American Economic Review* **91**(1): pp. 149–166.
- BOUAKEZ, H., E. CARDIA, & F. J. RUGE-MURCIA (2005): “Habit formation and the persistence of monetary shocks.” *Journal of Monetary Economics* **52**(6): pp. 1073–1088.
- BRODEUR, A., M. LE, M. SANGNIER, & Y. ZYLBERBERG (2016): “Star Wars: The Empirics Strike Back.” *American Economic Journal: Applied Economics* **8**(1): pp. 1–32.
- CAMPBELL, J. Y. & J. COCHRANE (1999): “Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior.” *Journal of Political Economy* **107**(2): pp. 205–251.
- CARD, D. & A. B. KRUEGER (1995): “Time-Series Minimum-Wage Studies: A Meta-analysis.” *American Economic Review* **85**(2): pp. 238–43.
- CARROLL, C. D. (2001): “Death to the Log-Linearized Consumption Euler Equation! (And Very Poor Health to the Second-Order Approximation).” *Advances in Macroeconomics* **1**(1): pp. 1–38.
- CARROLL, C. D., J. OVERLAND, & D. N. WEIL (1997): “Comparison Utility in a Growth Model.” *Journal of Economic Growth* **2**(4): pp. 339–67.
- CARROLL, C. D., J. SLACALEK, & M. SOMMER (2011): “International Evidence on Sticky Consumption Growth.” *The Review of Economics and Statistics* **93**(4): pp. 1135–1145.
- CHETTY, R., A. GUREN, D. MANOLI, & A. WEBER (2011): “Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins.” *American Economic Review* **101**(3): pp. 471–75.

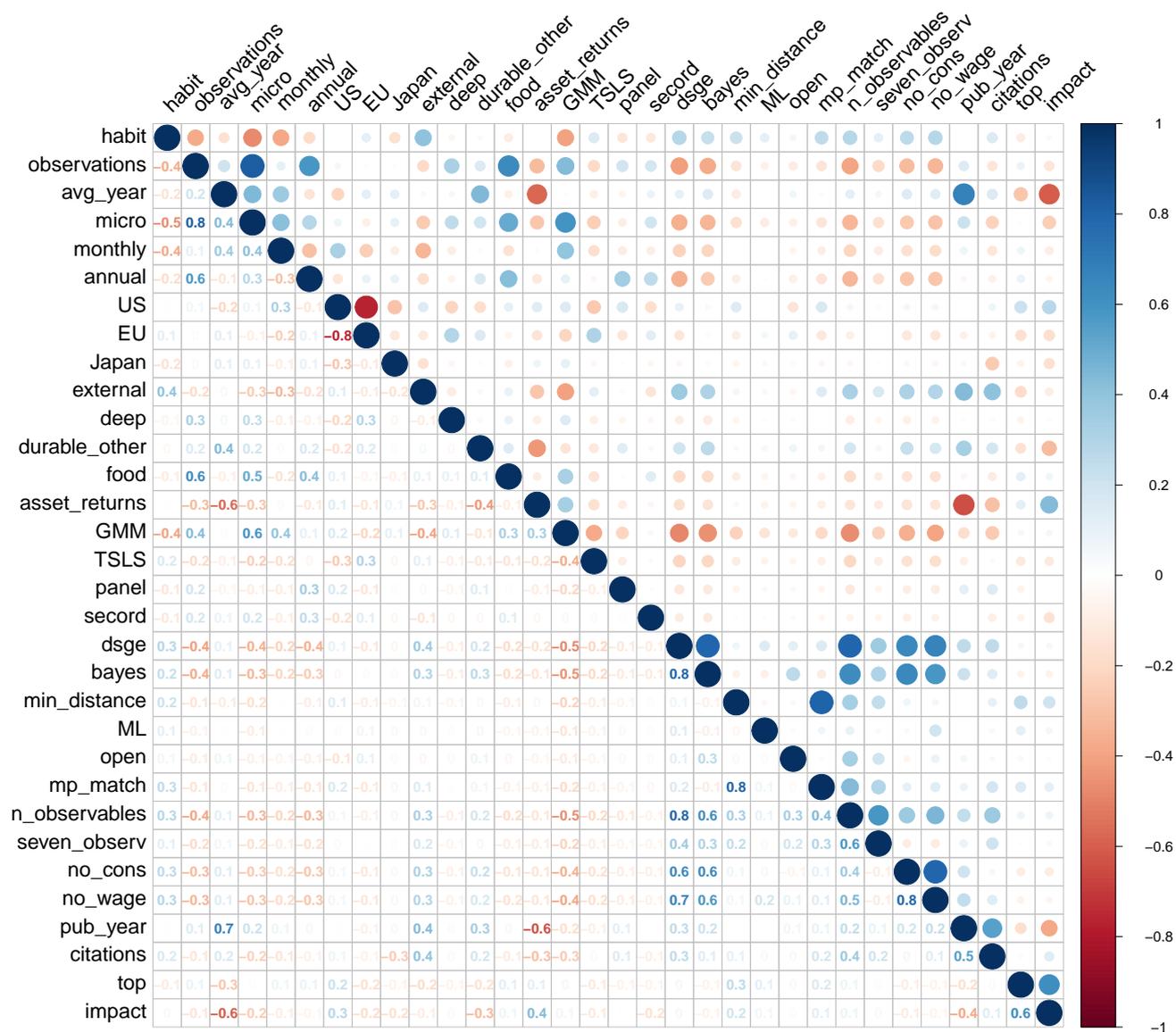
- CHRISTIANO, L. J., M. EICHENBAUM, & C. L. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* **113**(1): pp. 1–45.
- COLLADO, M. D. & M. BROWNING (2007): "Habits and heterogeneity in demands: a panel data analysis." *Journal of Applied Econometrics* **22**(3): pp. 625–640.
- CONSTANTINIDES, G. M. (1990): "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy* **98**(3): pp. 519–43.
- DEL NEGRO, M., F. SCHORFHEIDE, F. SMETS, & R. WOUTERS (2007): "On the Fit of New Keynesian Models." *Journal of Business & Economic Statistics* **25**: pp. 123–143.
- DENNIS, R. (2009): "Consumption Habits in a New Keynesian Business Cycle Model." *Journal of Money, Credit and Banking* **41**(5): pp. 1015–1030.
- DISDIER, A.-C. & K. HEAD (2008): "The Puzzling Persistence of the Distance Effect on Bilateral Trade." *The Review of Economics and Statistics* **90**(1): pp. 37–48.
- DUNN, K. B. & K. J. SINGLETON (1986): "Modeling the term structure of interest rates under non-separable utility and durability of goods." *Journal of Financial Economics* **17**(1): pp. 27–55.
- DYNAN, K. E. (2000): "Habit Formation in Consumer Preferences: Evidence from Panel Data." *American Economic Review* **90**(3): pp. 391–406.
- EGGER, M., G. SMITH, M. SCHNEIDER, & C. MINDER (1997): "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal* **316**(7129): pp. 469–471.
- EICHENBAUM, M. & L. P. HANSEN (1990): "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data." *Journal of Business & Economic Statistics* **8**(1): pp. 53–69.
- 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.
- EVERAERT, G. & L. POZZI (2014): "The Predictability Of Aggregate Consumption Growth In OECD Countries: A Panel Data Analysis." *Journal of Applied Econometrics* **29**(3): pp. 431–453.
- FELDKIRCHER, M. & S. ZEUGNER (2009): "Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging." *IMF Working Papers 09/202*.
- FERNANDEZ, C., E. LEY, & M. F. J. STEEL (2001): "Benchmark priors for Bayesian model averaging." *Journal of Econometrics* **100**(2): pp. 381–427.
- FERSON, W. E. & G. M. CONSTANTINIDES (1991): "Habit persistence and durability in aggregate consumption: Empirical tests." *Journal of Financial Economics* **29**(2): pp. 199–240.
- FUHRER, J. C. (2000): "Habit Formation in Consumption and Its Implications for Monetary-Policy Models." *American Economic Review* **90**(3): pp. 367–390.
- GRUBER, J. W. (2004): "A present value test of habits and the current account." *Journal of Monetary Economics* **51**(7): pp. 1495–1507.
- GUARIGLIA, A. (2002): "Consumption, habit formation, and precautionary saving: evidence from the British Household Panel Survey." *Oxford Economic Papers* **54**(1): pp. 1–19.
- GUERRON-QUINTANA, P. A. (2010): "What you match does matter: the effects of data on DSGE estimation." *Journal of Applied Econometrics* **25**(5): pp. 774–804.
- HANSEN, B. E. (2007): "Least Squares Model Averaging." *Econometrica* **75**(4): pp. 1175–1189.
- HAVRANEK, T. (2015): "Measuring Intertemporal Substitution: The Importance Of Method Choices And Selective Reporting." *Journal of the European Economic Association* **13**(6): pp. 1180–1204.
- HAVRANEK, T., R. HORVATH, Z. IRSOVA, & M. RUSNAK (2015a): "Cross-Country Heterogeneity in Intertemporal Substitution." *Journal of International Economics* **96**(1): pp. 100–118.
- HAVRANEK, T. & Z. IRSOVA (2011): "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics* **85**(2): pp. 234–244.
- HAVRANEK, T. & Z. IRSOVA (2012): "Survey Article: Publication Bias in the Literature on Foreign Direct Investment Spillovers." *Journal of Development Studies* **48**(10): pp. 1375–1396.
- HAVRANEK, T. & Z. IRSOVA (2017): "Do Borders Really Slash Trade? A Meta-Analysis." *IMF Economic Review* (forthcoming).
- HAVRANEK, T., Z. IRSOVA, & K. JANDA (2012): "Demand for gasoline is more price-inelastic than commonly thought." *Energy Economics* **34**(1): pp. 201–207.
- HAVRANEK, T., Z. IRSOVA, K. JANDA, & D. ZILBERMAN (2015b): "Selective reporting and the social cost of carbon." *Energy Economics* **51**(C): pp. 394–406.
- HAVRANEK, T. & A. SOKOLOVA (2016): "Do Consumers Really Follow a Rule of Thumb? Three Thousand Estimates from 130 Studies Say 'Probably Not'." *Working paper 8/2016*, Czech National Bank.
- HE, Y., J. S. DYER, & J. C. BUTLER (2013): "On the axiomatization of the satiation and habit formation utility models." *Operations Research* **61**(6): pp. 1399–1410.
- HEATON, J. (1995): "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference

- Specifications.” *Econometrica* **63(3)**: pp. 681–717.
- IOANNIDIS, J. P. A., T. D. STANLEY, & H. DOUCOULIAGOS (2017): “The Power of Bias in Economics Research.” *Economic Journal* (**forthcoming**).
- IRSOVA, Z. & T. HAVRANEK (2013): “Determinants of Horizontal Spillovers from FDI: Evidence from a Large Meta-Analysis.” *World Development* **42(C)**: pp. 1–15.
- JEFFREYS, H. (1961): *Theory of Probability*. Oxford: Clarendon Press.
- KANO, T. (2009): “Habit formation and the present-value model of the current account: Yet another suspect.” *Journal of International Economics* **78(1)**: pp. 72–85.
- KANO, T. & J. M. NASON (2014): “Business Cycle Implications of Internal Consumption Habit for New Keynesian Models.” *Journal of Money, Credit and Banking* **46(2-3)**: pp. 519–544.
- KASS, R. & A. RAFTERY (1995): “Bayes Factors.” *Journal of the American Statistical Association* **90(430)**: pp. 773–795.
- KOOP, G. (2003): *Bayesian Econometrics*. John Wiley & Sons.
- KOOPMANS, T. C. (1960): “Stationary ordinal utility and impatience.” *Econometrica* **28(2)**: pp. 287–309.
- KORNIOTIS, G. M. (2010): “Estimating Panel Models With Internal and External Habit Formation.” *Journal of Business & Economic Statistics* **28(1)**: pp. 145–158.
- LEVIN, A. T., A. ONATSKI, J. WILLIAMS, & N. M. WILLIAMS (2005): “Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models.” In “NBER Macroeconomics Annual 2005, Volume 20,” NBER Chapters, pp. 229–312.
- 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.
- LUBIK, T. A. & W. L. TEO (2014): “Deep Habits in the New Keynesian Phillips Curve.” *Journal of Money, Credit and Banking* **46(1)**: pp. 79–114.
- LUCAS, Robert E, J. (1978): “Asset Prices in an Exchange Economy.” *Econometrica* **46(6)**: pp. 1429–45.
- MADIGAN, D. & J. YORK (1995): “Bayesian graphical models for discrete data.” *International Statistical Review* **63(2)**: pp. 215–232.
- MAGNUS, J. R., O. POWELL, & P. PRUFER (2010): “A comparison of two model averaging techniques with an application to growth empirics.” *Journal of Econometrics* **154(2)**: pp. 139–153.
- MEHRA, R. & E. C. PRESCOTT (1985): “The equity premium: A puzzle.” *Journal of Monetary Economics* **15(2)**: pp. 145–161.
- MORAL-BENITO, E. (2015): “Model Averaging In Economics: An Overview.” *Journal of Economic Surveys* **29(1)**: pp. 46–75.
- NAIK, N. Y. & M. J. MOORE (1996): “Habit Formation and Intertemporal Substitution in Individual Food Consumption.” *The Review of Economics and Statistics* **78(2)**: pp. 321–28.
- NECKER, S. (2014): “Scientific misbehavior in economics.” *Research Policy* **43(10)**: p. 1747–1759.
- NICKELL, S. J. (1981): “Biases in Dynamic Models with Fixed Effects.” *Econometrica* **49(6)**: pp. 1417–26.
- RAFTERY, A., D. MADIGAN, & J. HOETING (1997): “Bayesian Model Averaging for Linear Regression Models.” *Journal of the American Statistical Association* **92(437)**: pp. 179–191.
- RAVN, M., S. SCHMITT-GROHÉ, & M. URIBE (2006): “Deep Habits.” *Review of Economic Studies* **73(1)**: pp. 195–218.
- ROZEN, K. (2010): “Foundations of Intrinsic Habit Formation.” *Econometrica* **78(4)**: pp. 1341–1373.
- RUSTICHINI, A. & P. SICONOLFI (2014): “Dynamic theory of preferences: Habit formation and taste for variety.” *Journal of Mathematical Economics* **55(C)**: pp. 55–68.
- SALA-I-MARTIN, X., G. DOPPELHOFER, & R. I. MILLER (2004): “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach.” *American Economic Review* **94(4)**: pp. 813–835.
- SMETS, F. & R. WOUTERS (2007): “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *American Economic Review* **97(3)**: pp. 586–606.
- SOMMER, M. (2007): “Habit Formation and Aggregate Consumption Dynamics.” *The B.E. Journal of Macroeconomics* **7(1)**: pp. 1–25.
- STANLEY, T. D. (2001): “Wheat from Chaff: Meta-analysis as Quantitative Literature Review.” *Journal of Economic Perspectives* **15(3)**: pp. 131–150.
- WIELAND, V., T. CCIK, G. J. MÜLLER, S. SCHMIDT, & M. WOLTERS (2012): “A new comparative approach to macroeconomic modeling and policy analysis.” *Journal of Economic Behavior & Organization* **83(3)**: pp. 523–541.

Appendix A: Supplementary Statistics and Analysis

A.1 Correlation of the Variables

Figure A1: Correlation matrix



Notes: A description of the variables is available in Table 3.

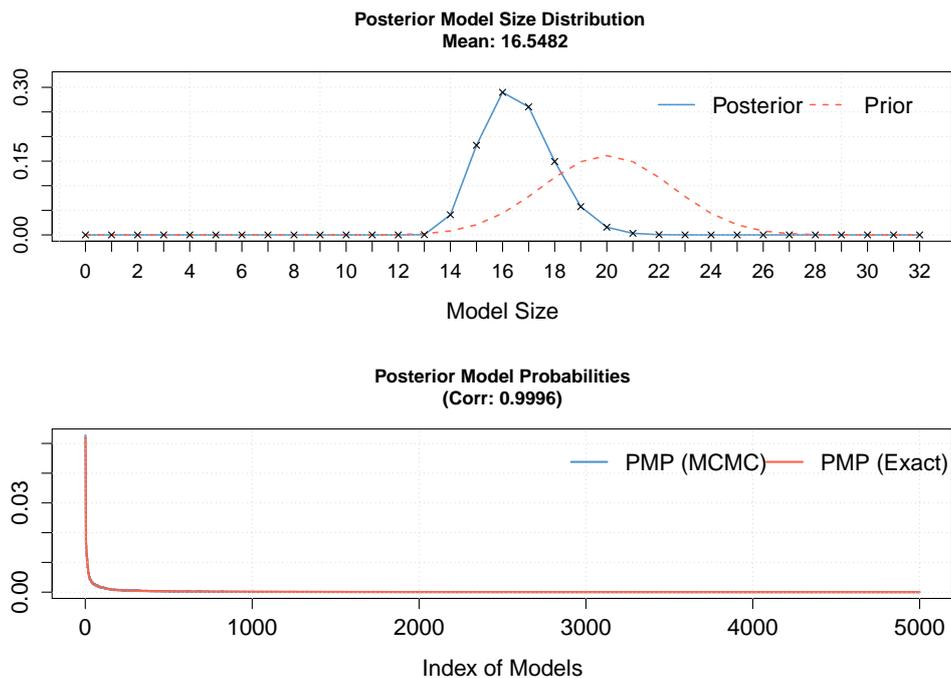
A.2 Diagnostics of BMA

Table A1: Summary of BMA estimation

<i>Mean no. regressors</i>	<i>Draws</i>	<i>Burn-ins</i>	<i>Time</i>
16.5482	$3e + 06$	$1e + 06$	10.33624 minutes
<i>No. models visited</i>	<i>Modelspace</i>	<i>Visited</i>	<i>Topmodels</i>
749,088	$4.3e + 09$	0.017%	92%
<i>Corr PMP</i>	<i>No. Obs.</i>	<i>Model Prior</i>	<i>g-Prior</i>
0.9996	597	uniform	UIP
<i>Shrinkage-Stats</i>			
Av= 0.9983			

Notes: In this specification we employ the priors suggested by Eicher *et al.* (2011) based on the predictive performance: the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation of the data).

Figure A2: Model size and convergence



A.3 Alternative Priors for BMA

Table A2: Explaining the differences in the estimates of habit formation (alternative priors)

Response variable: Estimate of habit formation	Bayesian model averaging		
	Post. mean	Post. std. dev.	PIP
Precision (1/SE)	0.272	0.144	0.878
<i>Data characteristics</i>			
No. of obs.	0.000093	0.0002	1.000
Average year	0.006	0.002	1.000
Micro	-0.572	0.072	1.000
Monthly	-0.357	0.119	1.000
Annual	0.226	0.067	1.000
<i>Countries examined</i>			
US	0.201	0.047	1.000
EU	-0.002	0.021	1.000
Japan	-0.112	0.069	1.000
<i>Variable definition</i>			
External	0.073	0.093	0.437
Deep	0.000	0.004	0.020
Durable	-0.001	0.011	0.035
Food	0.012	0.047	0.079
Asset returns	0.006	0.039	0.062
<i>Estimation approach</i>			
GMM	-0.002	0.017	0.029
TSLS	0.055	0.109	0.238
Panel	-0.528	0.064	1.000
Second-order approx.	-0.392	0.112	0.983
DSGE	-0.011	0.045	0.086
Bayes	-0.002	0.017	0.038
Minimum distance	0.058	0.134	0.192
ML	-0.005	0.032	0.123
<i>DSGE specification</i>			
Open	0.216	0.100	0.907
Matching IR to mon. policy	-0.021	0.050	0.210
No. of observables	0.000	0.004	0.033
Seven observables from SW	-0.002	0.018	0.021
No consumption	0.001	0.012	0.037
No wage	0.000	0.009	0.021
<i>Publication characteristics</i>			
Publication year	0.002	0.005	0.124
Citations	0.012	0.037	0.111
Top journal	-0.563	0.117	1.000
Impact	0.119	0.062	0.901
Constant	-2.753	NA	1.000
Studies	81		
Observations	597		

Notes: PIP = posterior inclusion probability. We use an alternative to the unit information prior, the BRIC prior suggested by Fernandez *et al.* (2001), which takes into account the number of explanatory variables for the determination of the weight of the zero prior for the regression parameters. In this set of priors we also employ the random beta-binomial model prior (Ley & Steel, 2009), which implies that each *model size* has the same prior probability.

A.4 Publication Bias

The mean and median reported estimates may represent a biased reflection of the underlying research results if some estimates are more likely than others to be selected for publication. For this reason, most meta-analyses test—and, if necessary, correct—for so-called publication bias. Brodeur *et al.* (2016) collect 50,000 p-values reported in economics and document widespread publication bias. A recent survey among the members of the European Economic Association, Necker (2014), reveals that a third of economists in Europe admit that they have engaged in presenting empirical findings selectively so they confirm their arguments and in searching for control variables until they get a desired result. Ioannidis *et al.* (2017) survey meta-analyses conducted in economics and find that most fields suffer from the bias, as editors, referees, or authors themselves prefer statistically significant results that have an intuitive sign.

Frequently discussed cases of publication bias include the literature on productivity spillovers from FDI (Havranek & Irsova, 2012), price elasticity of gasoline demand (Havranek *et al.*, 2012), and social cost of carbon (Havranek *et al.*, 2015b). Havranek (2015) finds strong publication bias in the literature that uses consumption Euler equations to estimate the elasticity of intertemporal substitution (often the same specification used to estimate habit formation).⁵ Most economists believe that the elasticity of substitution should be positive because negative elasticity implies a convex utility function. Therefore, negative estimates of the elasticity are rarely reported in the literature, as are statistically insignificant estimates. The under-reporting of negative estimates and estimates that are positive but small and imprecise biases the means upward because it is not matched by corresponding under-reporting of large imprecise estimates.

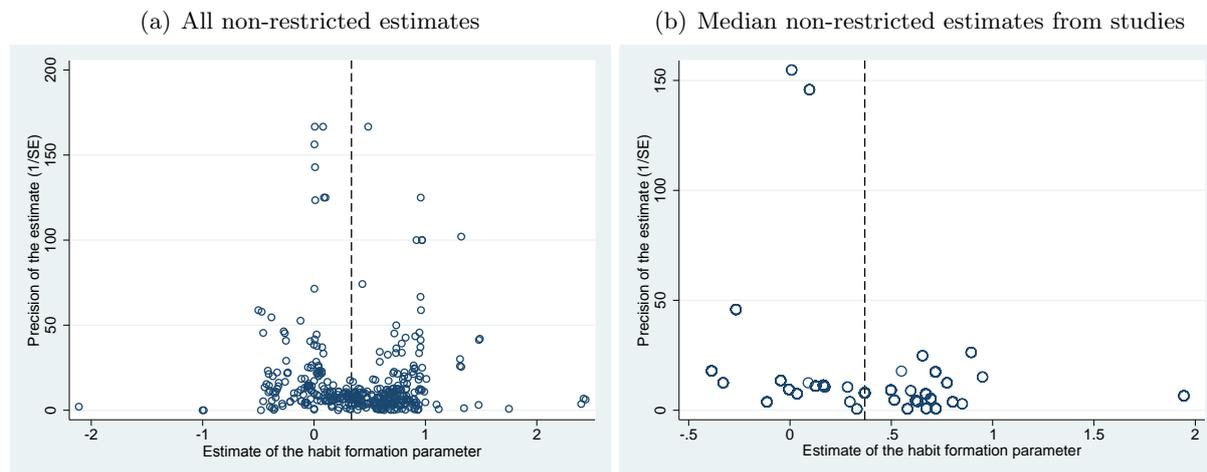
The empirical literature on habit formation differs from studies estimating the elasticity of intertemporal substitution in two major aspects. First, negative estimates of habit formation allow for intuitive interpretation: they may result from durability of the consumption measure used in the estimation—and may thus be more publishable than negative estimates of the elasticity of intertemporal substitution. Second, unlike large estimates of the elasticity, estimates of the habit formation parameter that exceed 1 are implausible because they imply non-stationary consumption growth. Figure 2, discussed in Section 2, suggests that the most common estimates lie close to the midpoint between the lower and upper boundaries of the 0–1 interval (consistent with habit formation), and that when an estimate surpasses either limit, its probability of being reported drops—in other words, both very small and very large estimates can sometimes be discarded by the researchers. This relative symmetry in decision rules on discarding implausible estimates implies that even if there is publication selection in the literature on habit formation, it does not necessarily lead to publication bias.

To test for potential publication bias researchers often evaluate so-called funnel plots (Egger *et al.*, 1997). A funnel plot is a scatter plot of the estimates (on the horizontal axis) against the inverse of their standard errors, the estimates’ precision (on the vertical axis). In the absence of publication bias the scatter plot forms an inverted funnel: the most precise estimates lie

⁵In a similar vein, Havranek & Sokolova (2016) identify publication bias in the literature on the excess sensitivity of consumption growth to anticipated income changes.

close to each other, while the less precise ones are more dispersed. The funnel plot should be symmetric because a feature of most estimation methods is that the ratio of the estimate to its standard error exhibits a symmetric distribution. Therefore all imprecise estimates, small and large, should have the same probability of being reported.

Figure A3: Funnel plots suggest little publication bias



Notes: In the absence of publication bias the funnel should be symmetrical around the most precise estimates of the habit formation parameter. The dashed vertical lines denote the mean of all the estimates in panel (a) and the mean of the median estimates reported in the studies in panel (b). Multiple peaks of the funnel suggest heterogeneity. For ease of exposition we exclude estimates with extreme precision values from the figure, but we use all the estimates in the statistical tests.

Figure A3 presents funnel plots for non-restricted estimates of the habit formation parameter. The left-hand panel depicts all estimates, while the right-hand panel plots median estimates reported in the studies against their precision. The plots show little signs of asymmetry (especially compared to other fields of empirical economics), but both 0 and 1 seem to be the boundaries that affect the probability of estimates being reported. In the online appendix at meta-analysis.cz/habits we test publication bias formally using several methods. While we find some indications of publication selection related to the 0 and 1 thresholds that define the range consistent with habit formation, we find little evidence of any systematic bias resulting from this selection. Our findings thus suggest that the effects of potential selection against negative estimates and potential selection against estimates larger than 1 cancel each other out, rendering the mean estimate reported in the habit formation literature unbiased.

Appendix B: BMA and Model Uncertainty in Meta-Analysis

This section discusses how our application of Bayesian model averaging departs from the standard approach employed by the literature. In Section 3 we identify 31 factors that we believe could contribute to the heterogeneity in the reported estimates $\hat{\gamma}$ of habit formation, and we would like to quantify their effects by estimating the following model (5), already featured in Subsection 3.2:

$$\hat{\gamma}_{ij} = \alpha_0 + \sum_{k=1}^{31} \theta_k Z_{k,ij} + \varepsilon_{ij}. \quad (5)$$

Among the 31 explanatory variables Z_k , eight describe the critical features of the data generating process, and some of the remaining factors may have effects that are small or insignificant. As *a priori* we do not know which of these remaining elements have a systematic effect on the estimates, we are facing a total of 2^{31-8} possible models we could use to describe the variation in the estimates. Our use of BMA aims to resolve this type of uncertainty: uncertainty over which features of study design affect the estimates of the habit formation parameter, conditional on the effects being linear and on there being only 31 possible explanatory variables.

Let $B \equiv [B_1, B_2, \dots, B_{2^{31-8}}]$ denote all possible combinations of explanatory variables Z_k that could be included in regression (5). Let θ denote a vector of the 31 effects θ_k associated with regressors Z_k . The posterior of θ can then be written as

$$p(\theta|\hat{\gamma}, B) = \sum_{m=1}^{2^{31-8}} p(\theta|\hat{\gamma}, B_m) p(B_m|\hat{\gamma}, B), \quad (6)$$

where $p(\theta|\hat{\gamma}, B_m)$ is obtained by estimating model B_m on the set $\hat{\gamma}$, and $p(B_m|\hat{\gamma}, B)$ is a posterior model probability associated with combination B_m that can be calculated via

$$p(B_m|\hat{\gamma}, B) = \frac{p(B_m|B)p(\hat{\gamma}|B_m)}{\sum_{m=1}^{2^{31-8}} p(B_m|B)p(\hat{\gamma}|B_m)}, \quad (7)$$

where $p(B_m|B)$ and $p(\hat{\gamma}|B_m)$ are prior probability of model B_m and its marginal likelihood. We follow the standard BMA approach and use (7) to identify posterior model probabilities. But in doing so we treat $\hat{\gamma}$ as data points, not estimates—this treatment ignores a portion of uncertainty attached to the choices of data and methodology made by the authors of the primary studies.

In Section 3 we state that the literature studying consumption habits is very diverse, employing different data sets and methods. To be more precise, all studies in our sample employ unique data sets, and in some cases there is even variation in data used within one study. Furthermore, the structural models that primary studies rely on vary. Even though many studies employ approximations similar to (4), the estimated specifications differ substantially: (4) can be estimated on its own, or as part of some DSGE model. In Section 3 we pinpoint key differ-

ences between modeling approaches; however, with this strategy we cannot hope to fully capture the diversity of structural models employed in the literature and the uncertainty attached to the modeling choices. Below we sketch a strategy that could, if successfully implemented, resolve these issues. We thank an anonymous referee for providing the underlying idea.

Let $A \equiv [A_1, A_2, \dots, A_n]$ denote the sequence of structural models used to obtain estimates $\hat{\gamma} \equiv [\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_n]$ where $n = 597$, and $Y \equiv [Y_1, Y_2, \dots, Y_n]$ be corresponding data sets. Let \mathcal{A} denote the set of structural models used within the literature. (The number of elements in \mathcal{A} is smaller than 597 because many studies apply same methodology to different data sets.) The probability of an auxiliary model B_m conditional on data Y and a collection of models A can then be expressed as

$$p(B_m|B, \mathcal{A}, Y) = \int p(B_m|B, \hat{\gamma})p(\hat{\gamma}|\mathcal{A}, Y)d\hat{\gamma}, \quad (8)$$

where $p(\hat{\gamma}|\mathcal{A}, Y)$ is probability of the set of habit parameter estimates conditional on estimating structural model set \mathcal{A} on data Y . To account for uncertainty with respect to structural models, we would need to further decompose this probability as follows:

$$\begin{aligned} p(\hat{\gamma}|\mathcal{A}, Y) &= p(\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_n|\mathcal{A}, Y) = \\ &= \prod_{i=1}^n [p(\hat{\gamma}_i|A_i, Y_i)p(A_i|\mathcal{A}, Y)]. \end{aligned} \quad (9)$$

where $p(A_i|\mathcal{A}, Y)$ is a probability attached to structural model A_i , conditional on model set \mathcal{A} and data Y .

A study wishing to fully account for uncertainty over which structural models should be used to evaluate Euler equations with habits would need to assess $p(A_i|\mathcal{A}, Y)$, facing a variety of obstacles, some of which may prove insurmountable. As discussed before, structural models in \mathcal{A} differ along many dimensions, which makes comparing their relative performance not straightforward. Furthermore, the data in Y have features that may affect the relative performance of each model. As we saw in Figure 2, it seems that estimates coming from macro and micro data and data of different frequencies are associated with distinct distributions. What is more, some structural models are meant to only be applied to certain types of data. For example, models that account for taste shifters are designed for micro studies, whereas DSGE models can only be estimated on aggregate data sets.

In our understanding these difficulties make complete Bayesian treatment of both sources model uncertainty infeasible. In this paper we address the model uncertainty associated with the choice of variables in the meta-analysis model. Since we do not address the other source of model uncertainty, related to structural models A , the resulting posterior standard deviations may be underestimated. For this reason, as a robustness check, we also estimate a frequentist model averaging specification, for which we can use the typical approach employed in estimated dependent variable models.

Appendix C: Studies Included in the Data Set

- ADOLFSON, M., S. LASEEN, J. LINDE, & M. VILLANI (2007): "Bayesian estimation of an open economy DSGE model with incomplete pass-through." *Journal of International Economics* **72(2)**: pp. 481–511.
- ADOLFSON, M., S. LASÉEN, J. LINDÉ, & M. VILLANI (2008): "Empirical Properties Of Closed- And Open-Economy Dsg Models Of The Euro Area." *Macroeconomic Dynamics* **12(S1)**: pp. 2–19.
- ALESSIE, R. & A. KAPTEYN (1991): "Habit Formation, Interdependent References and Demographic Effects in the Almost Ideal Demand System." *Economic Journal* **101(406)**: pp. 404–19.
- ALESSIE, R. & F. TEPPA (2010): "Saving and habit formation: evidence from Dutch panel data." *Empirical Economics* **38(2)**: pp. 385–407.
- ALLAIS, O. (2004): "Local Substitution and Habit Persistence: Matching the Moments of the Equity Premium and the Risk-Free Rate." *Review of Economic Dynamics* **7(2)**: pp. 265–296.
- ALTIG, D., L. CHRISTIANO, M. EICHENBAUM, & J. LINDE (2011): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle." *Review of Economic Dynamics* **14(2)**: pp. 225–247.
- ANDREASEN, M. M. (2012): "An estimated DSGE model: Explaining variation in nominal term premia, real term premia, and inflation risk premia." *European Economic Review* **56(8)**: pp. 1656–1674.
- ANDRÉS, J., J. DAVID LÓPEZ-SALIDO, & E. NELSON (2009): "Money and the natural rate of interest: Structural estimates for the United States and the euro area." *Journal of Economic Dynamics and Control* **33(3)**: pp. 758–776.
- AURAY, S. & P. FEVE (2008): "On Sunspots, Habits and Monetary Facts." *Macroeconomic Dynamics* **12(01)**: pp. 72–96.
- BALTAGI, B. H., G. BRESSON, J. M. GRIFFIN, & A. PIROTE (2003): "Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from French regional gasoline consumption." *Empirical Economics* **28(4)**: pp. 795–811.
- BARTOLOMEO, G. D., L. ROSSI, & M. TANCIONI (2010): "Monetary policy, rule-of-thumb consumers and external habits: a G7 comparison." *Applied Economics* **43(21)**: pp. 2721–2738.
- BATINI, N., A. JUSTINIANO, P. LEVINE, & J. PEARLMAN (2006): "Robust inflation-forecast-based rules to shield against indeterminacy." *Journal of Economic Dynamics and Control* **30(9-10)**: pp. 1491–1526.
- BEKAERT, G., S. CHO, & A. MORENO (2010): "New Keynesian Macroeconomics and the Term Structure." *Journal of Money, Credit and Banking* **42(1)**: pp. 33–62.
- BJORNLAND, H., K. LEITEMO, & J. MAIH (2011): "Estimating the natural rates in a simple New Keynesian framework." *Empirical Economics* **40(3)**: pp. 755–777.
- BOIVIN, J. & M. P. GIANNONI (2006): "Has Monetary Policy Become More Effective?" *The Review of Economics and Statistics* **88(3)**: pp. 445–462.
- BOUAKEZ, H., E. CARDIA, & F. J. RUGE-MURCIA (2005): "Habit formation and the persistence of monetary shocks." *Journal of Monetary Economics* **52(6)**: pp. 1073–1088.
- BOVER, O. (1991): "Relaxing Intertemporal Separability: A Rational Habits Model of Labor Supply Estimated from Panel Data." *Journal of Labor Economics* **9(1)**: pp. 85–100.
- BRAUN, P. A., G. M. CONSTANTINIDES, & W. E. FERSON (1993): "Time nonseparability in aggregate consumption: International evidence." *European Economic Review* **37(5)**: pp. 897–920.
- CAMPBELL, J. Y. & N. G. MANKIW (1991): "The response of consumption to income: A cross-country investigation." *European Economic Review* **35(4)**: pp. 723–756.
- CARROLL, C. D., J. SLACALEK, & M. SOMMER (2011): "International Evidence on Sticky Consumption Growth." *The Review of Economics and Statistics* **93(4)**: pp. 1135–1145.
- CASTELNUOVO, E. & S. NISTICO (2010): "Stock market conditions and monetary policy in a DSGE model for the U.S." *Journal of Economic Dynamics and Control* **34(9)**: pp. 1700–1731.
- CHIB, S. & S. RAMAMURTHY (2010): "Tailored randomized block MCMC methods with application to DSGE models." *Journal of Econometrics* **155(1)**: pp. 19–38.
- CHRISTIANO, L. J., M. EICHENBAUM, & C. L. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* **113(1)**: pp. 1–45.
- CHRISTOFFEL, K., K. KUESTER, & T. LINZERT (2009): "The role of labor markets for euro area monetary policy." *European Economic Review* **53(8)**: pp. 908–936.
- COLLADO, M. D. & M. BROWNING (2007): "Habits and heterogeneity in demands: a panel data analysis." *Journal of Applied Econometrics* **22(3)**: pp. 625–640.
- COOLEY, T. F. & M. OGAKI (1996): "A Time Series Analysis of Real Wages, Consumption, and Asset Returns." *Journal of Applied Econometrics* **11(2)**: pp. 119–34.
- DE GRAEVE, F. (2008): "The external finance premium

- and the macroeconomy: US post-WWII evidence.” *Journal of Economic Dynamics and Control* **32(11)**: pp. 3415–3440.
- DEL NEGRO, M. & F. SCHORFHEIDE (2008): “Forming priors for DSGE models (and how it affects the assessment of nominal rigidities).” *Journal of Monetary Economics* **55(7)**: pp. 1191–1208.
- DEL NEGRO, M., F. SCHORFHEIDE, F. SMETS, & R. WOUTERS (2007): “On the Fit of New Keynesian Models.” *Journal of Business & Economic Statistics* **25**: pp. 123–143.
- DENNIS, R. (2009): “Consumption Habits in a New Keynesian Business Cycle Model.” *Journal of Money, Credit and Banking* **41(5)**: pp. 1015–1030.
- DYNAN, K. E. (2000): “Habit Formation in Consumer Preferences: Evidence from Panel Data.” *American Economic Review* **90(3)**: pp. 391–406.
- EDGE, R. M., M. T. KILEY, & J.-P. LAFORTE (2008): “Natural rate measures in an estimated DSGE model of the U.S. economy.” *Journal of Economic Dynamics and Control* **32(8)**: pp. 2512–2535.
- EICHENBAUM, M. & L. P. HANSEN (1990): “Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data.” *Journal of Business & Economic Statistics* **8(1)**: pp. 53–69.
- EVERAERT, G. & L. POZZI (2014): “The Predictability Of Aggregate Consumption Growth In OECD Countries: A Panel Data Analysis.” *Journal of Applied Econometrics* **29(3)**: pp. 431–453.
- FERNÁNDEZ-VILLAVERDE, J. (2010): “The econometrics of DSGE models.” *SERIEs - Journal of the Spanish Economic Association* **1(1)**: pp. 3–49.
- FERSON, W. E. & G. M. CONSTANTINIDES (1991): “Habit persistence and durability in aggregate consumption: Empirical tests.” *Journal of Financial Economics* **29(2)**: pp. 199–240.
- FERSON, W. E. & C. R. HARVEY (1992): “Seasonality and Consumption-Based Asset Pricing.” *Journal of Finance* **47(2)**: pp. 511–52.
- FLAVIN, M. & S. NAKAGAWA (2008): “A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence.” *American Economic Review* **98(1)**: pp. 474–95.
- FUHRER, J. C. (2000): “Habit Formation in Consumption and Its Implications for Monetary-Policy Models.” *American Economic Review* **90(3)**: pp. 367–390.
- FUSARO, M. A. & D. H. DUTKOWSKY (2011): “What explains consumption in the very short-run? Evidence from checking account data.” *Journal of Macroeconomics* **33(4)**: pp. 542–552.
- GERALI, A., S. NERI, L. SESSA, & F. M. SIGNORETTI (2010): “Credit and Banking in a DSGE Model of the Euro Area.” *Journal of Money, Credit and Banking* **42(s1)**: pp. 107–141.
- GRUBER, J. W. (2004): “A present value test of habits and the current account.” *Journal of Monetary Economics* **51(7)**: pp. 1495–1507.
- GUARIGLIA, A. (2002): “Consumption, habit formation, and precautionary saving: evidence from the British Household Panel Survey.” *Oxford Economic Papers* **54(1)**: pp. 1–19.
- GUERRON-QUINTANA, P. A. (2008): “Refinements on macroeconomic modeling: The role of non-separability and heterogeneous labor supply.” *Journal of Economic Dynamics and Control* **32(11)**: pp. 3613–3630.
- GUERRON-QUINTANA, P. A. (2010): “What you match does matter: the effects of data on DSGE estimation.” *Journal of Applied Econometrics* **25(5)**: pp. 774–804.
- HEATON, J. (1993): “The Interaction between Time-Nonseparable Preferences and Time Aggregation.” *Econometrica* **61(2)**: pp. 353–385.
- HEATON, J. (1995): “An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications.” *Econometrica* **63(3)**: pp. 681–717.
- HIROSE, Y. (2008): “Equilibrium Indeterminacy and Asset Price Fluctuation in Japan: A Bayesian Investigation.” *Journal of Money, Credit and Banking* **40(5)**: pp. 967–999.
- HIROSE, Y. & S. NAGANUMA (2010): “Structural Estimation Of The Output Gap: A Bayesian Dsge Approach.” *Economic Inquiry* **48(4)**: pp. 864–879.
- IACOVIELLO, M. (2004): “Consumption, house prices, and collateral constraints: a structural econometric analysis.” *Journal of Housing Economics* **13(4)**: pp. 304–320.
- IACOVIELLO, M. & S. NERI (2010): “Housing Market Spillovers: Evidence from an Estimated DSGE Model.” *American Economic Journal: Macroeconomics* **2(2)**: pp. 125–64.
- IWAMOTO, K. (2013): “Habit formation in household consumption: evidence from Japanese panel data.” *Economics Bulletin* **33(1)**: pp. 323–333.
- JUSTINIANO, A. & B. PRESTON (2010): “Can structural small open-economy models account for the influence of foreign disturbances?” *Journal of International Economics* **81(1)**: pp. 61–74.
- KANO, T. (2009): “Habit formation and the present-value model of the current account: Yet another suspect.” *Journal of International Economics* **78(1)**: pp. 72–85.
- KILEY, M. T. (2010): “Habit Persistence, Nonseparability between Consumption and Leisure, or Rule-of-Thumb Consumers: Which Accounts for the Predictability of Consumption Growth?” *The Review of Economics and Statistics* **92(3)**: pp. 679–683.
- KORNIOTIS, G. M. (2010): “Estimating Panel Models With Internal and External Habit Formation.” *Jour-*

- nal of Business & Economic Statistics* **28(1)**: pp. 145–158.
- LAFORTE, J.-P. (2007): “Pricing Models: A Bayesian DSGE Approach for the U.S. Economy.” *Journal of Money, Credit and Banking* **39(s1)**: pp. 127–154.
- LEVIN, A. T., A. ONATSKI, J. WILLIAMS, & N. M. WILLIAMS (2005): “Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models.” In “NBER Macroeconomics Annual 2005, Volume 20,” NBER Chapters, pp. 229–312.
- LEVINE, P., J. PEARLMAN, G. PERENDIA, & B. YANG (2012): “Endogenous Persistence in an estimated DSGE Model Under Imperfect Information.” *Economic Journal* **122(565)**: pp. 1287–1312.
- LUBIK, T. A. & F. SCHORFHEIDE (2004): “Testing for Indeterminacy: An Application to U.S. Monetary Policy.” *American Economic Review* **94(1)**: pp. 190–217.
- LUBIK, T. A. & W. L. TEO (2014): “Deep Habits in the New Keynesian Phillips Curve.” *Journal of Money, Credit and Banking* **46(1)**: pp. 79–114.
- MATHESON, T. (2010): “Assessing the fit of small open economy DSGEs.” *Journal of Macroeconomics* **32(3)**: pp. 906–920.
- MAURER, J. & A. MEIER (2008): “Smooth it Like the ‘Joneses’? Estimating Peer-Group Effects in Intertemporal Consumption Choice.” *Economic Journal* **118(527)**: pp. 454–476.
- MEHRA, Y. P. & E. W. MARTIN (2003): “Why does consumer sentiment predict household spending?” *Economic Quarterly* **89(4)**: pp. 51–67.
- MERTENS, K. & M. O. RAVN (2011): “Understanding the Aggregate Effects of Anticipated and Unanticipated Tax Policy Shocks.” *Review of Economic Dynamics* **14(1)**: pp. 27–54.
- MILANI, F. (2007): “Expectations, learning and macroeconomic persistence.” *Journal of Monetary Economics* **54(7)**: pp. 2065–2082.
- NAIK, N. Y. & M. J. MOORE (1996): “Habit Formation and Intertemporal Substitution in Individual Food Consumption.” *The Review of Economics and Statistics* **78(2)**: pp. 321–28.
- PAGANO, P. (2004): “Habit persistence and the marginal propensity to consume in Japan.” *Journal of the Japanese and International Economies* **18(3)**: pp. 316–329.
- RABANAL, P. (2007): “Does inflation increase after a monetary policy tightening? Answers based on an estimated DSGE model.” *Journal of Economic Dynamics and Control* **31(3)**: pp. 906–937.
- RAVN, M. O., S. SCHMITT-GROHE, M. URIBE, & L. UUSKULA (2010): “Deep habits and the dynamic effects of monetary policy shocks.” *Journal of the Japanese and International Economies* **24(2)**: pp. 236–258.
- RHEE, W. (2004): “Habit formation and precautionary saving: evidence from the Korean household panel studies.” *Journal of Economic Development* **29(2)**: pp. 1–19.
- SAHUC, J.-G. & F. SMETS (2008): “Differences in Interest Rate Policy at the ECB and the Fed: An Investigation with a Medium-Scale DSGE Model.” *Journal of Money, Credit and Banking* **40(2-3)**: pp. 505–521.
- SCHMITT-GROHÉ, S. & M. URIBE (2012): “What’s News in Business Cycles.” *Econometrica* **80(6)**: pp. 2733–2764.
- SLANICAY, M. & O. VAŠÍČEK (2011): “Habit Formation, Price Indexation and Wage Indexation in the DSGE Model: Specification, Estimation and Model Fit.” *Review of Economic Perspectives* **11(2)**: pp. 71–91.
- SMETS, F. & R. WOUTERS (2003): “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area.” *Journal of the European Economic Association* **1(5)**: pp. 1123–1175.
- SMETS, F. & R. WOUTERS (2007): “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *American Economic Review* **97(3)**: pp. 586–606.
- SOMMER, M. (2007): “Habit Formation and Aggregate Consumption Dynamics.” *The B.E. Journal of Macroeconomics* **7(1)**: pp. 1–25.
- STOCK, J. H. & J. WRIGHT (2000): “GMM with Weak Identification.” *Econometrica* **68(5)**: pp. 1055–1096.
- SUGO, T. & K. UEDA (2008): “Estimating a dynamic stochastic general equilibrium model for Japan.” *Journal of the Japanese and International Economies* **22(4)**: pp. 476–502.
- TRIGARI, A. (2009): “Equilibrium Unemployment, Job Flows, and Inflation Dynamics.” *Journal of Money, Credit and Banking* **41(1)**: pp. 1–33.
- WOUTERS, R. & F. SMETS (2005): “Comparing shocks and frictions in US and euro area business cycles: a Bayesian DSGE Approach.” *Journal of Applied Econometrics* **20(2)**: pp. 161–183.