

A Direct Estimate of Rule of Thumb Behavior using the Method of Simulated Quantiles and Cross Validation

Nathan M. Palmer^{*†}

February 13, 2019

Preliminary and incomplete.

Abstract

Campbell and Mankiw (1989, 1990) demonstrated that aggregate data supported a model of household consumption in which roughly 50% of agents followed an optimizing strategy while the other 50% followed a "rule of thumb" strategy, consuming their current income. This paper revisits that hypothesis using structural semi-parametric estimation and microeconomic data. I directly estimate both the intensive and extensive margins on rule-of-thumb behavior in data from the Survey of Consumer Finances. A formal model selection exercise finds strong evidence supporting a generalization of the original Campbell and Mankiw (1989, 1990) result: roughly 50% of the population behaves in a way similar to rule of thumb consumers, even when the data is allowed to dictate the intensity of the rule of thumb behavior. In addition, this paper demonstrates the usefulness and flexibility of both the Method of Simulated Quantiles and k-fold cross validation for selecting between non-trivial models of agent behavior. This type of estimation and model selection is crucial for creating detailed heterogeneous agent and agent-based macro-financial models.

1 Introduction

This paper presents a method of formally selecting between models of economic agent behavior in the dynamic stochastic decision problem that is the basis of asset pricing, macro-finance, international finance, and macroeconomics more broadly. The formal selection is conducted using microeconomic data and structural estimation.

Specifically, I propose using k-fold cross validation to select between different structural microeconomic models that are estimated using a simulated method of moment estimator. I find strong evidence supporting the conclusions of Campbell and Mankiw (1989, 1990): that roughly 50% of the population follows something like a "consume everything" rule of thumb. Whereas Campbell and Mankiw (1989, 1990) approach this using aggregate data and aggregate estimation techniques, I approach this using microeconomic data and structural micro-econometric estimation, and formally select between models using k-fold cross validation.

^{*}Federal Reserve Board of Governors

[†]The views expressed are my own and not necessarily those of the Federal Reserve Board of Governors.

Typical macroeconomic and macro-financial modeling do not require a means to select between different types of household behavior from microeconomic data. The rational, optimizing framework for modeling households, banks, financial institutions, firms, and other macro-financial actors is taken as a given. A side effect of this modeling choice is that complicated macroeconomic and financial systems must be modeled relatively simply – if they are not, the model cannot be solved. Ten years have passed since the start of the 2007-2008 Financial Crisis. However, as described in Yellen (2016), one of the key areas in which macroeconomic research still lacking is "financial linkages to the real economy." This is still a pressing need in macroeconomic research a decade after the financial crisis occurred.¹

Advocates for agent-based modeling often point out an alternative: assign agents rules of thumb, which then make complicated general equilibrium models tractable.² This only pushes the issue further down the road. There are two important questions that immediately arise: first, where does agent behavior come from, if not from solving an optimization problem? Second, even assuming we have identified a set of possible candidate behaviors, how do we choose between them? As Sims (1980) famously opined, there is only one way for economic agents to be "rational" but infinitely many ways to be non-rational.

This second selection problem is particularly thorny. Typically the types of behavior suggested by agent-based modeling and learning literatures (discussed further below) exhibit a number of properties that make them very difficult to compare and select between in any formal way:

- There is often no simple, closed-form descriptions of agent behavior.
- Behavioral models are typically not nested, which eliminates a wide swath of selection procedures.
- Likelihood surfaces of agent behavior are very hard to compute, if they exist.
- For household behavior, easily available data on life-cycle choices is lacking in a number of ways that make selection difficult.

One solution to many of these issues is to use semi-parametric estimation such as Simulated Method of Moments. However as a non-likelihood-based estimation framework, many formal selection tests simply cannot be applied.

This paper proposes a solution to this problem: estimate different models of agent behavior using a semi-parametric estimator³, and then select between these models using k-fold cross validation, a very popular and flexible model selection method from the statistics literature.

This is demonstrated by selecting between models of household behavior. In particular, I conduct a structural, microeconomic replication of Campbell and Mankiw (1989, 1990): I directly estimate the fraction of households which act as rule of thumb consumers. In fact I go a step further and generalize their question. Instead of asking what fraction have either a "consume everything" rule of thumb or fully optimal behavior, I allow the data to jointly estimate both the fraction of agents who are near rule of thumb, as well as the *degree* to which agents are rule of thumb consumers.

I choose to execute this selection experiment on a model of household behavior for a number of reasons. First, the basic household consumption-savings problem, as described here, is the foundation of consumption-

¹See also Kiley (2016) for a more detailed view of developments in this area and work still to be done.

²See for example the discussion in Turrell (2016).

³As described below, a version of the Method of Simulated Quantiles of Dominic and Veredas (2013).

based asset pricing, macro-finance, international finance, and macroeconomics more broadly.⁴ Methods which are applicable to the consumer problem outlined in Section 3 can, in theory, be applied to a range of problems across a number of fields relevant to macro-finance and financial stability.

Second, the consumption-savings response of households, particularly when balance sheet positions are fragile, may have been a prime contributor to the severity and depth of the 2007 financial crisis. As Mian and Sufi (2015) argue, household debt growth was a key dynamic leading up to the crisis. Households in general were in weak balance sheet positions just before the crisis, and their reactions as the crisis developed exacerbated the situation. A particularly large portion of household balance sheets is housing. Leamer et al. (2007); Leamer (2015) and research cited therein forcefully claim that housing is a prime leading indicator of the business cycle and related crises. There may be many mechanisms for this. Glaeser (2016), for example argues that passive investors, and banks in particular, prefer housing and other real estate as a foundational part of their own balance sheet. Thus system-wide, correlated shocks to housing markets can be particularly damaging to bank balance sheets leading to housing playing a key role in the development and continuation of the crises. Finally, Geanakoplos et al. (2012) argues that the distributional properties of household balance sheet positions, combined with household behaviors, belief, and the broader economic conditions in general equilibrium, can lead to a wide range of dynamics in housing markets that are not revealed when examining a homogeneous-agent framework. These dynamics can include a coordinating equilibrium in which household beliefs and behaviors interact to create both booms and busts in house prices, which in turn can drive a mortgage crisis.

Central to all of these stories, implicitly or explicitly, is the role played by the household consumption-savings decision in general equilibrium. As pointed out in Carroll et al. (2017), Carroll (2012a), and works cited therein, even when households are traditionally rational, the heterogeneity in their behavior is very important for capturing general equilibrium dynamics, particularly during crisis and recessionary periods. This paper directly addresses these questions by calculating what fraction of the population may react quite strongly to wealth shocks.

Finally, agent-based models that are relevant to financial stability nearly always require reasonable models of household behavior to be trusted by policymakers. Geanakoplos et al. (2012) requires reasonable household behavior if the housing market dynamics are to be believed. Even a model such as Chan-Lau (2017), which is largely a model of the banking sector and its reliance on various funding sources, ultimately requires reasonable behavior on the part of the household depositor agents for the conclusions from the model to be trustworthy to policy makers. This paper contributes to agent-based models that may be used for financial stability purposes, by proposing a method for choosing behavior for a key class of agents in these models.

The rest of the paper is structured as follows: Section 2 outlines related literatures briefly. Section 3 outlines the household problem and its solution, which is the foundation of the estimation and selection process. Section 4 discusses the semi-parametric estimation method used. Section 5 discusses how the selection process will be applied to the models I estimate. Section 6 provides brief background on the selection process, k-fold cross validation. Section 7 discusses results, and Section 8 summarizes and concludes. Finally, there is an appendix at the end of the paper that visualizes the results in alternative ways to aid understanding.

⁴The literatures which support this claim are extensive, and I cite only a few illustrative examples in the literature review.

2 Related Literature

This paper contributes to a number of literatures. First and foremost, it contributes to the literature on model selection, particularly when applied to agent-level dynamic behavior. There are a few papers along these lines. At the aggregate level, Campbell and Mankiw (1989, 1990) spawned an entire literature asking whether rule of thumb agents can be recognized and even estimated in the data. See for example Carroll et al. (2011) and the literatures cited therein. At the microeconomic level, Carroll (2001) is an early example of specifically testing relevance of different models of micro-level consumer behavior, particularly for the problem formulation I use. Very recently, Ganong and Noel (2017) conduct a selection exercise similar to mine here. The difference between that work and my own is that I jointly re-estimate the key behavioral parameters of each model I consider, which allows me to apply the selection technique I use.⁵

Li (2009) is a structural estimation and selection paper that is likely closest to my approach. The author uses a measure of predictive capability of the model in a way similar to my own. It is unclear if the method can be directly applied to the semi-parametric estimation approach I use. Determining whether this approach is flexible enough to apply to the types of behavioral models I would like to select between is a promising avenue for future research. Related to this is a large literature on selecting between non-nested models which have likelihood representations; Vuong (1989) is an extremely popular approach for this class of models. The estimation method I use does not have a likelihood surface, and thus cannot use any of these methods.⁶

As described in the introduction, this paper contributes to the wide literature that relies on the household consumption-savings problem, including consumption-based asset pricing, macro-finance, international finance, and macroeconomics more broadly. The literatures which support this claim are extensive, and I cite only a few illustrative examples here. For consumption-based asset pricing, Cochrane (2017) is an excellent example; others include Gabaix (2012) and Gourio (2012) on rare disasters, He and Krishnamurthy (2013) on intermediary-based asset pricing, and the classic text Cochrane (2009). For international finance, Obstfeld et al. (1996) is a foundational text and Lane (2001) provides a nice, if dated, overview. There is no single citation for macroeconomics broadly, although Sargent and Stachurski (2017) provides an extensive overview.

Finally, the reason this paper exists is to provide a method of agent-level behavioral selection for agent-based models in the areas of asset pricing, macro-finance, and macroeconomics more broadly. This class of papers consists of both models of individual learning, and the large-scale models in which this learning might be applied. The types of agent-level learning I have in mind, particularly as applied to the household consumption problems, are represented by Howitt and Özak (2014), Özak (2014), Allen and Carroll (2001), Başçı and Orhan (2000)⁷. Tesfatsion and Judd (2006) provides an excellent outline of early uses of reinforcement learning in economics, and Sinitskaya and Tesfatsion (2014) demonstrates how these may be applied in a macroeconomic setting for potential policy use. Palmer (2012) and Palmer (2016) are models I would like to apply this to. I believe the work of Individual Evolutionary Learning, as described in Arifovic and Ledyard (2012) is very promising for macro-financial agent-based models. A very productive potential direction of research is to apply this same selection technique to data obtained from lab experiments combined with

⁵This is, of course, an extremely time-intensive process.

⁶See also Gourieroux and Monfort (1994) for a slightly dated but very nice review of this topic.

⁷Which revisits the very interesting results of Lettau and Uhlig (1999), which examines ways in which households might very easily learn the “wrong” rules even when the optimal rule is in their choice set.

agent-based models. For an excellent implementation and overview of this idea, see Houser et al. (2004) and more recently Cotla (2016) and references therein.

In the more traditional learning space, Gabaix (2014) developed a sparsity-based dynamic programming model that seeks to capture the idea that agents do not re-evaluate their behavior unless they are prompted by “big enough” events in their world, resulting in an (s, S) style behavioral rule which may be used in large-scale models. Evans and McGough (2015) address a similar problem as that of Allen and Carroll (2001), but in a general equilibrium context. Like Howitt and Özak (2014), their agents must know something about the first-order conditions of their optimization problem, and use these to learn about the shadow prices of their choices. As with Krusell and Smith (1996), they find that the aggregate dynamics, particularly the transition dynamics, are greatly affected by learning.

The types of aggregate agent-based models for which I want to choose learning behavior include Geanakoplos et al. (2012), the CRISIS project under Doyne Farmer,⁸ the EURACE project as described in Deissenberg et al. (2008), Fischer and Riedler (2014), Henry et al. (2013), and many models cited in the excellent survey Turrell (2016).

3 The Household Problem

To conduct a formal model selection process, each of the following things requires the previous item in the list: the agent problem, a solution method, an estimation procedure for key behavioral parameters, and finally a selection method to choose between best estimates for different models. If there is no solution method, for example, the estimation step cannot be carried out. If there is no appropriate estimation technique, the selection process cannot select between best-fit models. This and the following sections lay out these components of the paper in this order: agent problem, solution method, estimation method, and finally, the selection method. Brief review is provided of each step when necessary.

3.1 Finite Horizon Consumer Problem

Consider the following finite horizon consumption-under-uncertainty problem. At time $T + 1$, the consumer dies with certainty. The problem is to allocate consumption appropriately from $t = 0$ to $t = T$. The full problem can be described as follows, with notation that closely follows that of Carroll (2012c). This setup can describe a wide range of consumer circumstances, including retirement and fixed income over final years of life, by properly defining the members of Γ_t , ψ_t , and ξ_t discussed below:

⁸See <https://github.com/crisis-economics/CRISIS> for more details.

$$\begin{aligned}
& \max_{\{\mathbf{c}_t\}_{t=0}^T} \mathbb{E}_0 \left[\sum_{t=0}^T \hat{\beta}^t \mathcal{D}_t u(\mathbf{c}_t) \right] && (1) \\
& \text{s.t.} \\
& \mathbf{a}_t = \mathbf{m}_t - \mathbf{c}_t \\
& \mathbf{b}_{t+1} = \mathbf{a}_t \mathbf{R} \\
& \mathbf{p}_{t+1} = \mathbf{p}_t \Gamma_{t+1} \psi_{t+1} = \mathbf{p}_t \tilde{\Gamma}_{t+1} \\
& \mathbf{m}_{t+1} = \mathbf{b}_{t+1} + \mathbf{y}_{t+1} \\
& \mathbf{y}_{t+1} = \mathbf{p}_{t+1} \xi_{t+1} \\
& \mathbf{c}_t \geq 0 \\
& \mathbf{m}_t \geq -\bar{\mathbf{q}}_t \\
& \mathbf{m}_0 \text{ given}
\end{aligned}$$

where

- $\hat{\beta}$ is the discount rate,
- \mathcal{D}_t is the independently distributed probability of staying alive through period t , with $\mathcal{D}_{T+1} = 0.0$
- \mathbf{a}_t is end-of-period assets,
- \mathbf{m}_t is beginning-of-period total market resources (“cash on hand”), $\mathbf{m} \in M \subset \mathbb{R}$,
- \mathbf{c}_t is consumption in period t ,
- \mathbf{R} is a constant return factor on assets, $\mathbf{R} = (1 + r)$,
- \mathbf{y}_t is income in period t ,
- \mathbf{p}_t is permanent non-asset income,
- Γ_t is deterministic permanent income growth factor,
- $\tilde{\Gamma}_t$ is a combination of Γ_t and ψ_t for notational convenience,
- $-\bar{\mathbf{q}}_t$ is a borrowing constraint, $-\bar{\mathbf{q}}_t \in [0, \infty) \forall_t$,
- ψ_t is a mean-1 iid permanent shock to income, and
- ξ_t is a mean-1 iid transitory shock income, composed as

$$\xi_t = \begin{cases} 0 & \text{with prob } \wp_t > 0 \\ \frac{\zeta_t}{\wp_t} & \text{with prob } \wp_t \equiv (1 - \wp_t) \end{cases}, \text{ where}$$

- \wp_t is a small probability that income will be zero
- ζ_t is a mean-1 iid transitory shock to income

The utility function $u(\cdot)$ if the objective function (1) is of the Constant Relative Risk Aversion (CRRA) form with risk-aversion parameter ρ :

$$u(c) = \frac{c^{(1-\rho)}}{1-\rho}.$$

The consumer dies with certainty at period T but may die with probability D_t in any other period t , where we assume that the arrival of death is independent each period. This implies that the probability of being alive in a period s can be calculated:

$$D_s = \prod_{j=0}^s (1 - D_j)$$

or when $D_j = D \ \forall j$, we can restate the expression in simpler terms:

$$\begin{aligned} D &\equiv (1 - D) \\ \rightarrow D_s &= (1 - D)^s \ \forall s. \end{aligned}$$

When death occurs all subsequent period utility functions are 0. Note that the finite horizon problem could be equivalently re-written with an infinite horizon but a utility function defined as

$$u(c_t) = \begin{cases} \frac{c^{(1-\rho)}}{1-\rho} & \text{for } t \leq T \\ 0 & \text{for } t > T. \end{cases}$$

This also indicates why D_s shows up in a clean form in the objective function: the expectation each period can be decomposed into the convex combination of expectations under either staying alive or dying in the following period. However, the summation of payoffs following death is 0. This term falls out of the expression leaving only D_s in the objective.

As in Carroll (2012b), this problem can be normalized by permanent income \mathbf{p}_t to produce a simplified version of the full problem with a reduced number of state variables. The bold symbols used above indicate non-normalized variables, while the regular non-bold symbols used below indicate variables normalized by permanent income. The normalized problem can be written in Bellman form:

$$v_t(m_t) = \max_{c_t} u(c_t) + \hat{\beta} D_{t+1} \mathbb{E}_t \left[\tilde{\Gamma}_{t+1}^{1-\rho} v_{t+1}(m_{t+1}) \right] \quad (2)$$

s.t.

$$a_t = m_t - c_t$$

$$b_{t+1} = \left(\frac{R}{\tilde{\Gamma}_{t+1}} \right) a_t = \mathcal{R}_{t+1} a_t$$

$$m_{t+1} = b_{t+1} + \xi_{t+1}$$

$$c_t \geq 0$$

$$m_t \geq -\bar{q}$$

m_0 given,

or simplified further:

$$\begin{aligned}
v_t(m_t) &= \max_{c_t} u(c_t) + \hat{\beta} \mathcal{D}_{t+1} \mathbb{E}_t \left[\tilde{\Gamma}_{t+1}^{1-\rho} v_{t+1}(m_{t+1}) \right] & (3) \\
&\text{s.t.} \\
m_{t+1} &= \mathcal{R}_{t+1}(m_t - c_t) + \xi_{t+1} \\
c_t &\geq 0 \\
m_t &\geq -\bar{q} \\
m_0 &\text{ given.}
\end{aligned}$$

3.2 Solution

The general solution method is as follows: in the final period T , the value function in the following period is $v_{T+1}(m) = 0 \forall m$, and the value function in period T is simply the utility of consuming all available resources, that is $v_T(m) = u(m)$. This makes the problem in period $T-1$ straightforward to solve numerically for both the consumption function and value functions $c_{T-1}^*(m)$ and $v_{T-1}^*(m)$:

$$\begin{aligned}
c_{T-1}^*(m) &= \underset{c \in [0, \bar{m}]}{\operatorname{argmax}} u(c) + \hat{\beta} \mathcal{D}_T \mathbb{E}_{T-1} \left[\Gamma_T^{1-\rho} u(\mathcal{R}_T(m - c) + \xi_{t+1}) \right] \\
&\text{and} \\
v_{T-1}^*(m) &= u(c_{T-1}^*(m)) + \hat{\beta} \mathcal{D}_{t+1} \mathbb{E}_{T-1} \left[\Gamma_T^{1-\rho} u(m_T) \right]
\end{aligned}$$

where \bar{m} is a self-imposed liquidity constraint.⁹

With these numerical solutions in hand, the solution method is now simply recursive: step back one more period to $T-2$ and solve for optimal consumption and value functions using $c_{T-1}^*(m)$ and $v_{T-1}^*(m)$. This process can be continued back until the first period $t=0$. The result is a set of policy functions (consumption functions) for each date t of life. This solution process is outlined in greater detail in Carroll (2012b).

4 Structural Estimation

Denote the behavioral parameters β, ρ , (discounting and risk aversion, respectively) as

$$\phi = \{\beta, \rho\}$$

and denote the structural problem parameters as

$$\begin{aligned}
\varrho &= \{\varrho_t\}_{t=0}^T, \text{ where} \\
\varrho_t &= \{\Gamma, \psi_t, \xi_t, \wp_t, \theta_t\}, \forall t.
\end{aligned}$$

Given an arbitrary behavioral parameter set $\phi = \{\beta, \rho\}$, and choosing the values and data-generating processes for the structural problem parameters ϱ to match consumer experiences described in Carroll (2012b),

⁹Carroll (2012b) demonstrates the reasoning behind this derivation. In a model with positive probability of a zero-income event, $\bar{m} = m$.

I can solve for the set of consumption functions that are optimal under these conditions, $\{c_t^*(m)\}_{t=0}^T$.

Now I can use the calibrated parameters ϱ to generate N different simulated consumer experiences (vectors of income shocks) for all periods $t = 0, 1, \dots, T$. Applying the consumption functions $\{c_t^*(m)\}_{t=0}^T$ to this set of simulated experiences generates a panel of size $T \times N$ simulated wealth, where wealth is defined as the variable b_t in the mathematical description of the consumer problem above. Moments of these cross-sectional distributions of wealth can then be compared to the equivalent moments in appropriately constructed empirical data from the Survey of Consumer Finance (SCF). The "textbook" structural empirical estimation of this problem using the SCF is described in depth in Carroll (2012b), wherein the author matches to the median quantile alone. As described in Section 4.1 below, I find the median alone does not capture the change in dispersion of the data over age cohorts, and thus I add the interquartile range to the minimization objective.

Using appropriately constructed SCF data, create empirical distributions of wealth-to-income ratios for seven age groups, in the 5-year windows used by the SCF (finer-grained ages are not available in this dataset): 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, and 56-60. Label these age groups as $\tau = 1, \dots, 7$. This produces our seven empirical distributions against which we will conduct our semi-parametric estimation. To create comparable distributions from the simulation described above, pool simulated data to create wealth distributions that match the age groups of SCF data.

Calculate quantiles on both the empirical and simulated data:

$$\begin{aligned} \text{Empirical: } & \hat{q}_{\tau,25}, \hat{q}_{\tau,50}, \hat{q}_{\tau,75} \quad \text{for } \tau = 1, 2, \dots, 7 \\ \text{Simulated: } & q_{\tau,25}^\phi, q_{\tau,50}^\phi, q_{\tau,75}^\phi \quad \text{for } \tau = 1, 2, \dots, 7 \end{aligned}$$

Construct functions of quantiles:

$$\text{Empirical: } \hat{\varphi}_\tau = \begin{pmatrix} \hat{q}_{\tau,75} - \hat{q}_{\tau,25} \\ \hat{q}_{\tau,50} \end{pmatrix}$$

$$\text{Simulated: } \varphi_{\phi,\tau}^N = \begin{pmatrix} q_{\tau,75}^\phi - q_{\tau,25}^\phi \\ q_{\tau,50}^\phi \end{pmatrix}$$

Use these to construct vectors of functions of quantiles:

$$\text{Empirical: } \hat{\varphi} = (\hat{\varphi}_1^\top, \hat{\varphi}_2^\top, \dots, \hat{\varphi}_7^\top)^\top$$

$$\text{Simulated: } \varphi_\phi^N = (\varphi_{\phi,1}^\top, \varphi_{\phi,2}^\top, \dots, \varphi_{\phi,7}^\top)^\top$$

Form the following objective function, which compares functions of population quantiles between empirical wealth-to-income ratio from the SCF and its simulated equivalent:

$$\varpi_\varrho(\phi) = (\hat{\varphi} - \varphi_\phi^N) \mathbf{W} (\hat{\varphi} - \varphi_\phi^N)$$

where \mathbf{W} is a positive definite matrix of weights.

Here $\varpi_\varrho(\phi)$ represents the objective value for the distance between functions of quantiles of the two populations. This estimation approach follows the Method of Simulated Quantiles as proposed by Dominicy and Veredas (2013). Due to computational time constraints, the initial results calculated here are based on a static positive definite weighting matrix \mathbf{W} ; see Dominicy and Veredas (2013) for more discussion of the form of this objective function.¹⁰

Once this value has been constructed as a function of ϕ , the estimation occurs by simply minimizing ϕ numerically:

$$\min_{\phi} \varpi_\varrho(\phi).$$

This is accomplished in code by handing the expression $\varpi_\varrho(\phi)$ to a numerical minimization process.¹¹ The standard error on the resulting estimation of $\{\beta^*, \rho^*\}$ may be found by bootstrapping the empirical data and repeating the above estimation process the desired $N_{bootstrap}$ number of times. The estimation conducted for this paper was executed using an extension of the HARK modeling framework (Carroll et al., 2018).

We can summarize a single "step" in the structural estimation process as follows:

1. Given behavioral parameter set $\phi = \{\beta, \rho\}$ and calibration set ϱ solve for optimal consumption functions $\{c_t^*(m)\}_{t=0}^T$.
2. Use the calibrated parameters ϱ to generate N different vectors of income shocks from $t = 0, 1, \dots, T$, producing an $N \times T$ panel of income shocks.
3. Apply the consumption functions $\{\tilde{c}_{n,t}(m)\}_{t=0}^T$, to the set of simulated income experiences and generate an $N \times T$ distribution of simulated wealth holdings.
4. Construct the functions of quantiles of empirical and simulated wealth data and score the parameter set $\phi = \{\beta, \rho, x\}$.

This produces the "score" for this model, which is minimized in Equation (4).

4.1 Empirical Motivation

We use semi-parametric estimation for this model because we know the finite-horizon optimization framework systematically matches the behavior of very wealthy households inadequately; see Carroll (2002) and citations therein. Semi-parametric estimation approaches allow us to explicitly exclude very wealthy households from the objective function without distorting the rest of the procedure. This is the reason that Cagetti (2003) and

¹⁰Robustness checks with estimation methods used in Carroll (2012b), which employ absolute-value semi-parametric estimators, have been conducted and demonstrate qualitatively similar results.

¹¹The nonlinear numerical optimization approach I employ uses a suite of fast gradient-free optimizers, started from points on a grid over the optimization space. The two smallest minima points selected by the fast optimizers were used as initial points in slower optimizers to end up at final results. This was repeated from many starting points and with many algorithms to ensure that the final points were trustworthy minima. The numerical algorithms used were: BOBYQA, Powell (2009); Sbplx (based on Subplex), Rowan (1990); PRAXIS, Brent (2013); NEWUOA, Powell (2006); and NELDERMEAD Nelder and Mead (1965), Box (1965), and Richardson and Kuester (1973). Finally, each of these algorithms was individually nested inside the Augmented Lagrangian method described in Conn et al. (1991) and Birgin and Martínez (2008) to allow for generalized imposition of restrictions and bounds.

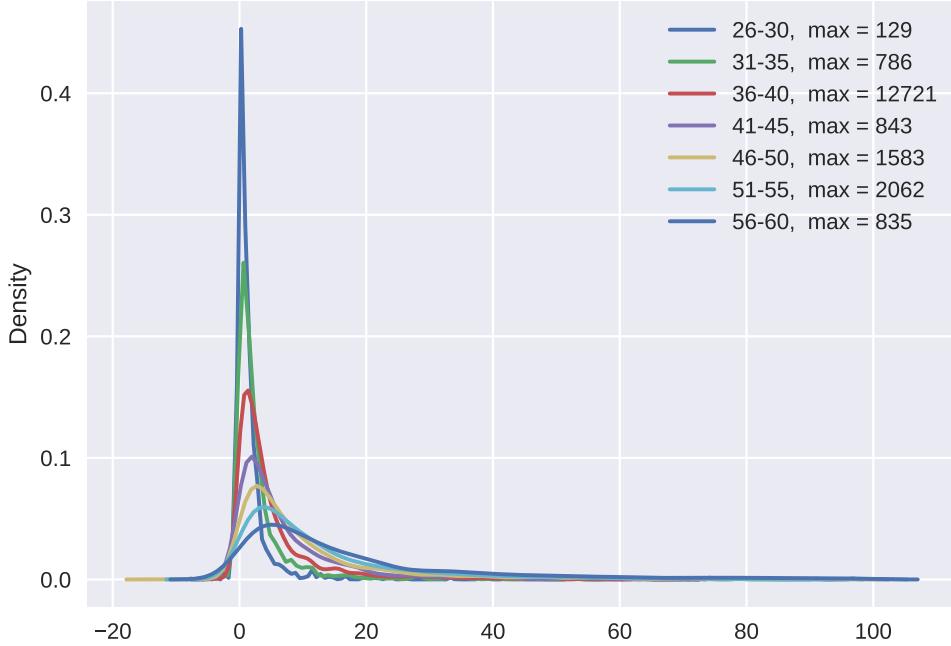


Figure 1: Empirical Kernel Density: Wealth to permanent income b_t for 7 SCF Age Groups
Author's analysis. Survey of Consumer Finance, Federal Reserve Board.

Carroll (2012b) use medians of the seven SCF wealth distributions and the Simulated Method of Moments (SMM) estimator to estimate behavior parameters. Their estimation approach targets the non-parametric location of the distribution as represented by the median to avoid the corruption which may theoretically occur if estimation includes measures of rich agent behavior in the tails of the data.

The primary difference between the estimation conducted in Cagetti (2003) and Carroll (2012b), and the estimation I conduct, is that I explicitly include a non-parametric measurement of the dispersion of the distributions, as well as a measurement of the median. To see why this may be important, consider Figure 1.

Figure 1 displays the seven age-group distributions of wealth normalized by permanent income from the SCF that are used in the estimation process (this is variable b_t in the full problem description). When the MSQ estimation described in Section 4 is conducted using only medians, or the SMM employed by Cagetti (2003) and Carroll (2012b) is used, the medians of these distributions are matched quite well. However, the dispersion of the distributions is matched quite poorly. The best fit simulated wealth data that results will have dispersion as narrow as the first age group on the far left, 26-30, for all age groups.

However even when the quantiles-based measure of dispersion is added to the estimation process as I describe in Section 4, the fit to dispersion improves only modestly. The later-life distributions are simply too wide to be fit by agents following optimal plans. The simulated agents following optimal plans simply always save too much early on to produce the wealth dispersion seen later in the data. However, the data fit improves tremendously in this regard when we allow some fraction of the population to be near-hand-to-mouth. In fact, we will see this emerge from the data itself in the following sections.

5 Selecting Between Models of Agent Behavior

The final goal of the methodology explored in this paper is to use microeconomic data, structural estimation, and k-fold cross validation to select between different models of non-optimizing agent behavior, such as described in Palmer (2016, 2012), Howitt and Özak (2014), Yıldızoglu et al. (2014), or Evans and McGough (2015), Gabaix (2014), or even the variety of alternative models suggested by Ganong and Noel (2017). This is a lofty goal, and one that is reserved for future effort. This paper instead lays the foundation for such work by illustrating how this selection may be done under a familiar framework.

Consider the following question, posed by Campbell and Mankiw (1989, 1990): is the population of households composed of all agents who optimize, or does some fraction follow a rule of thumb, such as "consume all income?" Famously they found that a model in which roughly 50% of households optimized, while 50% followed a spendthrift, "consume everything" rule of thumb, was most consistent with the data.

One could ask the same question following the estimation procedure described in Section 4 above. Instead of estimating the (β, ρ) pair of behavioral parameters, the same estimation procedure could be employed to jointly re-estimate the parameters (β, ρ, x) , where now x would represent the fraction of agents who are rule of thumb consumers.

The extension of the estimation method is straightforward. As before, we follow the estimation steps outlined in Section 4. Note that this time, step (2) is new:

1. Given behavioral parameter set $\phi = \{\beta, \rho\}$ and calibration set ϱ solve for optimal consumption functions $\{c_t^*(m)\}_{t=0}^T$.
2. Use the fraction x to randomly determine which of the $\tilde{N} \subset \{1, 2, \dots, N\}$ simulated consumers will be assigned to use a spendthrift consumption rule instead of optimal consumption functions. Call the new set of consumption functions $\{\tilde{c}_{n,t}(m)\}_{t=0}^T$; note that the subscript n determines whether the agent is using a rule of thumb or an optimal consumption function for their entire life.¹²
3. Use the calibrated parameters ϱ to generate N different vectors of income shocks from $t = 0, 1, \dots, T$, producing an $N \times T$ panel of income shocks.
4. Apply the consumption functions $\{\tilde{c}_{n,t}(m)\}_{t=0}^T$, to the set of simulated income experiences and generate an $N \times T$ distribution of simulated wealth holdings.
5. Construct the functions of quantiles of empirical and simulated wealth data and score the parameter set $\phi = \{\beta, \rho, x\}$.

This process is repeated as described in Section 4 until the minimal score is identified.

Denote the model produced by the estimation process in Section 4 as "Model I," which has all agents following the optimal consumption policy, and denote the model produced by steps 1-5 directly above in Section 4 as "Model II." Conveniently, Model I has one type of agent in it (optimal consumers), and Model II has two types (both optimal and rule of thumb consumers.)

Because we are applying the same estimation process to both Model I and Model II, we can use k-fold cross validation process described in the following section to formally select between these models. Importantly,

¹²Care must be taken to ensure that random number generation seeds are set appropriately for this process.

because we jointly re-estimate the parameter set $\phi = \{\beta, \rho, x\}$ for Model II, we will be selecting between the two best fit versions of each model. This is of key importance. The question of the trade-off between improved fit and over-fitting cannot be answered if we do not re-estimate both sets of parameters. For example, adding the additional "x" may make the model so flexible that it can fit anything, but is poor at prediction. We must control for this in the selection process.

Note, however, that I can capture rule of thumb consumption even in the standard approach: simply set $\beta = 0.0$. This observation allows an even further refinement and generalization on the selection question asked above. We can actually let the data itself tell us two things: first, whether there are "hand-to-mouth" style consumers, second, *to what degree* there are hand-to-mouth consumers in the data.

This is done quite naturally: as described for Model II above, jointly re-estimate $\{\beta_1, \beta_2, \rho, x\}$, where now x is the fraction of agents who have $\{\beta_2, \rho\}$ preferences. If there are truly hand-to-mouth consumers in the data, this model should be able to capture that by estimating a non-trivial fraction of the population with $\beta_1 = 0$ or $\beta_2 = 0$. If there are in fact hand-to-mouth consumers in the model, they should emerge naturally here as agents with very low β . In addition, this approach will let the data itself tell us what the best "low β " looks like for representing such consumption behavior in this framework. Note that ρ is assumed constant across all agents to improve identification in the model. This is now a 4-parameter estimation, $\{\beta_1, \beta_2, \rho, x\}$, with all parameters jointly re-estimated.

We can take one additional step. Observe that we are essentially selecting between the number of "types of β " that the data supports in the structural model we will use. This is motivated by the question, "what fraction of agents might be rule of thumb consumers." However our formulation allows us to take this question one step further: we can also simply ask, "under this structural estimation, how many types of agents best describes the population?"

We can define a "Model J" estimation process for $J \geq 2$, which generalizes the process for Model II. Note again that only step (2) is changed:

1. Given behavioral parameter set $\phi = \{\beta_1, \beta_2, \dots, \beta_J, x_1, \dots, x_{J-1}, \rho\}$ and calibration set ϱ solve for optimal consumption functions $\{c_{t,j}^*(m)\}_{t=0}^T$.
2. Use the fractions x_j to randomly determine which of the $\tilde{N} \subset \{1, 2, \dots, N\}$ simulated consumers will be assigned to use which consumption rules. Call the new set of consumption functions $\{\tilde{c}_{n,t}(m)\}_{t=0}^T$; note that the subscript n determines which consumption rule the agent will use for their entire life.¹³
3. Use the calibrated parameters ϱ to generate N different vectors of income shocks from $t = 0, 1, \dots, T$, producing an $N \times T$ panel of income shocks.
4. Apply the consumption functions $\{\tilde{c}_{n,t}(m)\}_{t=0}^T$, to the set of simulated income experiences and generate an $N \times T$ distribution of simulated wealth holdings.
5. Construct the functions of quantiles of empirical and simulated wealth data and score the parameter set $\phi = \{\beta_1, \beta_2, \dots, \beta_J, x_1, \dots, x_{J-1}, \rho\}$.

¹³Care must be taken to ensure that random number generation seeds are set appropriately for this process.

For this paper I test up to $J = 6$ different types of consumers. Listed below, the six models estimated and selected between are:

- Model I: 1 type of optimizing agent: estimate 2 parameters: $\{\beta, \rho\}$
- Model II: 2 types of optimizing agent: jointly estimate 4 parameters: $\{\beta_1, \beta_2, x_1, \rho\}$
- Model III: 3 types of optimizing agent: estimate 6 parameters: $\{\beta_1, \beta_2, \beta_3, x_1, x_2, \rho\}$
- Model IV: 4 types of optimizing agent: estimate 8 parameters: $\{\beta_1, \beta_2, \beta_3, \beta_4, x_1, x_2, x_3, \rho\}$
- Model V: 5 types of optimizing agent: estimate 10 parameters: $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, x_1, x_2, x_3, x_4, \rho\}$
- Model VI: 6 types of optimizing agent: estimate 12 parameters: $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, x_1, x_2, x_3, x_4, x_5, \rho\}$

Here the importance of model selection becomes clear: as the number of types of agents increases (i.e. as the number of estimation parameters increases), the fit should improve mechanically. If we do not penalize that in some way we will be led astray by the ever-improving fit of increased model size. In a number of model frameworks, we can use selection criteria such as BIC and AIC to assess whether increased model size "improves" the model. Because the estimation framework used here does not have a likelihood surface, we cannot use typical selection approaches. Instead I employ k-fold cross validation, a common and flexible framework which can be applied here. This is described in the next section.¹⁴

6 K-Fold Cross Validation Review

This section briefly reviews the selection mechanism I use to choose between these models, and describes how I apply it in this context. The review presented here is a very brief treatment of k-fold cross validation and follows the discussion in Shalizi (2017).¹⁵

K-fold cross validation uses a version of resampling to estimate the out-of-sample predictive capability of a fitted model. A brief example, following Chapter 3 of Shalizi (2017), illustrates the use quite well.

Consider data generated by adding random $N(0, 1)$ noise to a second-order polynomial. Specifically, Figure (2) plots this equation:

$$y = -0.3x + 5.5x^2 + \epsilon, \text{ where } \epsilon \sim N(0, 1).$$

¹⁴An aside on terminology: the parameter β is often described as "patience," however we must handle this term with extreme care here. In common usage, "patience" is often a moralizing description of whether an individual "plans for the future enough," conditional on the complex real-world environment they actually live and make decisions in. β is used here mechanically to generalize a particular model. However its technical definition here is far different than the common usage of the word patience. Here it simply allows selection between different possible models. As in Carroll et al. (2017), β can stand in for any number of things, the simplest of which is demographic variables such as age. In addition, Gabaix and Laibson (2017) make an extremely important point. They show that the complexity of the computation forecast problem for individual agents when agents are fully rational, completely patient ($\beta = 1$), but they exist in a more complex setup than the traditional household savings problem described in Section 3.1, exhibit "as-if discounting." They appear as if they are solving a problem such as described in Section 3.1 but with $\beta \leq 1$. Thus lower β itself may simply reflect a different complexity and difficulty of solving real-world problems, separate from the common usage of the word "patience." Finally, Banerjee and Duflo (2011) makes this exact point again from a development perspective: economic agents with the least resources must work extra hard to smooth consumption and plan for the future, because the circumstances in which they solve their household problem is far more difficult than for the "median" economic agent.

¹⁵For a more thorough exposition, please reference one of these excellent introductory sources: Shalizi (2017), Friedman et al. (2009), Efron and Tibshirani (1994), and for a survey overview, see Arlot et al. (2010).

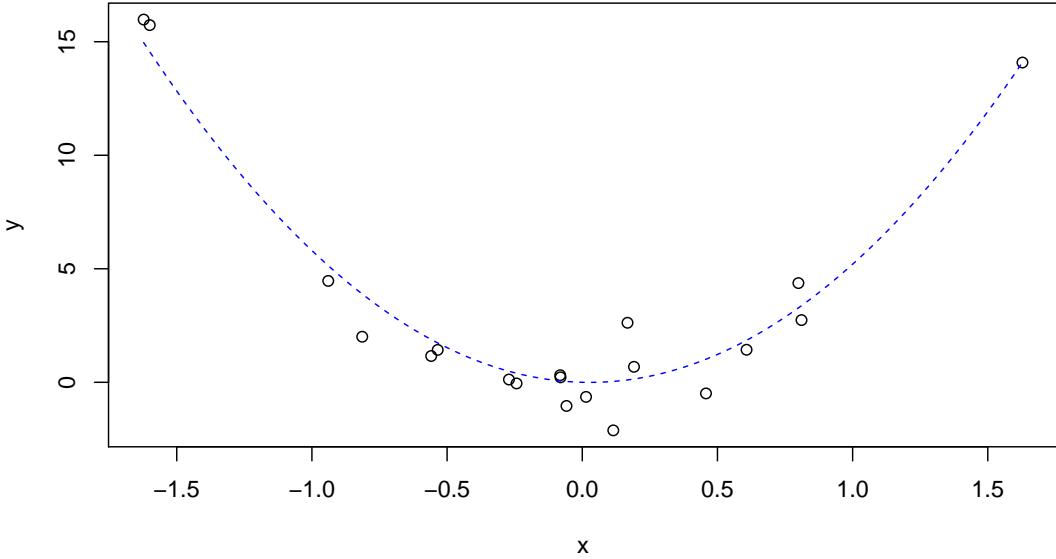


Figure 2: Noisy Data Around 2nd Degree Polynomial, 25 points
Dashed line is the true line; dots are noisy $N(0, 1)$ data around the line.
Author's reproduction of Shalizi (2017).

It is straightforward to fit linear regression models to this data that include increasing polynomial degrees. Figure (3) displays the results of fitting "0th" (simply the sample average, the flat blue line) up to 9th degree polynomial regression. That is, the regression of polynomial degree 10 is:

$$y = \gamma_0 + \gamma_1 x + \gamma_2 x^2 + \gamma_3 x^3 + \dots + \gamma_9 x^9.$$

As expected, Figure (4) demonstrates that the R^2 and even adjusted R^2 for each of these polynomial regression models increases and remains high as the polynomial degree increases.

The in-sample prediction as measured by the mean squared error (MSE), displayed in Figure (5), tells much the same story. As we increase the degree of the polynomial, the MSE drops monotonically. The models fit better and better. However, as Figure (3) might indicate to the observant eye, this increased "quality of fit" is spurious. As the polynomial degree increases, the model is overfitting this specific set of data.

The increased quality of fit as measured by the R^2 and MSE comes at the cost of lost flexibility. If we were to take the best fit model, of polynomial degree 9, re-draw many new data points from the data generating process (DGP), and measure the MSE of the best fit model on the *new* data, we might find a different story. Figure (6) displays the MSE of the best fit models on the new data, when 30,000 new data points are drawn. It can be seen plainly that the models of polynomial degree greater than 3 begin to fit the data increasingly poorly; note that the y-axis is a logarithmic scale. This fact is not reflected by the in-sample MSE or R^2 measures; it is detected here only when the best-fit models are applied to out-of-sample data.

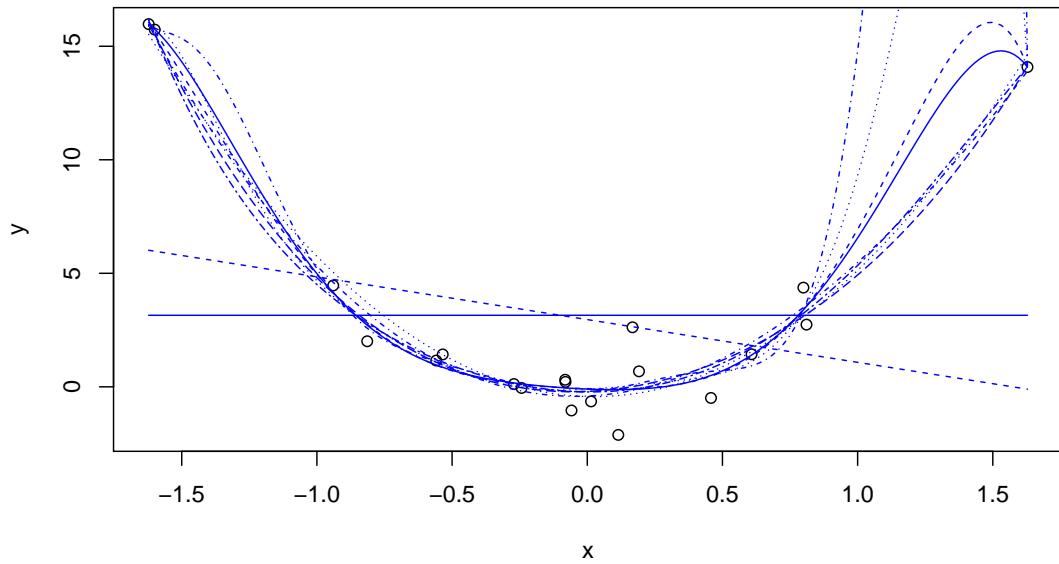


Figure 3: Fitting 0th through 9th Degree Polynomial Regression
Author's reproduction of Shalizi (2017).

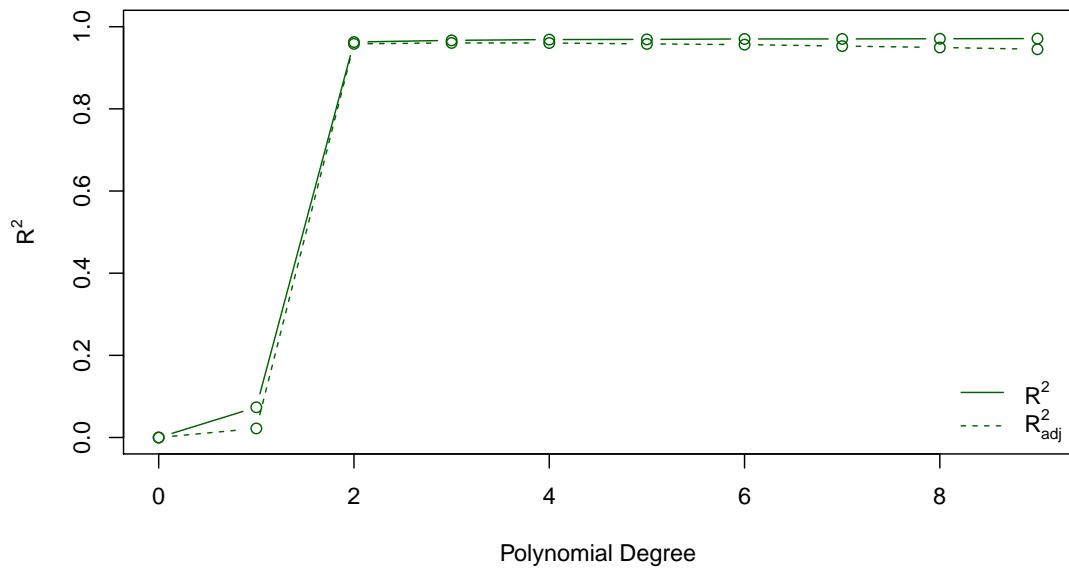


Figure 4: R^2 and Adjusted R^2 for 0th through 9th Degree Polynomial Regressions
Author's reproduction of Shalizi (2017).

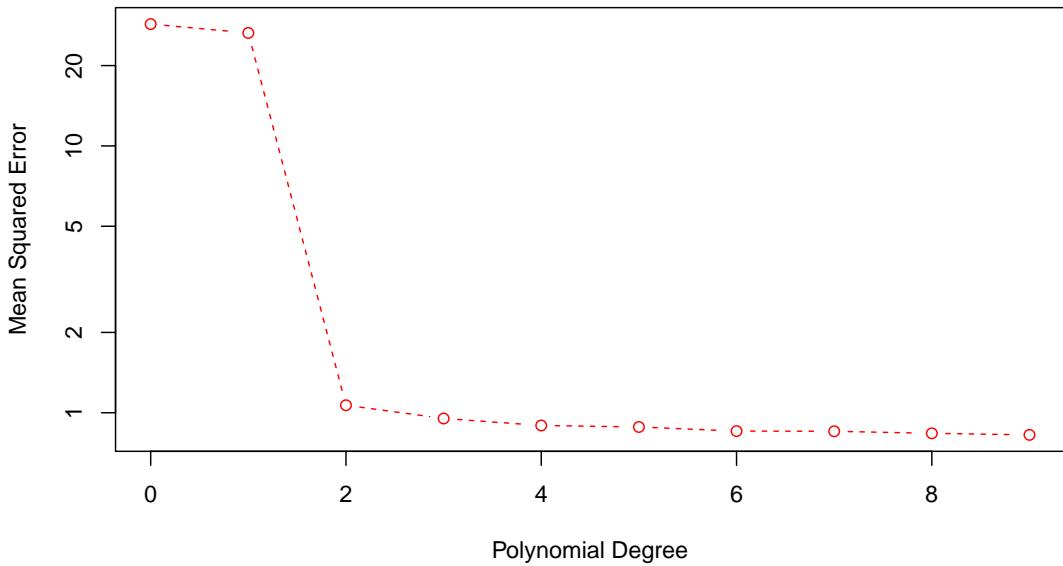


Figure 5: Mean Squared Error for 0^{th} through 9^{th} Degree Polynomial Regressions
Author's reproduction of Shalizi (2017).

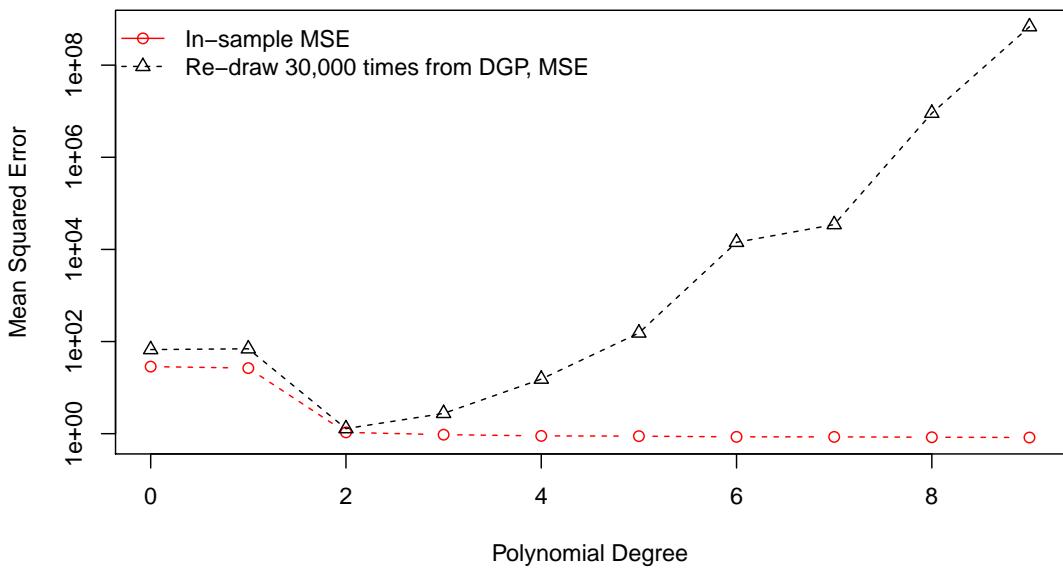


Figure 6: Mean Squared Error for 0^{th} through 9^{th} Degree Polynomial Regressions
Author's reproduction of Shalizi (2017).

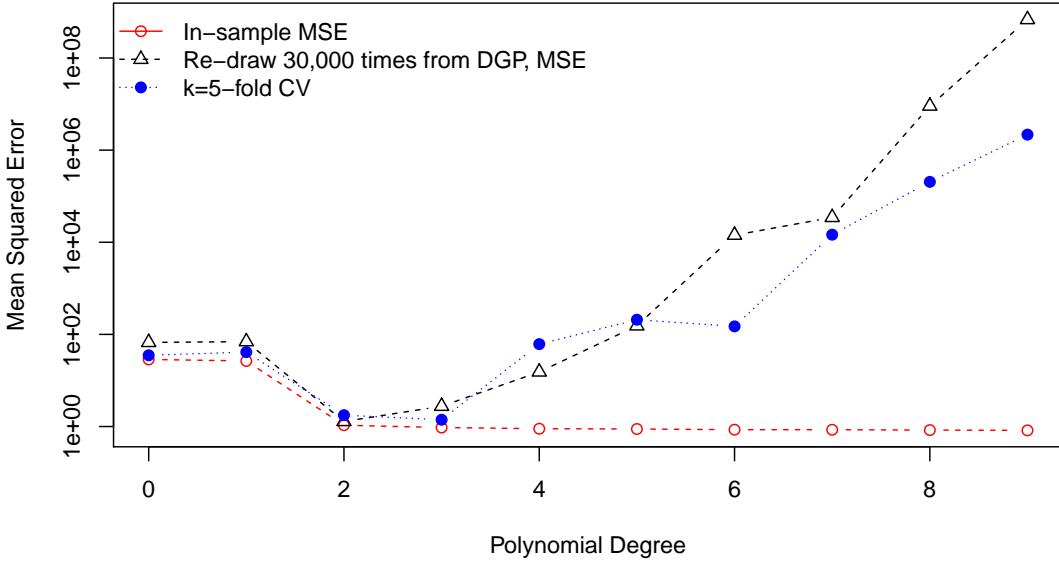


Figure 7: Mean Squared Error for 0^{th} through 9^{th} Degree Polynomial Regressions, with k=5-fold CV

Author's reproduction of Shalizi (2017).

The difficulty, of course, is that we cannot simply re-run the data generating process for the "real world" and generate new out-of-sample data to test the fit of models. However we can use resampling of the data we do have to approximate this effect. K-fold cross validation does just that. Here is a very simple description of how it is conducted; for mode details please see the references noted at the beginning of this section.

Choose a number $k = \{2, 3, \dots, 5, \dots, 10, \dots, N_{data}\}$ and divide the data randomly and evenly (as evenly as possible) into k subsets.¹⁶ For each subset $k = 1, 2, \dots, K$, set aside that subset as the "pseudo out-of-sample" testing data, and poll the remaining $k - 1$ datasets and fit model on this pooled dataset. Then take the testing dataset and measure the goodness of fit of the estimated model against this dataset. Save the result, and repeat the process for all k data subsets. The result will be k goodness of fit scores. Take the average and this is the k-fold cross validation (CV) score for this model. Repeat this process for all models you want to compare, being careful to always use the same sub-setting of the data for the entire process. That is, do not re-divide the data randomly into k subsets for each model, but retain the data division for all models.

The results of conducting this with $k = 5$ for our polynomial regression example is displayed in Figure (7). Impressively, the $k = 5$ -fold CV tracks relatively closely to the out-of-sample scoring accomplished by genuinely re-drawing from the DGP.

¹⁶Common values used in practice are $k = 5, 10$, and $k = N_{data}$. The last choice in that list leads to "leave-on-out" cross validation, which is theoretically equivalent to Akaike information criterion for a certain class of models and as the data goes to infinity. See Claeskens et al. (2008) for more details.

6.1 K-Fold Cross Validation and Semi-parametric Structural Micro-estimation

This paper conducts the exact same analysis, but using the estimation procedure described in Section 4. Recall the list of models described at the end of Section 5, repeated here for convenience:

- Model I: 1 type of optimizing agent: estimate 2 parameters: $\{\beta, \rho\}$
- Model II: 2 types of optimizing agent: jointly estimate 4 parameters: $\{\beta_1, \beta_2, x_1, \rho\}$
- Model III: 3 types of optimizing agent: estimate 6 parameters: $\{\beta_1, \beta_2, \beta_3, x_1, x_2, \rho\}$
- Model IV: 4 types of optimizing agent: estimate 8 parameters: $\{\beta_1, \beta_2, \beta_3, \beta_4, x_1, x_2, x_3, \rho\}$
- Model V: 5 types of optimizing agent: estimate 10 parameters: $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, x_1, x_2, x_3, x_4, \rho\}$
- Model VI: 6 types of optimizing agent: estimate 12 parameters: $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, x_1, x_2, x_3, x_4, x_5, \rho\}$

To run the simulation-based estimation and selection, I used a population of $N_{agents} = 10,000$ and selected $k = 5$ as the cross validation number of subsets.¹⁷ I then did the following:

1. Randomly divide the SCF wealth data into five subsets. Do this such that 1/5 of each of the seven age distributions is assigned to each of the subsets. Denote these as $k \in \{1, 2, 3, 4, 5\}$.
2. For each Model I, II, ..., VI listed above, do the following:
 - (a) For each $k \in \{1, 2, 3, 4, 5\}$, retain the k^{th} subset as the "test" set and pool the remaining subsets by age as the "estimation" dataset, denoted $-k$.
 - (b) Run the estimation described in Section 4 on this pooled $-k$ dataset.
 - (c) Using the best fit obtained from estimating the model on the $-k$ data, calculate the quadratic objective score expressed in Equation (4). This will produce 5 objective scores for each model.
 - (d) Take the average of the five objective scores, producing a single CV score for each model.

The preliminary results of this process are described in the next section.

7 Results

Preliminary results indicate that the k-fold cross validation process strongly selected away from Model I, the model in which there is only 1 type of agent. In addition, Models II-VI, the models with multiple types of agents, all displayed relatively large fractions of the population with β values below typical discount factors at an annual frequency for this type of model of roughly 0.9 – 0.99.

The results of the selection process are split across two tables, Tables 1 and 2. The two header rows of each table display the Model number and the corresponding number of types of agent, N_β . The next three rows on each table, labeled ρ , β , and $frac$, show the estimation results for each model. As defined in Section 3, ρ is risk aversion for all agents, β is the set of discount factors estimated for all agents, and $frac$ is the fraction of the population estimated to be associated with each discount factor for each model. Thus Model I

¹⁷ $k=5$ is chosen due to the extensive computational time required to execute this estimation process.

Table 1: Estimation and Selection Results, $N_\beta \in \{1, 2, 3, 4\}$

Model	I	II	III	IV
N_β	1	2	3	4
ρ	1.65	4.65	4.94	4.24
β	1.01	0.25, 1.04	0.29, 0.99, 1.05	0.01, 0.41, 0.81, 1.04
$frac$	n.a.	0.46, 0.54	0.26, 0.35, 0.38	0.17, 0.19, 0.09, 0.54
CV score	12.0	0.83	0.82	0.81
CV score stdev	3.1	0.77	0.78	0.74

Table 2: Estimation and Selection Results, $N_\beta \in \{5, 6\}$

Model	V	VI
N_β	5	6
ρ	4.70	4.74
β	0.00, 0.01, 0.52, 0.79, 1.04	0.11, 0.16, 0.24, 0.28, 0.45, 1.04
$frac$	0.09, 0.07, 0.15, 0.17, 0.53	0.03, 0.17, 0.09, 0.08, 0.1, 0.54
CV score	0.81	0.84
CV score stdev	0.73	0.78

has a risk aversion parameter of $\rho = 1.65$ and discount factor estimate of $\beta = 1.01$.¹⁸ By comparison, Model II has risk aversion of $\rho = 4.65$, and two discount factor estimates of $\beta_1, \beta_2 = \{0.25, 1.04\}$, with population fractions of $frac = \{0.46, 0.54\}$, respectively. Nearly 50% of agents in Model II have a discount factor of 0.25. We will return to this again in further discussion below.

The final two rows in Tables 1 and 2 display the k-fold cross validation scores for each model, as well as the standard deviation in each score for each model. It is immediately clear that the CV scores for Models II-VI are two orders of magnitude lower than the CV score for Model I. This is plotted in Figure 1, along with +/- 1 standard deviation.

Figure 9 focuses on Models II-VI alone, and at first observation appears to display the same pattern as Figure 7. However when we observe these scores with +/- 1 standard deviation, as displayed in Figure 8, it is unclear that we can appropriately distinguish these results. Further examination is required: selection under different values of k may be in order, or bootstrapping these estimates itself may be in order.

What is clear, however, is that the model with only one type of agent is strongly selected against using k-fold cross validation. To visualize what this means, Figure 10 displays the consumption function of the agents from Model I and Figures 11 and 12 plot the "high β " and "low β " consumption functions that arise in Model II.

A little explanation is in order. These functions map cash on hand, on the x-axis, to spending on consumption goods on the y-axis.¹⁹ Conversely, these also represent the behavior of these agents with respect to savings – cash on hand minus this value would give us the agent's savings. Thus for very "flat" consumption functions, such as all the black consumption functions in Figures 10 and 11, a large change in cash on hand along the x-axis will not change consumption or savings much – their consumption functions

¹⁸As discussed in Carroll (2012b), the calibration of the microeconomic structural model incorporates the probability of death for each period. This probability is not represented in the β estimated here, and thus this estimated discount factor is slightly higher than 0.99. Please see Carroll (2012b) and Carroll (2012c) for further discussion.

¹⁹These values are normalize by permanent income, hence the units.

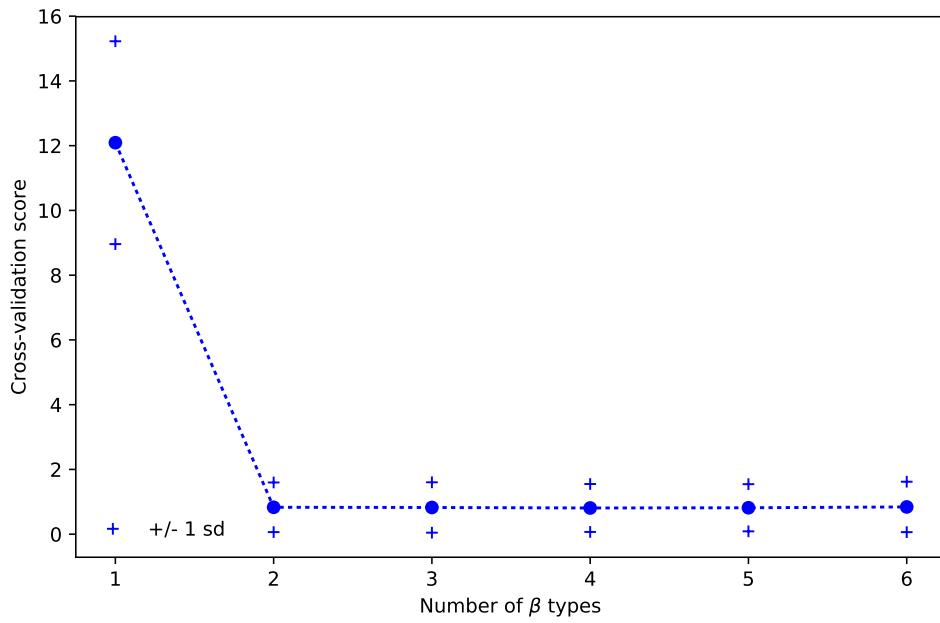


Figure 8: Mean Cross Validation Score per Model, with $+/- 1$ Stdev, for k=5-fold CV

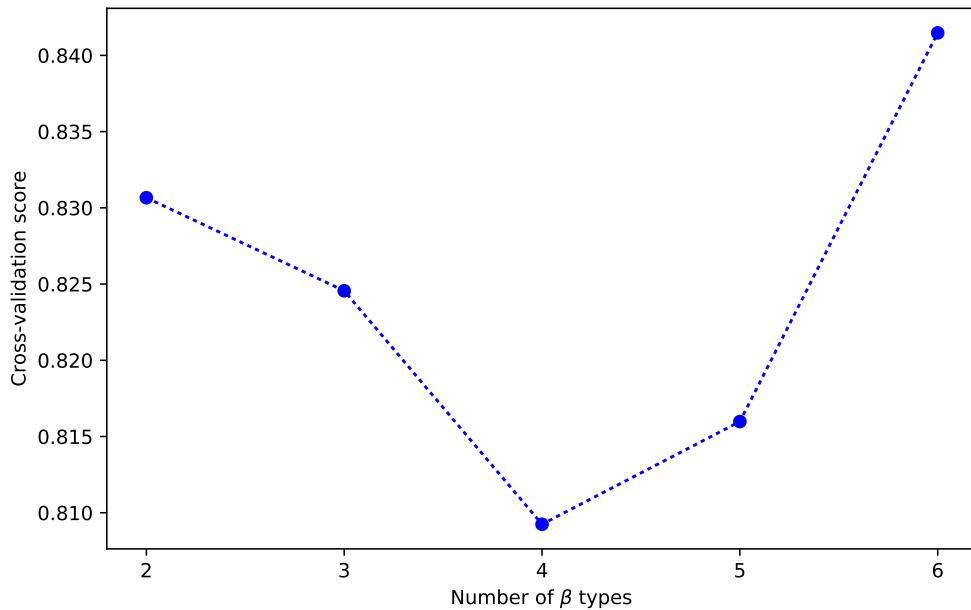


Figure 9: Mean Cross Validation Score, Models II-VI, for k=5-fold CV

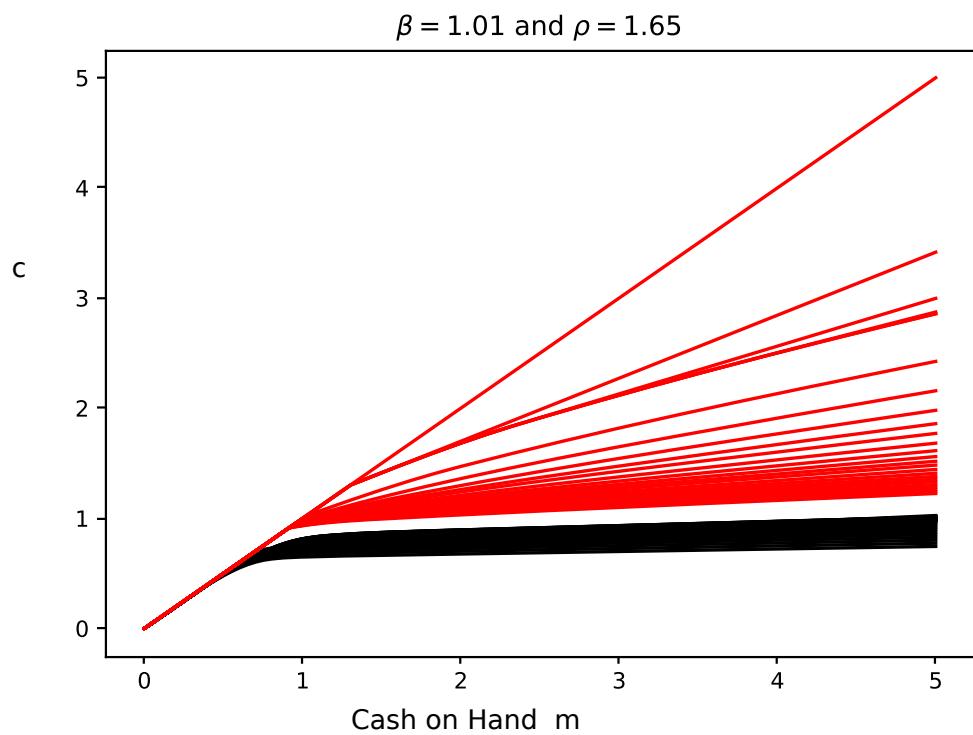


Figure 10: Consumption Function, Model I
 Black consumption functions are behavior before retirement at age 65. Red consumption functions are consumption after retirement. The red 45° line is consumption in the final period of life.

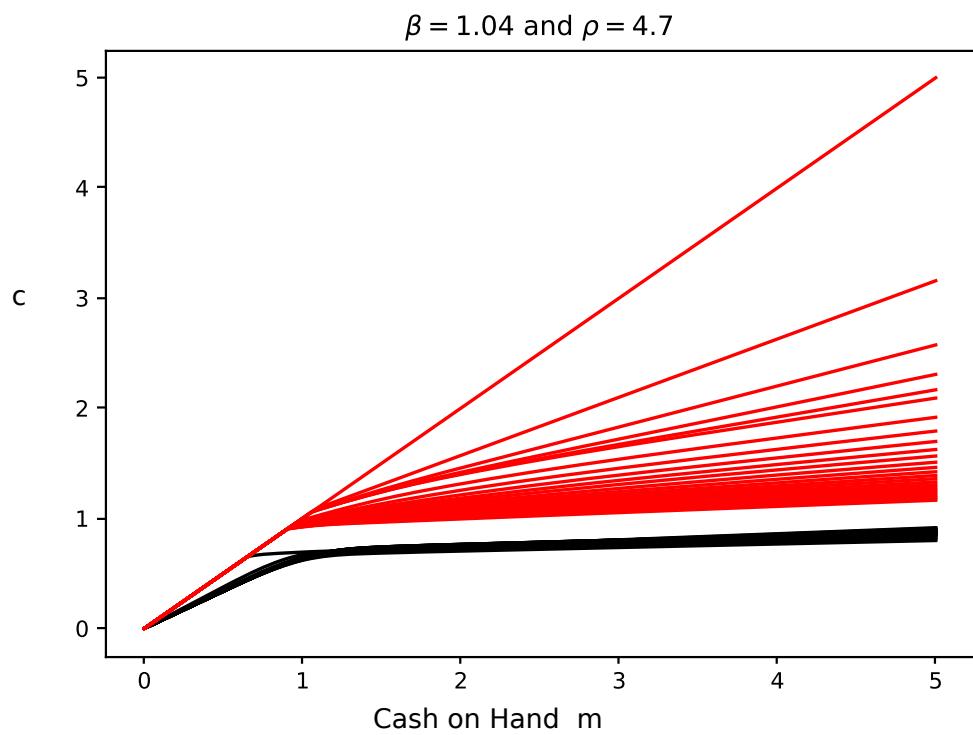


Figure 11: Consumption Function, Model II, β_{high}
 Black consumption functions are behavior before retirement at age 65. Red consumption functions are consumption after retirement. The red 45° line is consumption in the final period of life.

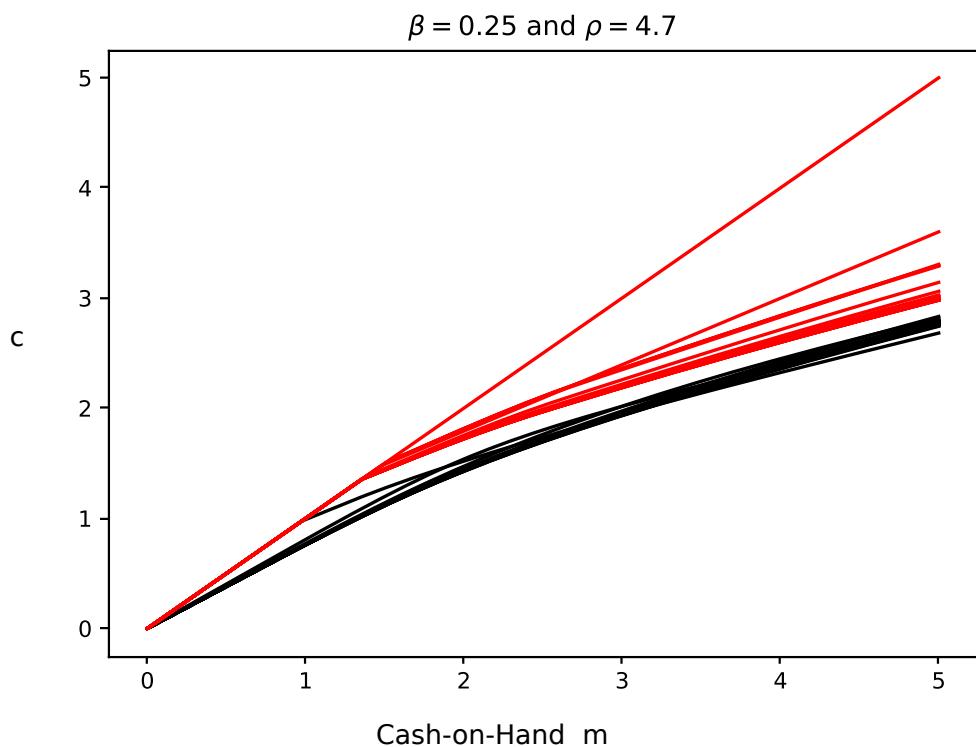


Figure 12: Consumption Function, Model II, β_{low}
 Black consumption functions are behavior before retirement at age 65. Red consumption functions are consumption after retirement. The red 45° line is consumption in the final period of life.

are quite flat for much of the relevant region.²⁰ Conversely, for all consumption functions in Figure 12, a large shock to wealth will correspond to a large change in consumption and savings behavior.

In these plots, the black consumption functions are behavior before retirement at age 65, and red consumption functions represent consumption after retirement. The 45° line is consumption in the final period of life. The consumption function associated with the youngest age is for age 25 and is the black consumption function closest to the x-axis. Consumption functions monotonically increase until the end of life at the 45° line, which occurs at age 90 in this model. Thus these plots show the entire lifetime of consumption functions for an agent who lives from 25 to 90 and follows the optimal lifetime consumption-savings behavior consistent with the stated ρ, β parameters.

The consumption functions for Model I are quite similar to the "high β " consumption functions in Model II. The "low β " consumption functions are qualitatively different from both previous consumption functions. More important is the fraction of agents who possess the "low β " consumption functions in Model II: 46% of the population of Model II possess these low- β consumption functions, and thus will react sharply to shocks to their wealth. The estimation process puts their fraction at nearly 50% of the population. This is in line with Campbell and Mankiw (1989, 1990). In addition, this is stylistically in line with the story Mian and Sufi (2015) tell regarding why the 2007 Financial Crisis was so severe.

Thus far we have discussed the relationship between Model I and Model II because this is the easiest story to communicate. The relationship between Model I and Models III - VI are similar, but much more difficult to visualize and communicate. Please see the Appendix for more discussion and visualization of the fraction of populations that are represented in Models II through VI.²¹

8 Summary and Conclusion

There are two primary contributions of this paper: first, I demonstrate the application of a formal model selection approach to a semi-parametric, structural econometric model of household behavior. The ability to conduct such selection is of critical importance when constructing agent-based models of macro-financial systems, such as Geanakoplos et al. (2012), the CRISIS project under Doyne Farmer²², the EURACE project as described in Deissenberg et al. (2008), Fischer and Riedler (2014), Henry et al. (2013), and many models cited in the excellent survey Turrell (2016). The flexibility of the approach I use may be used for a wide range of agent behaviors, without requiring detailed (and often difficult to obtain) microeconomic data.

Second, this paper provides structural micro-econometric evidence in line with the conclusions of Campbell and Mankiw (1989, 1990). Using aggregate data and aggregate methods, they find that roughly 50% of the population follows something like a "consume everything" rule of thumb. Using microeconomic data and micro-econometric methods, I directly select between a model of optimizing households and a model in which agents may differ, and I allow the data to speak for itself regarding the degree to which households may be "consume everything" rule of thumb agents. I find evidence for a minimum of two types of agent behavior: traditional-looking lifetime optimization behavior, and behavior which looks quite close to a "consume everything" rule of thumb. To my knowledge, this is the first paper to apply formal model selection

²⁰Most agent experience, about 90%, will be spent roughly between 0.5 and 2.5 on the x-axis.

²¹Note that Models IV and V actually estimate a non-trivial portion of the population with β very close to 0.0.

²²See <https://github.com/crisis-economics/CRISIS> for more details.

techniques to structural estimation of microeconomic dynamic stochastic optimization behavior in this life cycle consumption-saving context.

This paper demonstrates that selection between competing models of consumption-savings behavior can be accomplished using k-fold cross validation. There is much more work to be done. These include more detailed application of the same process to subsets of the microeconomic data, as well as to alternative data sources. A key next step is applying the process again to a number of learning models of agent behavior, applied to the same problem. A further future step is formally integrating such agent behavior into large-scale agent-based models for macroprudential and macro-financial monitoring and analysis.

References

- Allen, T. W. and Carroll, C. D. (2001). Individual learning about consumption. *Macroeconomic Dynamics*, 5(02):255–271.
- Arifovic, J. and Ledyard, J. (2012). Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism. *Journal of Public Economics*, 96(9):808–823.
- Arlot, S., Celisse, A., et al. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79.
- Banerjee, A. V. and Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs.
- Başçı, E. and Orhan, M. (2000). Reinforcement learning and dynamic optimization. *Journal of Economic and Social Research*, 2(1):39–57.
- Birgin, E. G. and Martínez, J. M. (2008). Improving ultimate convergence of an augmented lagrangian method. *Optimization Methods and Software*, 23(2):177–195.
- Box, M. (1965). A new method of constrained optimization and a comparison with other methods. *The Computer Journal*, 8(1):42–52.
- Brent, R. P. (2013). *Algorithms for minimization without derivatives*. Dover Publications.
- Cagetti, M. (2003). Wealth accumulation over the life cycle and precautionary savings. *Journal of Business & Economic Statistics*, 21(3):339–353.
- Campbell, J. Y. and Mankiw, N. G. (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER macroeconomics annual*, 4:185–216.
- Campbell, J. Y. and Mankiw, N. G. (1990). Permanent income, current income, and consumption. *Journal of Business & Economic Statistics*, 8(3):265–279.
- 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).
- Carroll, C. D. (2002). *Portfolios of the Rich*. Household Portfolios: Theory and Evidence. Mit Press.

- Carroll, C. D. (2012a). Implications of wealth heterogeneity for macroeconomics.
- Carroll, C. D. (2012b). Solving microeconomic dynamic stochastic optimization problems. *Lecture Notes, The Johns Hopkins University, Department of Economics*.
- Carroll, C. D. (2012c). Theoretical foundations of buffer stock saving. *Mimeo, The Johns Hopkins University, Department of Economics*.
- Carroll, C. D., Kaufman, A. M., Kazil, J. L., Palmer, N. M., and White, M. N. (2018). The econ-ark and hark: Open source tools for computational economics. *SciPy 2018, 17th Python in Science Conference, Austin, Texas, July 9-15, 2018*.
- Carroll, C. D., Slacalek, J., and Sommer, M. (2011). International evidence on sticky consumption growth. *Review of Economics and Statistics*, 93(4):1135–1145.
- Carroll, C. D., Slacalek, J., Tokuoka, K., and White, M. N. (2017). The distribution of wealth and the marginal propensity to consume. *Quantitative Economics*.
- Chan-Lau, J. A. (2017). Abba: An agent-based model of the banking system. Technical report, International Monetary Fund.
- Claeskens, G., Hjort, N. L., et al. (2008). *Model selection and model averaging*, volume 330. Cambridge University Press.
- Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.
- Cochrane, J. H. (2017). Macro-finance. *Review of Finance*, 21(3):945–985.
- Conn, A. R., Gould, N. I., and Toint, P. (1991). A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28(2):545–572.
- Cotla, C. R. (2016). *Social Preferences, Learning, and the Dynamics of Cooperation in Networked Societies: A Dialogue Between Experimental and Computational Approaches*. PhD thesis, Department of Computational Social Sciences, George Mason University.
- Deissenberg, C., Van Der Hoog, S., and Dawid, H. (2008). Eurace: A massively parallel agent-based model of the european economy. *Applied Mathematics and Computation*, 204(2):541–552.
- Dominicy, Y. and Veredas, D. (2013). The method of simulated quantiles. *Journal of Econometrics*, 172(2):235–247.
- Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Evans, G. W. and McGough, B. (2015). Learning to optimize. *Mimeo, The University of Oregon, Department of Economics*.
- Fischer, T. and Riedler, J. (2014). Prices, debt and market structure in an agent-based model of the financial market. *Journal of Economic Dynamics and Control*, 48:95–120.

- Friedman, J., Hastie, T., and Tibshirani, R. (2009). *The elements of statistical learning*. Springer.
- Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly journal of economics*, 127(2):645–700.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.
- Gabaix, X. and Laibson, D. (2017). Myopia and discounting. Technical report, National Bureau of Economic Research.
- Ganong, P. and Noel, P. (2017). Consumer spending during unemployment: Positive and normative implications. *Working Paper. The University of Chicago Booth School of Business*.
- Geanakoplos, J., Axtell, R., Farmer, D. J., Howitt, P., Conlee, B., Goldstein, J., Hendrey, M., Palmer, N. M., and Yang, C.-Y. (2012). Getting at systemic risk via an agent-based model of the housing market. *The American Economic Review*, 102(3):53–58.
- Glaeser, E. L. (2016). Real estate bubbles and urban development.
- Gourieroux, C. and Monfort, A. (1994). Testing non-nested hypotheses. *Handbook of econometrics*, 4:2583–2637.
- Gourio, F. (2012). Disaster risk and business cycles. *The American Economic Review*, 102(6):2734–2766.
- He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2):732–70.
- Henry, J., Kok, C., Amzallag, A., Baudino, P., Cabral, I., Grodzicki, M., Gross, M., Halaj, G., Kolb, M., Leber, M., et al. (2013). A macro stress testing framework for assessing systemic risks in the banking sector.
- Houser, D., Keane, M., and McCabe, K. (2004). Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. *Econometrica*, 72(3):781–822.
- Howitt, P. and Özak, Ö. (2014). Adaptive consumption behavior. *Journal of Economic Dynamics and Control*, 39:37–61.
- Kiley, M. T. (2016). Macroeconomic modeling of financial frictions for macroprudential policymaking: A review of pressing challenges. Technical report, Board of Governors of the Federal Reserve System (US).
- Krusell, P. and Smith, A. A. (1996). Rules of thumb in macroeconomic equilibrium: A quantitative analysis. *Journal of Economic Dynamics and Control*, 20(4):527–558.
- Lane, P. R. (2001). The new open economy macroeconomics: a survey. *Journal of international economics*, 54(2):235–266.
- Leamer, E. E. (2015). Housing really is the business cycle: what survives the lessons of 2008–09? *Journal of Money, Credit and Banking*, 47(S1):43–50.

- Leamer, E. E. et al. (2007). Housing IS the business cycle. In *Proceedings-Economic Policy Symposium-Jackson Hole*, pages 149–233. Federal Reserve Bank of Kansas City.
- Lettau, M. and Uhlig, H. (1999). Rules of thumb versus dynamic programming. *American Economic Review*, 89(1):148–174.
- Li, T. (2009). Simulation based selection of competing structural econometric models. *Journal of Econometrics*, 148(2):114–123.
- Mian, A. and Sufi, A. (2015). *House of debt: How they (and you) caused the Great Recession, and how we can prevent it from happening again*. University of Chicago Press.
- Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313.
- Obstfeld, M., Rogoff, K. S., et al. (1996). *Foundations of international macroeconomics*, volume 30. MIT press Cambridge, MA.
- Özak, Ö. (2014). Optimal consumption under uncertainty, liquidity constraints, and bounded rationality. *Journal of Economic Dynamics and Control*, 39:237–254.
- Palmer, N. M. (2012). Learning to consume: Individual vs. social learning. *Working Paper. Presented at Computing in Economics and Finance Conference, Society of Computational Economics 2012*.
- Palmer, N. M. (2016). Learning to dynamically optimize from individual experience. *Working Paper. Presented at Computing in Economics and Finance Conference, Society of Computational Economics 2016*.
- Powell, M. J. (2006). The NEWUOA software for unconstrained optimization without derivatives. In *Large-scale nonlinear optimization*, pages 255–297. Springer.
- Powell, M. J. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*.
- Richardson, J. A. and Kuester, J. (1973). Algorithm 454: the complex method for constrained optimization [e4]. *Communications of the ACM*, 16(8):487–489.
- Rowan, T. (1990). *Functional Stability Analysis of Numerical Algorithms*. PhD thesis, Department of Computer Sciences, University of Texas at Austin.
- Sargent, T. and Stachurski, J. (2017). Quantitative economics. URL <http://quant-econ.net>.
- Shalizi, C. (2017). Advanced data analysis from an elementary point of view.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.
- Sinitskaya, E. and Tesfatsion, L. (2014). Macroeconomies as constructively rational games. *Mimeo, Iowa State University, Department of Economics*.

- Tesfatsion, L. and Judd, K. L. (2006). *Handbook of computational economics: Agent-based computational economics*, volume 2. Elsevier.
- Turrell, A. (2016). Agent-based models: understanding the economy from the bottom up. *Bank of England Quarterly Bulletin*, 56(4):173.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333.
- Yellen, J. L. (2016). Macroeconomic research after the crisis. Speech 915, Board of Governors of the Federal Reserve System.
- Yıldızoglu, M., Sénégas, M.-A., Salle, I., and Zumpe, M. (2014). Learning the optimal buffer-stock consumption rule of carroll. *Macroeconomic Dynamics*, 18(04):727–752.

Appendix

Each of Models II-VI has a significant portion of the population with β below typical discounting values at an annual frequency, for example 0.9 – 0.99. This is difficult to visualize in Tables 1 and 2.

Here in the Appendix we visualize the above statement directly. The following plots, Figure 13 through Figure 17, display two values for each Model II - VI: first, the fraction of the population which has a β which falls below the β_{cutoff} chosen for that chart, and second, the population-weighted average β for the population which falls below that β_{cutoff} . This collapses some of the information in Tables 1 and 2 into a more easily digestible format. The β_{cutoff} values used are 0.9, 0.67, 0.5, 0.33, and 0.1, respectively, for Figure 13 through Figure 17.

For a cutoff of $\beta_{cutoff} = 0.9$, Figure 13 shows that 40%-50% of the population in Model II and Models IV-VI fall below that cutoff, while Model III has about 25% of its population falling below $\beta_{cutoff} = 0.9$. For all models, the population-weighted β is from 0.25-0.45. This is very low for a large fraction of the population. Most agents represented in this plot will have consumption functions similar to Figure 12, displaying very high reaction to changes in cash-on-hand. The fact that nearly all of these models display roughly half of the population having β values this low is quite significant for determining sensitivity to shocks to household balance sheets.

Figures 14 through 17 display similar stories, however as the β_{cutoff} value is lowered, mechanically all values in the charts must be pushed lower monotonically. Interestingly enough, even for $\beta_{cutoff} < 0.1$, displayed in the final Figure 17, two models have non-trivial fractions of the population with essentially $\beta = 0$ discount factors.

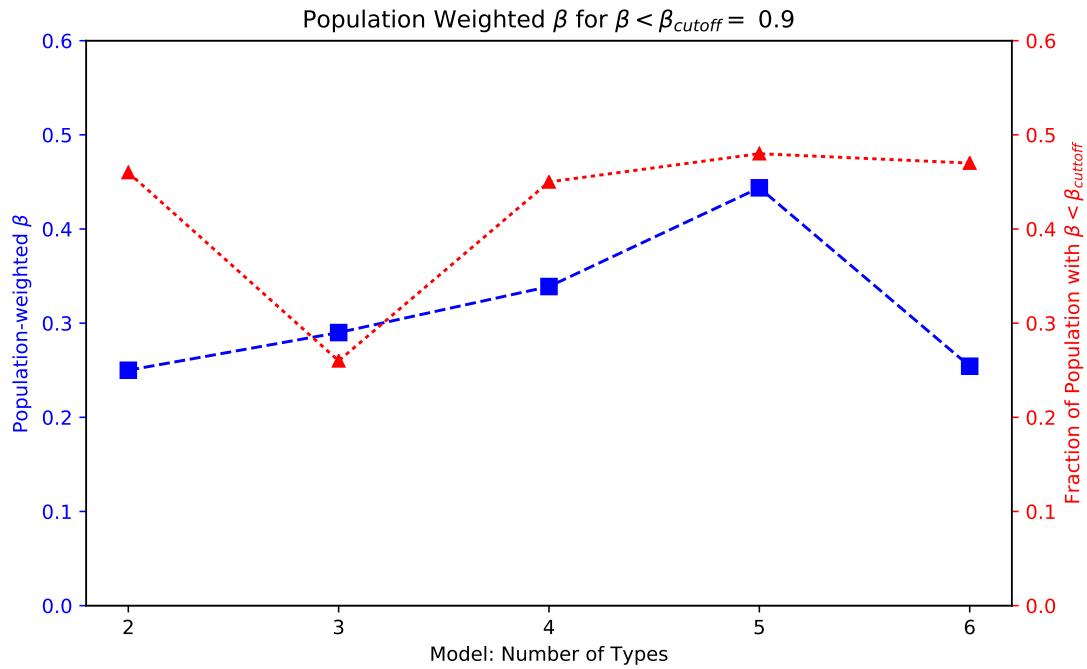


Figure 13: Population-Weighted β and Fraction of Population $< \beta_{cutoff} = 0.9$

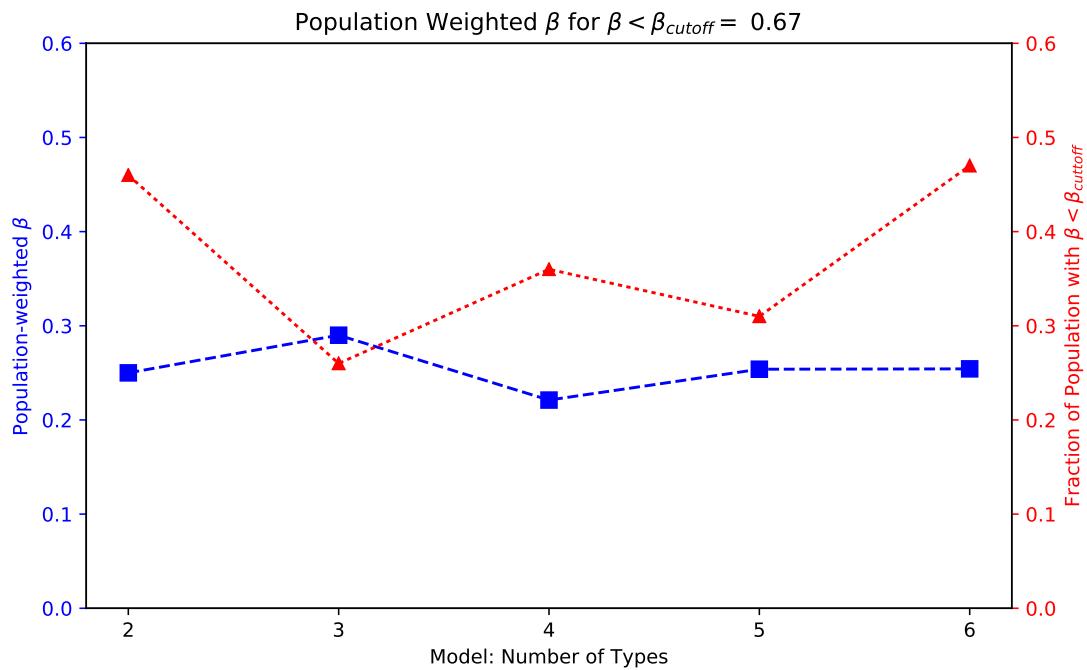


Figure 14: Population-Weighted β and Fraction of Population $< \beta_{cutoff} = 0.67$

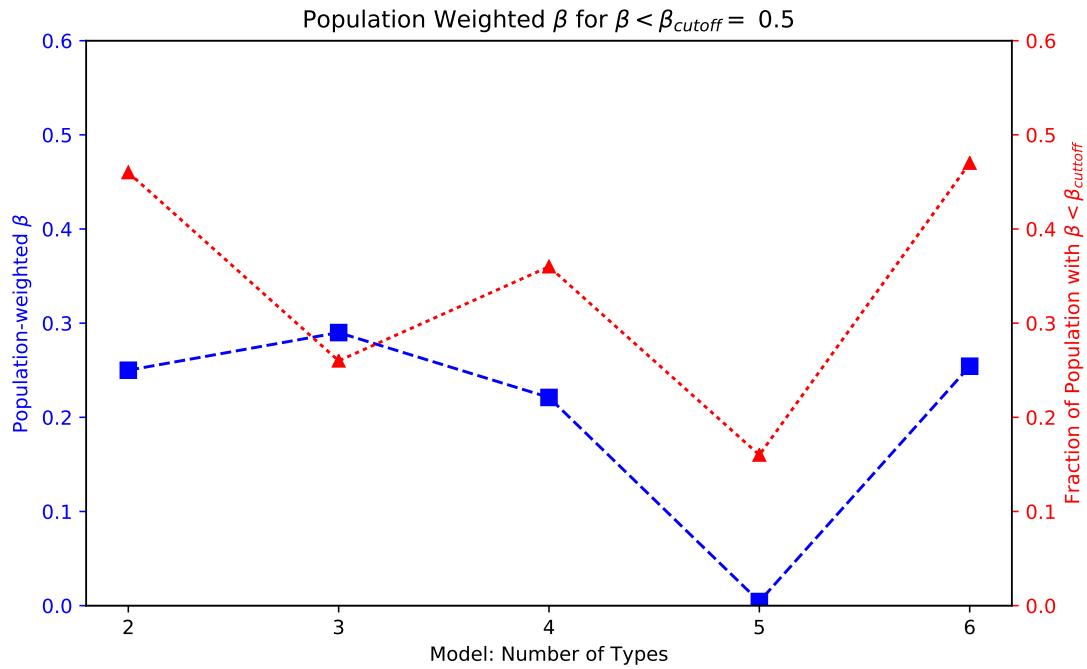


Figure 15: Population-Weighted β and Fraction of Population $< \beta_{cutoff} = 0.5$

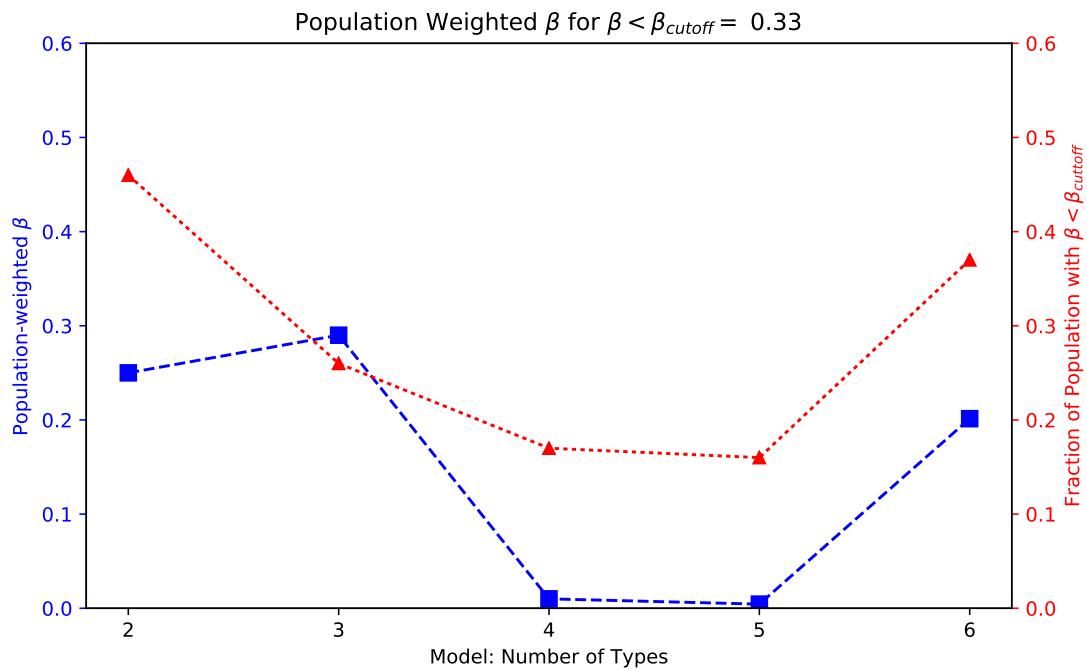


Figure 16: Population-Weighted β and Fraction of Population $< \beta_{cutoff} = 0.33$

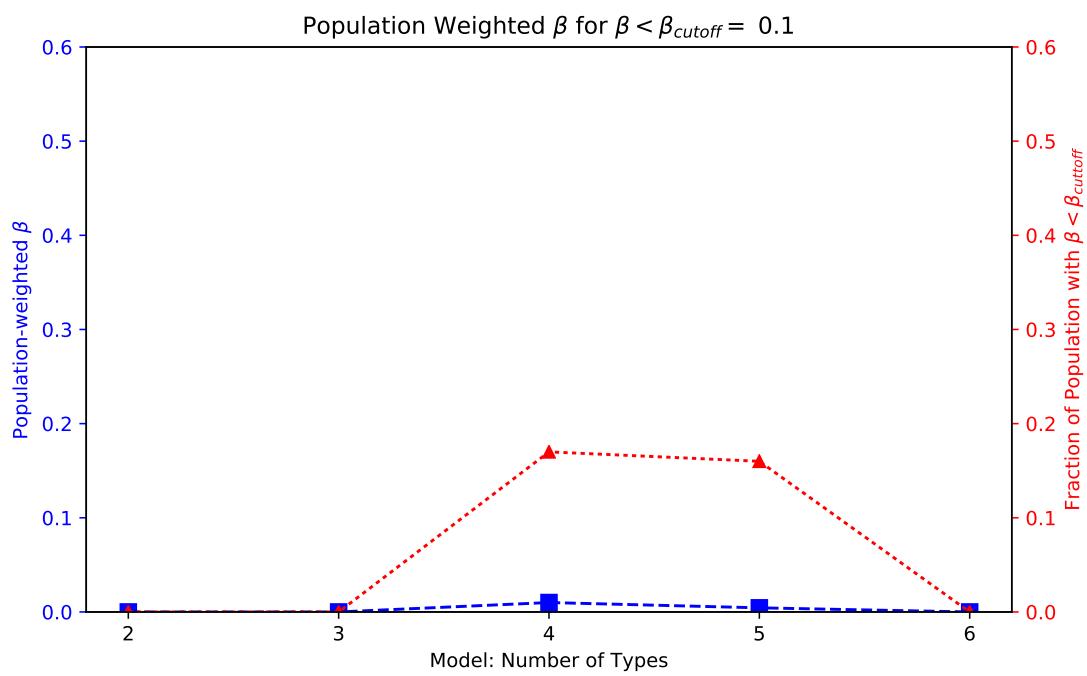


Figure 17: Population-Weighted β and Fraction of Population $< \beta_{cutoff} = 0.1$