Testing Consumption Optimality using Aggregate Data

Fábio Augusto Reis Gomes, João Vítor Issler

Maio de 2014

URL: http://hdl.handle.net/10438/11805
Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação Getulio Vargas.

ESCOLA DE PÓS-GRADUAÇÃO EM ECONOMIA
Diretor Geral: Rubens Penha Cysne
Vice-Diretor: Aloisio Araujo
Diretor de Ensino: Carlos Eugênio da Costa
Diretor de Pesquisa: Humberto Moreira
Vice-Diretores de Graduação: André Arruda Villela & Luís Henrique Bertolino Braido
Testing Consumption Optimality using Aggregate Data

Fábio Augusto Reis Gomes
Department of Economics, USP-RP
fabigomes@fearp.usp.com.br

João Victor Issler†
Graduate School of Economics – EPGE
Getulio Vargas Foundation
Praia de Botafogo 190 s. 1100
Rio de Janeiro, RJ 22250-900
Brazil
jissler@fgv.br

May, 2014.

*We are especially grateful to Caio Almeida, Marco Bonomo, Luis Braido, Carlos E. Costa, Russell Davidson, Pedro C. Ferreira, Karolina Goraus, Anwar Khyat, Naércio A. Menezes Filho, for their comments and suggestions of improvement on earlier versions of this paper. We also benefited from comments given by the participants of ISCEF conferences in Paris, 2014, where this paper was presented. The usual disclaimer applies. Fabio Augusto Reis Gomes and João Victor Issler gratefully acknowledge support given by CNPq-Brazil. Issler also acknowledges the support given by CAPES, Pronex, FAPERJ, and INCT. We gratefully acknowledge research assistance given by Rafael Burjack, Marcia Valeria Machado and Marcia Marcos.

†Corresponding Author.
Abstract

The objective of this paper is to test for optimality of consumption decisions at the aggregate level (representative consumer) taking into account popular deviations from the canonical CRRA utility model – rule of thumb and habit. First, we show that rule-of-thumb behavior in consumption is observational equivalent to behavior obtained by the optimizing model of King, Plosser and Rebelo (Journal of Monetary Economics, 1988), casting doubt on how reliable standard rule-of-thumb tests are. Second, although Carroll (2001) and Weber (2002) have criticized the linearization and testing of euler equations for consumption, we provide a deeper critique directly applicable to current rule-of-thumb tests. Third, we show that there is no reason why return aggregation cannot be performed in the nonlinear setting of the Asset-Pricing Equation, since the latter is a linear function of individual returns. Fourth, aggregation of the nonlinear euler equation forms the basis of a novel test of deviations from the canonical CRRA model of consumption in the presence of rule-of-thumb and habit behavior.

We estimated 48 euler equations using GMM, with encouraging results vis-a-vis the optimality of consumption decisions. At the 5% level, we only rejected optimality twice out of 48 times. Empirical-test results show that we can still rely on the canonical CRRA model so prevalent in macroeconomics: out of 24 regressions, we found the rule-of-thumb parameter $\lambda$ to be statistically significant at the 5% level only twice, and the habit parameter $\gamma$ to be statistically significant on four occasions.

The main message of this paper is that proper return aggregation is critical to study intertemporal substitution in a representative-agent framework. In this case, we find little evidence of lack of optimality in consumption decisions, and deviations of the CRRA utility model along the lines of rule-of-thumb behavior and habit in preferences represent the exception, not the rule.

Keywords: Consumption, Intertemporal Substitution, Risk Aversion, Aggregate Return, Rule of Thumb Behavior.

JEL Classification: C22; D91; E21.
1 Introduction

For the U.S. economy, there has been a large early literature using time-series data rejecting optimizing behavior in consumption which generated some relevant puzzles; see Hall (1978), Flavin (1981), Hansen and Singleton (1982, 1983, 1984), Mehra and Prescott (1985), Mark (1985), Campbell and Deaton (1989), and Weil (1989). Most of these studies employed the constant-relative-risk-aversion (CRRA) utility function with exponential discounting of future utility in defining welfare. These rejections have led to two different strands of the consumption literature. The first investigated whether changing preferences could accommodate optimizing behavior; see Abel (1990) and Campbell and Cochrane (1999) for research on habit, and Epstein and Zin (1989, 1991) for research on non-expected utility. The second strand introduced explicit forms of non-optimizing behavior for consumption decisions, testing whether they fit the data. On that regard, the most influential study is that of Campbell and Mankiw (1989, 1990), who extended the basic optimizing model incorporating what they have labelled rule-of-thumb behavior: there are two types of consumers, the first type consumes according to optimizing behavior but the second consumes only his/her current income\(^1\). In this setup, changes in aggregate consumption respond solely to expected changes in aggregate income, and the response is a function of the importance of rule-of-thumb consumers.

In the context above, rejecting optimizing behavior using aggregate data (time series) is an important setback in macroeconomics, where it is commonly assumed an optimizing representative-consumer framework with a CRRA utility function. Moreover, this rejection has far-reaching implications: it raises the issue of whether or not we can postulate optimizing behavior in economics – if one cannot defend optimizing behavior at the aggregate level, one can question whether it is applicable at all.

This paper has three original contributions to the literature on consumption optimality at the aggregate level. Our setup starts by preserving the standard CRRA framework for the representative consumer, where the generalized method of moment (GMM) is used in estimation and testing. We employ an encompassing model that simultaneously allows for the existence of rule-of-thumb behavior and habit in preferences. These two departures from the standard model are tested as exclusion restrictions for specific parameters. Optimality, on the other hand, is tested using over-identifying-restriction tests. Below, we outline in detail these three contributions.

\(^1\)Campbell and Mankiw test no rule-of-thumb behavior using a first order log-linearized version of the euler equation for the optimizing agent. For the U.S. economy, they conclude that about 50% of total income belongs to rule-of-thumb consumers.
First, using the dynamic stochastic general equilibrium (DSGE) model of King, Plosser and Rebelo (1988), we show that there exists a linear combination of consumption and income growth which is unpredictable (Issler and Vahid (2001)). This is exactly what is being tested in the basic rule-of-thumb test. Observational equivalence implies that, even if one finds that income and consumption growth have similar short-run co-movement, one cannot conclude in favor of the presence of rule-of-thumb (and against optimality) as the previous literature has done. Despite that, we will still consider its possibility and test for the presence of rule-of-thumb behavior in a broader setting.

Second, employing a generalized taylor expansion of the optimizing consumer’s euler equation (Araujo and Issler (2011)), we consider what are the econometric consequences of ignoring high-order terms once a naïve first-order expansion is fitted to data. At least since Carroll (2001), its is well-known that ignoring higher-order terms yields inconsistent estimates of the respective euler equation, invalidating hypothesis testing. This happens because past observed values do not constitute valid instruments in this context, but these are exactly the instruments the previous literature has used. As shown below, this critique applies directly to linear or log-linear rule-of-thumb tests.

Third, we circumvent the problem of lack of instruments in linear equations by using a nonlinear setup for estimation and testing. Our approach has two main ingredients. The first is to exploit the nonlinearity of the euler equation of the optimizing agent, where, under rule of thumb, her/his consumption is a linear combination of consumption and income (Weber (2002)). The second is to aggregate returns in the euler equation for the optimizing agent. This is possible because the latter is linear on gross returns, although it is nonlinear on consumption and preference parameters.

Aggregating returns has several benefits: (i) from a theoretical point-of-view, we know that only pervasive variation of returns affects intertemporal substitution in consumption.\(^2\)

\(^2\)For the habit specification, it is also a function of a linear combination of lagged consumption and income.

\(^3\)The quote in Mulligan (2002) summarizes well why aggregating returns is a good strategy:

"If we were interested, say, in the willingness of consumers to substitute food for other goods, then we should look at the correlation between food expenditure and a food price index. This correlation would have little relation with the correlation between food expenditure and the price of carrots, unless there were a perfect correlation between the price of carrots and the price of all other foods. There may be theories of food demand implying that the price of carrots is always in the same proportion to other food prices, but if in fact there were something moving the price of carrots apart from the prices of other foods, then a price index for all foods is needed. By analogy, my paper compares consumption growth with the return on a large
Due to the law-of-large numbers, aggregation preserves pervasive variation of returns, throwing away idiosyncratic variation. This allows a representative-consumer interpretation of utility parameters, where aggregate consumption is matched with aggregate returns – not with individual returns – which usually do not capture the pervasive variation of returns; (ii) estimating euler equations for several assets requires knowledge of participation – what assets are used to transfer wealth across time in every period; see Vissing-Jørgensen (2002), and Attanasio, Banks, and Tanner (2002)\(^4\). Although this may be a problem for panel-data studies, participation at the aggregate level is readily available from financial markets, wealth surveys, and national accounts; (iii) standard GMM estimation employing a large number of returns is usually infeasible because the number of time periods is small vis-a-vis the number of assets. Return aggregation preserves the pervasive portion of return variation. On the other hand, if one focuses on a subset of returns in empirical tests, as it is commonly done in the literature, asset-return information is thrown away – which is sub-optimal.

Our empirical implementation for testing optimality in consumption decisions, rule-of-thumb behavior, and habit in preferences, requires the use of an aggregate return measure for the economy as a whole – what Mulligan (2002) labelled the *return to aggregate capital*. Here, we employ proxies of the return to aggregate capital in two different frequencies: the annual measures computed by Mulligan (2002) and Mulligan and Threinen (2010), and the quarterly measures computed by Mulligan and Threinen (2010). When these measures are used in estimation and testing, we provide unequivocal evidence from diagnostic tests that linearization or log-linearization of the representative-consumer euler equation is problematic. Beyond linearization, we provide strong evidence against rule-of-thumb behavior for U.S. consumers and against habits in consumer preferences. Our results are in sharp contrast to those in Campbell and Mankiw regarding rule of thumb and to those in Weber regarding habit. Indeed, we show that we can appropriately represent preferences for the U.S. representative consumer using a constant relative-risk-aversion (CRRA) utility function. The estimates of the annual discount factor and the relative risk-aversion coefficient are significant almost everywhere. Our estimates show that we can describe reasonably well the U.S. representative consumer with an annual discount rate of 0.95 and a relative-risk-aversion coefficient roughly between 1 and 2, depending on whether we employ consumption

\[\text{portfolio of capital assets, rather than the return on a particular asset.}\]

\(^4\)For instance, Vissing-Jørgensen (2002) and Gross and Souleles (2002) have split households in groups according to their asset holdings, reaching large estimates of the intertemporal elasticities of substitution in consumption - around 0.8 - when stock return was used for stockholders or credit card interest rate for credit card debtors.
of nondurables or consumption of nondurables and services in estimation.

Regarding the main objective of this paper – test consumption optimality at the aggregate level – we found very little evidence of rejection of optimality in consumption decisions when over-identifying-restriction tests are employed, although, on occasion, there were rejections. Our key result is that, once proper models and econometric techniques are applied to aggregate consumption, income, and aggregate return data, there is no reason to challenge optimizing behavior in consumption, as was the case with previous rule-of-thumb tests. We also show that augmented models for preferences such as consumption with habit formation are unnecessary to characterize intertemporal substitution. In that sense, our evidence reduces the fear that optimizing behavior is the exception, not the rule.

The paper proceeds as follow. Section 2 presents the consumption models, the linear and the nonlinear consumer euler equation, as well as the asset returns aggregation. Section 3 presents the econometric methodology. Section 4 presents the empirical results, and Section 5 concludes.

2 Consumption Models

2.1 The Standard Approach in Macroeconomics – CRRA Utility with Aggregate Data

The standard approach in macroeconomics consists of a single-good economy of identical consumers, whose utility functions are of the CRRA type:

$$u(C_t) = \frac{C_t^{1-\phi} - 1}{1 - \phi}$$

where $C_t$ is consumption in period $t$, and $\phi$ is the constant relative risk-aversion coefficient. Subject to a budget constraint and transversality conditions, consumers choose consumption and asset holdings to maximize the lifetime utility, given by:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(C_t)$$

where $\beta \in (0, 1)$ is the intertemporal discount factor, and the mathematical expectation operator $\mathbb{E}_t$ ($\cdot$) is formed conditional on information available to the consumer up to period $t$. The representative agent can transfer wealth from one period to the next by buying individual assets, indexed by $i, i = 1, 2, \cdots, N$, whose returns are defined as:

$$R_{i,t} = \frac{P_{i,t} + D_{i,t}}{P_{i,t-1}}.$$
where \( P_{i,t} \) and \( D_{i,t} \) are respectively its price and dividend. In this setup, the well-known nonlinear euler equation is given by:

\[
\mathbb{E}_t \left\{ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\phi} R_{i,t+1} \right\} = 1, \quad i = 1, 2, \ldots, N;
\]

(2)


### 2.2 Testing Consumption Rule of Thumb in a Linear Framework

If one assumes a CRRA utility function, arriving at the euler equation (2), further assuming joint conditional log-Normality and homoskedasticity of \( \left( \frac{C_{t+1}}{C_t}, R_{1,t}, R_{2,t}, \ldots, R_{N,t} \right)^T \), the usual time series log-linear representation of consumption growth rate is obtained:

\[
\Delta \ln C_t = \alpha + \frac{1}{\phi} r_{i,t} + \mu_{i,t},
\]

(3)

where \( r_{i,t} \equiv \ln R_{i,t} \), \( \alpha \equiv \frac{(\ln \beta + \frac{1}{2} \sigma^2)}{\phi} \), and \( \sigma^2 = VAR[\Delta \ln C_t - \frac{1}{\phi} r_{i,t}] \). The the error term \( \mu_{i,t} \) is unpredictable, since it is an innovation regarding the optimizing agent’s information set. The coefficient of the rate of return, \( 1/\phi \), is the elasticity of intertemporal substitution (EIS), being the reciprocal of the constant relative risk-aversion coefficient.

The conditions under which equation (3) is derived are very stringent: \( \mu_{i,t} | \Omega_{t-1} \sim N \left( 0, \sigma^2_{\mu_i} \right) \) for all \( i \), with \( \Omega_{t-1} \) representing the information set of the optimizing agent. The fact that \( \mu_{i,t} \) is conditionally Gaussian and uncorrelated with elements of the conditioning set \( \Omega_{t-1} \) implies that \( \mu_{i,t} \) and \( \mu_{i,t-s}, s > 0 \) are independent. Moreover, \( \mu_{i,t} \) must be independent of any function of the variables in \( \Omega_{t-1} \). In principle, residual-based tests of normality, conditional homoskedasticity, and serial correlation can be used to ensure that these restrictions apply to \( \mu_{i,t} \).

Campbell and Mankiw (1989, 1990) proposed rule-of-thumb behavior for consumers at the aggregate level. The whole setup is either linear or log-linear. There are two types of consumers: type 1 consumes according to optimizing behavior. Type 2, on the other hand, is restricted to consume her/his current income (\( y_{2,t} \)). Income of the non-optimizing agent holds a fixed proportion (\( \lambda \)) to aggregate income as follows, \( \lambda = \frac{y_{2,t}}{y_t} \), leading to \( C_{2,t} = y_{2,t} = \lambda y_t \), where \( C_{2,t} \) is agent 2 consumption.

For the optimizing agent, they considered two different benchmark cases. The first imposes Hall’s (1978) quadratic utility setup, where consumption of the optimizing agent follows a martingale process, i.e.,

\[
\mathbb{E}_t (\Delta C_{1,t+1}) = 0,
\]
where $C_{1,t}$ is type 1 consumption. The second benchmark imposes CRRA utility for the optimizing agent, leading to (2).

Under Hall’s specification, it is straightforward to derive a linear setup\(^5\) for testing $H_0 : \lambda = 0$:

$$\Delta C_t = \delta + \lambda \Delta y_t + (1 - \lambda) \eta_t,$$

or, to a logarithmic approximation,

$$\Delta \ln (C_t) = \delta + \lambda \Delta \ln (y_t) + (1 - \lambda) \eta_t,$$

where $\eta_t$ is unpredictable.

As stressed in Issler and Vahid (2001), in the (optimizing) dynamic stochastic general equilibrium (DSGE) model of King, Plosser, and Rebelo (1988), output, consumption and investment share short-run co-movement. Under suitable conditions\(^6\), the closed-form solutions for the logarithms of output or income ($y_t$), consumption ($C_t$), and investment ($I_t$), are, respectively:

\[
\begin{align*}
\ln (y_t) &= \ln (X_t^p) + \overline{y} + \pi_{yk} \tilde{k}_t \\
\ln (C_t) &= \ln (X_t^p) + \overline{c} + \pi_{ck} \tilde{k}_t \\
\ln (I_t) &= \ln (X_t^p) + \overline{i} + \pi_{ik} \tilde{k}_t,
\end{align*}
\]

where $\ln (X_t^p) = \mu + \ln (X_{t-1}^p) + \varepsilon_t^p$ is the random-walk productivity process, $\overline{y}$, $\overline{c}$ and $\overline{i}$ are the steady-state values of $\ln (y_t/X_t^p)$, $\ln (C_t/X_t^p)$, and $\ln (I_t/X_t^p)$ respectively, and $\pi_{jk}$, $j = y, c, i$ is the elasticity of variable $j$ with respect to deviations of the capital stock from its stationary value $\left(\tilde{k}_t\right)$.

Focusing on the first two equations in (6), it is straightforward to verify that:

$$\Delta \ln (C_t) = \frac{\pi_{ck}}{\pi_{yk}} \Delta \ln (y_t) + \frac{\pi_{ck} - \pi_{yk}}{\pi_{yk}} (\mu + \varepsilon_t^p).$$

where $\varepsilon_t^p$ is the innovation of the productivity process $\log (X_t^p)$.

It is straightforward to conclude that (7) and (5) are observationally equivalent, with $\frac{\pi_{ck}}{\pi_{yk}} = \lambda$, $\frac{\pi_{ck} - \pi_{yk}}{\pi_{yk}} = \delta$, and $\frac{\pi_{ck} - \pi_{yk}}{\pi_{yk}} \varepsilon_t^p = (1 - \lambda) \eta_t$. Thus, validating (5) with data cannot be used to support the existence of rule-of-thumb behavior, since it also validates the optimizing behavior behind (6).

---

\(^5\)In some specifications the regressor is $\mathbb{E}_{t-1} (\Delta y_t)$ instead of $\Delta y_t$. Since we can always write $\Delta y_t = \mathbb{E}_{t-1} (\Delta y_t) + \xi_t$, where $\xi_t$ is a martingale difference, what changes between these two specifications is the nature of the error term.

\(^6\)They are: log-utility, full-depreciation of the capital stock, and Cobb-Douglas technology.
Perhaps, because of this observational-equivalence result with a basic model à la Hall, the literature has also focused on a broader model incorporating CRRA utility for the optimizing agent. Under the conditions that allowed arriving at (3) from (2), we obtain a log-linear equation for testing $H_0 : \lambda = 0$ in this context:

$$\Delta \ln (C_t) = \lambda \Delta \ln (y_t) + (1 - \lambda) \left( \alpha + \frac{1}{\phi} r_{i,t} \right) + \tilde{\mu}_{i,t},$$

where $\tilde{\mu}_{i,t} = (1 - \lambda) \mu_{i,t}$, being unpredictable.

### 2.3 A Generalized Taylor-Expansion Approach to Test Optimality in Consumption

The first modern study to focus on approximations to the euler equation of consumption decisions was Carroll (2001). He states that:

“In principle, the theoretical problems with Euler equation estimation stem from approximation error. The standard procedure has been to estimate a log-linearized, or first-order approximated, version of the Euler equation. This paper shows, however, that the higher order terms are endogenous with respect to the first-order terms (and also with respect to omitted variables), rendering consistent estimation of the log-linearized Euler equation impossible. Unfortunately, the second-order approximation fares only slightly better.”

Araujo and Issler (2011) generalized this results, showing that estimation of approximations that omit higher order terms do not have standard valid instruments, which consist of lagged values of observables. Their starting point is the Asset-Pricing Equation (or Pricing Equation, for short):

$$\mathbb{E}_t \{ M_{t+1} R_{i,t+1} \} = 1, \quad i = 1, 2, \ldots, N,$$

where $M_t$ is a general stochastic discount factor. Consider a generalized taylor expansion (not an approximation) of the exponential function around $x$, with increment $h$, as follows:

$$e^{x+h} = e^x + he^x + \frac{h^2 e^{x+\lambda(h)\cdot h}}{2}, \text{ with } \lambda(h) : \mathbb{R} \rightarrow (0, 1).$$

Dividing (10) by $e^x$:

$$e^h = 1 + h + \frac{h^2 e^{\lambda(h)\cdot h}}{2},$$

From an empirical point-of-view, these conditions will be investigated below.
shows that (11) does not depend on $x$. Let $h = \ln(M_t R_{i,t})$ to obtain:

$$M_t R_{i,t} = 1 + \ln(M_t R_{i,t}) + \frac{[\ln(M_t R_{i,t})]^2 e^{\lambda \ln(M_t R_{i,t}) \ln(M_t R_{i,t})}}{2}.$$  (12)

The behavior of $M_t R_{i,t}$ is governed solely by that of $\ln(M_t R_{i,t})$. It is useful to define a random variable collecting the higher-order term in (12):

$$z_{i,t} \equiv \frac{1}{2} \times [\ln(M_t R_{i,t})]^2 e^{\lambda \ln(M_t R_{i,t}) \ln(M_t R_{i,t})}.$$  

Notice that $z_{i,t}$ is a function of $\ln(M_t R_{i,t})$ alone and that $z_{i,t} \geq 0$ for all $(i, t)$. Taking the conditional expectation of both sides of (12), imposing the Pricing Equation and rearranging terms gives:

$$\mathbb{E}_{t-1} (z_{i,t}) = -\mathbb{E}_{t-1} \{\ln(M_t R_{i,t})\}.$$  (13)

To ensure that $\mathbb{E}_{t-1} (z_{i,t})$ exists we constrain the process $\{\ln(M_t R_{i,t})\}$ so that its conditional expectation is well defined. Weak-stationarity of $f \ln(M_t R_{i,t})g$ is a sufficient condition for that. Let $\varepsilon_t \equiv (\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{N,t})'$ stack the conditional mean of the individual forecast errors $\varepsilon_{i,t} = \ln(M_t R_{i,t}) - \mathbb{E}_{t-1} \{\ln(M_t R_{i,t})\}$. From the definition of $\varepsilon_t$ we have:

$$\ln(M_t R_t) = \mathbb{E}_{t-1} \{\ln(M_t R_t)\} + \varepsilon_t.$$  (14)

Starting with (14), denoting $m_t = \ln(M_t)$, and using $-\mathbb{E}_{t-1} (z_{i,t}) = \mathbb{E}_{t-1} \{\ln(M_t R_{i,t})\}$, we get:

$$r_{i,t} = m_t - \mathbb{E}_{t-1} (z_{i,t}) + \varepsilon_{i,t}, \quad i = 1, 2, \ldots, N.$$  (15)

Letting $u(C_t)$ be the CRRA utility function gives,

$$\Delta \ln (C_t) = \ln(\frac{\beta}{\phi}) + \frac{1}{\phi} r_{i,t} + \frac{\mathbb{E}_{t-1} (z_{i,t})}{\phi} + \mu_{i,t}, \quad i = 1, 2, \ldots, N,$$  (16)

where $\mu_{i,t} \equiv -\frac{\varepsilon_{i,t}}{\phi}$.

It is important to stress that $\frac{\mathbb{E}_{t-1} (z_{i,t})}{\phi}$ captures the effect of the higher-order terms of the taylor expansion, in general it will be a function of the variables in the conditioning set used by the econometrician to compute $\mathbb{E}_{t-1} (\cdot)$. Therefore, omission of $\frac{\mathbb{E}_{t-1} (z_{i,t})}{\phi}$ (or of parts of it) in estimating (16) will generate an omitted-variable bias. This will turn out to be a major problem for versions of (16)\textsuperscript{8}. Note that, the only reason why this term is present in (16) is because we use a log-linear approximation of the Pricing Equation (2). Thus, we can circumvent the problem if we do not try to log-linearize the Pricing Equation.

\textsuperscript{8}Notice that (16) is a generalized version of (3) when we consider several assets instead of just one. We do not impose any distributional assumptions on $\mu_{i,t}$ and the higher order term $\frac{\mathbb{E}_{t-1} (z_{i,t})}{\phi}$ shows naturally in it. Homoskedasticity in deriving (3) conveniently imposes that $\frac{\mathbb{E}_{t-1} (z_{i,t})}{\phi}$ is constant, which is very restrictive.
2.4 Nonlinear Euler Equation and Return Aggregation

Using a nonlinear instrumental variable estimator – e.g., a generalized method-of-moment (GMM) estimator – we can estimate the following system and test hypotheses of interest:

\[ \mathbb{E}_{t-1} \{ M_t R_{i,t} \} = \mathbb{E}_{t-1} \left\{ \beta \left( \frac{C_t}{C_{t-1}} \right)^{-\phi} R_{i,t} \right\} = 1, \quad i = 1, 2, \ldots, N, \tag{17} \]

which is valid for \( N \) assets in the economy.

Efficient estimation of preference and other parameters in system (17) requires estimating the whole system instead of just a portion of it. This happens for the same reason why single-equation OLS estimation is less efficient than SUR estimation in the context of a system of linear regressions. However, system estimation may pose a problem in this context, since, in principle, the number of traded assets \( (N) \) in a real economy is large relative to the number of time-series observations \( (T) \) one finds in practice. Most of the literature opted to limit the size of \( N \), e.g., \( N = 2 \): a risky and a “riskless” asset or, at most, a handful of assets or portfolios with a limited asset coverage. Of course, this solution is sub-optimal at the cost of efficiency.

In an interesting paper, Mulligan (2002) shows that an alternative to estimating the system as a whole is cross-sectional aggregation, where we do not throw away useful information contained in \( R_{i,t}, i = 1, 2, \ldots, N \), but rather aggregate returns across \( i \) to isolate the common component of asset returns; see also the alternative approach in Araujo and Issler (2011). Support for cross-sectional aggregation in this context is based on the idea that idiosyncratic risk, uncorrelated with \( M_t \), must be irrelevant for intertemporal substitution, and cross-sectional aggregation naturally eliminates idiosyncratic risk. If \( N \) is sufficiently large, return aggregation will deliver the common component of returns associated with intertemporal substitution, allowing matching aggregate consumption with aggregate return.

In what follows we present a stylized version of Mulligan’s approach to be able to discuss the problems related with approximating (9). Consider the sequence of deterministic weights \( \{ \omega_i \}_{i=1}^N \), such that \( |\omega_i| < \infty \) uniformly on \( N \), with \( \sum_{i=1}^N \omega_i = 1 \) or \( \lim_{N \to \infty} \sum_{i=1}^N \omega_i = 1 \), depending on whether we allow or not the existence of an infinite number of assets. Cross-sectional aggregation of (16) implies:

\[ \Delta \ln (C_t) = \frac{1}{\phi} \ln (\beta) + \frac{1}{\phi} r_t + \frac{1}{\phi} \mathbb{E}_{t-1} (z_t) + \mu_t, \tag{18} \]

where \( r_t = \sum_{i=1}^N \omega_i r_{i,t} \) is the logarithm of the return to the geometric average of aggregate capital, \( z_t = \sum_{i=1}^N \omega_i z_{i,t} \), and \( \mu_t = \sum_{i=1}^N \omega_i \mu_{i,t} \). Notice that we can specialize \( \omega_i = 1/N \) to
use equal weights in aggregation. Despite aggregating returns, Mulligan omits the term $\frac{1}{\phi} E_{t-1} (z_t)$ in estimating (18), potentially leading to an omitted-variable bias.

From an econometric point-of-view, the cross-sectional aggregation leading to (18) is very similar to the theoretical approach of Driscoll and Kraay (1998), although the latter is more general because it also applies if the setup is non-linear. They use orthogonality conditions of the form $E (h (\theta, w_{i,t})) = 0, i = 1, 2, \ldots, N$. If $N$ is large relative to $T$, GMM estimation is not feasible, since we cannot estimate consistently the variance-covariance matrix of the sample moments. Despite that, since for all $i$ the orthogonality conditions hold, we can form a cross-sectional average $\bar{h} (\theta, w_t) = \frac{1}{N} \sum_{i=1}^{N} h (\theta, w_{i,t})$, and estimate $\theta$ by GMM from $E (\bar{h} (\theta, w_t)) = 0$.

Under a set of standard assumptions, Driscoll and Kraay prove consistency and asymptotic normality for the GMM estimates of $\theta$. That happens whether $N$ is fixed or $N \to \infty$ at any rate.

It is important to stress that, although the approach in Mulligan is inappropriate if the log-linear approximation of (9) is invalid, it is a clever way of preserving information on all returns that would otherwise be lost if $N$ is large relative to $T$.

Using aggregate returns, we next show how to construct an encompassing consumption model that allows simultaneously for the existence of rule-of-thumb behavior and habit in preferences. These two departures from the standard CRRA model can then be tested as exclusion restrictions for the parameters of the encompassing model.

As in Weber, the key issue is to note that the optimizing agent – type 1 – obeys the euler equation. Since we want to allow for habit and rule of thumb, we start with preferences with habit for the optimizing agent:

$$u(C_{1,t}, C_{1,t-1}) = \frac{(C_{1,t} - \gamma C_{1,t-1})^{1-\phi} - 1}{1 - \phi}, \quad \phi \neq 1. \quad (19)$$

The optimizing agent euler equation is:

$$E_{t-1} \left\{ (C_{1,t-1} - \gamma C_{1,t-2})^{-\phi} - \beta (C_{1,t} - \gamma C_{1,t-1})^{-\phi} [\gamma + R_{i,t}] \right. \left. + \gamma \beta (C_{1,t+1} - \gamma C_{1,t})^{-\phi} R_{i,t} \right\} = 0, \quad \text{all } i. \quad (20)$$

Recall that aggregate consumption must be the sum of the consumption of the two types. Thus, $C_{1,t} = C_t - \lambda y_t$. Substituting the latter into (20):

$$E_{t-1} \left\{ \left[ (C_{t-1} - \lambda y_{t-1}) - \gamma (C_{t-2} - \lambda y_{t-2}) \right]^{-\phi} - \beta \left[ (C_t - \lambda y_t) - \gamma (C_{t-1} - \lambda y_{t-1}) \right]^{-\phi} [\gamma + R_{i,t}] \right. \left. + \gamma \beta \left[ (C_{t+1} - \lambda y_{t+1}) - \gamma (C_t - \lambda y_t) \right]^{-\phi} R_{i,t} \right\} = 0, \quad \text{all } i. \quad (21)$$

12
Notice that (21) is a general model that encompasses rule of thumb and habit, under CRRA. It has three special cases: habit alone ($\lambda = 0$), rule of thumb alone ($\gamma = 0$), and neither habit nor rule of thumb ($\lambda = \gamma = 0$), which is the case of CRRA utility. Equation (21) only depends on observables, although $C - \lambda y$ is highly persistent, which can be dealt with transformations using ratios: $C_t/C_{t-1}$, $C_t/y_t$, etc. Although (21) is nonlinear on consumption, it is linear on $R_{i,t}$. Thus it allows cross-sectional aggregation leading to GMM estimation as long as instruments are not indexed by $i$. Start with:

$$
\mathbb{E}_{t-1} \left\{ \frac{\beta}{w^t} \frac{u'(C_t - \lambda y_t, C_{t-1} - \lambda y_{t-1})}{u'(C_{t-1} - \lambda y_{t-1}, C_{t-2} - \lambda y_{t-2})} R_{i,t} \right\} = 1, \quad \text{all } i. \tag{22}
$$

Center, post-multiply by instruments $X_{t-1}$, and use the law of iterated expectations:

$$
\mathbb{E} \left\{ \left[ \frac{\beta}{w^t} \frac{u'(C_t - \lambda y_t, C_{t-1} - \lambda y_{t-1})}{u'(C_{t-1} - \lambda y_{t-1}, C_{t-2} - \lambda y_{t-2})} R_{i,t} - 1 \right] \otimes X_{t-1} \right\} = 0, \quad \text{all } i. \tag{23}
$$

From Driscoll and Kraay, cross-sectionally aggregate (23), using weights $w_i$, $0 \leq w_i \leq 1$, $\sum_{i=1}^{N} w_i = 1$, with $R_t = \sum_{i=1}^{N} w_i R_{i,t}$. Denote the terms in brackets of (23) by $h(\theta, w_{i,t})$. Then their aggregate version is:

$$
\tilde{h}(\theta) = \sum_{i=1}^{N} w_i h(\theta, w_{i,t}) = \left[ \frac{\beta}{w^t} \frac{u'(C_t - \lambda y_t, C_{t-1} - \lambda y_{t-1})}{u'(C_{t-1} - \lambda y_{t-1}, C_{t-2} - \lambda y_{t-2})} R_t - 1 \right] \otimes X_{t-1},
$$

where it becomes clear that we can estimate $\theta = (\beta, \phi, \gamma, \lambda)'$ by GMM using:

$$
\mathbb{E} \left\{ \tilde{h}(\theta) \right\} = \mathbb{E} \left\{ \left[ \frac{\beta}{w^t} \frac{u'(C_t - \lambda y_t, C_{t-1} - \lambda y_{t-1})}{u'(C_{t-1} - \lambda y_{t-1}, C_{t-2} - \lambda y_{t-2})} R_t - 1 \right] \otimes X_{t-1} \right\} = 0. \tag{24}
$$

The euler equation behind the moment restrictions (24) is interpretable and can be viewed as that of the optimizing agent who holds a portfolio $R_t = \sum_{i=1}^{N} \omega_i R_{i,t}$ in every period\(^{10}\). A natural way to construct weights ($\omega_i$ or $\omega_{i,t}$) is to look at participation of different assets on the portfolio of aggregate wealth in every period. This is motivated by the fact that euler equations of the form in (21) only hold as an equality in $t$ if asset $i$ is being used to transfer wealth from $t-1$ to $t$. This is a crucial issue when testing for optimality, since the euler equation must hold under the null hypothesis.

\(^9\)A list of transformed models, using these ratios, for each type – habit alone ($\lambda = 0$), rule of thumb alone ($\gamma = 0$), and neither habit nor rule of thumb ($\lambda = \gamma = 0$) – is available upon request.

\(^{10}\)For simplicity of notation we use weights that do not depend on time $t, \omega_i$. However, there is no problem with having $\omega_{i,t}$ as weights.
Participation is discussed by Vissing-Jørgensen (2002), and Attanasio, Banks, and Tanner (2002) in a panel-data context. There, the main problem is that we do not possess the information on specific assets used to smooth out consumption across time for every individual. However, for the representative consumer, one has the information on the composition of aggregate wealth. Mulligan referred to the composite return $R_t$ in the following terms: “the interest rate in aggregate theory is not the promised yield on a Treasury Bill or Bond, but should be measured as the expected return on a representative piece of capital.” In our view, this is the return that should be used to recover interpretable preference parameters for the representative consumer. For that reason, optimality tests here will be conducted using the encompassing model (24) in the form of a J-test (Sargan test).

3 Econometric Methodology

3.1 Data

The critical series used in this study is the aggregate real interest rate represented by $R_t$ above, which is used to uncover (or identify) the structural preference parameters of the representative consumer. $R_t$ here is measured in different forms and by different authors. The first measure is the capital rental rate after income and property taxes in the U.S., as computed by Mulligan (2002)\(^{11}\), and its updated version measured as the annual and quarterly estimates of the net marginal product of capital in the U.S., as computed by Mulligan and Threinen (2010)\(^{12}\). These first two measures are identical, in a context where aggregate capital exists and its marginal product is net of depreciation.

It is straightforward to show that $M_t$ and $R_t$ are inversely related under suitable conditions. Assume log-utility for preferences – where $M_t = \beta \frac{C_{t+1}}{C_t}$ – as well as no production in the economy. Dividends are equal to consumption in every period, and the price of the portfolio representing aggregate capital $P_t$ can be computed with the usual expected present-value formula:

$$P_t = \mathbb{E}_t \left\{ \sum_{i=1}^{\infty} M_{t+i} C_{t+i} \right\} = \mathbb{E}_t \left\{ \sum_{i=1}^{\infty} \beta^i \frac{C_t}{C_{t+i}} C_{t+i} \right\} = \frac{\beta}{1 - \beta} C_t.$$ 

Hence, the return on aggregate capital $R_t$ is given by:

$$R_t = \frac{P_t + C_t}{P_{t-1}} = \frac{\beta C_t + (1 - \beta) C_t}{\beta C_{t-1}} = \frac{C_t}{\beta C_{t-1}} = \frac{1}{M_t}. \quad (25)$$

\(^{11}\)We thank Casey Mulligan for providing this series to us. \(^{12}\)We thank Casey Mulligan for providing this series to us as well.
The capital rental rate after income and property taxes (Mulligan, 2002) in the U.S. is available in annual frequency from 1947 to 1997, whereas the net marginal product of capital (Mulligan and Threinen, 2010) in the U.S. is available from 1930 to 2009 on an annual frequency – although we focus on post-war data from 1950 onwards for annual data and from 1950:1 onwards on a quarterly frequency.

The rest of the data used here were extracted from the U.S. National Income and Product Account (NIPA) and from the U.S. Census Bureau. From NIPA, we extracted annual data for real disposable personal income, nominal consumption of nondurables and its price index, and nominal services consumption and its price index. We used two measures of consumption in this paper, following almost all of the consumption literature: real consumption of nondurables and real consumption of nondurables and services. Unfortunately, there is no deflator for nondurables plus services. Thus, we aggregated nondurables and services using Irving Fisher’s ideal price index – an equally weighted geometric average of the Laspeyres and Paasche price indices. Intuitively, by employing Fisher’s method, we allow rebalancing the weights of the parts on the sum of the components. Simply summing up the deflated parts implies keeping these weights fixed throughout the whole post-war sample, which is obviously inappropriate. To obtain per capita series we used population data from the US Census Bureau.

3.2 Estimating and Testing Log-Linear Models

Two linear models are estimated using a measure of the aggregate interest rate $R_t$. The first one based on log-normality and homoskedasticity assumptions, represented by equation (3). The second model is the extension to include rule of thumb behavior, represented by equation (16). In both cases, the higher order terms are omitted. These models were estimated by two-stage least squares (TSLS) and the specific instrument set includes lags of observables in each equation being estimated\(^\text{13}\).

In order to investigate the consequences of the omission of higher order terms, the following diagnostic procedures are employed: RESET, homoskedasticity, serial-correlation, and normality tests. These tests were applied to verify the conditions under which log-linearization is valid, i.e., it is not an approximation. This requires the error term to be Normally distributed with constant variance. Since the error is also a martigale difference by construction, it must also be independent of any function of the conditioning set used by the agent.

\(^{13}\)Due to aggregation problems, the use of lags of variables no closer than the second lag has been recommended by Hall (1988).
The usual version of Ramsey’s regression specification-error test, RESET, is based on low order polynomials of the predicted value of the dependent variable, i.e., $\hat{y}^2$, $\hat{y}^3$ and $\hat{y}^4$. The relevance of any polynomial term indicates mispecification. Without endogeneity, the residual terms should not be correlated with $\hat{y}^j$, $j > 1$. However, under endogeneity, $\hat{y}$ is a function of the endogenous covariates and the usual procedure is no longer appropriated. Pagan and Hall (1983) overcome this problem using the predicted values of $y$ from a regression of $y$ on the instruments. Pesaran and Taylor (1999) used the “optimal forecast” values defined as $\hat{y} = \hat{X}\hat{\beta}_{IV}$, where $\hat{X}$ is the reduced form predicted values of the endogenous regressors plus the exogenous regressors and $\hat{\beta}_{IV}$ is the IV estimate of the coefficients. We perform both tests considering three cases: $\hat{y}^2; \hat{y}^3$ and $\hat{y}^4$.

Heteroskedasticity tests (null of Homoskedasticity) verify if there is a connection between the squared residuals of the regression and indicator variables that are hypothesized to be related to the heteroskedasticity. As a result, a smaller set of indicator variables may be unable to detect heteroskedasticity. Under endogeneity, the indicator variables must be functions of exogenous variables (instruments) only. First, we use levels of IVs alone – fitted values of the dependent variable and their square, where fitted values are calculated as a linear combination of the instruments. For both cases, we employ four tests: i) Pagan and Hall’s (1983) general test statistic for heteroskedasticity in an IV regression; ii) Pagan and Hall’s (1983) version based on normal distributed error term; iii) White/Koenker $T R^2$ test statistic and, iv) Breusch-Pagan/Godfrey/Cook-Weisberg test.

Serial correlation of the error term is investigated by means of the Cumby and Huizinga (1992) test, which can be used under endogeneity and/or conditional heteroskedasticity of the regression error term. We used this test to investigate four cases: $AR(p)$, $p = 1, 2, 3, 4$. Finally, we employ Shapiro-Wilk, Jarque-Bera, and Shapiro-Francia Normality tests.

### 3.3 Estimating and Testing Nonlinear Models under Rule-of-Thumb and Habit

Since consumption and income are known to have roots of the autoregressive polynomial equal (or nearly equal to) to unity, we transform euler equations to achieve stationarity. In the context of rule-of-thumb tests, Weber (2002) discusses this issue at some length, dividing euler equations by specific powers of $y_t$ to generate non-integrated terms. These powers depend on preferences. Here, we opted for a slightly different route. For the preferences in (1) or (19), dividing euler equations by $C_{t-1}$ generates terms that have the following formats: gross growth rates of consumption or income, income-to-consumption or consumption-to-income ratios, or products of them.
We list below the four euler equations in untransformed and transformed formats, highlighting the transformations performed in order to achieve stationarity.

1. Optimizing agent with external habit and rule-of-thumb. The untransformed model is:

\[
\begin{aligned}
\mathbb{E}_{t-1} \left\{ \begin{array}{l}
\beta [R_t + \gamma] \left[ (C_t - \lambda y_t) - \gamma (C_{t-1} - \lambda y_{t-1}) \right]^{-\phi} \\
-R_t \beta^2 \gamma \left[ (C_{t+1} - \lambda y_{t+1}) - \gamma (C_t - \lambda y_t) \right]^{-\phi} \\
- \left[ (C_{t-1} - \lambda y_{t-1}) - \gamma (C_{t-2} - \lambda y_{t-2}) \right]^{-\phi}
\end{array} \right\} = 0.
\end{aligned}
\]

Collecting terms and multiplying the above equation by \(\frac{1}{C_t}\) gives the following transformed model\(^{14}\):

\[
\begin{aligned}
\mathbb{E}_{t-1} \left\{ \begin{array}{l}
\beta [R_t + \gamma] \left[ \frac{C_t}{C_{t-1}} - \lambda \frac{y_t}{C_{t-1}} - \gamma \left( 1 - \lambda \frac{y_{t-1}}{C_{t-1}} \right) \right]^{-\phi} \\
-R_t \beta^2 \gamma \left[ \frac{C_{t+1}}{C_{t-1}} - \lambda \frac{y_{t+1}}{C_{t-1}} - \gamma \left( \frac{C_t}{C_{t-1}} - \lambda \frac{y_t}{C_{t-1}} \right) \right]^{-\phi} \\
- \left[ 1 - \lambda \frac{y_{t-1}}{C_{t-1}} - \gamma \left( \frac{C_{t-1}}{C_{t-2}} \right)^{-\phi} - \lambda \left( \frac{C_{t-1}}{y_{t-2}} \right)^{-\phi} \right]
\end{array} \right\} = 0. \tag{26}
\end{aligned}
\]

2. Optimizing agent with external habit and no rule-of-thumb. The untransformed model is:

\[
\begin{aligned}
\mathbb{E}_{t-1} \left\{ \begin{array}{l}
R_t \beta (C_t - \gamma C_{t-1})^{-\phi} - R_t \beta^2 \gamma (C_{t+1} - \gamma C_t)^{-\phi} - (C_{t-1} - \gamma C_{t-2})^{-\phi} \\
+ \gamma \beta (C_t - \gamma C_{t-1})^{-\phi}
\end{array} \right\} = 0.
\end{aligned}
\]

Collecting terms and multiplying the above equation by \(\frac{1}{C_t}\) gives the following transformed model:

\[
\begin{aligned}
\mathbb{E}_{t-1} \left\{ \begin{array}{l}
\beta [R_t + \gamma] \left( \frac{C_t}{C_{t-1}} - \gamma \right)^{-\phi} - R_t \beta^2 \gamma \left( \frac{C_{t+1}}{C_{t-1}} - \gamma \frac{C_t}{C_{t-1}} \right)^{-\phi} \\
- \left( 1 - \gamma \frac{C_{t-2}}{C_{t-1}} \right)^{-\phi} 
\end{array} \right\} = 0. \tag{27}
\end{aligned}
\]

\(^{14}\)Notice that we can write:

\[
\begin{aligned}
\frac{C_t}{y_{t-1}} &= \frac{C_t}{C_{t-1}} \times \frac{C_{t-1}}{y_{t-1}}, \\
\frac{y_t}{C_{t-1}} &= \frac{y_t}{C_t} \times \frac{C_t}{C_{t-1}}, \\
\frac{C_t}{C_{t-2}} &= \frac{C_t}{C_{t-1}} \times \frac{C_{t-1}}{C_{t-2}}, \text{ and,} \\
\frac{y_t}{C_{t-2}} &= \frac{y_t}{C_t} \times \frac{C_t}{C_{t-1}} \times \frac{C_{t-1}}{C_{t-2}}.
\end{aligned}
\]
3. Optimizing agent with CRRA utility, rule-of-thumb, and no habit. The untransformed model is:

\[ E_{t-1} \left[ \beta \left( \frac{C_t - \lambda y_t}{C_{t-1} - \lambda y_{t-1}} \right)^{-\phi} R_t - 1 \right] = 0. \]

Dividing the numerator and denominator of \( \frac{C_t - \lambda y_t}{C_{t-1} - \lambda y_{t-1}} \) by \( C_{t-1} \), gives,

\[ E_{t-1} \left\{ \beta \left( \frac{C_t}{C_{t-1}} - \frac{\lambda y_t}{C_{t-1}} \right) \right\}^{-\phi} R_t - 1 = 0 \] (28)

4. Optimizing agent with CRRA utility – no rule-of-thumb and no habit. The model is already in stationary form:

\[ E_{t-1} \left[ \beta \left( \frac{C_t}{C_{t-1}} \right)^{-\phi} R_t - 1 \right] = 0. \] (29)

These models will be estimated by GMM, using the continuously updating method of Hansen, Heaton and Yaron (1996), which has shown superior properties vis-a-vis alternative methods in empirical simulations. As instruments, we employ only lags of observables in each equation being considered\(^\text{15}\).

### 4 Empirical Results

First of all, we present the results for log-linear models. Table 1 reports the estimation of log-linear models of the form in (3) for consumption of nondurables and consumption of nondurables and services. First, notice that the model is not rejected by the J-test on any occasion, at the 5% level. However, in additional mispecification tests it is rejected on every occasion by at least one test, with two exceptions – annual frequency, 1950-1997, with nondurables and nondurables and services, with lags 2 and 3 as instruments. Overall, most rejections occur in Normality and Serial-correlation tests, followed by rejections in RESET tests.

In Table 2 we test the log-linear model for rule-of-thumb under the same two alternative measures of consumption – equation (8). With one exception – consumption of nondurables and services, annual frequency from 1950-1997, lags 2 and 3 as instruments – for every regression run, there is a rejection on at least one of the specification tests discussed above:

\(^{15}\)Due to time aggregation problems, lags start at order two, as recommended by Hall (1988).
RESET, Homoskedasticity, Serial-correlation and Normality, casting a serious doubt on results of rule-of-thumb tests under log-linearization. Given the poor performance of the log-linearized model so far, our next step is to focus on nonlinear estimation results, where we have an encompassing model able test jointly rule of thumb and habit, as well as optimality.

Table 3 presents GMM estimation of the encompassing model allowing for habit and rule of thumb – equation (26). We first look at annual data collected by Mulligan (2002) and Mulligan and Threinen (2010). Regardless of whether one uses consumption of nondurables or of nondurables and services, there are no rejections using Hansen’s (1982) J-test of over-identifying restrictions. Moreover, in no occasion we rejected either $\gamma = 0$ or $\lambda = 0$ using robust t-ratios. Evidence with quarterly data is not so overwhelming: we still find no rejections of optimality using J-tests. However, when we employ nondurables and services, there is evidence that the habit parameter is statistically significant at the 5% level, with $\hat{\gamma} \approx 0.95$ on two occasions\footnote{It is interesting no note that, on these two rejections of $\gamma = 0$, the estimates of $\phi$ jumped from the interval $[1, 3.5]$ to 7.1 and 15.2, respectively. Empirically, the continuously updating estimator of Hansen, Heaton and Yaron (1996) displays fat tails, which may be the case here on these occasions.}. Changing the measure of consumption to nondurables take us back to the same results with annual data. We still find no rejection on J-tests and neither $\gamma$ nor $\lambda$ are statistically significant anywhere. Taking the whole evidence into account points toward simplifying the encompassing model in both dimensions, one at a time (equation (27) or (28)).

We consider next restricting the encompassing model with $\lambda = 0$, resulting in a pure habit model – equation (27). The results are displayed in Table 4. For annual data, we find no rejections for over-identifying-restriction tests (optimality), as well as no rejection of $\gamma = 0$ with robust t-ratios at 5%. For quarterly data, we rejected $\gamma = 0$ on two occasions when we employ nondurables and services\footnote{These are the exact same data and instrument set that had significant $\gamma$s on Table 3.}. In one of them, we also rejected optimality using the J-test. Still, when we used consumption of nondurables alone, we neither found $\gamma$ statistically significant nor we rejected optimality when J-test are employed.

Table 5 presents GMM estimation of rule of thumb models for the constrained agent. J-tests never reject the over-identifying restrictions implied by the model. Moreover, in all but two cases, we did no find the rule-of-thumb parameter $\lambda$ to be statistically significant at the 5% level. Significant rule-of-thumb parameter $\lambda$ occurred for nondurable consumption with annual data (1950-2009), instruments lags 2 and 3, and quarterly data with instruments lagged twice. It is worth noting that, on these two instances, estimated values of $\lambda$ are close to 0.2, well below the 0.5 values found by Campbell and Makiw. Still on Table 5, the relative-risk-aversion coefficient $\phi$ and the discount factor $\beta$ are significant almost everywhere with
plausible values: between 1 and 2 for the former and around 0.95 (annually) for the latter. So, our next step is to examine the CRRA case.

Finally, Table 6 presents GMM estimation of the basic CRRA model. J-tests only reject the restrictions implied by over-identifying restrictions once: lag 2 instruments with quarterly frequency. Still, the relative-risk-aversion coefficient $\phi$ and the discount factor $\beta$ are significant almost everywhere with plausible values: $\phi$ is not statistically different from 1 or 2, depending on the consumption measure used, and $\beta$ is statistically equal to 0.95 (annually) mostly everywhere.

All in all, we estimated 48 euler equations using GMM, with encouraging results vis-a-vis the optimality of consumption decisions – the title of this paper. If we take the level of significance to be 5%, we only rejected optimality twice out of 48 times. Regarding the issue of whether we can still rely on the canonical CRRA model, our opinion that the evidence here supports its use with a few caveats: after all, out of 24 regressions testing the significance of habit or rule of thumb, we found the rule-of-thumb parameter $\lambda$ to be statistically significant at the 5% level only twice, and the habit parameter $\gamma$ to be statistically significant on four occasions. So, the overall evidence points toward accepting optimality under CRRA utility whenever an aggregate return is used\(^{18}\).

\section{Conclusions}

This paper has the following original contributions to the literature on consumption optimality. First, we show that rule-of-thumb behavior in consumption is observational equivalent to behavior obtained by the optimizing model of King, Plosser and Rebelo (1988), casting doubt on how reliable standard (linear or log-linear) rule-of-thumb tests are. Second, we show that the omission of the higher-order term in the log-linear approximation yields inconsistent estimates of the structural parameters when lagged observables are used as instruments. Although Carroll (2001) and Weber (2002) also criticized the linearization and testing of euler equations on similar grounds, our setup offers a much deeper critique of current rule-of-thumb tests. Third, we show that the nonlinear estimation of a system of $N$ Asset-Pricing Equations can be done efficiently even if the number of asset returns ($N$) is high vis-a-vis the number of time-series observations ($T$), where system estimation is infeasible. We argue that efficiency can be restored by aggregating returns into a single returns. \footnote{This raises the question of whether we would reach a similar conclusion if we have had a large sample of time periods and returns with feasible GMM estimation, i.e., if we perform system estimation with a large number of returns. We leave this to future research.}
measure that fully captures intertemporal substitution. Indeed, we show that there is no reason why return aggregation cannot be performed in the nonlinear setting of the Pricing Equation, since the latter is a linear function of individual returns. Fourth, aggregation of the nonlinear euler equation forms the basis of a novel test of deviations from the canonical CRRA model of consumption in the presence of rule-of-thumb and habit behavior. Indeed, it can be viewed as testing for the importance of rule-of-thumb consumers and habit when the optimizing agent holds an equally-weighted portfolio of returns or a weighted portfolio of returns.

What we view as our main contribution is the use of a nonlinear setup for estimation and testing, where returns are properly aggregated. It allows matching aggregate consumption growth with aggregate returns in a proper way, which delivers identification and testing of structural preference parameters for the representative consumer. Aggregating returns has several benefits: only pervasive variation of returns affects intertemporal substitution in consumption and this is preserved under aggregation; the issue of participation is taken into account while aggregating returns; feasibility of standard GMM estimation is preserved without the waste of information entailed when we throw out individual returns and focus on a subset of returns for estimation and testing.

All in all, we estimated 48 euler equations using GMM, with encouraging results vis-a-vis the optimality of consumption decisions. At the 5% level, we only rejected optimality twice out of 48 times. Empirical-test results show that we can still rely on the canonical CRRA model so prevalent in macroeconomics: out of 24 regressions, we found the rule-of-thumb parameter \( \lambda \) to be statistically significant at the 5% level only twice, and the habit parameter \( \gamma \) to be statistically significant on four occasions.

The main message of this paper is that proper return aggregation is critical to study intertemporal substitution in a representative-agent framework. This involves aggregating the nonlinear euler equation and performing estimation and testing using it. In this case, we find no evidence of lack of optimality, as well as no evidence of either rule-of-thumb behavior or habit in preferences.

References


A Appendix: Tables
Table 1 - Instrumental-variable estimation for consumption and capital aggregate return

\[ \Delta \ln C_t = \alpha + \frac{1}{\sigma} r_t + \mu_t \]

<table>
<thead>
<tr>
<th>Aggregate Return</th>
<th>Consumption of Nondurables</th>
<th>Consumption of Nondurables and Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>Annual Annual Quarterly</td>
<td>Annual Annual Quarterly</td>
</tr>
<tr>
<td>Instruments</td>
<td>Lag 2 Lags 2, 3 Lag 2 Lags 2, 3</td>
<td>Lag 2 Lags 2, 3 Lag 2 Lags 2, 3 Lag 2 Lags 2, 3</td>
</tr>
<tr>
<td>( r_t )</td>
<td>0.749 0.779* 0.425 0.473* 0.479* 0.495*</td>
<td>0.658 0.646 0.370 0.399 0.515* 0.511*</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(0.423) (0.382) (0.239) (0.221) (0.231) (0.231)</td>
<td>(0.357) (0.320) (0.221) (0.196) (0.170) (0.169)</td>
</tr>
<tr>
<td>Diagnostic Tests</td>
<td>J test 0.977 1.379 0.656 4.076 1.327 4.478</td>
<td>J test 0.367 0.566 0.010 1.099 4.303* 6.158</td>
</tr>
<tr>
<td></td>
<td>(p-value) (0.323) (0.711) (0.418) (0.253) (0.249) (0.214)</td>
<td>(p-value) (0.545) (0.904) (0.921) (0.777) (0.038) (0.104)</td>
</tr>
<tr>
<td></td>
<td>Null hypothesis is rejected? Yes* No No No Yes*</td>
<td>Null Hypothesis is rejected? No No Yes** No Yes** Yes**</td>
</tr>
<tr>
<td></td>
<td>Homoskedasticity No No No No Yes* Yes**</td>
<td>No No No No Yes* Yes* Yes*</td>
</tr>
<tr>
<td></td>
<td>Serial Correlation No No Yes* Yes* No No</td>
<td>No No Yes* Yes* Yes* Yes*</td>
</tr>
<tr>
<td></td>
<td>Normality No No Yes* Yes* Yes** Yes**</td>
<td>No No No No Yes** Yes**</td>
</tr>
<tr>
<td>Note: MT (2010) refers to Mulligan and Threinen (2010). Regression estimated by two-stage least squares using Newey and West’s (1987) procedure for robust S.E. The instrument lists is composed of lags of the observables in the equation being estimated. ** and * means significant at 1% and 5%, respectively. RESET linearity tests used here are described in Pagan and Hall (1983) and Pesaran and Taylor (1999). Error serial correlation is investigated by means of the test in Cumby and Huizinga (1992). The null of Homoskedasticity is investigated by tests in Pagan and Hall (1983), the White-Koenker test, and Brensch-Pagan/Godfrey/Cook-Weisberg test. Finally, we employ Shapiro-Wilk, Jarque-Bera, and Shapiro-Francia Normality tests.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2 - Instrumental-variable estimation for consumption and capital aggregate return

\[ \Delta \ln (C_t) = \lambda \Delta \ln (y_t) + (1 - \lambda) \left( \alpha + \frac{1}{\bar{\sigma}} r_t \right) + \mu_t \]

<table>
<thead>
<tr>
<th>Aggregate Return</th>
<th>Consumption of Nondurables</th>
<th>Consumption of Nondurables and Services</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Annual</td>
<td>Quarterly</td>
</tr>
<tr>
<td></td>
<td>1950q1-2009q4</td>
<td>1950q1-2009q4</td>
</tr>
<tr>
<td>Instruments</td>
<td>Lag 2</td>
<td>Lags 2, 3</td>
</tr>
<tr>
<td></td>
<td>Lag 2</td>
<td>Lags 2, 3</td>
</tr>
<tr>
<td></td>
<td>Lag 2</td>
<td>Lags 2, 3</td>
</tr>
<tr>
<td></td>
<td>Lag 2</td>
<td>Lags 2, 3</td>
</tr>
<tr>
<td></td>
<td>Lag 2</td>
<td>Lags 2, 3</td>
</tr>
<tr>
<td></td>
<td>Lags 2, 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r_t )</td>
<td>-0.201</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>(s.e.)</td>
<td>(0.518)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.519)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.183)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.185)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.565)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.257)</td>
</tr>
<tr>
<td>( \Delta \ln Y_t )</td>
<td>0.943</td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td>(s.e.)</td>
<td>(0.642)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.406)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.377)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.302)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.948)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.395)</td>
</tr>
<tr>
<td>Diagnostic Tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J test</td>
<td>0.005</td>
<td>2.720</td>
</tr>
<tr>
<td></td>
<td>(p-value)</td>
<td>(0.941)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.606)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.781)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.774)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.664)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.073)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RESET</td>
<td>Yes*</td>
<td>Yes*</td>
</tr>
<tr>
<td>Homoskedasticity</td>
<td>Yes**</td>
<td>Yes*</td>
</tr>
<tr>
<td>Serial Correlation</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Normality</td>
<td>Yes*</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes*</td>
</tr>
<tr>
<td></td>
<td>Yes**</td>
<td>Yes**</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Yes**</td>
<td>Yes**</td>
</tr>
</tbody>
</table>

Note: See Note in Table 1.
Table 3 - GMM estimation for consumption, aggregate capital return and income

\[
\mathbb{E}_{t-1} \left\{ \beta [R_t + \gamma \left( \frac{C_t}{C_{t-1}} - \lambda \frac{y_t}{C_{t-1}} - \gamma \left( \frac{y_{t-1}}{C_{t-1}} \right) \right]^{\phi} - R_t \beta^2 \gamma \left( \frac{C_{t+1}}{C_{t-1}} - \lambda \frac{y_{t+1}}{C_{t-1}} - \gamma \left( \frac{C_{t-1}}{C_{t-2}} \right) \right]^{\phi} - \left[ 1 - \lambda \frac{y_t}{C_{t-1}} - \gamma \left( \frac{C_{t-1}}{C_{t-2}} \right)^{-1} - \lambda \left( \frac{C_{t-1}}{C_{t-2}} \right)^{-1} \right]^{\phi} \right\} = 0
\]

<table>
<thead>
<tr>
<th>Aggregate Return</th>
<th>Consumption of Nondurables</th>
<th>Consumption of Nondurables and Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>Lag 2</td>
<td>Lags 2, 3</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.944** (0.073)</td>
<td>0.952** (0.019)</td>
</tr>
<tr>
<td>( \phi )</td>
<td>1.135 (3.376)</td>
<td>1.207 (1.547)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.138 (0.325)</td>
<td>0.076 (0.094)</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.460 (0.487)</td>
<td>0.566 (0.563)</td>
</tr>
<tr>
<td>J-Test p-value</td>
<td>0.277 (0.451)</td>
<td>0.667 (0.627)</td>
</tr>
</tbody>
</table>


The instrument lists is composed of lags of the observables in the equation being estimated. ** and * means significant at 1% and 5%, respectively.
Table 4 - GMM estimation for consumption and aggregate capital return

\[
E_{t-1} \left\{ \beta \left[ R_t + \gamma \right] \left( \frac{C_t}{C_{t-1}} - \gamma \right)^{\phi} - R_t \beta^2 \gamma \left( \frac{C_{t+1}}{C_{t-1}} - \frac{C_t}{C_{t-1}} \right)^{\phi} - \left( 1 - \gamma \frac{C_t}{C_{t-1}} \right)^{\phi} \right\} = 0
\]

<table>
<thead>
<tr>
<th>Aggregate Return</th>
<th>Consumption of Nondurables</th>
<th>Consumption of Nondurables and Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>Lag 2  Lags 2, 3</td>
<td>Lag 2  Lags 2, 3</td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.959** 0.958**</td>
<td>0.982** 0.977**</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(0.006) (0.006)</td>
<td>(0.004) (0.011)</td>
</tr>
<tr>
<td>(\phi)</td>
<td>1.258* 1.121*</td>
<td>2.976 2.528</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(0.523) (0.436)</td>
<td>(1.817) (1.131)</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>-1.038 0.099</td>
<td>-1.047 0.053</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(12.533) (0.261)</td>
<td>(2.717) (0.215)</td>
</tr>
<tr>
<td>J-Test p-value</td>
<td>0.520 0.750</td>
<td>0.721 0.130</td>
</tr>
</tbody>
</table>

Note: See Note in Table 3.
Table 5 - GMM estimation for consumption, aggregate capital return and income

\[ \mathbb{E}_{t-1}\left\{ \beta \left( \frac{C_t - \lambda \frac{R_t}{1 - \lambda} \frac{C_{t-1}}{C_{t-1}}} {1 - \lambda \frac{R_t}{1 - \lambda} \frac{C_{t-1}}{C_{t-1}}} \right)^{-\phi} R_t - 1 \right\} = 0 \]

<table>
<thead>
<tr>
<th>Aggregate Return</th>
<th>Consumption of Nondurables</th>
<th>Consumption of Nondurables and Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>Annual</td>
<td>Annual</td>
</tr>
<tr>
<td>Instrument</td>
<td>Lag 2</td>
<td>Lags 2, 3</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.976**</td>
<td>0.981**</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(0.019)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>( \phi )</td>
<td>2.573</td>
<td>2.866*</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(1.413)</td>
<td>(1.250)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.003</td>
<td>-0.028</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(0.051)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>J-Test p-value</td>
<td>0.354</td>
<td>0.483</td>
</tr>
</tbody>
</table>

Note: See Note in Table 3.
Table 6 - GMM estimation for consumption and aggregate capital return

\( \mathbb{E}_{t-1} \left[ \beta \left( \frac{C_t}{C_{t-1}} \right)^{-\phi} R_t - 1 \right] = 0. \)

<table>
<thead>
<tr>
<th>Aggregate Return</th>
<th>Frequency</th>
<th>Period</th>
<th>Instrument</th>
<th>( \beta )</th>
<th>(s.e.)</th>
<th>( \phi )</th>
<th>(s.e.)</th>
<th>J-Test p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td></td>
<td>Annual</td>
<td>Lag 2</td>
<td>0.959**</td>
<td>(0.007)</td>
<td>1.225*</td>
<td>(0.542)</td>
<td>0.328</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annual</td>
<td>Lags 2, 3</td>
<td>0.957**</td>
<td>(0.005)</td>
<td>1.069**</td>
<td>(0.333)</td>
<td>0.735</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quarterly</td>
<td>Lag 2</td>
<td>0.932**</td>
<td>(0.015)</td>
<td>2.199*</td>
<td>(1.066)</td>
<td>0.425</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quarterly</td>
<td>Lags 2, 3</td>
<td>0.920**</td>
<td>(0.005)</td>
<td>1.362**</td>
<td>(0.358)</td>
<td>0.332</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lag 2</td>
<td>0.981**</td>
<td>(0.003)</td>
<td>2.003*</td>
<td>(0.791)</td>
<td>0.332</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lags 2, 3</td>
<td>0.982**</td>
<td>(0.003)</td>
<td>2.123**</td>
<td>(0.715)</td>
<td>0.234</td>
</tr>
</tbody>
</table>

| Frequency        |           | Annual       | Lag 2      | 0.973**    | (0.014) | 1.471*     | (0.722) | 0.555         |
|                  |           | Annual       | Lags 2, 3  | 0.974**    | (0.013) | 1.533*     | (0.665) | 0.891         |
|                  |           | Quarterly    | Lag 2      | 0.953**    | (0.031) | 2.634      | (1.546) | 0.928         |
|                  |           | Quarterly    | Lags 2, 3  | 0.941**    | (0.014) | 1.997**    | (0.656) | 0.761         |
|                  |           |              | Lag 2      | 0.984**    | (0.003) | 1.972**    | (0.546) | 0.035*        |
|                  |           |              | Lags 2, 3  | 0.986**    | (0.004) | 2.279**    | (0.630) | 0.120         |

Note: See Note in Table 3.