NBER WORKING PAPER SERIES

# MULTIVARIATE FRACTIONAL REGRESSION ESTIMATION OF ECONOMETRIC SHARE MODELS

John Mullahy

Working Paper 16354
http://www.nber.org/papers/w16354

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2010

I am indebted to participants at presentations of various aspects of this work at Catholic University of Rome, Michigan State University, the University of Arizona, the University of Coimbra, University College Dublin, and UW-Madison, as well as to Marguerite Burns, Ben Craig, Alberto Holly, Steve Koch, José Murteira, Stephanie Robert, Nilay Shah, João Santos Silva, and Dave Vanness for their thoughtful comments, suggestions, and discussions. In addition, Badi Baltagi and Jeff Wooldridge provided some helpful guidance with the literature. All these colleagues, of course, are absolved from any blame for the paper's shortcomings. Partial financial support from the Robert Wood Johnson Foundation Health & Society Scholars Program is acknowledged. Some of this work was completed as a visiting scholar at the UCD Geary Institute, which provided brilliant hospitality. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Multivariate Fractional Regression Estimation of Econometric Share Models
John Mullahy
NBER Working Paper No. 16354
September 2010
JEL No. C3,D12


## ABSTRACT

This paper describes and applies econometric strategies for estimating regression models of economic share data outcomes where the shares may take boundary values (zero and one) with nontrivial probability. The main focus of the paper is on the conditional mean structures of such data. The paper proposes an extension of the fractional regression methodology proposed by Papke and Wooldridge, 1996, 2008, in univariate cross-sectional and panel contexts. The paper discusses the stochastic aspects of share definition and measurement, and summarizes important features of the existing literature on econometric strategies for share model estimation. The paper then goes on to discuss the univariate fractional regression estimation strategies proposed by Papke and Wooldridge and to extend the fractional regression approach to estimation of and inference about regression models describing the multivariate share data. Some issues involving outcome aggregation/ disaggregation are considered, as is a full likelihood estimation approach based on Dirichlet-multinomial models. The paper demonstrates the workings of these various empirical strategies by estimating models of financial asset portfolio shares using data from the 2001, 2004, and 2007 U.S. Surveys of Consumer Finances.



John Mullahy
University of Wisconsin-Madison
Dept. of Population Health Sciences
787 WARF, 610 N. Walnut Street
Madison, WI 53726
and NBER
jmullahy@facstaff.wisc.edu

## Prologue

Displayed in table 1a is an extract of nine observations on nine mutually exclusive and exhaustive categories of healthcare expenditures from a multiyear sample of the U.S. Medical Expenditure Panel Survey (MEPS); the specific quantities reported are the shares of total healthcare expenditures contributed by each of the nine expenditure categories. Table 1b exhibits an extract of nine observations on seventeen mutually exclusive and exhaustive two-digit categories of time use from a multiyear sample of the American Time Use Survey (ATUS); the detailed figures are the number of minutes reported being spent in each category of time use during the one-day, or 1,440-minute, time diary recall period. Finally, table 1c displays an extract of nine observations on ten mutually exclusive and exhaustive categories of financial assets from a multiyear sample of the Survey of Consumer Finances (SCF); the detailed data presented in this table are the shares of total financial assets accounted for by each of the ten financial asset categories.

The samples from which these observations are extracted have structures that share two analytically important features. First, the multivariate outcomes are mutually exclusive and exhaustive shares of some total. Second, there is a nontrivial empirical frequency of shares that are realized at upper and lower boundary values.

## 1. Introduction

Multivariate outcomes measured as shares of some overall total arise in numerous contexts in applied microeconometrics. Whether the particular analysis focuses on time use (Mullahy and Robert, 2010), portfolio shares (e.g. Poterba and Samwick, 2001, 2002), consumer budgeting (see the references in section 3), market share analysis (Berry et al., 1995; Dubin, 2007), or some other topic, there will often arise -- as demonstrated by the above examples -- commonalities of the data structures under investigation with those exhibited in tables 1a-1c.

Letting $y_{ik}$ represent the k-th outcome for the i-th individual, M denote the number of outcomes, and $\mathbf{x}_i$, i=1,...,N, be a p-vector of exogenous covariates, such data are characterized formally by the following:

$$y_{ik} \in [0, B_i], \tag{1}$$

$$\Pr(y_{ik} = 0 | \mathbf{x}_i) > 0 \text{ and } \Pr(y_{ik} = B_i | \mathbf{x}_i) > 0, \tag{2}$$

and

$$\sum_{m=1}^{M} y_{im} = B_i \text{ for all i}, \tag{3}$$

or in vector notation $\mathbf{y}_i \in [0, B_i]^M$ and $\mathbf{1}'\mathbf{y}_i = B_i$. Here $B_i$ represents some finite total or upper bound: $B_i$=total annual healthcare spending in table 1a; $B_i$=B=1,440 minutes per day in the time use data in table 1b; and $B_i$=total financial assets in the data underlying the financial asset share data in table 1c. The $B_i$ may or may

not vary across i and may or may not be exogenous (more on this below).[1] Considerations (1) and (3) are standard in the econometric share equation literature. Econometric strategies to handle of (2) in light of (1) and (3) are less studied and are the main focus of this paper.

This paper describes and applies econometric strategies for estimating regression models of various features of outcome data like those described above, with a main focus on conditional means. Specifically, for the analysis of the conditional mean structures of such data, this paper proposes an extension of the fractional regression methodology proposed by Papke and Wooldridge in univariate cross-sectional (Papke and Wooldridge, 1996) and panel (Papke and Wooldridge, 2008) contexts.[2]

The emphasis on conditional first-moment structures is central. The working premise is that the parameters of concern to the analyst are the set of conditional means $E\left[y_k|\mathbf{x}\right]$, k=1,...,M, which are specified to satisfy

$$E\left[y_k|\mathbf{x}\right] \in (0,B) \quad , \quad k=1,...,M \tag{4}$$

and

$$\sum_{m=1}^{M} E\left[y_m|\mathbf{x}\right] = B . \tag{5}$$

Note that in (4) the conditional means are assumed to span the open interval (0,B) rather than the closed interval [0,B] which the $y_k$ can occupy. While probably a reasonable assumption in general, this could be restrictive in some instances where for some values of $\mathbf{x}$ $\Pr\left(y_k = 0|\mathbf{x}\right) = 1$ or $\Pr\left(y_k = B|\mathbf{x}\right) = 1$ might be possible.[3] The subsequent analysis accommodates either of these boundary probabilities being arbitrarily close to, but not identically, zero or one.

Henceforth the paper will work with the normalized outcomes or shares $s_k=y_k/B$ instead of the $y_k$ themselves, with the vector of shares $\mathbf{s}$ satisfying $\mathbf{s} \in \left[0,1\right]^M$ and $\mathbf{1}'\mathbf{s} = 1$. Moreover, the analysis will proceed under the assumption that the $E\left[s_k|\mathbf{x}\right]$ have a parametric structure, i.e. $E\left[s_k|\mathbf{x}\right] = \xi_k\left(\mathbf{x};\alpha\right)$, where the generic common parameter vector $\alpha = \left[\alpha_1,...,\alpha_M\right]$ will generally be shared across

---

[1] The "i" subscript indexing observations will be suppressed henceforth unless useful for clarity.

[2] Some applications may involve multivariate bounded outcome data that are not subject to adding-up restrictions like (5). The analysis of such "seemingly unrelated" outcomes might proceed generally by considering modifications of the framework proposed by Papke and Wooldridge, 2008, for panel data structures.

[3] If $\Pr\left(y_k = 0|\mathbf{x}\right) = 1$ or $\Pr\left(y_k = B|\mathbf{x}\right) = 1$ for all $\mathbf{x}$ values subsequent analysis would likely be uninteresting and/or trivial.

the M conditional mean parameters $\xi_k(\mathbf{x};\boldsymbol{\alpha})$ to enforce condition (5).

The plan for the remainder of the paper is as follows. Section 2 discusses the stochastic aspects of share definition and measurement. Section 3 summarizes salient features of the existing literature on econometric strategies for share model estimation. Section 4 highlights various features of the fractional regression estimation strategies proposed by Papke and Wooldridge (1996, 2008). Sections 5-8 are the methological core of the paper: section 5 extends the fractional regression approach to the multivariate share model context; section 6 considers several issues involving inference and specification testing; section 7 offers some ideas on testing aggregation or disaggregation of outcome categories; and section 8 presents a Dirichlet-multinomial likelihood-based approach to estimating multivariate share models that accommodates important features of the observed outcomes. Section 9 presents an empirical example of the proposed methodologies using data on financial asset portfolio shares from the 2001, 2004, and 2007 U.S. Surveys of Consumer Finances. Section 10 concludes.

## 2. Share Definition and Stochastic Characteristics

The foundation of the empirical analysis is the joint distribution $\phi(\mathbf{y},\mathbf{x})$ of an M-vector of outcomes $\mathbf{y} \geq \mathbf{0}$ and covariates $\mathbf{x}$. From this, share measures may arise naturally in at least two ways that may imply different stochastic structures for the resulting econometric share models.

Suppose there are M quantities $y_k = g_k(\mathbf{x},\boldsymbol{\alpha}_k) + u_k$, k=1,...,M, that arise from some constrained optimization problem (utility maximization; cost minimization; portfolio composition optimization; etc.), where $E[u_k|\mathbf{x}] = 0$, k=1,...,M, so that the $g_k(.)$ are conditional means. The corresponding shares are given by

$$s_k = \frac{y_k}{\sum_{m=1}^{M} y_m} = \frac{g(\mathbf{x},\boldsymbol{\alpha}_k) + u_k}{\sum_{m=1}^{M} g(\mathbf{x},\boldsymbol{\alpha}_m) + u_m} = \frac{g(\mathbf{x},\boldsymbol{\alpha}_k) + u_k}{Y}. \qquad (6)$$

In some cases (e.g. time use) there is a nonstochastic exogenous constraint B (e.g. 1,440 minutes per day) such that $\sum_{m=1}^{M} y_m = Y = B$. If the nonstochastic adding up restriction $\sum_{m=1}^{M} g(\mathbf{x},\boldsymbol{\alpha}_m) = B$ is required or enforced, then $\sum_{m=1}^{M} u_m = 0$ and

$$s_k = \frac{g(\mathbf{x},\boldsymbol{\alpha}_k) + u_k}{\sum_{m=1}^{M} g(\mathbf{x},\boldsymbol{\alpha}_m)} = \frac{g(\mathbf{x},\boldsymbol{\alpha}_k)}{\sum_{m=1}^{M} g(\mathbf{x},\boldsymbol{\alpha}_m)} + v_k = \xi_k(\mathbf{x},\boldsymbol{\alpha}) + v_k, \qquad (7)$$

where the $v_k$ are conditionally mean-zero heteroskedastic errors so that $E[s_k|\mathbf{x}] = \xi_k(\mathbf{x},\boldsymbol{\alpha})$.

Alternatively, the total Y may simply be defined as the sum of the M stochastic quantities $y_k$ whose measurements share a common metric (e.g. currency units) without being subject to any analytically relevant exogenous constraint,[4] in which case the share equations have the more general form $s_k = h(\mathbf{x}, \alpha_k, u_k)/H(\mathbf{x}, \alpha, \mathbf{u})$. In this instance, stochastic elements appear in both numerator and denominator of the share functions. As such, the derivation of $E[s_k|\mathbf{x}]$ is no longer straightforward, requiring integration over the joint distribution of the entire vector of residuals **u**.

By making primitive first-moment assumptions along the lines of $E[s_k|\mathbf{x}] = \xi_k(\mathbf{x}, \alpha)$, this paper proceeds for the most part under the assumption that the simpler structure (7) holds, although conceiving of $E[s_k|\mathbf{x}] = \xi_k(\mathbf{x}, \alpha)$ as a first-order approximation via an expansion of $h(\mathbf{x}, \alpha_k, u_k)/H(\mathbf{x}, \alpha, \mathbf{u})$ around **u**=**0** may also be reasonable.

## 3. Approaches to Econometric Share Model Estimation

This section provides a brief survey of approaches to econometric share model estimation that have been prominent in the literature.

*Econometric Share Model Estimation*

Much but not all of the econometric share equation literature focuses on the relationship between empirical share models and underlying constrained optimization behaviors yielding outcomes (e.g. commodity category expenditures or patterns of time use) that are shares of some particular total (e.g. money or time budgets). Early contributions to this literature include Christensen et al., 1975, and Wales and Woodland, 1977, who examine consumer demands and corresponding expenditure shares in utility maximization contexts. Subsequent studies have approached share equation estimation from theoretically motivated optimization models in which stochastic components are embedded to play particular roles (preference heterogeneity; technical or allocative inefficiency; etc.) in the optimization framework rather than being appended additively to nonstochastic share functions in what might be an ad hoc manner; such examples include Brown and Walker, 1989, Chavas and Segerson, 1987, Kooreman and Kapteyn, 1987, and McElroy, 1987. Considine and Mount, 1984, derive a specification in which the set of share equations has a multinomial logit functional form; this is noteworthy because a multinomial logit form is at the core of the specifications proposed below.

[4] This may be a reasonable characterization of the shares that are analyzed in the empirical analysis reported in section 9. Here the shares are the fractions of overall financial assets in each of ten financial asset categories. Even were the overall level of assets to be characterized as nonstochastic, the split between financial and nonfinancial assets and, therefore, total financial assets, would presumably be stochastic.

Dubin, 2007, uses nested multinomial logit market share models to estimate valuations of intangible assets. Fry et al., 1996, based in part on ideas developed in Aitchison, 1982, apply to the estimation of share models methods from compositional data analysis (CODA), which involve essentially modeling logs of ratios of shares.[5]

*Estimation of Share Models with Boundary or Corner Solutions*

Fewer studies have tackled the thorny empirical problems that arise when observed shares take on corner or boundary solutions with nontrivial probabilities.[6] Appealing to Kuhn-Tucker conditions and corresponding virtual or support prices, Lee and Pitt, 1986, propose a general empirical structure for multivariate share systems when corner solutions at zero are prominent in the data. Morey et al., 1995, explore a variety of statistical models to accommodate boundary outcomes, among these multinomial models for discretely measured (count) outcomes that share some features with the multivariate fractional regression models that are proposed below.

Since boundary solutions are a prominent feature of the share data of concern here, providing a general strategy for analyzing such outcomes is of some interest. Ad hoc fixes are not an appealing approach to the boundary solution phenomenon, particularly when the probability of such boundary outcomes is nontriviall.

*Dirichlet Share Models*

Woodland, 1979, proposed the Dirichlet distribution as a direct statistical model for shares without particular consideration of any underlying economic optimization framework. The Dirichlet density conditional on $\mathbf{x}$ is given by[7]

$$D\left(\mathbf{s}\middle|\mathbf{x};\boldsymbol{\psi}\right)=\left(\frac{\Gamma\left(\sum_{m=1}^{M}z_m\left(\mathbf{x},\boldsymbol{\psi}\right)\right)}{\prod_{m=1}^{M}\Gamma\left(z_m\left(\mathbf{x},\boldsymbol{\psi}\right)\right)}\right)\prod_{m=1}^{M}s_m^{\left(z_m(\mathbf{x},\boldsymbol{\psi})-1\right)}. \qquad (8)$$

[5] See also Billheimer et al., 2001.

[6] For instance, Kooreman and Kapteyn's elegant analysis of time use demands notes the potential for boundary solutions but then goes on to comment: "We will ignore the [boundary condition constraint equations in the theoretical model], which are binding for only a limited number of observations."

[7] Woodland specifies a general functional form for the Dirichlet regression parameters but then uses a linear (not exponential) specification in his empirical analysis. Nonetheless, since the Dirichlet parameters must be positive, an exponential specification is appealing. (José Murteira, in a private communication, pointed out that the expression for the Dirichlet density in Woodland's equation (4) contains a typographical error: The product in the denominator of the correct expression is displayed as a summation in Woodland's paper.)

Note that if any $s_k=0$ then the density $D(.)=0$ and if any $s_k=1$ then necessarily all the other $s_k=0$. As such, the density is undefined in either event, thus precluding direct application of the Dirichlet model to the kinds of share data examined here where boundary values are prominent.

Yet it is useful for purposes of this paper to note that for the Dirichlet model, the conditional first moments are

$$E\left[s_k \left| \mathbf{x}\right.\right] = \frac{z_k\left(\mathbf{x},\boldsymbol{\psi}\right)}{\sum_{m=1}^{M} z_m\left(\mathbf{x},\boldsymbol{\psi}\right)}, \quad k=1,\ldots,M \tag{9}$$

or

$$E\left[s_k \left| \mathbf{x}\right.\right] = \frac{\exp\left(\mathbf{x}\boldsymbol{\psi}_k\right)}{\sum_{m=1}^{M} \exp\left(\mathbf{x}\boldsymbol{\psi}_m\right)} = \frac{\exp\left(\mathbf{x}\left(\boldsymbol{\psi}_k - \boldsymbol{\psi}_M\right)\right)}{1+\sum_{m=1}^{M-1} \exp\left(\mathbf{x}\left(\boldsymbol{\psi}_m - \boldsymbol{\psi}_M\right)\right)}, \quad k=1,\ldots,M, \tag{10}$$

using a natural exponential-with-linear-index parameterization for the z(.). Note that this corresponds to the standard functional form for multinomial logit probabilities even though for the Dirichlet model all M of the $\boldsymbol{\psi}_m$ are identified. Noteworthy for present purposes is that this conditional first-moment structure coincides with that of the multivariate fractional share model whose specification and estimation is discussed below.

## 4. Fractional Regression Estimation

Before exploring the multivariate estimation strategies that are the main focus of this paper, a brief overview of fractional regression (FREG) methods is worthwhile.[8] The FREG model was proposed initially by Papke and Wooldridge (PW), 1996, in their study of voluntary individual contributions to retirement accounts in which the univariate dependent variable of interest is the fraction $s \in \left[0,1\right]$ of allowable contributions made by individuals in their sample. The key result in the Papke-Wooldridge paper is that even when the outcomes take on values at the extremes of the bounded range they occupy -- i.e. s=0 or s=1 -- with nonzero probability, the FREG method provides consistent estimates of the parameters of a univariate conditional mean function so long as it is specified with the correct functional form and embedded in a suitable quasi-ML estimator or M-estimator. Specifically, PW, 1996, consider the case of univariate fractional outcome data (s) and conditional means $E\left[s \left| \mathbf{x}\right.\right]$ with M=1, while PW, 2008, consider the panel data context with the added j-dimension, j=1,...,J, giving outcomes for the i-th individual at the j-th time period $s_{ijm}=s_{ij}$ that are multivariate (J>1) in the j-dimension but univariate (M=1) in the m-dimension. In particular, in this case there are no implied adding-up restrictions of the form (5) to accommodate.

[8] See Ramalho et al., 2010, for an excellent survey of fractional regression model estimation.

The basic idea underlying FREG estimation in the cross-sectional univariate outcome context is that if focus is exclusively on conditional first moments, then quasi-ML estimation when a correct parametric specification of the conditional first-moment structure is embedded in an exponential-family quasi-likelihood will yield consistent estimates of the first-moment parameters regardless of whether the nominal quasi-likelihood is true or not (Gourieroux et al., 1984a,b). In light of these arguments, PW suggest that the key property of a conditional first-moment model for a fractional outcome is that it obeys the boundary restrictions, i.e. $E\left[s\left|\mathbf{x}\right.\right] = \xi\left(\mathbf{x}\right) \in \left(0,1\right)$. PW then suggest that a general class of parametric functional forms that satisfy this restriction are distribution functions $G\left(.\right)$ of continuous random variables, i.e. $E\left[s\left|\mathbf{x}\right.\right] = \xi\left(\mathbf{x};\boldsymbol{\omega}\right) = G\left(\mathbf{x};\boldsymbol{\omega}\right) \in \left(0,1\right)$.

Thus direct specification of such a conditional mean structure embedded in an exponential-family quasi-likelihood whose maximization estimates such a distribution function should provide consistent estimates of $\boldsymbol{\omega}$ so long as $G\left(\mathbf{x};\boldsymbol{\omega}\right)$ is a correct specification of the conditional first moment. Bernoulli quasi-likelihoods $G\left(\mathbf{x};\boldsymbol{\omega}\right)^{s} \times \left(1 - G\left(\mathbf{x};\boldsymbol{\omega}\right)\right)^{1-s}$ for the fractional (not binary) outcome measures $s \in \left[0,1\right]$ are the obvious choice, with the particular functional form for $G\left(\mathbf{x};\boldsymbol{\omega}\right)$ specified as a logit, probit, or other cumulative distribution function, typically with a linear index argument $\mathbf{x}\boldsymbol{\omega}$. Consistent inferences are straightforward, but will generally involve using robust sandwich or bootstrap covariance estimators since the share data will be underdispersed relative to the nominal Bernoulli model (see section 6).

The fractional logit ("FLOGIT") version of the FREG model with

$$E\left[s \mid \mathbf{x}\right] = G\left(\mathbf{x};\boldsymbol{\omega}\right) = \frac{\exp\left(\mathbf{x}\boldsymbol{\omega}\right)}{\left(1 + \exp\left(\mathbf{x}\boldsymbol{\omega}\right)\right)} \tag{11}$$

is the univariate foundation for the multivariate FREG estimator discussed now.

## 5. Multivariate Fractional Logit: Estimation

The central goal of this paper is to provide consistent estimation strategies to estimate properties of the conditional distribution of share data that enforce (12) and (13) and accommodate (14) and (15):

$$E\left[s_k \left|\mathbf{x}\right.\right] = \xi_k\left(\mathbf{x};\boldsymbol{\beta}\right) \in \left(0,1\right), \quad k=1,\ldots,M, \tag{12}$$

$$\sum_{m=1}^{M} E\left[s_m \left|\mathbf{x}\right.\right] = 1, \tag{13}$$

$$\Pr\left(s_k = 0 \left|\mathbf{x}\right.\right) \geq 0, \quad k=1,\ldots,M, \tag{14}$$

$$\Pr\left(s_k = 1 \left|\mathbf{x}\right.\right) \geq 0, \quad k=1,\ldots,M, \tag{15}$$

where $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_M\right]$. The main concern in this section and the next is with estimation of the conditional first-moment structure of such data, i.e. $\boldsymbol{\xi}\left(\mathbf{x};\boldsymbol{\beta}\right)$. Section 8 extends this inquiry to other features of the joint conditional probability models $\phi\left(\mathbf{s}\middle|\mathbf{x}\right)$.

The extension of the PW approach to a multivariate fractional logit ("MFLOGIT") setting is straightforward.[9] This and the following three sections offer detailed exposition of the estimator and its properties. PW and Gourieroux et al., 1984a,b provide the fundamental arguments to establish consistency for the multivariate/multinomial version of the univariate PW quasi-ML approach. Assume that the sample are independent draws from the (M+C)-variate distribution $\phi\left(\mathbf{s},\mathbf{x}\right)$. Based on (10), specify the M conditional means to have a multinomial logit functional form in linear indexes as

$$E\left[s_k\middle|\mathbf{x}\right] = \xi_k\left(\mathbf{x};\boldsymbol{\beta}\right) = \frac{\exp\left(\mathbf{x}\boldsymbol{\beta}_k\right)}{\sum_{m=1}^{M}\exp\left(\mathbf{x}\boldsymbol{\beta}_m\right)}, \quad k=1,\ldots,M. \tag{16}$$

Note that this specification enforces both (12) and (13). Alternative specifications, e.g. $\xi_k\left(\mathbf{x};\boldsymbol{\varphi}\right) = \mathbf{x}\boldsymbol{\varphi}_k \Big/ \sum_{m=1}^{M}\mathbf{x}\boldsymbol{\varphi}_m$, are estimable -- indeed, this is the conditional first-moment functional form implied in the Dirichlet simulations conducted by Woodland, 1979 -- but they admit the possibility of predicted shares falling outside the $\left[0,1\right]$ interval at some values of $\mathbf{x}$. This paper thus focuses on the specification (16) although the merits of competing first-moment specifications could be adjudicated empirically by conditional-moment or related tests.

As with the familiar multinomial logit estimator, some normalization is required since all M of the $\boldsymbol{\beta}_k$ will not be separately identified in the multinomial quasi-likelihood; $\boldsymbol{\beta}_M = \mathbf{0}$ is used henceforth, giving

$$\xi_k\left(\mathbf{x};\boldsymbol{\beta}\right) = \frac{\exp\left(\mathbf{x}\boldsymbol{\beta}_k\right)}{1+\sum_{m=1}^{M-1}\exp\left(\mathbf{x}\boldsymbol{\beta}_m\right)}, \quad k=1,\ldots,M\text{-}1 \tag{17}$$

and

$$\xi_M\left(\mathbf{x};\boldsymbol{\beta}\right) = \frac{1}{1+\sum_{m=1}^{M-1}\exp\left(\mathbf{x}_i\boldsymbol{\beta}_m\right)}. \tag{18}$$

Owing in part to the normalization, interpretation of the signs and magnitudes of the $\boldsymbol{\beta}_k$ is generally not straightforward.[10] Typically much more interesting and

[9] This estimation strategy has been applied in Mullahy, 2004, Mullahy and Robert, 2010, Koch, 2010, as well as in the transportation research literature (Sivakumar and Bhat, 2002; Ye and Pendyala, 2005).

[10] See Crawford et al., 1998.

useful in applications are the corresponding average partial effects (APEs) that are invariant with respect to the particular normalization selected. These are described in detail in Appendix 1.

Appealing to the quasi-ML estimation methods described by PW for the univariate case, one can define a multinomial logit quasi-likelihood function Q(.) that embeds the functional form (16) and that uses the observed shares $s_{ik} \in [0,1]$ in place of the binary indicators that would typically populate a multinomial logit likelihood function, i.e.

$$Q(\boldsymbol{\beta}) = \prod_{i=1}^{N} \prod_{m=1}^{M} \xi_m(\mathbf{x}_i;\boldsymbol{\beta})^{s_{im}} . \tag{19}$$

The log quasi-likelihood is

$$J(\boldsymbol{\beta}) = \log\left(Q(\boldsymbol{\beta})\right) = \sum_{i=1}^{N} \sum_{m=1}^{M} s_{im} \times \log\left(\xi_m(\mathbf{x}_i;\boldsymbol{\beta})\right), \tag{20}$$

with the corresponding $p \times (M-1)$ estimating or score equations

$$\frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k} = \sum_{i=1}^{N} \mathbf{x}_i^T \left[ s_{ik} - \left( \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_k)}{1 + \sum_{m=1}^{M-1} \exp(\mathbf{x}_i \boldsymbol{\beta}_m)} \right) \right], \quad k=1,\ldots,M-1, \tag{21}$$

which are the same score equations as those corresponding to a standard multinomial logit estimator except that the $s_{ik}$ are, in general, nonbinary.[11] Consistency of the resulting $\hat{\boldsymbol{\beta}}$ follows from the arguments in PW and Gourieroux et al., 1984a, in particular that $E\left[\partial J(\boldsymbol{\beta})/\partial \boldsymbol{\beta}_k\right] = E_{\mathbf{x}} E\left[\partial J(\boldsymbol{\beta})/\partial \boldsymbol{\beta}_k \middle| \mathbf{x}\right] = \mathbf{0}$, $k=1,\ldots,M-1$, given standard full-rank assumptions that ensure a unique maximizer.[12]

---

[11] Some canned multinomial logit estimation packages, e.g. Stata's *mlogit*, do not accommodate nonbinary $s_m$. The estimates presented here are obtained using a procedure written in Stata's Mata language, which is available on request.

[12] In closing this section it might be noted that it would be feasible and straightforward to ignore the share system nature of the outcome data and estimate a first-moment structure with M independent binary FLOGIT models (11) yielding M corresponding $\widehat{\boldsymbol{\omega}}_k$ estimates and APEs. While straightforward, such an approach ultimately implies an unrealistic form of aggregation across the M categories (see section 7 for related discussion). Yet whether such an approach would have substantive implications for the estimates of parameters of interest like APEs is an open question; preliminary results using the SCF data (not reported here) show considerable similarity of the APE estimates from the MFLOGIT and the independent binary FLOGIT estimators.

## 6. Multivariate Fractional Logit: Underdispersion, Inference, and Specification Testing

*Underdispersion*

Overdispersion (resp. underdispersion) is typically characterized in a univariate outcome context as a situation where the empirical variance of the distribution of some outcome y is greater than (resp. less than) the variance that would obtain if y followed a reference or nominal distribution $\phi_{nom}$, possibly conditioned on covariates, i.e. $\text{Var}_{emp}\left(y|\mathbf{x}\right) > \text{Var}_{nom}\left(y|\mathbf{x}\right)$ (resp. $\text{Var}_{emp}\left(y|\mathbf{x}\right) < \text{Var}_{nom}\left(y|\mathbf{x}\right)$), possibly enforcing the restriction $E_{emp}\left[y|\mathbf{x}\right] = E_{nom}\left[y|\mathbf{x}\right]$. In the multivariate outcome context with outcome $\mathbf{y}$ an $M \times 1$ vector, a natural extension is to define overdispersion (resp. strict overdispersion) as the situation where the matrix difference $\text{Cov}_{emp}\left(\mathbf{y}|\mathbf{x}\right)$ - $\text{Cov}_{nom}\left(\mathbf{y}|\mathbf{x}\right)$ is positive semidefinite (resp. positive definite), and underdispersion (resp. strict underdispersion) as the situation where the $\text{Cov}_{nom}\left(\mathbf{y}|\mathbf{x}\right) - \text{Cov}_{emp}\left(\mathbf{y}|\mathbf{x}\right)$ is positive semidefinite (resp. positive definite).

The nature of the multivariate share data analyzed here is such that the data necessarily manifest (conditional) underdispersion relative to their nominal multinomial counterparts in this matrix-definiteness sense. This can be seen as follows. Given the quasi-ML first-moment assumptions and multinomial nominal likelihood, it follows that $\text{Pr}_{nom}\left(s_k = 1|\mathbf{x}\right) = E_{nom}\left[s_k|\mathbf{x}\right] = E_{emp}\left[s_k|\mathbf{x}\right] = \xi_k\left(\mathbf{x}\right)$. Thus for the share vector $\mathbf{s}$, $\text{Cov}_{nom}\left(\mathbf{s}|\mathbf{x}\right)$ is given by the multinomial covariance matrix

$$\text{Cov}_{nom}\left(\mathbf{s}|\mathbf{x}\right) = \begin{bmatrix} \xi_1\left(\mathbf{x}\right)\left(1-\xi_1\left(\mathbf{x}\right)\right) & -\xi_1\left(\mathbf{x}\right)\xi_2\left(\mathbf{x}\right) & \cdots & -\xi_1\left(\mathbf{x}\right)\xi_M\left(\mathbf{x}\right) \\ -\xi_1\left(\mathbf{x}\right)\xi_2\left(\mathbf{x}\right) & \xi_2\left(\mathbf{x}\right)\left(1-\xi_2\left(\mathbf{x}\right)\right) & & \vdots \\ \vdots & & & \\ -\xi_1\left(\mathbf{x}\right)\xi_M\left(\mathbf{x}\right) & \cdots & & \xi_M\left(\mathbf{x}\right)\left(1-\xi_M\left(\mathbf{x}\right)\right) \end{bmatrix}, \qquad (22)$$

while $\text{Cov}_{emp}\left(\mathbf{s}|\mathbf{x}\right)$ is given by

$$\text{Cov}_{emp}\left(\mathbf{s}|\mathbf{x}\right) = \begin{bmatrix} E\left[\left(s_1 - \xi_1\left(\mathbf{x}\right)\right)^2|\mathbf{x}\right] & E\left[\left(s_1 - \xi_1\left(\mathbf{x}\right)\right)\left(s_2 - \xi_2\left(\mathbf{x}\right)\right)|\mathbf{x}\right] & \cdots & E\left[\left(s_1 - \xi_1\left(\mathbf{x}\right)\right)\left(s_M - \xi_M\left(\mathbf{x}\right)\right)|\mathbf{x}\right] \\ E\left[\left(s_1 - \xi_1\left(\mathbf{x}\right)\right)\left(s_2 - \xi_2\left(\mathbf{x}\right)\right)|\mathbf{x}\right] & E\left[\left(s_2 - \xi_2\left(\mathbf{x}\right)\right)^2|\mathbf{x}\right] & & \vdots \\ \vdots & & & \\ E\left[\left(s_1 - \xi_1\left(\mathbf{x}\right)\right)\left(s_M - \xi_M\left(\mathbf{x}\right)\right)|\mathbf{x}\right] & \cdots & & E\left[\left(s_M - \xi_M\left(\mathbf{x}\right)\right)^2|\mathbf{x}\right] \end{bmatrix}. \qquad (23)$$

Thus

$$\Delta(\mathbf{x}) = \mathrm{Cov}_{nom}(\mathbf{s}|\mathbf{x}) - \mathrm{Cov}_{emp}(\mathbf{s}|\mathbf{x}) = \begin{bmatrix} \xi_1(\mathbf{x}) - E[s_1^2|\mathbf{x}] & -E[s_1 s_2|\mathbf{x}] & \cdots & -E[s_1 s_M|\mathbf{x}] \\ -E[s_1 s_2|\mathbf{x}] & \xi_2(\mathbf{x}) - E[s_2^2|\mathbf{x}] & & \vdots \\ \vdots & & & \\ -E[s_1 s_M|\mathbf{x}] & \cdots & & \xi_M(\mathbf{x}) - E[s_M^2|\mathbf{x}] \end{bmatrix}. \quad (24)$$

If $\Pr(s_k \in (0,1)|\mathbf{x}) > 0$ for all k then each of the diagonal elements of $\Delta(\mathbf{x})$ is positive (since $z > z^2$ for any $z \in (0,1)$). Note too that $\sum_{\substack{m=1 \\ m \neq k}}^{M} E[s_k s_m|\mathbf{x}] = E\left[\left(s_k \sum_{\substack{m=1 \\ m \neq k}}^{M} s_m\right)\middle|\mathbf{x}\right] = E[s_k(1-s_k)|\mathbf{x}]$, so that the absolute value of the row sum of the off-diagonal elements of any row in $\Delta(\mathbf{x})$ equals the diagonal element in that row.

*Definition*

A symmetric $M \times M$ matrix $\mathbf{D}$ is weakly diagonally dominant if $D_{kk} \geq \sum_{\substack{m=1 \\ m \neq k}}^{M} |D_{km}|$ for all k=1,...,M, and is diagonally dominant if the inequality holds strictly.

It follows that the matrix $\Delta(\mathbf{x})$ is weakly diagonally dominant. Furthermore, a diagonally dominant (resp. weakly diagonally dominant) matrix is positive definite (resp. positive semidefinite).[13] As such, the empirical distribution of the share vector $\mathbf{s}$ conditional on $\mathbf{x}$ manifests underdispersion relative to the nominal multinomial quasi-likelihood. The implications of this for inference are considered below.

*Inference*

The asymptotic distribution of $\hat{\boldsymbol{\beta}}$ follows from the arguments in PW and in Gourieroux et al., 1984a,b. Specifically, given correct specification of the conditional first moments $\boldsymbol{\xi}(\mathbf{x}) = E[\mathbf{s}|\mathbf{x}]$, $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically $N(\mathbf{0}, \mathbf{V}_{MFLOGIT})$ where

$$\mathbf{V}_{MFLOGIT} = \mathbf{A}_{\mathbf{P}}^{-1} \mathbf{A}_{\mathbf{Q}} \mathbf{A}_{\mathbf{P}}^{-1} \quad (25)$$

with

$$\mathbf{A}_{\mathbf{P}} = E_{\mathbf{x}}\left[(\mathbf{I}_{M-1} \otimes \mathbf{x})' \mathbf{P}(\mathbf{x})(\mathbf{I}_{M-1} \otimes \mathbf{x})\right], \quad (26)$$

[13] See Graybill, 1983, Theorem 12.2.16, and Intriligator, 1971, equation B.7.10.

$$\mathbf{A_Q} = \mathrm{E}_{\mathbf{x}}\left[\left(\mathbf{I}_{M-1} \otimes \mathbf{x}\right)' \mathbf{Q}(\mathbf{x})\left(\mathbf{I}_{M-1} \otimes \mathbf{x}\right)\right], \tag{27}$$

$$\mathbf{P}(\mathbf{x}) = \begin{bmatrix} \xi_1(\mathbf{x})(1-\xi_1(\mathbf{x})) & -\xi_1(\mathbf{x})\xi_2(\mathbf{x}) & \cdots & -\xi_1(\mathbf{x})\xi_{(M-1)}(\mathbf{x}) \\ -\xi_1(\mathbf{x})\xi_2(\mathbf{x}) & \xi_2(\mathbf{x})(1-\xi_2(\mathbf{x})) & & \vdots \\ \vdots & & \ddots & \\ -\xi_1(\mathbf{x})\xi_{(M-1)}(\mathbf{x}) & \cdots & & \xi_{(M-1)}(\mathbf{x})\left(1-\xi_{(M-1)}(\mathbf{x})\right) \end{bmatrix}, \tag{28}$$

and

$$\mathbf{Q}(\mathbf{x}) = \begin{bmatrix} \mathrm{Var}(s_1|\mathbf{x}) & \mathrm{Cov}(s_1, s_2|\mathbf{x}) & \cdots & \mathrm{Cov}\left(s_1, s_{(M-1)}|\mathbf{x}\right) \\ \mathrm{Cov}(s_1, s_2|\mathbf{x}) & \mathrm{Var}(s_2|\mathbf{x}) & & \vdots \\ \vdots & & \ddots & \\ \mathrm{Cov}\left(s_1, s_{(M-1)}|\mathbf{x}\right) & \cdots & & \mathrm{Var}\left(s_{(M-1)}|\mathbf{x}\right) \end{bmatrix}. \tag{29}$$

Although estimating the model using quasi-ML methods will provide consistent estimates of the $\boldsymbol{\beta}_k$ parameters, the corresponding inverse Hessian MNL covariance matrix will not be a consistent estimator of the true covariance matrix so long as $\Pr\left(s_k \in (0,1) \mid \mathbf{x}\right) > 0$. Note that if the data were truly conditionally multinomially distributed, then $\sqrt{N}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$ would have asymptotic covariance matrix $\mathbf{V}_{MNL} = \mathbf{A_P^{-1}}\mathbf{A_P}\mathbf{A_P^{-1}} = \mathbf{A_P^{-1}}$. The difference between $\mathbf{V}_{MNL}$ and $\mathbf{V}_{MFLOGIT}$ is

$$\mathbf{V}_{MNL} - \mathbf{V}_{MFLOGIT} = \mathbf{A_P^{-1}}\left(\mathbf{A_P} - \mathbf{A_Q}\right)\mathbf{A_P^{-1}}. \tag{30}$$

In turn, the matrix difference $\mathbf{A_P} - \mathbf{A_Q}$ can be written as

$$\mathbf{A_P} - \mathbf{A_Q} = \mathrm{E}_{\mathbf{x}}\left[\left(\mathbf{I}_{M-1} \otimes \mathbf{x}\right)'\left(\mathbf{P}(\mathbf{x}) - \mathbf{Q}(\mathbf{x})\right)\left(\mathbf{I}_{M-1} \otimes \mathbf{x}\right)\right]. \tag{31}$$

Note that the matrix difference $\mathbf{P}(\mathbf{x}) - \mathbf{Q}(\mathbf{x})$ equals the matrix $\Delta(\mathbf{x})$ defined in eq. (24) with the M-th row and M-th column deleted. As such $\mathbf{P}(\mathbf{x}) - \mathbf{Q}(\mathbf{x})$ will in general be strictly diagonally dominant and, therefore, positive definite. Being quadratics in positive definite matrixes, it thus follows that $\mathbf{A_P} - \mathbf{A_Q}$ and, therefore, $\mathbf{A_P^{-1}}\left(\mathbf{A_P} - \mathbf{A_Q}\right)\mathbf{A_P^{-1}}$, will themselves be positive definite. As such, $\mathbf{V}_{MNL} - \mathbf{V}_{MFLOGIT}$ will in general be positive definite. As will be seen below in section 9, inverse Hessian estimates of $\mathrm{Cov}\left(\hat{\boldsymbol{\beta}}\right)$ based on the standard multinomial logit quasi-likelihood yield estimated parameter t-statistics for the individual $\hat{\boldsymbol{\beta}}_{mk}$ that range in this application from about 1.1 to 2.4 times *smaller* than those obtained using a robust sandwich estimator.

*Specification Testing*

To assess the quality of the MFLOGIT first-moment specification fit, conditional moment tests can be conducted that are based on cross-products of a vector of functions of $\mathbf{x}$ and the estimated MFLOGIT residuals, $\boldsymbol{\Lambda}(\mathbf{x}) \times \left(\mathbf{s} - \hat{\mathrm{E}}\left[\mathbf{s}|\mathbf{x}\right]\right)$.

The power of such tests to detect poorness of fit depends on the specification of $\boldsymbol{\Lambda}(\mathbf{x})$. The particular form of $\boldsymbol{\Lambda}(\mathbf{x})$ used here is suggested by the Hosmer-Lemeshow test strategy used commonly in the evaluation of binary logit models, i.e. $\boldsymbol{\Lambda}(\mathbf{x})$ is specified as a vector of indicator functions based on L sample quantiles of the $\hat{\mathrm{E}}\left[s_{im}|\mathbf{x}_i\right]$, i.e.

$$\lambda_{mq} = N^{-1}\sum_{i=1}^{N} 1\left(\xi_m\left(\mathbf{x}_i;\hat{\boldsymbol{\beta}}\right) \in J_{mq}\right) \times \left(s_{im} - \xi_m\left(\mathbf{x}_i;\hat{\boldsymbol{\beta}}\right)\right), \quad q=1,\ldots,L \tag{32}$$

where the $J_{mq}$ denote intervals on the real line defined by the sample quantiles of the $\hat{\mathrm{E}}\left[s_{im}|\mathbf{x}_i\right]$.

## 7. Multivariate Fractional Logit: Aggregation and Disaggregation of Outcome Categories

In empirical contexts where MFLOGIT-type estimation strategies might be applied it may sometimes be of interest to determine whether subsets of the outcome measures $s_k$ might sensibly be aggregated or pooled (to reduce dimensionality) or, if such data are available, disaggregated to (to refine detail). Depending on the purpose of the analysis, aggregability of outcome categories could be characterized in a variety of ways that include: similarity of corresponding category-parameter vectors; similarity of corresponding category-partial effects; and others. This discussion will focus on aggregation characterized by similarity of parameter vectors; as such, given the MFLOGIT's first-moment structure, aggregation and disaggregation of outcome measures are tantamount to summation and proper subsetting, respectively.

Likelihood-based strategies for testing for aggregation or disaggregation of categories in multinomial or nested logit models with discrete outcomes (e.g. likelihood-ratio tests) are well established in the multinomial logit literature (Cramer and Ridder, 1991; Hill, 1983). These approaches are not directly applicable in the MFLOGIT's first-moment/quasi-likelihood context, however. Instead this discussion will consider straightforward approaches based on robust Wald tests (a "bottom up" approach based on estimation of models for disaggregated outcomes) as opposed to Lagrange multiplier tests (a "top down" approach that would be more challenging to implement in the quasi-likelihood context) or tests based on criteria like mean squared error reduction. The merits of the approaches suggested here relative to such alternatives is for future research to assess.

There are at least two fundamentally distinct circumstances in which aggregation considerations may arise in the analysis of share data. The first ("structured aggregation") occurs when the outcome data follow some natural and/or predefined tree structure. At each level of such a tree the share outcome categories at that level -- which are outcome subcategories by reference to the next-higher level -- are mutually exclusive and exhaustive and sum to one.[14] The considerations raised above notwithstanding, testing for aggregation with such tree structures can more or less proceed using classical testing approaches. The second circumstance ("unstructured aggregation") arises when there are again mutually exclusive and exhaustive subcategory share outcomes that sum to one but in which there is either no natural and/or predefined tree structure or in which the analyst for some reason elects to ignore a given tree structure (e.g. to consider whether subcategories in different branches can be pooled with each other). In these circumstances, alternative testing strategies will generally be required. Prototypical examples of each type of aggregation are displayed in figure 1.[15]

Only structured aggregation will be considered here. Moreover, while the basic ideas generalize to multiple levels, this section considers only a simple two-level aggregation context in which there are M outcome categories or aggregates (e.g. the m=1,...,M $s_m$ measures) and R>M outcome subcategories or disaggregates (denoted $v_r$, r=1,...,R) with $\sum_{m=1}^{M} s_m = \sum_{r=1}^{R} v_r = 1$. Define implicitly M index sets $C_m$ via $\sum_{r \in C_m} v_r = s_m$, m=1,...,M, with $\bigcup_{m=1}^{M} C_m = \{1,\ldots,R\}$ and $\bigcap_{m=1}^{M} C_m = \varnothing$. Thus, in the example depicted in the top panel of figure 1, $C_1 = \{1,2,3\}$, $C_2 = \{4\}$, ... , and $C_M = \{R-2, R-1, R\}$.

As such, and ignoring for now a necessary identifying parameter normalization,

$$E\left[s_m \middle| \mathbf{x}\right] = E\left[\sum_{n \in C_m} v_n \middle| \mathbf{x}\right] = \sum_{n \in C_m} E\left[v_n \middle| \mathbf{x}\right] = \sum_{n \in C_m} \left( \frac{\exp\left(\mathbf{x}\boldsymbol{\theta}_n\right)}{\sum_{r=1}^{R} \exp\left(\mathbf{x}\boldsymbol{\theta}_r\right)} \right) \qquad (33)$$

$$= \sum_{n \in C_m} \left( \frac{\exp\left(\theta_{n0} + \mathbf{x}_1\boldsymbol{\theta}_{n1}\right)}{\sum_{r=1}^{R} \exp\left(\theta_{r0} + \mathbf{x}_1\boldsymbol{\theta}_{r1}\right)} \right) = \sum_{n \in C_m} \left( \frac{\exp\left(\theta_{n0} + \mathbf{x}_1\boldsymbol{\theta}_{n1}\right)}{\sum_{m=1}^{M} \sum_{r \in C_m} \exp\left(\theta_{r0} + \mathbf{x}_1\boldsymbol{\theta}_{r1}\right)} \right),$$

for m=1,...,M. Suppose for some k it holds that all elements of the set of slope

[14] Well-known examples are two-to-six-digit NAICS/SIC industry definition codes, one-to-three-level expenditure hierarchies used in the Consumer Expenditure Survey, and the American Time Use Survey's two-, four-, and six-digit time-use categories.

[15] While their purpose was different that this paper's, Cotterman and Peracchi, 1992 offer a useful conceptual discussion of structured vs. unstructured aggregation.

parameter vectors $\{\boldsymbol{\theta}_{n1} | n \in C_k\}$ are identical and equal to (say) $\boldsymbol{\Theta}_k$. Then

$$E[s_k|\mathbf{x}] = \frac{\exp\left(\ln\left(\sum_{q\in C_k}\exp(\theta_{q0})\right) + \mathbf{x}_1\boldsymbol{\Theta}_k\right)}{\exp\left(\ln\left(\sum_{q\in C_k}\exp(\theta_{q0})\right) + \mathbf{x}_1\boldsymbol{\Theta}_k\right) + \sum_{\substack{m=1\\m\neq k}}^{M}\sum_{r\in C_m}\exp(\theta_{r0} + \mathbf{x}_1\boldsymbol{\theta}_{r1})} \quad (34)$$

$$= \frac{\exp\left(\theta_0^k + \mathbf{x}_1\boldsymbol{\Theta}_k\right)}{\exp\left(\theta_0^k + \mathbf{x}_1\boldsymbol{\Theta}_k\right) + \sum_{\substack{m=1\\m\neq k}}^{M}\sum_{r\in C_m}\exp(\theta_{r0} + \mathbf{x}_1\boldsymbol{\theta}_{r1})},$$

where $\theta_0^k = \ln\left(\sum_{q\in C_k}\exp(\theta_{q0})\right)$. That is, the subcategory outcomes $v_n$, $n \in C_k$, aggregate in the sense that they share common slope coefficient vectors. While this is perhaps an obvious characterization of aggregation in the fractional share outcome setting, its deeper implications are less obvious. For instance, aggregation in this sense would imply that the aggregated subcategories all have the same conditional $\mathbf{x}_1$-elasticities but not the same conditional $\mathbf{x}_1$-partial effects.

Given considerations of structured aggregation in this slope-coefficient sense, at least two testing strategies are suggested.[16] The first entails testing jointly the entirety of the equality restrictions implied if the subcategories under all categories having at least two subcategories simultaneously aggregate thusly. Note that if *all* outcome subcategories branching from the aggregated outcome categories aggregated in the slope-coefficient sense, it would follow that

$$E\left[\sum_{n\in C_k} v_n | \mathbf{x}\right] = E[s_k|\mathbf{x}] = \frac{\exp\left(\theta_0^k + \mathbf{x}_1\boldsymbol{\Theta}_k\right)}{\sum_{m=1}^{M}\exp\left(\theta_0^m + \mathbf{x}_1\boldsymbol{\Theta}_m\right)}, \quad k=1,\ldots,M. \quad (35)$$

Given estimates of the subcategory model, this test could be conducted as a straightforward Wald test of the implied parameter restrictions on the (p-1)-vectors $\boldsymbol{\theta}_{r1}$. Under a null hypothesis of slope-coefficient aggregation, such a test statistic would follow a large-sample $\chi^2_{(p-1)(N-M)}$ distribution.[17]

A second and likely more useful approach involves testing separately each candidate subcategory aggregation. For each k this entails testing jointly the equality of the elements in $\{\boldsymbol{\theta}_{n1} | n \in C_k\}$ via Wald tests. In isolation, such test statistics would follow null $\chi^2_{(\#C_k-1)\times(p-1)}$ distributions. However, simultaneous

[16] These testing strategies follow the basic approach of Cramer and Ridder, 1991, except that Cramer and Ridder: are concerned with discrete outcomes; rely on likelihood-ratio tests; and do not address issues of multiple testing that are raised below.

[17] Computational details for these Wald tests are provided in Appendix 2.

testing of such restrictions across as many as M aggregate categories presents a multiple comparisons problem. Using the Benjamini-Yekutieli, 2001, conservative false discovery rate control approach (or related methods) to modify the set of reference rejection p-values may provide some protection against false positives.[18]

## 8. Dirichlet-Multinomial Estimation

Given the discrepancy between the empirical and QMLE pseudo-multinomial second moment structures, it should in principle be possible to improve estimator efficiency if reasonable conditional second-moment assumptions can be made (Gourieroux et al., 1984a). While the share structure of the data provides some guidance on such specifications (e.g. that, like the first moments, the second moments must themselves be bounded) there appears to be little additional general guidance about second-moment specification offered by the data themselves.

A more-structured alternative approach is to postulate a probability model for the data that can describe the important features of the data recognizing, of course, that there is an inconsistency cost that may be incurred if such a probability model is incorrectly specified. Of course, circumstances may arise when the entire conditional probability structure of the multivariate outcomes is of interest, in which case the first-moment estimates offered by MFLOGIT will not be adequately informative.

One working probability model that exhibits underdispersion relative to a multinomial structure and that also accommodates positive probability mass[19] for shares $s_k=0$ and $s_k=1$ is based on a Dirichlet mixture of multinomials ("DM") or multivariate negative hypergeometric (Johnson et al., 1997, pp. 80ff), which is the multivariate version of the beta-binomial distribution (Heckman and Willis, 1977).[20] Imagine that some underlying multivariate T-trial counts $\mathbf{n} = \left[ n_k \right]$, k=1,...,M, follow a conditional DM probability model

[18] One minor detail to be considered is the normalization used to estimate the R-category disaggregated-outcome model. If this normalization is on (say) the R-th category $v_R$, i.e. $\boldsymbol{\theta}_R = 0$, and if R is an element of a multi-element index set $C_k$ (e.g. as depicted in the top panel of figure 1, where $C_M = \{R-2, R-1, R\}$), then the Wald tests for aggregation described above would entail testing the other slope parameters in this branch -- e.g. $\boldsymbol{\theta}_{(R-2)1}$ and $\boldsymbol{\theta}_{(R-1)1}$, with $C_M = \{R-2, R-1, R\}$ -- against the nonstochastic zero vector $\mathbf{0}_{(p-1)}$.

[19] See Vanness and Hanmer, 2010, for a related discussion in a univariate and Bayesian context.

[20] See Guimarães and Lindrooth, 2007, for an interesting econometric application of the DM distribution.

$$DM(\mathbf{n}|\mathbf{x};T) = \frac{\Gamma(T+1)\Gamma\left(\sum_{m=1}^{M} a_m(\mathbf{x})\right)\prod_{m=1}^{M}\Gamma(n_m + a_m(\mathbf{x}))}{\Gamma\left(T + \sum_{m=1}^{M} a_m(\mathbf{x})\right)\prod_{m=1}^{M}\{\Gamma(n_m)\Gamma(a_m(\mathbf{x}))\}}, \qquad (36)$$

with $\mathbf{n} \in \{0,1,\ldots,T\}^M$, $\mathbf{1}'\mathbf{n} = T$, and $a_k(\mathbf{x}) = \exp(\mathbf{x}\boldsymbol{\zeta}_k)$ a natural parameterization. Letting $A(\mathbf{x}) = \sum_{m=1}^{M} a_m(\mathbf{x})$, the conditional marginal mean and marginal variance of the $n_k$ are given by

$$E[n_k|\mathbf{x};T] = \frac{T a_k(\mathbf{x})}{A(\mathbf{x})} \quad \text{and} \quad Var(n_k|\mathbf{x};T) = T\left(\frac{T + A(\mathbf{x})}{1 + A(\mathbf{x})}\right)\left(\frac{a_k(\mathbf{x})}{A(\mathbf{x})}\right)\left(\frac{1}{A(\mathbf{x})}\right), \qquad (37)$$

where the latter expression is seen to be $(T + A(\mathbf{x}))/(1 + A(\mathbf{x}))$ times the conditional marginal variance of the underlying T-trial multinomial distribution, taking $a_k(\mathbf{x})/A(\mathbf{x})$ as the multinomial probabilities.

Given these counts, the shares $s_k \in [0,1]$ -- or, more precisely, $s_k \in \{0,1/T,2/T,\ldots,1\}$ -- are given by $s_k = n_k / T$, so that, in particular,

$$E[s_k|\mathbf{x}] = \frac{a_k(\mathbf{x})}{A(\mathbf{x})} \quad \text{and} \quad Var(s_k|\mathbf{x};T) = \frac{1}{T}\left(\frac{T + A(\mathbf{x})}{1 + A(\mathbf{x})}\right)\left(\frac{a_k(\mathbf{x})}{A(\mathbf{x})}\right)\left(\frac{1}{A(\mathbf{x})}\right). \qquad (38)$$

Since $(T + A(\mathbf{x}))/(T(1 + A(\mathbf{x})))$ is less than one for T>1, $Var(s_k|\mathbf{x};T)$ is smaller than the conditional variance of a one-trial multinomial distribution having a corresponding conditional first-moment or probability structure, which in turn is the quasi-likelihood model for the MFLOGIT. It is also easily shown that $Var(s_k|\mathbf{x};T)$ is decreasing in T.

Since T does not vanish from the probability model for or the conditional variance functions of the $s_j$, the application of the DM(.) model in cases where T does not have a natural interpretation is clearly as an approximation to the true probability model. In some instances, a particular specification for T might be suggested naturally by the nature of the data's measures (e.g. 1,440 integer-measured minutes in a day or some integer-measured number of currency units in a budget). In other instances, however, specifying a value for T will be ad hoc. Moreover, when the measures of the observed share data do not follow a natural lattice structure (i.e. $s_k \in \{0,1/r,2/r,\ldots,r/r\}$) but instead are "continuously" measured, some coarsening of the data that maps $s_k$ into $s_j^c$ will be required for $n_k^c = Ts_k^c$ to satisfy $n_k^c \in \{0,1,\ldots,T\}$ as necessary for the DM distribution.

Two DM models are estimated here for comparison with the MFLOGIT results discussed above. For present purposes, two coarsenings of the data that imply

different values for T are considered. For T=10 and T=100, these are given by $n_{kT}^{c} = \text{floor}\left(Ts_j\right)$ (i.e. $\lfloor Ts_j \rfloor$) for all shares except the share with the largest empirical marginal frequency while the coarsened measure corresponding to that share (k) is given by $n_{jT}^{c} = T - \sum_{\substack{m=1 \\ m \neq j}}^{M} n_{mT}^{c}$ to ensure proper adding up. The resulting parameter and APE estimates are then compared with each other and with those obtained from the MFLOGIT estimator. Given the specification $E\left[s_k \left| \mathbf{x} \right.\right] = a_k\left(\mathbf{x}\right) / A\left(\mathbf{x}\right) = \exp\left(\mathbf{x}\boldsymbol{\zeta}_k\right) / \sum_{m=1}^{M} \exp\left(\mathbf{x}\boldsymbol{\zeta}_m\right)$, and normalizing post-estimation on $\boldsymbol{\zeta}_M$ (since all M $\boldsymbol{\zeta}_M$'s are identified, unlike the MFLOGIT QMLE), some comparability of the point estimates of $\boldsymbol{\zeta}_M$ and $\boldsymbol{\beta}_k$ and, in particular, of the corresponding APE estimates might be expected if the DM likelihood is a reasonable approximation to the true probability model.

## 9. Modeling Financial Asset Portfolio Shares

*Data and Estimation Sample*

This section demonstrates some of the properties of the share model estimators and tests described above by estimating regression models of financial asset portfolio shares. Estimation of portfolio share models has been considered by Heaton and Lucas, 2000, and by Poterba and Samwick, 2001, 2002, among others. The data used are from the combined public use 2001, 2004, and 2007 U.S. Surveys of Consumer Finances (SCF). The SCF is a triennial sample whose collection is sponsored by the U.S. Federal Reserve to provide information on the financial circumstances of U.S. families (for details on the SCF, see Bucks et al, 2009). This combined sample comprises 13,379 household-level observations. Household-level sampling weights provided in the SCF are used in the computation of the APE estimates (note that the SCF weights are designed for within-year but not necessarily across-year weighting). This analysis focuses specifically on financial assets and the ten major subcategories of financial assets defined in the SCF listed in the top panel of table 2. Additional details on the data are provided in Appendix 3.

These data are summarized in figure 2 and in the top panel of table 2. In the sample of 12,723 household observations with defined shares, it might be noted that only six households (.047%) have strictly positive shares for all ten financial asset categories, whereas for 2,348 (18.5%) of the sample's households some financial asset category share is 1.0. Covariates used in the analysis are: age in years of household head (*Age*); a dummy for race of survey respondent (*White*); a dummy for marital status of household head (*Married*); number of children in the household (*Number of Kids*); dummies for educational attainment (*High School Graduate*, *Some College*, and *College Graduate*); and survey year dummies (*Year 2004*, *Year 2007*).[21] Descriptive statistics for these covariates are presented in the

[21] These specifications do not include Total Financial Assets as a covariate. Whether it is appropriate to do so entails issues akin to those involved in whether (cont.)

bottom panel of table 2.

*MFLOGIT Parameter Estimates and Inference*

The MFLOGIT parameter estimates, normalized on the M-th (Other Financial Assets) category, are reported in table 3. Owing to the normalization, it is not straightforward to interpret the signs or magnitudes of these estimates. For purposes of hypothesis testing, however, such relative magnitudes may be informative, so asymptotic standard errors based on the robust sandwich estimator are presented in the table. Due to the large number of parameters ($p \times (M-1) = 90$) estimated, a multiple comparisons situation may arise for hypothesis testing. As such, and to accommodate the mutual dependence parameter estimates, the conservative false discovery rate (FDR) rejection criteria suggested by Benjamini and Yekutieli, 2001, are computed and point estimates with sufficiently small standard errors to meet these criteria are shaded in the table.

The extent of underdispersion in the data (relative to the nominal multinomial quasi-likelihood) can be appreciated in several ways. As suggested in section 6, the difference between the empirical multinomial logit (inverse Hessian) and robust sandwich parameter covariance matrixes is positive definite in the sample (the smallest eigenvalue of the matrix difference is positive). The ratios of non-robust to robust standard errors range from 1.12 to 2.37 over the 90 estimated parameters, suggesting a nontrivial degree of underdispersion.[22]

*Average Partial Effects*

Table 4 presents the weighted APE estimates and bootstrap 95% CIs (based on 500 bootstrap replications) across the M=10 outcomes. Recall that by construction the row sum of the APEs for each covariate will be zero. In this exercise, the schooling attainment and the year indicator variables are treated as groupwise dummies as discussed in Appendix 1.

Overall the estimated patterns of partial effects appear reasonable. The *Age* effects are consistent with the bulk of the estimation sample being of pre-retirement ages (70% are under age 60). Estimated patterns for *Married* and *Number of Kids* accord with the kinds of investment behaviors one would anticipate

---

(cont.)
to include a measure of total expenditure as a covariate in a consumer expenditure share model.

[22] Given the considerable size of the estimated parameter covariance matrix (4,095 unique elements), a comparison bootstrap covariance matrix was estimated using 1,000 bootstrap replicates to check the performance of the standard robust sandwich parameter covariance estimator. Across the 90 parameters estimated in this specification, the mean and median ratios of robust to bootstrap standard errors were .993 and .999, respectively, with a range of 0.86 to 1.06. To the extent that this result generalizes, use of the sandwich estimator in empirical applications of the MFLOGIT estimator may be reasonable.

for such household structures for many of the shares (e.g. *Liquid Assets*, *Quasi-Liquid Retirement Accounts*, *Directly Held Pooled Funds*, *Cash Value Whole Life*). For many of the shares (e.g. *Liquid Assets*, *Quasi-Liquid Retirement Accounts*, *Directly Held Pooled Funds*, *Directly Held Stocks*), the estimated schooling attainment APEs are particularly large and precisely estimated, likely reflecting the schooling variables proxying for the most important human capital and wealth effects in these models. Finally the estimates for the year dummies show some interesting time trends for several of the share outcomes (e.g. *Quasi-Liquid Retirement Accounts*, *Directly Held Stocks*, *Cash Value Whole Life*, *Other Managed Assets*).

*MFLOGIT Estimator Performance*

For the conditional moment tests described in section 6, this application specifies the $\mathbf{\Lambda}(\mathbf{x})$ based on the vingtiles of each of the $\hat{E}\left[s_{ik} \,|\mathbf{x}_i\right]$ resulting in L=20 test indicators for each $s_k$ outcome. The sampling variation of these test indicators is estimated using 500 bootstrap replications, resulting in 95% CIs for each indicator-outcome combination as well as an overall $\chi^2_{LM-1}$ goodness-of-fit test statistic for the full multivariate model.

The results are summarized in figure 3, which depicts for each of the M=10 share outcomes and at each of the L=20 vingtiles the test statistic point estimate ($N \times \lambda_{mq}$; dark line) and its bootstrap 95% CI (shaded area). Multiple comparisons concerns notwithstanding, the results depicted in figure 3 show only relatively few instances where the 95% CIs fail to cover zero, and these are typically in the tails with the exception of (*Liquid Assets*, *Directly Held Stocks*). In a few cases (*Quasi-Liquid Retirement Accounts*, *CDs*, *Directly Held Stocks*, *Other Financial Assets*) there is at least a suggestion of a U- or inverse U-shaped pattern across the vingtiles (underprediction in the tails and overprediction in the center, or vice-versa). Finally, the overall conditional moment $\chi^2_{199}$ test statistic is 674.2 (p-value effectively zero).

Several general model performance statistics are summarized in table 5. For this exercise, the MFLOGIT estimator was compared with a set of M=10 univariate linear regression (estimated by OLS) and univariate Tobit estimators that include the same covariate vector as used in the MFLOGIT models. The first performance criteria were out-of-sample MPE and MSE as assessed by an 80/20 cross-validation averaged over 100 replications. For this exercise, the linear model dominates MFLOGIT and Tobit on the MPE criterion while MFLOGIT generally dominates the linear model and Tobit on the MSE criterion. One obvious potential drawback of the linear model is that its predictions are not restricted to obey the [0,1] interval bounds. The rightmost columns of table 5 summarize the extent to which this is of concern in this empirical context. The two rightmost columns provide the cross-validation (averaged over replicates) and the in-sample frequencies with which the linear model predictions are less than zero (predictions greater than one were not observed). The cross-validation and in-sample frequencies are quite similar across the share categories, and suggest that the out-of-interval prediction problem is most severe for share categories with the smallest marginal empirical frequencies

(e.g. *Directly Held Bonds*, *Other Managed Assets*).

*Aggregation Testing*

Figure 4 depicts the tree structure of the SCF categories. R=20 subcategories are specified for this analysis of structured aggregation as described in section 7, with the aggregation tests described in that section applied to the three branches with $\#C_k>1$ (Liquid Assets, Quasi-Liquid Retirement Funds, and Directly Held Bonds).[23] The MFLOGIT point estimates of the $\boldsymbol{\theta}_n$ parameters corresponding to those branches and the $\chi^2$ aggregation test statistics are presented in table 6.

The overall aggregation test for all three branches strongly rejects slope-parameter aggregation across all three branches. The individual category tests for aggregation also strongly suggest that aggregation is not reasonable for any of these three categories.[24] Indeed, casual inspection of the individual slope-parameter point estimates indicates considerable variability within each main category.

*Dirichlet-Multinomial Estimates*

Two variants of the DM model were estimated here, these reflecting different degrees of data coarsening as discussed in section 8. Specifically, models for T=10 and T=100 were estimated. A useful, direct comparison between the DM and MFLOGIT estimators of concern here is in terms of their performance in estimating the conditional first-moment structure of the data, with these summarized most straightforwardly by comparing the point estimates of the estimated APEs. This comparison is presented in table 7. In most cases (the exceptions being the shaded cells) the MFLOGIT and DM APE estimates have the same signs. Quite broadly, the magnitudes of the point APE estimates roughly comparable, but typically larger for MFLOGIT than for either the T=10 or T=100 DM estimators. One consideration beyond the comparison of the APEs is the possible efficiency gain from using a full-likelihood estimation approach (DM) over a first-moment estimation approach (MFLOGIT). It turns out in this application that the efficiency gains are small at best. For the $p\times(M-1)=90$ normalized parameters, the median of the ratio of MFLOGIT to DM robust standard errors is 1.05 for the T=100 model and 0.99 for the T=10 specification.

One test of overall goodness of fit for likelihood-based models like the DM is the information matrix (IM) test proposed by White, 1982 (see also Chesher, 1983, Lancaster, 1984, and Orme, 1990). With such a large model as that estimated

[23] The subcategories under the Quasi-Liquid Retirement Accounts and Other Managed Assets categories were ignored for computational reasons.

[24] With only three subaggregate branches specified and in light of the large values of the realized aggregation test statistics, the multiple comparisons issues discussed above in section 7 are effectively irrelevant and are ignored here.

here, it is not obvious whether the asymptotic properties of the IM test can be invoked given the available sample size, as the full model IM test has 4,095 degrees of freedom with a sample size of 12,723.[25] Such considerations notwithstanding, the IM test statistic is computed for the T=10 and T=100 specifications using the method suggested by Lancaster, 1984. The DM model's $\chi^2_{4095}$ test statistics are 11,261 (T=10) and 10,575 (T=100) which, while suggesting a slightly better fit for T=100 than for T=10, still both have effective p-values of zero. For comparison, however, the MFLOGIT quasi-likelihood IM test statistic based on the coarsened T=10 data is 1.72E+07. More concretely perhaps, for all $p \times M = 100$ parameter estimates, the ratio of robust to inverse-Hessian standard error estimates ranges from .69 to 1.12 (median 1.02) for the T=10 specification and from .62 to 1.20 (median 1.03) for the T=100 specification.

It is also possible (though not undertaken here) to compute an overall $\chi^2$ goodness-of-fit test statistic (see Andrews, 1988) for the coarsened outcome cells or interesting aggregates thereof. In this spirit, the performance of the DM estimator in modeling the overall conditional probability structure of the data is summarized in figures 5 and 6 and in table 8. Figure 5 depicts for three of the share outcomes the coarsened marginal empirical frequency distribution juxtaposed with the estimated marginal empirical frequencies from the T=10 and T=100 specifications (the latter computed as the sample averages of the conditional frequencies). Figure 6 shows Lorenz curve summaries in which are plotted the T=100 marginal cumulative probability estimates against the corresponding cumulative 101 coarsened cell frequencies in the data. For *Liquid Assets*, *Quasi-Liquid Retirement Accounts*, *Directly Held Pooled Funds*, and *Directly Held Stocks* there are some noteworthy fit problems. Finally, table 8 summarizes the quality of fit of the DM estimators at the s=0 and s=1 endpoints. For most of the share outcomes, the T=100 estimator provides a much closer fit to the marginal empirical frequencies than does the T=10 estimator.

Overall, then, the merits of estimating and conducting inference using DM models in a full likelihood context are mixed. One presumably trades off some robustness relative to approaches like MFLOGIT in estimating first-moment models, with the differences in corresponding APE point estimates between the two approaches perhaps being nontrivial, at least in this empirical exercise. However, if one is interested in estimating the full conditional probability structure of such models (perhaps at the cost of using coarsened data), the DM approach may be a useful strategy to consider.

## 10. Summary and Discussion

With a central focus on estimation of conditional means, this paper has

[25] For comparison with the MFLOGIT quasi-likelihood, the parameter estimates are normalized against the *Other Financial Assets* baseline, thus reducing the dimensionality of the test.

proposed econometric strategies for estimating regression models of economic share data in cases where shares assume values of zero and one with nontrivial probability. The main contribution has been to explore properties of an extension to share models of the fractional regression methodologies proposed by Papke and Wooldridge.

Several outstanding issues should be important items on the future research agenda. First involves further considerations of share category aggregation or disaggregation beyond those offered in section 7. A second consideration involves "covariate adjustment." For instance, in applications where understanding the determinants of shares net of the influence of some conditioning covariates is a prominent issue, the manner in which covariates are netted out is critical. How to effect this in a framework that involves bounded shares that obey adding-up restrictions is an open question.

Third, in some empirical contexts (e.g. like the portfolio share example presented here) there is necessarily selection on subsamples for which shares are defined (by nonzero denominators). This raises an important issue regarding for which populations inferences drawn from the estimated share models are relevant. While in some sense a garden variety selection problem, the issue of how to address this in the MFLOGIT or related estimation contexts remains unresolved.

Finally, a possible extension of this line of work would be to consider analogs to conditional logit (Hausman and McFadden, 1984) estimation that would fit into the fractional outcome data setting. While analogies to the discrete outcome random utility (RUM) framework are not obvious, the implied moment structures of such models might offer statistical tools for analyzing data where outcome-specific covariates or attributes are available. For instance, one might imagine a time use study in which time prices or wage rates for each outcome are available. Briefly, consider a situation where a vector of attributes for the k-th outcome is given by $\mathbf{w}_k$ with the vector $\mathbf{w} = [\mathbf{w}_k, \mathbf{w}_{-k}]$ describing the entirety of such attributes over all outcomes. Then the first-moment share structure corresponding to a standard RUM model (with normalization $\mathbf{w}_M$=0) would be given by:

$$E[s_k | \mathbf{w}_k, \mathbf{w}_{-k}] = E[s_k | \mathbf{w}] = \frac{\exp(\mathbf{w}_k\boldsymbol{\delta})}{1 + \sum_{m=1}^{M-1} \exp(\mathbf{w}_m\boldsymbol{\delta})}, \quad k=1,\ldots,M, \tag{39}$$

which could be extended to accommodate $\mathbf{x}$ in a mixed-logit structure,

$$E[s_k | \mathbf{x}, \mathbf{w}] = \frac{\exp(\mathbf{x}\boldsymbol{\beta}_k + \mathbf{w}_k\boldsymbol{\delta})}{1 + \sum_{m=1}^{M-1} \exp(\mathbf{x}\boldsymbol{\beta}_m + \mathbf{w}_m\boldsymbol{\delta})}, \quad k=1,\ldots,M. \tag{40}$$

Of course, in the absence of an underlying RUM structure, the influences of such outcome-specific attributes could also be captured in a standard MFLOGIT conditional mean model with

$$E\left[s_k \,|\, \mathbf{x}, \mathbf{w}\right] = \frac{\exp\left(\mathbf{x}\boldsymbol{\beta}_k + \mathbf{w}\boldsymbol{\delta}_k\right)}{1 + \sum_{m=1}^{M-1} \exp\left(\mathbf{x}\boldsymbol{\beta}_m + \mathbf{w}\boldsymbol{\delta}_m\right)}, \quad k=1,\ldots,M, \tag{41}$$

in which the $\boldsymbol{\delta}_k$ would describe different patterns of own-**w** vs. cross-**w** effects across k=1,...,M.


## Acknowledgements

I am indebted to participants at presentations of various aspects of this work at Catholic University of Rome, Michigan State University, the University of Arizona, the University of Coimbra, University College Dublin, and UW-Madison, as well as to Marguerite Burns, Ben Craig, Alberto Holly, Steve Koch, José Murteira, Stephanie Robert, Nilay Shah, João Santos Silva, and Dave Vanness for their thoughtful comments, suggestions, and discussions. In addition, Badi Baltagi and Jeff Wooldridge provided some helpful guidance with the literature. All these colleagues, of course, are absolved from any blame for the paper's shortcomings. Partial financial support from the Robert Wood Johnson Foundation Health & Society Scholars Program is acknowledged. Some of this work was completed as a visiting scholar at the UCD Geary Institute, which provided brilliant hospitality.

## References


Aitchison, J. 1982. "The Statistical Analysis of Compositional Data." *JRSS-B* 44: 139-177.

Andrews, D.W.K. 1988. "Chi-Square Diagnostic Tests for Econometric Models: Theory." *Econometrica* 56: 1419-1453.

Benjamini, Y. and D. Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing under Dependency." *Annals of Statistics* 29: 1165-1188.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63: 841-890.

Billheimer, D., P. Guttorp, and W.F. Fagan. 2001. "Statistical Interpretation of Species Composition." *JASA* 96: 1205-1214.

Brown, B.W. and M.B. Walker. 1989. "The Random Utility Hypothesis and Inference in Demand Systems." *Econometrica* 57: 815-829.

Bucks, B.K., A.B. Kennickell, T.L. Mach, and K.B. Moore. 2009. "Changes in U.S. Family Finances from 2004 to 2007: Evidence from the Survey of Consumer Finances." *Federal Reserve Bulletin* 95: A1-A55.

Chavas, J.-P. and K. Segerson. 1987. "Stochastic Specification and Estimation of Share Equation Systems." *Journal of Econometrics* 35: 337-358.

Chesher, A. 1983. "The Information Matrix Text: Simplified Calculation via a Score Test Implementation." *Economics Letters* 13: 45-48.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau. 1975. "Transcendental Logarithmic Utility Functions." *American Economic Review* 65: 367-383.

Considine, T.J. and T.D. Mount. 1984. "The Use of Linear Logit Models for Dynamic Input Demand Systems." *Review of Economics and Statistics* 66: 434-443.

Cotterman, R. and F. Peracchi. 1992. "Classification and Aggregation: An Application to Industrial Classification in CPS Data." *Journal of Applied Econometrics* 7: 31-51.

Cramer, J.S. and G. Ridder. 1991. "Pooling States in the Multinomial Logit Model." *Journal of Econometrics* 47: 267-272.

Crawford, D.L., R.A. Pollak, and F. Vella. 1998. "Simple Inference in Multinomial and Ordered Logit." *Econometric Reviews* 17: 289-299.

Dubin, J.A. 2007. "Valuing Intangible Assets with a Nested Logit Market Share Model." *Journal of Econometrics* 139: 285-302.

Fry, J.M., T.R.L. Fry, and K.R. McLaren. 1996. "The Stochastic Specification of Demand Share Equations: Restricting Budget Shares to the Unit Simplex." *Journal of Econometrics* 73: 377-385.

Gourieroux, C., A. Monfort, and A. Trognon. 1984a. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica* 52: 681-700.

Gourieroux, C., A. Monfort, and A. Trognon. 1984b. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica* 52: 701-720.

Graybill, F.A. 1983. *Matrixes with Applications in Statistics*. Belmont, CA: Wadsworth Publishing.

Guimarães, P. and R.C. Lindrooth. 2007. "Controlling for Overdispersion in Grouped Conditional Logit Models: A Computationally Simple Application of Dirichlet-Multinomial Regression." *Econometrics Journal* 10: 439-452.

Hansen, B. 2010. *Econometrics*. Textbook manuscript, Dept. of Economics, UW-Madison.

Hausman, J. and D. McFadden. 1984. "Specification Tests for the Multinomial Logit Model." *Econometrica* 52: 1219-1240.

Heaton, J. and D. Lucas. 2000. "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk." *Journal of Finance* 55: 1163-1198.

Heckman, J.J. and R.J. Willis. 1977. "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women." *Journal of Political Economy* 85: 27-58.

Hill, M.A. 1983. "Female Labor Force Participation in Developing and Developed Countries-Consideration of the Informal Sector." *Review of Economics and Statistics* 65: 459-468.

Intriligator, M.D. 1971. *Mathematical Optimization and Economic Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Johnson, N.L., S. Kotz, and N. Balakrishnan. 1997. *Discrete Multivariate Distributions*. New York: Wiley.

Koch, S.F. 2010. "Fractional Multinomial Response Models with an Application to Expenditure Shares." Mimeo.

Kooreman, P. and A. Kapteyn. 1987. "A Disaggregated Analysis of the Allocation of Time within the Household." *Journal of Political Economy* 95: 223-249.

Lancaster, T. 1984. "The Covariance Matrix of the Information Matrix Test." *Econometrica* 52: 1051-1053.

Lee, L.-F. and M.M. Pitt. 1986. "Microeconometric Demand System with Binding Nonnegativity Constraints: The Dual Approach." *Econometrica* 54: 1237-1242.

McElroy, M.B. 1987. "Additive General Error Models for Production, Cost, and Derived Demand or Share Systems." *Journal of Political Economy* 95: 737-757.

Morey, E.R., D. Waldman, D. Assane. and D. Shaw. 1995. "Searching for a Model of Multiple-Site Recreation Demand That Admits Interior and Boundary Solutions." *American Journal of Agricultural Economics* 77: 129-140.

Mullahy, J. 2004. "Squandering Time? Economic Aspects of Children's Time Use." Working Paper, University of Wisconsin.

Mullahy, J. and S.A. Robert. 2010. "No Time to Lose: Time Constraints and Physical Activity in the Production of Health." *Review of Economics of the Household* (in press).

Orme, C. 1990. "The Small-Sample Performance of the Information-Matrix Test." *Journal of Econometrics* 46: 309-331.

Papke, L.E. and J.M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates." *Journal of Applied Econometrics* 11: 619-632.

Papke, L.E. and J.M. Wooldridge. 2008. "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates." *Journal of Econometrics* 145: 121-133.

Poterba, J.M. and A.A. Samwick. 2001. "Household Portfolio Allocation over the Life Cycle." Chapter 2 in S. Ogura, T. Tachibanaki, and D.A. Wise, eds., *Aging Issues in the United States and Japan*. Chicago: University of Chicago Press for NBER.

Poterba, J.M. and A.A. Samwick. 2002. "Taxation and Household Portfolio Composition: U.S. Evidence from the 1980s and 1990s." *Journal of Public Economics* 87: 5-38.

Ramalho, E.A., J.J.S. Ramalho, and J.M.R. Murteira. 2010. "Alternative Estimating and Testing Empirical Strategies for Fractional Regression Models." *Journal of Economic Surveys* (forthcoming).

Sivakumar, A. and C. Bhat. 2002. "Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis." *Transportation Research Record* 1790: 80-88.

Vanness, D.J. and J. Hanmer. 2010. "Health Utility Crosswalks: A Bayesian Beta Regression Approach." Presented at the 3rd Biennial Conference of the American Society of Health Economists, Cornell University.

Wales, T.J. and A.D. Woodland. 1977. "Estimation of the Allocation of Time for Work, Leisure, and Housework." *Econometrica* 45: 115-132.

White, H. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50: 1-25.

Woodland, A.D. 1979. "Stochastic Specification and the Estimation of Share Equations." *Journal of Econometrics* 10: 361-383.

Ye, X. and R.M. Pendyala. 2005. "A Model of Daily Time Use Allocation Using Fractional Logit Methodology." In H.S. Mahmassani, Ed. *Transportation and Traffic Theory: Flow, Dynamics, and Human Interaction*. Oxford: Pergamon, Elsevier Science Ltd.

## Appendix 1: Average Partial Effects for MFLOGIT Models

The general formula for the APE is

$$\widehat{APE}_{mk} = \sum_{i=1}^{N}\left(\frac{w_i}{\sum_{n=1}^{N} w_n}\right)\widehat{PE}_{mki} = \sum_{i=1}^{N}\left(\frac{w_i}{\sum_{n=1}^{N} w_n}\right)\frac{\delta\hat{E}\left[y_{im}\left|\mathbf{x}_i\right.\right]}{\delta x_{ik}},$$

where "$\delta$" denotes either "$\Delta$" or "$\partial$" and where the $w_i$ are nonnegative weights that may be used to estimate, for instance, population average APEs ($w_i=1$ for all i gives constant weight 1/N). Note that $\sum_{m=1}^{M}\widehat{APE}_{mk} = 0$ due to the adding-up restriction. In the case where $x_{ik}$ is a dummy variable, $APE_{mk}$ is computed as the (perhaps weighted) sample average, evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, of the difference

$$PE_{mki} = \frac{\Delta E\left[s_{im}\left|\mathbf{x}_i\right.\right]}{\Delta x_{ik}} = \frac{\exp\left(\mathbf{x}_{-k,i}\boldsymbol{\beta}_{m,-k} + \beta_{mk}\right)}{1+\sum_{j=1}^{M-1}\exp\left(\mathbf{x}_{-k,i}\boldsymbol{\beta}_{j,-k} + \beta_{jk}\right)} - \frac{\exp\left(\mathbf{x}_{-k,i}\boldsymbol{\beta}_{m,-k}\right)}{1+\sum_{j=1}^{M-1}\exp\left(\mathbf{x}_{-k,i}\boldsymbol{\beta}_{j,-k}\right)},$$

or the derivative

$$PE_{mki} = \frac{\partial E\left[s_{im}\left|\mathbf{x}_i\right.\right]}{\partial x_{ik}} = \exp\left(\mathbf{x}_i\boldsymbol{\beta}_m\right)\times\frac{\left(1+\sum_{j=1}^{M-1}\exp\left(\mathbf{x}_i\boldsymbol{\beta}_j\right)\right)\times\beta_{mk} - \sum_{j=1}^{M-1}\exp\left(\mathbf{x}_i\boldsymbol{\beta}_j\right)\times\beta_{jk}}{\left(1+\sum_{j=1}^{M-1}\exp\left(\mathbf{x}_i\boldsymbol{\beta}_j\right)\right)^2},$$

where $\mathbf{x}_{-k,i}$ is the vector $\mathbf{x}_i$ for the i-th observation with the k-th element excluded.[26]

Hansen, 2010, gives the lower and upper limits of the $C_2$ $(1-\alpha)$-CI for arbitrary statistic $\hat{\theta}$ as:

$$CI_L\left(\alpha,\hat{\theta},\left\{\hat{\theta}_b\right\}\right) = \hat{\theta} - q_{\left\{\hat{\theta}_b\right\}}\left(1-.5\alpha\right) \text{ and } CI_U\left(\alpha,\hat{\theta},\left\{\hat{\theta}_b\right\}\right) = \hat{\theta} - q_{\left\{\hat{\theta}_b\right\}}\left(.5\alpha\right),$$

where $q_{\left\{\hat{\theta}_b\right\}}(\tau)$ is the $\tau$-th quantile of the bootstrap sampling distribution $\left\{\hat{\theta}_b\right\}$. When $\hat{\theta} = \widehat{APE}_{mk}$, there are at least three possibilities for bootstrapping to estimate the $(1-\alpha)$-CI:

[26] When dummy variables are included in **x** as mutually exclusive and exhaustive (save an "omitted" category) members of sets of indicators -- e.g. race/ethnicity groups, educational attainment indicators -- setting up the discrete APE to capture the proper counterfactual is accomplished by zeroing out all of the group's dummy variables at baseline (i.e. setting all group dummies for all observations equal to the omitted category) and then setting the $x_{ik}$ the variable in question equal to one for all observations.

(a) Accommodate variation in $\mathbf{x}$ and $\hat{\boldsymbol{\beta}}$: Bootstrap the APEs via bootstrap draws $(\mathbf{y}_b, \mathbf{x}_b) = \left[\left(\mathbf{y}_{i(b)}, \mathbf{x}_{i(b)}\right)\right]$ from $(\mathbf{y}, \mathbf{x})$ that estimate $\hat{\boldsymbol{\beta}}_b$ and, correspondingly, $APE_{mkb}\left(\mathbf{x}_b; \hat{\boldsymbol{\beta}}_b\right) = \frac{1}{N}\sum_{i(b)=1}^{N} PE_{mki(b)}\left(\mathbf{x}_{i(b)}; \hat{\boldsymbol{\beta}}_b\right)$; accumulate the $APE_{mkb}\left(\mathbf{x}_b; \hat{\boldsymbol{\beta}}_b\right)$ in the B-vector $\left[APE_{mkb}\left(\mathbf{x}_b; \hat{\boldsymbol{\beta}}_b\right)\right]$, and base CIs on suitable percentiles of $\left[APE_{mkb}\left(\mathbf{x}_b; \hat{\boldsymbol{\beta}}_b\right)\right]$.

(b) Accommodate variation in $\hat{\boldsymbol{\beta}}$ only: Bootstrap the APEs via bootstrap draws $(\mathbf{y}_b, \mathbf{x}_b)$ from $(\mathbf{y}, \mathbf{x})$ that estimate $\hat{\boldsymbol{\beta}}_b$ and, correspondingly, $APE_{mkb}\left(\mathbf{x}; \hat{\boldsymbol{\beta}}_b\right) = \frac{1}{N}\sum_{i=1}^{N} PE_{mki}\left(\mathbf{x}_i; \hat{\boldsymbol{\beta}}_b\right)$; accumulate the $APE_{mkb}\left(\mathbf{x}; \hat{\boldsymbol{\beta}}_b\right)$ in the B-vector $\left[APE_{mkb}\left(\mathbf{x}; \hat{\boldsymbol{\beta}}_b\right)\right]$; and base CIs on suitable percentiles of $\left[APE_{mkb}\left(\mathbf{x}; \hat{\boldsymbol{\beta}}_b\right)\right]$.

(c) Accommodate variation in $\hat{\boldsymbol{\beta}}$ only, with weighting: Bootstrap the APEs via *unweighted* bootstrap replicates $(\mathbf{y}_b, \mathbf{x}_b)$ from $(\mathbf{y}, \mathbf{x})$ that estimate $\hat{\boldsymbol{\beta}}_b$ and, correspondingly, $APE^{w}_{mkb}\left(\mathbf{x}; \hat{\boldsymbol{\beta}}_b\right) = \sum_{i=1}^{N}\left(\frac{w_i}{\sum_{n=1}^{N} w_n}\right) PE_{mki}\left(\mathbf{x}_i; \hat{\boldsymbol{\beta}}_b\right)$; accumulate the $APE^{w}_{mkb}\left(\mathbf{x}; \hat{\boldsymbol{\beta}}_b\right)$ in the B-vector $\left[APE^{w}_{mkb}\left(\mathbf{x}; \hat{\boldsymbol{\beta}}_b\right)\right]$; and base CIs on suitable percentiles of $\left[APE^{w}_{mkb}\left(\mathbf{x}; \hat{\boldsymbol{\beta}}_b\right)\right]$.

The estimates presented in tables 4 and 7 are based on approach (c), but in this application it turns out that the CIs estimated using (a), (b), or (c) are quite similar (tables showing alternatives are available on request).

## Appendix 2: Computation of Wald Test Statistics for Slope-Parameter Aggregation

Let $\boldsymbol{\theta} = \text{vec}\left(\left[\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{N-1}\right]\right)$ denote the $p(N-1)\times 1$ vector of estimable parameters and suppose N>M. Let $\boldsymbol{\theta}$ be defined such that: (a) disaggregated subcategory outcomes from the same multi-subcategory branch have corresponding $\boldsymbol{\theta}_k$ that are adjacent in $\boldsymbol{\theta}$; (b) that the first element of each $\boldsymbol{\theta}_k$ is the "intercept" parameter, i.e. $\boldsymbol{\theta}_k = \left[\theta_{k0}, \boldsymbol{\theta}_{k1}'\right]'$; and (c) the K ($0 \le K < M$) aggregate outcome categories that branch to only single outcome subcategories correspond to the bottom rows of $\boldsymbol{\theta}$. It is assumed that the normalization $\boldsymbol{\theta}_N = \mathbf{0}$ has been imposed and, for simplicity of exposition, that subcategory $v_N$ is not being considered for aggregation with other subcategories. Then the Wald test statistics described in section 7 are given by the standard formula

$$\text{wald}\left(\hat{\boldsymbol{\theta}}\right) = \left(\mathbf{R}\hat{\boldsymbol{\theta}}\right)' \left(\mathbf{R}\hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}\right)\mathbf{R}'\right)^{-1}\left(\mathbf{R}\hat{\boldsymbol{\theta}}\right),$$

where $\hat{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}\right)$ is the estimate of the robust MFLOGIT covariance estimator given in (25).

The test for the entirety of slope-parameter aggregations specifies $\mathbf{R}$ as follows. Let

$$\mathbf{J}_p = \left[\mathbf{0}_{(p-1)\times 1},\ \mathbf{I}_{(p-1)}\right], \quad \dim\left(\mathbf{J}_p\right) = (p-1)\times p,$$

and

$$\mathbf{R}_k = \left[\left[\text{blockdiag}_{(\#C_k - 1)}\left(\mathbf{J}_p\right)\right],\ \begin{bmatrix} -\mathbf{J}_p \\ -\mathbf{J}_p \\ \vdots \\ -\mathbf{J}_p \end{bmatrix}\right], \qquad k=1,\ldots,(M-K),$$

$$\dim\left(\mathbf{R}_k\right) = \left((p-1)(\#C_k - 1)\right)\times\left(p\,\#C_k\right).$$

Then

$$\mathbf{R} = \left[\text{blockdiag}_{(k=1,\ldots,(M-K))}\left(\left[\mathbf{R}_k\right]\right),\ \mathbf{0}_{(p-1)(N-M)\times p(K-1)}\right],$$

$$\dim\left(\mathbf{R}\right) = (p-1)(N-M)\times p(N-1).$$

In this case $\text{wald}\left(\hat{\boldsymbol{\theta}}\right)$ follows a $\chi^2_{(p-1)(N-M)}$ distribution under the null.

For testing subcategory aggregation under a single outcome aggregate (say the m-th), the corresponding specification of $\mathbf{R}$ is

$$\mathbf{R} = \left[\mathbf{0}_{(p-1)(\#C_m-1)\times p\left(\sum_{k<m}\#C_k\right)},\ \mathbf{R}_m,\ \mathbf{0}_{(p-1)(\#C_m-1)\times p\left(N-1-\sum_{k\le m}\#C_k\right)}\right],$$

with $\mathbf{R}_m$ specified as $\mathbf{R}_k$ above and $\dim(\mathbf{R}) = (p-1)(\#C_m - 1) \times p(N-1)$. In this case $\text{wald}(\hat{\boldsymbol{\theta}})$ follows a $\chi^2_{(p-1)(\#C_m-1)}$ distribution under the null.

## Appendix 3: Survey of Consumer Finances Data

Downloaded in February and March 2010, the public use data files are contained in:

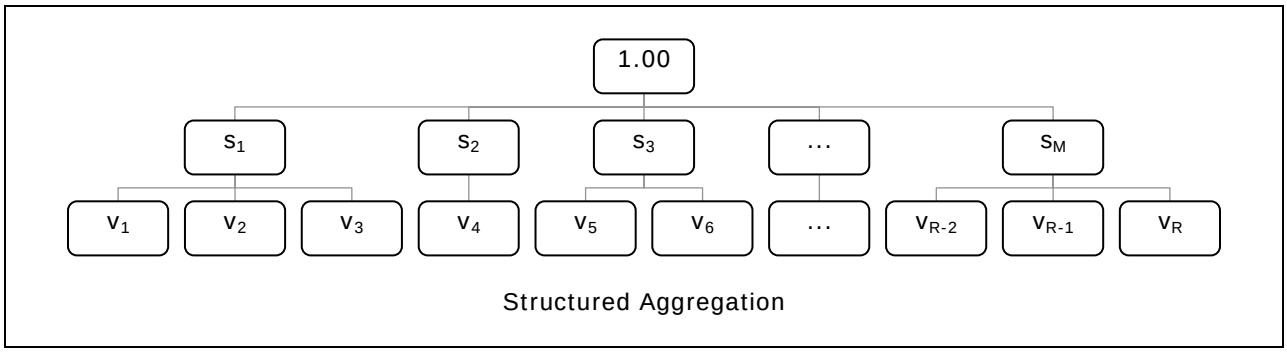
http://www.federalreserve.gov/pubs/oss/oss2/2007/scfp2007.zip
http://www.federalreserve.gov/pubs/oss/oss2/2004/scfp2004.zip
http://www.federalreserve.gov/pubs/oss/oss2/2001/scfp2001.zip

The financial asset data used in this analysis are derived at the household level by averaging over the five SCF "implicates" (imputed replicates) for each category, summing these household averages to obtain household total financial assets, and then obtaining the category shares as the ratios of the category implicate averages to this household total, so-defined. Specifically, for each survey year's sample, the Stata v.10 code used to compute the category shares is:

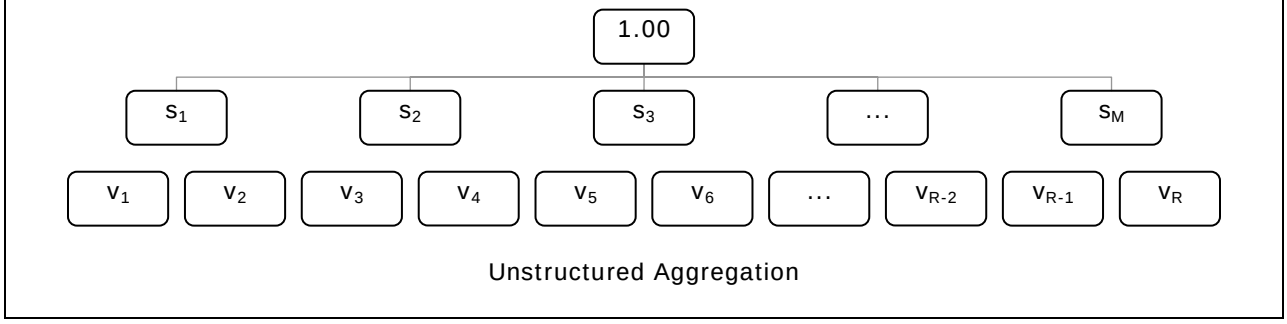
```
sort yy1
mac def nlfin "liq cds nmmf savb stoc bond cashli othma retq othf"
mac def nlfina "aliq acds anmmf asavb astoc abond acashli aothma aretq aothf"
foreach var of varlist $nlfin {
  by yy1: egen a`var'=mean(`var')
}
egen afin=rowtotal($nlfina)
foreach var of varlist $nlfin {
  gen sh`var'=a`var'/afin
}
```

This results in defined shares for 12,723 of the 13,379 total household-level observations.

Figure 1
Structured and Unstructured Aggregation: Examples

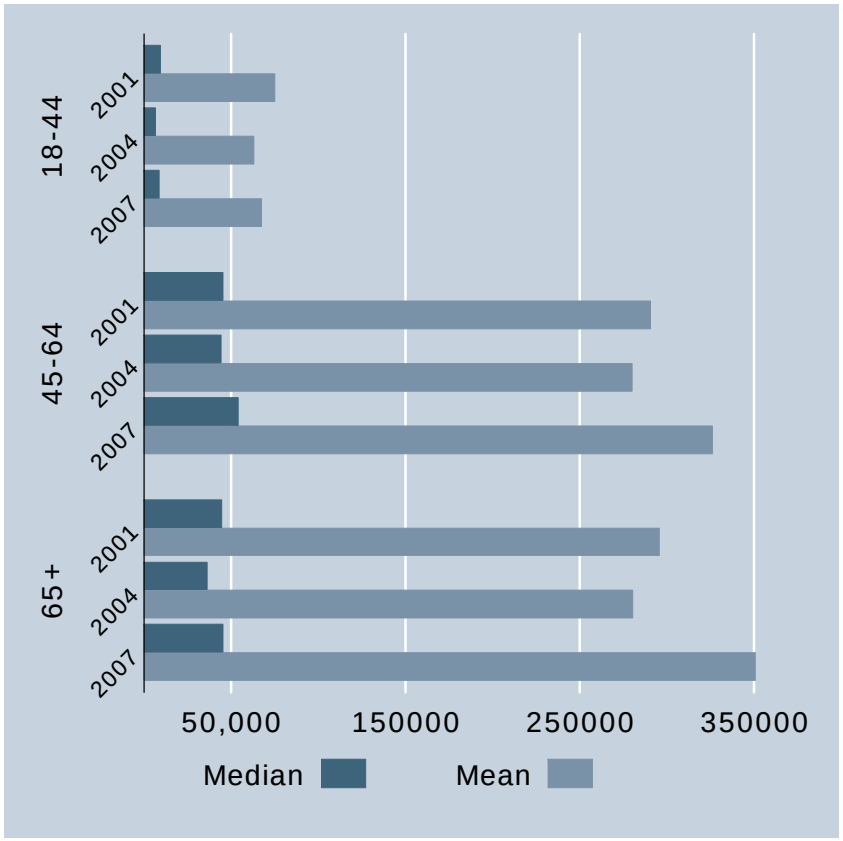



Structured Aggregation




Unstructured Aggregation

Figure 2
Survey of Consumer Finances, Combined 2001, 2004, 2007 Sample:
Financial Assets (N=13,379) and Financial Asset Shares (N=12,723), by Age and Year (Weighted)



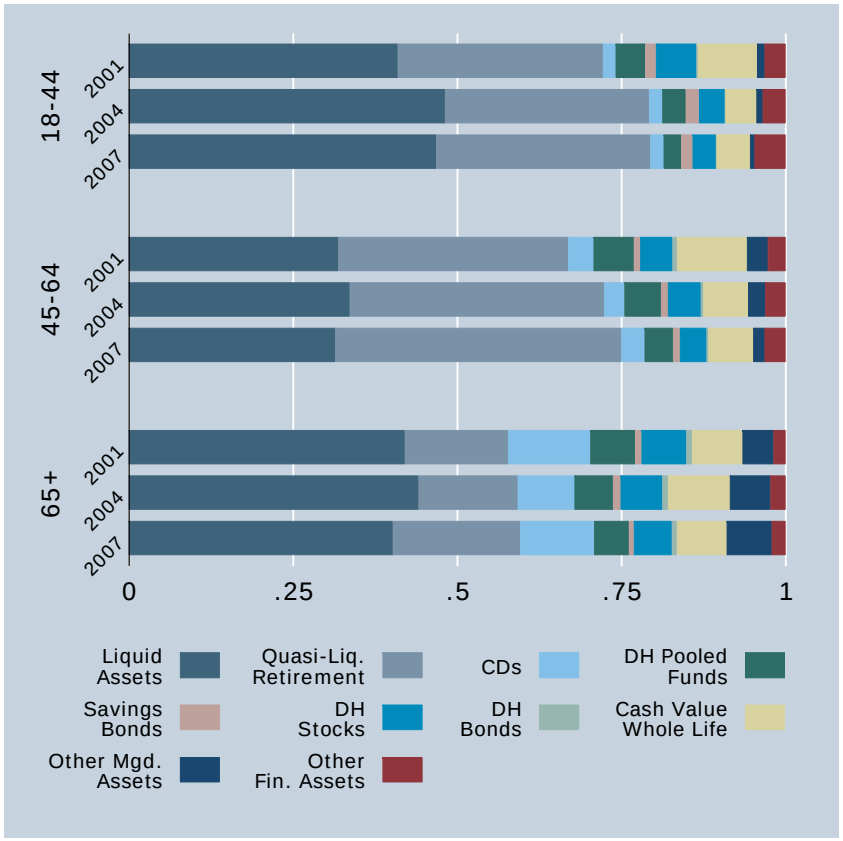



Figure 3

MFLOGIT Conditional Moment (CM) Tests Based on Percentiles of Estimated E[s|**x**]: Test Statistics $N \times \lambda_{mq}$ and Bootstrap 95%-CI Lower and Upper Bounds (CIs based on 500 Bootstrap Replications and Hansen $C_2$ Method)

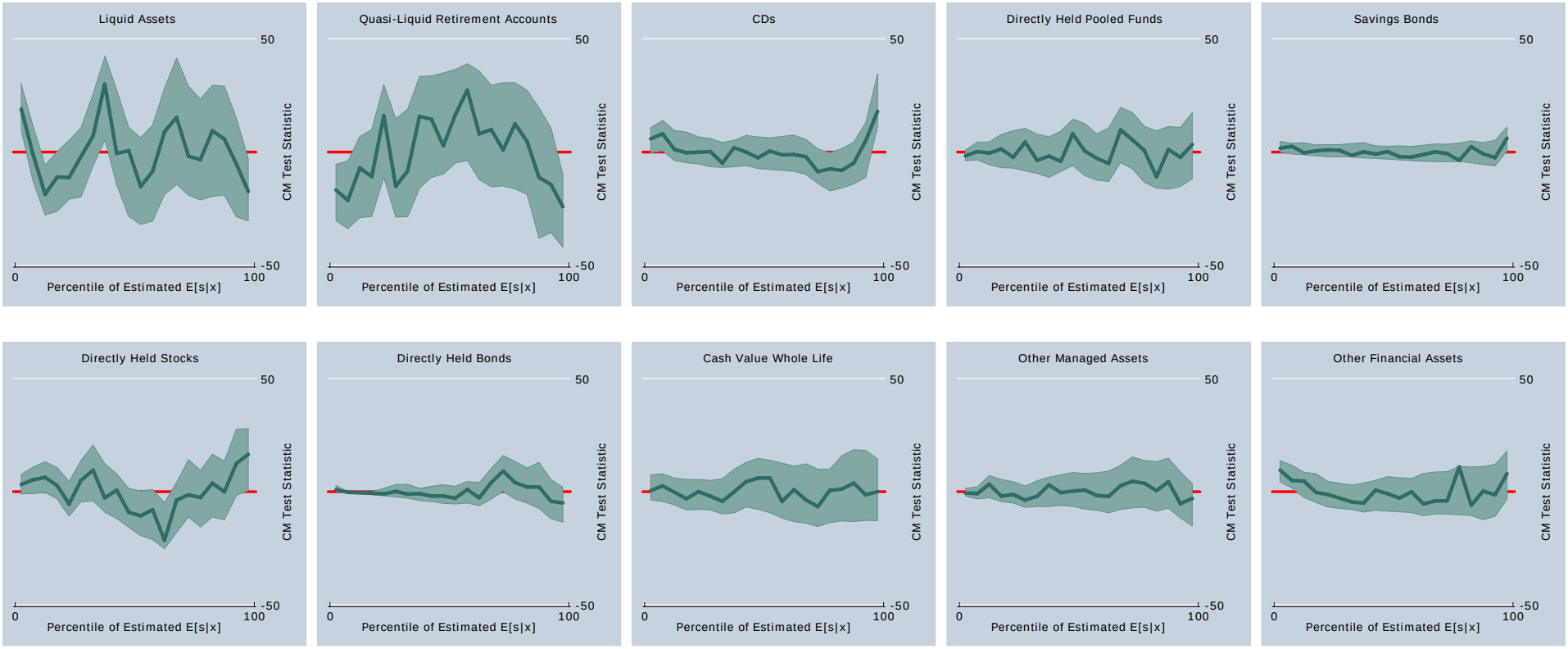


Figure 4
Surveys of Consumer Finances: Financial Asset Share Category Tree Structure
(Subcategories Shaded in Light Gray Are the N=20 $v_n$ Used in the Empirical Analysis;
Subcategories Shaded in Dark Gray Are Not Used in Empirical Analysis)

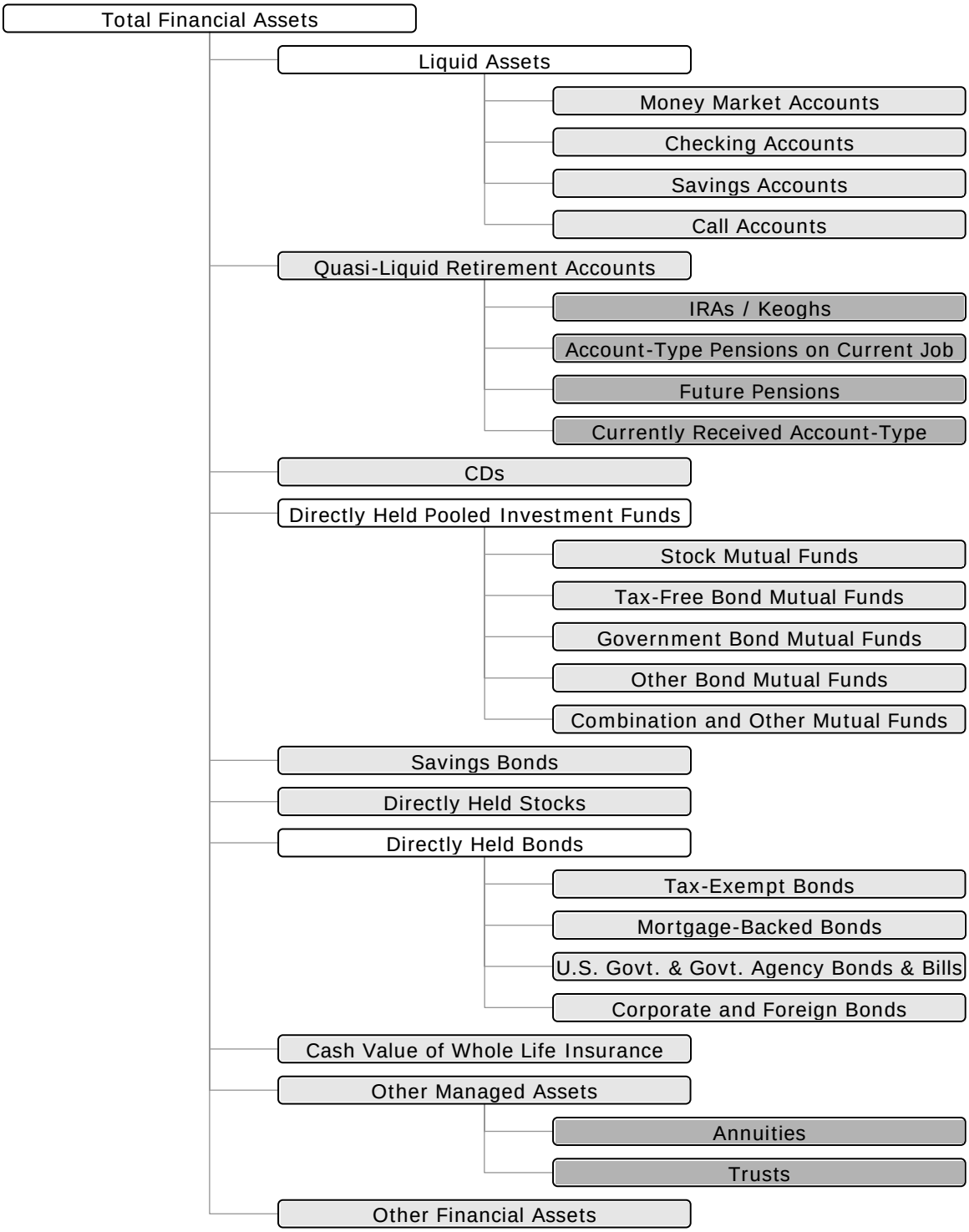


Figure 5
Comparison of Actual (Light Blue) and Estimated (Dark Blue) Marginal Distributions, Dirichlet-Multinomial Model, Selected Outcomes, T=10 and T=100 (Estimated Distributions are Unweighted Sample Averages of Estimated Cell Probabilities)

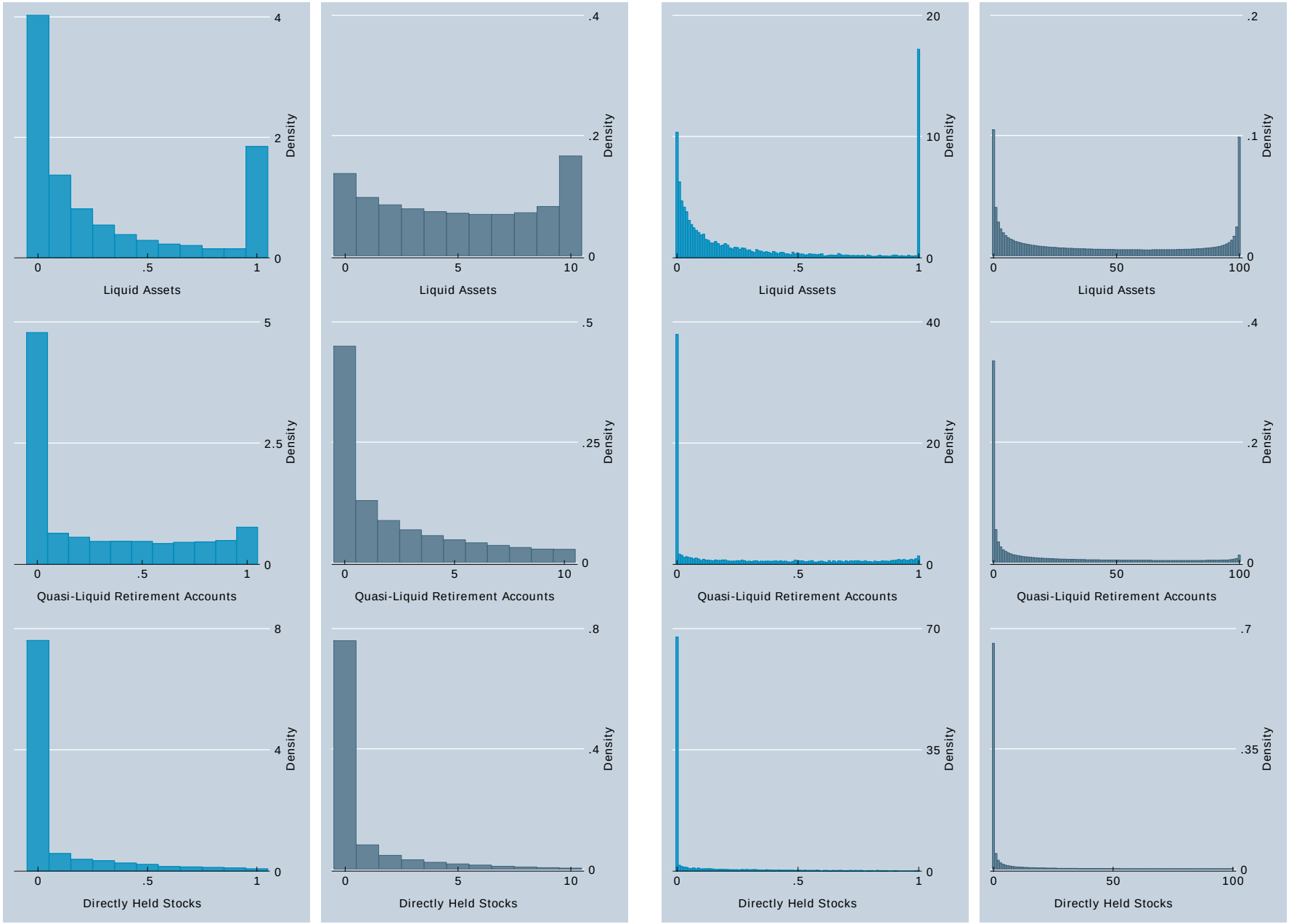


Figure 6
Lorenz Curve Plots of DM Estimates (T=100) against Corresponding Coarsened Data

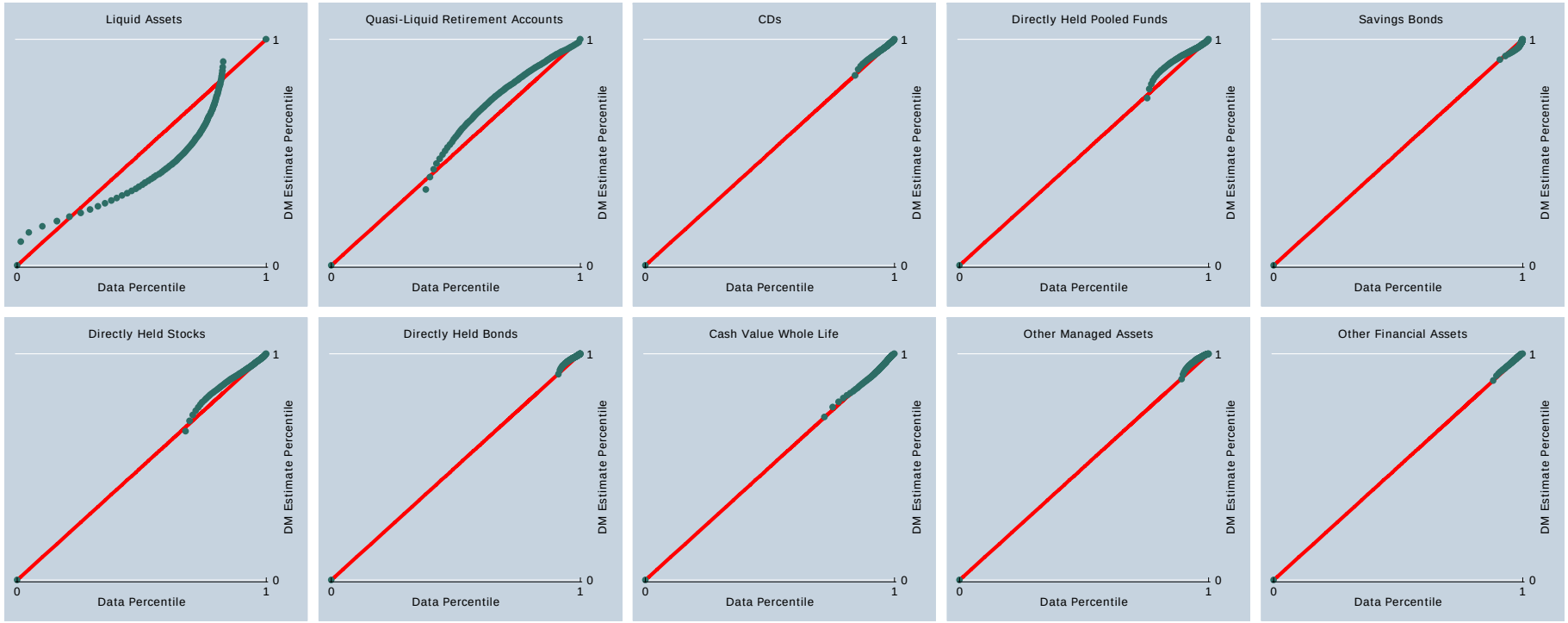


Table 1a
MEPS, Combined 1996-2007 Sample, Data Extract:
Healthcare Expenditure Shares, Nine Consecutive Observations with Defined Shares
(*dupersid* is MEPS Individual Case Identifier, with Data Presented in *dupersid* Sort Order; Boundary Solutions Shaded)

| Observation | | Healthcare Expenditure Category | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | *dupersid* | Office-Based Visits | Prescr. Drugs | Inpatient Stays | ER Visits | Out-patient Visits | Dental Visits | Home Health Care | Vision Aids | Other S & E | Total |
| 1997 | 00014013 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1.000 |
| 1997 | 00015011 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| 1997 | 00015015 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1.000 |
| 1997 | 00015022 | 0.045 | 0.126 | 0 | 0.828 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| 1997 | 00018036 | 0 | 0 | 0.903 | 0.097 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| 1997 | 00018059 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| 1997 | 00018073 | 0 | 0.001 | 0.952 | 0.006 | 0.041 | 0 | 0 | 0 | 0 | 1.000 |
| 1997 | 00019014 | 0.108 | 0 | 0 | 0 | 0.481 | 0.411 | 0 | 0 | 0 | 1.000 |
| 1997 | 00020011 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |

Table 1b
American Time Use Survey, Combined 2003-2008 Sample, Data Extract:
Two-Digit Time Use Categories, Nine Consecutive Observations
(*tucaseid* is ATUS Individual Case Identifier, with Data Presented in *tucaseid* Sort Order; Boundary Solutions Shaded)

| Observation | | Two-Digit Time Use Category | | | | | | | | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | *tucaseid* | Personal Care Activities | Household Activities | Caring for & Helping HH Members | Caring for & Helping Non-HH Members | Work & Work-Related Activities | Education | Consumer Purchases | Professional & Personal Care Svcs. | Household Services | Govt. Svcs. & Civic Obligations | Eating & Drinking | Socializing, Relaxing, & Leisure | Sports, Exercise, & Recreation | Religious & Spiritual Activities | Volunteer Activities | Telephone Calls | Travelling | |
| 2003 | 20030807033472 | 540 | 240 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 45 | 540 | 0 | 0 | 0 | 75 | 0 | 1440 |
| 2003 | 20030807033485 | 615 | 0 | 10 | 0 | 540 | 0 | 0 | 0 | 0 | 0 | 60 | 140 | 0 | 0 | 0 | 0 | 75 | 1440 |
| 2003 | 20030807033487 | 480 | 60 | 0 | 0 | 45 | 0 | 0 | 0 | 0 | 0 | 120 | 320 | 290 | 0 | 0 | 0 | 125 | 1440 |
| 2003 | 20030807033489 | 1440 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1440 |
| 2003 | 20030807033490 | 705 | 90 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 450 | 35 | 0 | 0 | 150 | 10 | 1440 |
| 2003 | 20030807033491 | 600 | 0 | 0 | 477 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 60 | 0 | 201 | 0 | 0 | 102 | 1440 |
| 2003 | 20030807033494 | 720 | 0 | 0 | 470 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 240 | 0 | 0 | 0 | 0 | 0 | 1440 |
| 2003 | 20030807033495 | 570 | 270 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 90 | 380 | 0 | 0 | 0 | 70 | 60 | 1440 |
| 2003 | 20030807033498 | 705 | 30 | 0 | 0 | 180 | 0 | 0 | 0 | 0 | 0 | 20 | 315 | 0 | 0 | 0 | 0 | 190 | 1440 |

Table 1c
Survey of Consumer Finances, Combined 2001, 2004, 2007 Sample, Data Extract:
Financial Asset Shares, Nine Consecutive Observations with Defined Shares
(*yy1* is SCF Household Case Identifier, with Data Presented in *yy1* Sort Order; Boundary Solutions Shaded)

| Observation | | Financial Asset Category | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | *yy1* | Liquid Assets | Quasi-Liquid Retir. | CDs | Dir. Held Pooled Funds | Savings Bonds | Directly Held Stocks | Directly Held Bonds | Cash Val. Whole Life | Other Mgd. Assets | Other Fin. Assets | Total |
| 2007 | 1531 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| 2007 | 1532 | .167 | .131 | 0 | 0 | 0 | 0 | 0 | .702 | 0 | 0 | 1.000 |
| 2007 | 1533 | .146 | .854 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| 2007 | 1534 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| 2007 | 1535 | .255 | .531 | .204 | 0 | .010 | 0 | 0 | 0 | 0 | 0 | 1.000 |
| 2007 | 1536 | .049 | .904 | .014 | 0 | 0 | .007 | 0 | .026 | 0 | 0 | 1.000 |
| 2007 | 1537 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1.000 |
| 2007 | 1538 | .359 | .513 | 0 | 0 | 0 | .128 | 0 | 0 | 0 | 0 | 1.000 |
| 2007 | 1539 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.000 |

Table 2
Survey of Consumer Finances, Combined 2001, 2004, 2007 Sample: Descriptive Statistics

| Financial Asset Aggregate Category Shares, Estimation Sample (N=12,723) | Mean | | Sample Percentages (Unweighted) | | |
|---|---|---|---|---|---|
| | Unwtd. | Weighted | $s_{im} = 0$ | $s_{im} = 1$ | $0 < s_{im} < 1$ |
| **Liquid Assets** | .332 | .399 | .018 | .170 | .812 |
| **Quasi-Liquid Retirement Accounts** | .292 | .312 | .355 | .005 | .640 |
| **CDs** | .039 | .044 | .823 | .001 | .176 |
| **Directly Held Pooled Investment Funds** | .074 | .047 | .745 | .0002 | .255 |
| **Savings Bonds** | .010 | .013 | .820 | .001 | .180 |
| **Directly Held Stocks** | .097 | .050 | .654 | .001 | .346 |
| **Directly Held Bonds** | .022 | .004 | .907 | 0 | .093 |
| **Cash Value of Whole Life Insurance** | .065 | .073 | .669 | .004 | .326 |
| **Other Managed Assets** | .037 | .026 | .885 | .0001 | .115 |
| **Other Miscellaneous Financial Assets** | .032 | .031 | .258 | .004 | .738 |

| Covariates, Full and Estimation Samples | Full Sample (N=13,379) | | | | Subsample with Defined Shares (N=12,723) | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | | | | Mean | | | |
| | Unwtd. | Wtd. | Min | Max | Unwtd. | Wtd. | Min | Max |
| **Age** | 50.9 | 49.5 | 18 | 95 | 51.4 | 50.0 | 18 | 95 |
| **White** | .78 | .73 | 0 | 1 | .80 | .75 | 0 | 1 |
| **Married** | .67 | .59 | 0 | 1 | .68 | .60 | 1 | 1 |
| **Number of Kids** | .85 | .82 | 0 | 10 | .84 | .80 | 0 | 10 |
| **High School Graduate** | .25 | .32 | 0 | 1 | .25 | .32 | 0 | 1 |
| **Some College** | .16 | .18 | 0 | 1 | .16 | .19 | 0 | 1 |
| **College Graduate** | .48 | .35 | 0 | 1 | .50 | .37 | 0 | 1 |
| **Year 2004** | .34 | -- | 0 | 1 | .34 | -- | 0 | 1 |
| **Year 2007** | .33 | -- | 0 | 1 | .33 | -- | 0 | 1 |

Table 3

Financial Asset Shares, MFLOGIT Point Estimates and Robust Asymptotic Standard Errors

(Normalization: $\boldsymbol{\beta}_{\text{Other Financial Assets}} = \mathbf{0}$; Shaded Entries Denote Conservative FDR Rejection Recommendation for $H_0$: $\beta_{km} = 0$)

| | **Financial Asset Category** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Liquid Assets** | **Quasi-Liquid Retir. Accts.** | **CDs** | **Dir. Held Pooled Funds** | **Savings Bonds** | **Directly Held Stocks** | **Directly Held Bonds** | **Cash Val. Whole Life** | **Other Managed Funds** |
| **Age** | .001 | .006 | .049 | .034 | -.007 | .039 | .069 | .020 | .058 |
| **s.e.** | .003 | .003 | .004 | .003 | .005 | .003 | .004 | .003 | .004 |
| | | | | | | | | | |
| **White** | -.264 | .097 | .203 | .796 | .508 | .796 | 1.952 | -.438 | .946 |
| **s.e.** | .102 | .104 | .139 | .133 | .192 | .125 | .288 | .115 | .181 |
| | | | | | | | | | |
| **Married** | .270 | .873 | .398 | .890 | .400 | 1.101 | 1.351 | .546 | .615 |
| **s.e.** | .086 | .088 | .108 | .100 | .152 | .097 | .145 | .101 | .117 |
| | | | | | | | | | |
| **Number of Kids** | -.033 | -.023 | -.073 | .001 | .099 | -.025 | .021 | .054 | .012 |
| **s.e.** | .039 | .039 | .052 | .044 | .054 | .043 | .059 | .044 | .055 |
| | | | | | | | | | |
| **High School Graduate** | -.024 | .972 | .672 | 1.405 | .920 | .945 | 1.287 | .445 | .888 |
| **s.e.** | .154 | .165 | .193 | .234 | .340 | .211 | .421 | .177 | .256 |
| | | | | | | | | | |
| **Some College** | -.280 | .898 | .415 | 1.736 | 1.197 | 1.388 | 2.108 | .271 | 1.113 |
| **s.e.** | .161 | .171 | .204 | .239 | .341 | .215 | .413 | .186 | .262 |
| | | | | | | | | | |
| **College Graduate** | -.296 | 1.384 | .678 | 2.681 | .787 | 2.341 | 3.181 | .161 | 1.923 |
| **s.e.** | .145 | .155 | .183 | .221 | .330 | .198 | .388 | .166 | .239 |
| | | | | | | | | | |
| **Year 2004** | -.041 | -.069 | -.208 | -.164 | -.014 | -.332 | -.034 | -.486 | -.331 |
| **s.e.** | .100 | .101 | .125 | .111 | .161 | .108 | .137 | .113 | .128 |
| | | | | | | | | | |
| **Year 2007** | -.249 | -.118 | -.296 | -.216 | -.336 | -.508 | -.394 | -.714 | -.656 |
| **s.e.** | .100 | .101 | .122 | .111 | .166 | .108 | .140 | .113 | .130 |
| | | | | | | | | | |
| **Constant** | 2.611 | .277 | -3.184 | -4.150 | -2.285 | -3.843 | -9.356 | -.199 | -5.313 |
| **s.e.** | .224 | .231 | .308 | .305 | .453 | .298 | .642 | .259 | .385 |

Table 4
Financial Asset Shares, MFLOGIT Estimates: Weighted APE Point Estimates and Bootstrap 95%-CI Lower and Upper Bounds (CIs based on 500 Bootstrap Replications and Hansen $C_2$ Method)

| | Financial Asset Category | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Liquid Assets** | **Quasi-Liquid Retir. Accts.** | **CDs** | **Dir. Held Pooled Funds** | **Savings Bonds** | **Directly Held Stocks** | **Directly Held Bonds** | **Cash Val. Whole Life** | **Other Managed Assets** | **Other Financial Assets** |
| **Age** | -.0037 | -.0020 | .0013 | .0010 | -.0002 | .0016 | .0007 | .0005 | .0012 | -.0004 |
| **CI-L** | -.0042 | -.0024 | .0011 | .0008 | -.0003 | .0014 | .0006 | .0003 | .0010 | -.0006 |
| **CI-U** | -.0033 | -.0017 | .0015 | .0012 | -.0001 | .0018 | .0009 | .0007 | .0014 | -.0002 |
| **White** | -.0999 | .0150 | .0059 | .0333 | .0050 | .0414 | .0149 | -.0355 | .0196 | .0003 |
| **CI-L** | -.1174 | -.0019 | -.0009 | .0270 | .0022 | .0348 | .0127 | -.0451 | .0146 | -.0069 |
| **CI-U** | -.0824 | .0304 | .0126 | .0398 | .0079 | .0490 | .0169 | -.0245 | .0253 | .0072 |
| **Married** | -.1030 | .0783 | -.0079 | .0144 | -.0019 | .0325 | .0094 | -.0004 | -.0008 | -.0206 |
| **CI-L** | -.1192 | .0648 | -.0141 | .0088 | -.0046 | .0267 | .0068 | -.0081 | -.0059 | -.0272 |
| **CI-U** | -.0879 | .0909 | -.0021 | .0210 | .0010 | .0380 | .0121 | .0068 | .0042 | -.0144 |
| **Number of Kids** | -.0054 | -.0014 | -.0022 | .0011 | .0013 | -.0006 | .0006 | .0051 | .0009 | .0007 |
| **CI-L** | -.0115 | -.0068 | -.0052 | -.0014 | .0005 | -.0033 | -.0007 | .0023 | -.0018 | -.0021 |
| **CI-U** | .0004 | .0037 | .0006 | .0035 | .0022 | .0021 | .0021 | .0081 | .0031 | .0033 |
| **High School Graduate** | -.1889 | .1267 | .0112 | .0238 | .0052 | .0177 | .0038 | .0066 | .0087 | -.0146 |
| **CI-L** | -.2153 | .1056 | .0022 | .0176 | .0009 | .0083 | .0007 | -.0082 | .0012 | -.0255 |
| **CI-U** | -.1607 | .1486 | .0216 | .0309 | .0096 | .0257 | .0065 | .0214 | .0156 | -.0010 |
| **Some College** | -.2495 | .1306 | .0026 | .0419 | .0106 | .0459 | .0128 | -.0002 | .0160 | -.0108 |
| **CI-L** | -.2785 | .1067 | -.0081 | .0343 | .0056 | .0351 | .0086 | -.0151 | .0080 | -.0238 |
| **CI-U** | -.2186 | .1534 | .0130 | .0509 | .0158 | .0559 | .0171 | .0134 | .0240 | .0039 |
| **College Graduate** | -.3540 | .1714 | -.0027 | .0819 | .0015 | .0968 | .0256 | -.0288 | .0300 | -.0217 |
| **CI-L** | -.3802 | .1513 | -.0110 | .0748 | -.0020 | .0879 | .0223 | -.0424 | .0237 | -.0326 |
| **CI-U** | -.3274 | .1917 | .0061 | .0904 | .0056 | .1050 | .0292 | -.0158 | .0361 | -.0102 |
| **Year 2004** | .0295 | .0159 | -.0027 | -.0013 | .0012 | -.0149 | .0020 | -.0275 | -.0062 | .0040 |
| **CI-L** | .0145 | .0041 | -.0087 | -.0078 | -.0015 | -.0221 | -.0014 | -.0361 | -.0111 | -.0025 |
| **CI-U** | .0443 | .0289 | .0035 | .0048 | .0043 | -.0084 | .0048 | -.0196 | -.0005 | .0100 |
| **Year 2007** | .0069 | .0429 | -.0003 | .0040 | -.0008 | -.0174 | -.0014 | -.0324 | -.0112 | .0096 |
| **CI-L** | -.0084 | .0280 | -.0066 | -.0016 | -.0036 | -.0243 | -.0038 | -.0413 | -.0163 | .0036 |
| **CI-U** | .0208 | .0565 | .0055 | .0102 | .0018 | -.0106 | .0014 | -.0247 | -.0063 | .0159 |

Table 5
Model Prediction Performance: 80/20 Cross-Validation for MPEs and MSEs, and Linear Model Predictions outside [0,1] Interval (Averages over 100 Replicates)

| | | Out of Sample Predictions (Best for Each Asset Category is Shaded) | | | | | | Linear Model: Fraction of Predictions < 0 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean Prediction Error | | | Mean Squared Error | | | | |
| | | MFLOGIT | Linear | Tobit | MFLOGIT | Linear | Tobit | Out of Sample | In-Sample |
| Financial Asset Category | **Liquid Assets** | .0117 | -.0002 | -.0666 | .1149 | .1141 | .1187 | 0 | 0 |
| | **Quasi-Liquid Retirement Accts.** | .0089 | .0003 | .0082 | .1084 | .1098 | .1109 | .0002 | .0001 |
| | **CDs** | .0014 | .0002 | -.0031 | .0180 | .0183 | .0184 | .0322 | .0292 |
| | **Directly Held Pooled Funds** | .0023 | .0004 | .0001 | .0293 | .0294 | .0296 | .0543 | .0538 |
| | **Savings Bonds** | .0004 | .0001 | -.0048 | .00387 | .00388 | .0040 | .0221 | .0193 |
| | **Directly Held Stocks** | .0018 | -.0005 | -.0034 | .03641 | .0370 | .03642 | .0750 | .0750 |
| | **Directly Held Bonds** | .0005 | .00003 | -.0007 | .0094 | .0095 | .0095 | .1707 | .1689 |
| | **Cash Value of Whole Life** | .0019 | -.0002 | -.0110 | .03149 | .03151 | .0325 | 0 | 0 |
| | **Other Managed Assets** | .0005 | -.0004 | -.0006 | .0196 | .0197 | .0197 | .0987 | .0984 |
| | **Other Financial Assets** | .0014 | .0003 | -.0043 | .01879 | .01879 | .0190 | .0007 | .0006 |

Table 6

Disaggregated Share Model: MFLOGIT Point Estimates, Robust Asymptotic Standard Errors, and $\chi^2$ Aggregation Tests

(Only Disaggregated Categories and Slope Parameters Shown; Normalization: $\boldsymbol{\theta}_{\text{Other Financial Assets}} = \mathbf{0}$)

| | **Financial Asset Category** | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Liquid Assets** | | | | | **Directly Held Pooled Funds** | | | | | | **Directly Held Bonds** | | | | |
| | **Money Market Accounts** | **Checking Accounts** | **Savings Accounts** | **Call Accounts** | **Category Aggregate (Reference)** | **Stock Mutual Funds** | **Tax-Free Bond Mutual Funds** | **Government Bond Mutual Funds** | **Other Bond Mutual Funds** | **Combination & Other Mutual Funds** | **Category Aggregate (Reference)** | **Tax-Exempt Bonds** | **Mortgage-Backed Bonds** | **U.S. Govt. & Govt. Agcy. Bonds & Bills** | **Corporate & Foreign Bonds** | **Category Aggregate (Reference)** |
| **Subcategory %** | **.167** | **.538** | **.281** | **.014** | | **.668** | **.118** | **.031** | **.045** | **.138** | | **.641** | **.059** | **.163** | **.137** | |
| **Age** | .021 | -.006 | .001 | .046 | **.001** | .029 | .054 | .053 | .051 | .040 | **.034** | .070 | .093 | .064 | .076 | **.069** |
| **s.e.** | .003 | .003 | .003 | .006 | **.003** | .003 | .005 | .008 | .008 | .005 | **.003** | .005 | .012 | .010 | .009 | **.004** |
| **White** | .087 | -.337 | -.313 | 1.411 | **-.264** | .751 | .846 | .766 | 1.405 | 1.017 | **.796** | 1.905 | 3.291 | 2.157 | 1.875 | **1.952** |
| **s.e.** | .124 | .105 | .110 | .309 | **.102** | .138 | .228 | .301 | .469 | .228 | **.133** | .348 | .525 | .540 | .428 | **.288** |
| **Married** | .718 | .101 | .283 | 1.116 | **.270** | .850 | 1.229 | .738 | .947 | 1.020 | **.890** | 1.543 | 1.177 | 1.009 | 1.254 | **1.351** |
| **s.e.** | .103 | .090 | .094 | .191 | **.086** | .104 | .164 | .236 | .249 | .161 | **.100** | .168 | .361 | .275 | .265 | **.145** |
| **Number of Kids** | .017 | -.041 | -.049 | .082 | **-.033** | -.009 | -.037 | .064 | -.067 | .106 | **.001** | .006 | -.028 | .041 | .095 | **.021** |
| **s.e.** | .046 | .041 | .043 | .072 | **.039** | .045 | .061 | .089 | .112 | .067 | **.044** | .065 | .155 | .099 | .120 | **.059** |
| **H.S. Graduate** | .257 | -.168 | .147 | .851 | **-.024** | 1.506 | .872 | 2.773 | 2.969 | 1.272 | **1.405** | .895 | 3.925 | 1.908 | 15.35 | **1.287** |
| **s.e.** | .196 | .159 | .166 | .571 | **.154** | .248 | .396 | .652 | .652 | .441 | **.234** | .462 | .905 | .689 | .401 | **.421** |
| **Some College** | .383 | -.495 | -.178 | 1.753 | **-.280** | 1.793 | 1.377 | 3.450 | 3.220 | 1.710 | **1.736** | 1.797 | 4.672 | 2.809 | 15.91 | **2.108** |
| **s.e.** | .203 | .167 | .175 | .547 | **.161** | .254 | .387 | .651 | .650 | .444 | **.239** | .449 | .879 | .628 | .380 | **.413** |
| **Coll. Graduate** | .917 | -.761 | -.240 | 2.684 | **-.296** | 2.734 | 2.284 | 4.173 | 4.523 | 2.782 | **2.681** | 2.880 | 5.737 | 3.876 | 17.13 | **3.181** |
| **s.e.** | .181 | .150 | .157 | .511 | **.145** | .235 | .358 | .609 | .591 | .410 | **.221** | .414 | .768 | .572 | .244 | **.388** |
| **Year 2004** | -.085 | .116 | -.296 | -.283 | **-.041** | -.266 | -.340 | -.081 | .183 | .408 | **-.164** | -.047 | .147 | -.402 | .279 | **-.034** |
| **s.e.** | .114 | .105 | .108 | .177 | **.100** | .114 | .155 | .245 | .236 | .176 | **.111** | .151 | .320 | .228 | .220 | **.137** |
| **Year 2007** | -.221 | -.137 | -.439 | -.650 | **-.249** | -.390 | -.113 | -.333 | -.130 | .582 | **-.216** | -.351 | -.861 | -.422 | -.420 | **-.394** |
| **s.e.** | .114 | .105 | .109 | .188 | **.100** | .114 | .157 | .227 | .248 | .173 | **.111** | .154 | .406 | .235 | .237 | **.140** |
| $\chi^2$ **Test Stats.** | | | | | | | | | | | | | | | | |
| **Within-Categ.** | **952.3 (d.f.=27, p<.0001)** | | | | | **153.5 (d.f.=36, p<.0001)** | | | | | | **1,591.2 (d.f.=27, p<.0001)** | | | | |
| **Overall** | **2,75.8 (d.f.=90, p<.0001)** | | | | | | | | | | | | | | | |

Table 7
Financial Asset Shares, Weighted APE Point Estimates: MFLOGIT and DM (T=10 and T=100) Estimator Comparison
(Shaded Cells Indicate Discordant Signs)

| | | Financial Asset Category | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Liquid Assets** | **Quasi-Liquid Retir. Accts.** | **CDs** | **Dir. Held Pooled Funds** | **Savings Bonds** | **Directly Held Stocks** | **Directly Held Bonds** | **Cash Val. Whole Life** | **Other Managed Assets** | **Other Financial Assets** |
| **Age** | **MFLOGIT** | -.0037 | -.0020 | .0013 | .0010 | -.0002 | .0016 | .0007 | .0005 | .0012 | -.0004 |
| | **DM, T=100** | -.0029 | -.0016 | .0010 | .0007 | -.0003 | .0012 | .0005 | .0010 | .0006 | -.0001 |
| | **DM, T=10** | -.0018 | -.0022 | .0009 | .0007 | -.0001 | .0012 | .0005 | .0003 | .0007 | -.0002 |
| **White** | **MFLOGIT** | -.0999 | .0150 | .0059 | .0333 | .0050 | .0414 | .0149 | -.0355 | .0196 | .0003 |
| | **DM, T=100** | -.0868 | .0236 | .0055 | .0220 | .0044 | .0304 | .0099 | -.0162 | .0101 | -.0030 |
| | **DM, T=10** | -.0679 | .0166 | .0016 | .0245 | .0029 | .0300 | .0092 | -.0259 | .0113 | -.0023 |
| **Married** | **MFLOGIT** | -.1030 | .0783 | -.0079 | .0144 | -.0019 | .0325 | .0094 | -.0004 | -.0008 | -.0206 |
| | **DM, T=100** | -.0978 | .0702 | -.0028 | .0077 | -.0032 | .0217 | .0050 | .0112 | .0010 | -.0132 |
| | **DM, T=10** | -.0736 | .0629 | -.0049 | .0084 | -.0040 | .0225 | .0055 | -.0010 | -.0007 | -.0150 |
| **N. of Kids** | **MFLOGIT** | -.0054 | -.0014 | -.0022 | .0011 | .0013 | -.0006 | .0006 | .0051 | .0009 | .0007 |
| | **DM, T=100** | -.0074 | -.0005 | -.0014 | .0020 | .0031 | -.0003 | .0007 | .0049 | .0003 | -.0015 |
| | **DM, T=10** | -.0031 | -.0014 | -.0024 | .0015 | .0011 | -.0002 | .0007 | .0039 | .0008 | -.0007 |
| **H.S. Grad.** | **MFLOGIT** | -.1889 | .1267 | .0112 | .0238 | .0052 | .0177 | .0038 | .0066 | .0087 | -.0146 |
| | **DM, T=100** | -.1410 | .0844 | .0071 | .0156 | .0116 | .0141 | .0018 | .0127 | .0020 | -.0083 |
| | **DM, T=10** | -.1298 | .0951 | .0055 | .0168 | .0046 | .0141 | .0029 | .0010 | .0028 | -.0130 |
| **Some Coll.** | **MFLOGIT** | -.2495 | .1306 | .0026 | .0419 | .0106 | .0459 | .0128 | -.0002 | .0160 | -.0108 |
| | **DM, T=100** | -.1911 | .0904 | .0001 | .0272 | .0147 | .0375 | .0093 | .0070 | .0088 | -.0037 |
| | **DM, T=10** | -.1666 | .0935 | -.0017 | .0283 | .0075 | .0364 | .0082 | -.0042 | .0103 | -.0120 |
| **Coll. Grad.** | **MFLOGIT** | -.3540 | .1714 | -.0027 | .0819 | .0015 | .0968 | .0256 | -.0288 | .0300 | -.0217 |
| | **DM, T=100** | -.2921 | .1465 | -.0057 | .0568 | .0026 | .0776 | .0173 | -.0085 | .0139 | -.0085 |
| | **DM, T=10** | -.2380 | .1302 | -.0107 | .0572 | -.0009 | .0737 | .0180 | -.0274 | .0154 | -.0176 |
| **Year 2004** | **MFLOGIT** | .0295 | .0159 | -.0027 | -.0013 | .0012 | -.0149 | .0020 | -.0275 | -.0062 | .0040 |
| | **DM, T=100** | .0209 | .0019 | -.0028 | -.0041 | .0017 | -.0053 | -.0006 | -.0117 | -.0027 | .0027 |
| | **DM, T=10** | .0129 | .0125 | -.0001 | -.0013 | .0012 | -.0082 | .0012 | -.0159 | -.0040 | .0016 |
| **Year 2007** | **MFLOGIT** | .0069 | .0429 | -.0003 | .0040 | -.0008 | -.0174 | -.0014 | -.0324 | -.0112 | .0096 |
| | **DM, T=100** | .0108 | .0203 | .0051 | -.0038 | -.0006 | -.0106 | -.0022 | -.0167 | -.0055 | .0033 |
| | **DM, T=10** | .0013 | .0304 | .0032 | .0024 | -.0003 | -.0123 | -.0008 | -.0197 | -.0077 | .0034 |

Table 8
DM Estimator Fit Performance: $\Pr(s_k = 0)$ and $\Pr(s_k = 1)$

| | | Financial Asset Category | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Liquid Assets** | **Quasi-Liquid Retir. Accts.** | **CDs** | **Dir. Held Pooled Funds** | **Savings Bonds** | **Directly Held Stocks** | **Directly Held Bonds** | **Cash Val. Whole Life** | **Other Managed Assets** | **Other Financial Assets** |
| **$\Pr(s_k = 0)$** | Sample Frequency | .018 | .355 | .823 | .745 | .820 | .654 | .907 | .669 | .885 | .856 |
| | DM Estimate, T=10 | .137 | .450 | .908 | .811 | .975 | .760 | .942 | .857 | .921 | .935 |
| | DM Estimate, T=100 | .105 | .335 | .840 | .740 | .910 | .658 | .910 | .721 | .889 | .882 |
| **$\Pr(s_k = 1)$** | Sample Frequency | .170 | .005 | .001 | .0002 | .001 | .001 | 0 | .004 | .0001 | .004 |
| | DM Estimate, T=10 | .166 | .028 | .003 | .003 | .001 | .004 | .0004 | .005 | .001 | .002 |
| | DM Estimate, T=100 | .099 | .012 | .002 | .001 | .001 | .002 | .0002 | .003 | .001 | .001 |