A Bayesian Approach to Imputing a Consumption-Income Panel Using the PSID and CEX*

Matthew E. Smith
Hutchin Hill Capital

Christopher Tonetti
Stanford GSB

September 8, 2014

Abstract

Many important questions in Macro and Labor economics can only be answered using a panel of income and consumption data. Since there is no existing joint panel in the U.S., previous papers have developed estimation strategies to impute consumption data from the CEX to the PSID. This paper develops an alternative Bayesian imputation algorithm that uses the full information from the likelihood of data contained in both the PSID and CEX and provides information about the uncertainty surrounding imputed values. Food consumption is available in both datasets, while total consumption is only available in the CEX. Following Blundell, Pistaferri, and Preston (2006) we specify a statistical relationship between food consumption, total consumption, and other (jointly available) covariates. We treat an individual agent's total consumption at time t as a latent variable and sample from the joint posterior distribution of latent states and parameters governing the food demand equation, given the data contained in the PSID and CEX simultaneously. This extremely high dimensional posterior distribution imposes sampling challenges that are overcome by exploiting derivatives of the posterior distribution using a variant of the Metropolis Adjusted Langevin Algorithm.

*Contact Information: matthew.e.smith@gmail.com; tonetti@stanford.edu. We thank Taisuke Nakata, Tom Sargent, Sam Schulhofer-Wohl, and Gianluca Violante for useful comments and suggestions and Eran Hoffmann for excellent research assistance.

1 Introduction

Many important questions in Macro and Labor economics can only be answered using a panel of income and consumption data. One recent example is Guvenen and Smith (2014), which elicits information about the properties and magnitude of labor income risk individuals face from their economic behavior. With only income data, it is impossible for an econometrician to distinguish between true "shocks" to income and anticipated changes. By using information on individual consumption and income, Guvenen and Smith (2014) attempt to disentangle the amount of increasing life-cycle inequality due to risk versus that due to heterogeneity in abilities. Another example is Blundell, Pistaferri, and Preston (2008), which attempts to provide a quantitative measure of market incompleteness in the U.S. by estimating the degree to which consumers are insured against shocks to their income. In order to identify these partial insurance coefficients, they exploit an individual specific statistical relationship between changes in consumption expenditure and changes in income, requiring panel data.

Unfortunately, for many countries, including the U.S., there is no existing joint panel of consumption and income. Historically, many papers needing a consumption panel have used the food expenditure measure available in the PSID. For example, the Hall and Mishkin (1982) and Zeldes (1989) tests of the permanent income hypothesis, the Cochrane (1991) and Hayashi, Altonji, and Kotlikoff (1996) tests of the consumption insurance hypotheses, and the Altonji (1986) test for intertemporal substitution in labor supply. This is obviously not ideal, as food is a neccesity, with implications for preference elasticities and expenditure volatility that are likely not generalizable to total nondurable consumption. One alternative is to use synthetic panels, like Attanasio and Davis (1996). There are advantages to the synthetic cohort approach, but there are also severe drawbacks, as the researcher is restricted to make observations about across group heterogeneity and cannot speak to within group heterogeneity. As Krueger and Perri (2006) document, much interesting variation over time occurs within groups.

An alternative to exclusively using the food expenditure data in the PSID, while still working with panel data, is to impute a measure of total nondurable expenditure in the PSID. This is the approach taken by the Guvenen and Smith (2014) and Blundell, Pistaferri, and Preston (2008) papers mentioned previously. One imputation strategy, originally suggested by Skinner (1987), is to impute total consumption by proposing a statistical model relating total consumption to a series of consumption components available in both data sets. Skinner (1987) estimates this relationship using CEX data and uses the model with these estimated parameters to impute total consumption in the PSID. Building on the Skinner (1987) methodology, Blundell, Pistaferri, and Preston (2006) (hereafter BPP) develop an estimation strategy to impute consumption data from the CEX to the PSID by using a standard food demand function for the statistical model. Food consumption is available in both datasets, while total consumption is only available in the CEX. BPP specify a statistical relationship between food consumption, total consumption, and other (jointly available) covariates. They impute total consumption in the PSID by estimating this food demand equation in the CEX and applying the inverted function to PSID food expenditure.[1]

This paper builds on the insights of Blundell, Pistaferri, and Preston (2006) by developing an alternative Bayesian imputation algorithm that uses the full information from the likelihood of data contained in both the PSID and CEX and provides information about the uncertainty surrounding point estimates. In the baseline model, we estimate essentially the same statistical relationship outlined by Blundell, Pistaferri, and Preston (2006), innovating on the estimation procedure and not the model to be estimated. We treat an individual's total consumption at time t as a latent variable and sample from the joint posterior of latent states and parameters governing the food demand equation, given the data contained in the PSID and CEX simultaneously. This extremely high dimensional posterior distribution imposes sampling challenges that are overcome by exploiting derivatives of the posterior distribution using a variant of the Metropolis Adjusted Langevin Algorithm. We establish the performance of the new estimator on simulated data in Section 5.3 and in Section 5.4 we perform the imputation on PSID and CEX data. A main finding, presented in Tables 3 and 8 is the significant degree of uncertainty around imputed "point estimates."

Since use of this BPP "data" is becoming increasingly popular, this uncertainty is important, as it must be acknowledged and accounted for in any second stage estimation that uses imputed consumption as "data." In Section 6 we demonstrate how to use the posterior distribution of consumption as imputed data in a simple example. To highlight the importance of imputation uncertainty, we analyze the evolution of consumption inequality over time, both accounting for and ignoring uncertainty in the imputed values. We document that ignoring imputation uncertainty leads to statistically and economically significant errors in statistics that rely on higher moments of the consumption distribution.

Finally, this estimation strategy is broadly applicable to any imputation problem and can handle more flexible statistical relationships, more data sources, and delivers more information on imputed variables than existing methods. In Section 7 we outline extensions to the baseline model and sampling algorithm that are designed to improve the quality of the imputed values.

2 Imputation: Model and Data

Imputation is the statistical process of replacing missing data with alternative values. There are many approaches to imputing missing data, with a correspondingly rich literature. In this paper, we focus on a particular structure for the imputation problem and provide a Bayesian approach to estimating the missing observations.[2] In our baseline scenario, there are two data sets and in data set 1, there are enough observations to estimate a useful statistical relationship. However, data set 2 is missing some of the variables in this statistical relationship. The problem is to extract information from the

[1] Blundell, Pistaferri, and Preston (2006) make additional contributions to the literature. They include relative prices in the regression and allow the budget elasticity to shift with observable characteristics. They allow for measurement error in consumption and for differences in food expenditure across data sources. Additionally, they examine conditions that ensure that the procedure matches trends in the mean and variance of the consumption distribution.

[2] For alternative strategies to imputing consumption see approaches that rely on "active savings" and the budget constraint identity, such as Ziliak (1998), Cooper (2010), and sections of Aguiar and Bils (2013).

first data set to impute the missing variables in the second data set, exploiting the common statistical relationship amongst the variables across datasets. To illustrate, separate all of the variables into three groups: x, y, and z. Data set 1 contains observations on all variables of interest, denoted x_1, y_1 and z_1. Data set 2 contains observations x_2 and y_2, but is missing observations on z_2. The statistical relationship between variables can be written as $x = F(y, z, \theta)$ and $z = G(x, y, \theta)$, with parameter vector θ controlling the relationship. Heuristically, since data set 1 contains data on x, y, and z, model F can be used to estimate θ. Since dataset 2 contains data on x and y, the estimated parameters, $\widehat{\theta}$, and model G can be used to impute values of z_2.

2.1 BPP

This is exactly the set up relevant for Blundell, Pistaferri, and Preston (2006), who develop a two step estimation strategy to impute consumption data from the CEX to the PSID. BPP specify a statistical relationship between food consumption, total nondurable consumption, and other covariates that are jointly available.

Let x index an observation from the CEX (the input data set) and let p index an observation from the PSID (the target data set). Assume, as constructed, that the CEX and PSID represent random samples drawn from the same underlying population. Consider the following food-demand equation:

$$\tau(f_{i,x}) = D'_{i,x}\beta + \gamma\eta(c_{i,x}) + e_{i,x}. \tag{1}$$

f is food expenditure which is available in both surveys. D contains individual specific attributes (such as race, education, and geographic region) and aggregate variables (such as relative prices) which are available in both data sets. c is total non-durable expenditure which is available only in the input data set (CEX). Finally, e captures the unobserved heterogeneity in the demand for food (including classical measurement error in food expenditure). Assume functions τ and η are known monotonic increasing functions (e.g., identity or ln). BPP estimate equation 1 yielding estimates $\widehat{\beta}$ and $\widehat{\gamma}$.[3] Using these estimated values, imputed consumption in the PSID is calculated by inverting the relationship (assuming $\gamma > 0$):

$$\widehat{c}_{i,p} = \eta^{-1}\left(\frac{\tau(f_{i,p}) - D'_{i,p}\widehat{\beta}}{\widehat{\gamma}}\right). \tag{2}$$

2.2 Bayesian Estimator

There are two major criteria by which to judge an imputation procedure: accuracy and efficiency. One major drawback of the BPP imputation methodology is that it does not provide useful information on the uncertainty surrounding the imputed values. Another problem is that the two-step procedure

[3] A major contribution of the BPP methodology is to augment standard OLS estimation of equation 1 to obtain consistent estimates in the presence of measurement error in total consumption and differences in food consumption across datasets.

does not make the most efficient use of available data. Instead of following the imputation approach of BPP, we develop a Bayesian estimator, treating imputation as a missing variable problem, that directly addresses these two concerns.

The goal of Bayesian estimation is to characterize the posterior distribution of parameters, which is defined as the distribution of parameters conditional on data, $p(\vartheta|Y)$. The posterior distribution is related to two objects, the prior distribution and the likelihood function, by Bayes Theorem:

$$p(\vartheta|Y) \propto L(Y|\vartheta)p(\vartheta),$$

where $p(\vartheta)$ is the prior distribution of parameters, which the researcher specifies. When the prior is uniform, the posterior is the same as the likelihood (as a function of parameters).[4]

To characterize the posterior distribution, the Bayesian estimator draws a large sample from the posterior, which can be used to compute statistics about the parameters of interest. Algorithms for sampling posterior distributions, an essential component of this paper, are discussed in detail in Section 4.

This Bayesian approach is related to the multiple imputation literature (see Rubin (1987, 1988, 1996)). The key to multiple imputation is to provide multiple imputed values, so that imputation uncertainty, which is displayed in the differences across imputed values, can be incorporated into any analysis that uses the imputed values. Quoting from Rubin (1988), "Multiple imputations ideally should be drawn according to the following general scheme. For each model being considered, the M imputations of the missing values, Y_{mis}, are M repetitions from the posterior predictive distribution of Y_{mis}, each repetition being an independent drawing of the parameters and missing values under appropriate Bayesian models for the data and the posited response mechanism." A main goal of this paper is to develop such a sampling procedure well-suited to the high-dimensional posterior associated with the consumption imputation problem.

2.2.1 Efficiency

To highlight the potential efficiency gains, a useful analogy is to frame the estimation procedure as an optimization problem. Consider the following three maximization problems:

$$\begin{aligned}
A &= \max_{X_p} p(X_p|Y_p,\theta) \max_{\theta} p(\theta|Y_x), \\
B &= \max_{(X_p,\theta)} p(X_p|Y_p,\theta)p(\theta|Y_x), \\
C &= \max_{(X_p,\theta)} p(X_p,\theta|Y_p,Y_x).
\end{aligned}$$

It should be immediately clear that $B \geq A$. The BPP procedure mimics problem A, in that first parameter values are estimated using only CEX data and then the best imputed value is chosen using

[4] See Geweke (2005) or Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin (2014) for a Bayesian statistics reference.

only PSID data, taking parameter values as given. Since the statistical relationship is restricted to be the same across datasets, there may be efficiency gains in choosing imputed values and parameter values jointly. However, going beyond the change in the optimization procedure, problem C changes the objective function to be a proper posterior, allowing for a natural interpretation of the exercise. Problem C is to choose values for all the unknowns to maximize their probability, using all available data to do so.[5]

2.2.2 Parameter Uncertainty

Parameters and latent states are unknowns. Following the Bayesian tradition, they are treated as random variables, with associated distributions. There are two major advantages of using Bayesian methods associated with parameter uncertainty. Most importantly, given draws from the posterior distribution, we have information about the uncertainty surrounding the imputed values (e.g., the variance of the marginal posterior distribution). This is essential information, as the entire reason for the exercise is to obtain accurate and precise estimates of missing values, making information about precision key to assessing the quality of the imputed data. Furthermore, with the Bayesian approach, unobserved consumption is inferred while simultaneously acknowledging and formally accounting for uncertainty in the values of the parameters governing the food demand equation.[6]

2.2.3 Statistical Model: Food Demand Equation

While we develop a Bayesian estimator designed to be flexible enough to estimate complicated statistical models, the goal of this section is to keep the statistical model close to that formulated by BPP in order to emphasize differences in estimation methods for a given statistical relationship and not differences in the assumed statistical relationship itself. Thus, we consider the following statistical model, which stacks BPP-style food demand equations for the CEX and PSID, while explicitly modeling measurement error in total nondurable consumption.[7]

[5] Note that we can factor problem C as follows,

$$\begin{aligned} C &= \max_{(X_p,\theta)} p(X_p,\theta|Y_p,Y_x) \\ &= \max_{(X_p,\theta)} p(X_p|\theta,Y_p,Y_x)p(\theta|Y_p,Y_x). \end{aligned}$$

If we assume $p(X_p|\theta,Y_p,Y_x) = p(X_p|\theta,Y_p)$, i.e., conditional on the data in data set p and a vector of parameters shared across datasets, the distribution of X_p is independent of the data in data set x, we have

$$C = \max_{(X_p,\theta)} p(X_p|\theta,Y_p)p(\theta|Y_p,Y_x)$$

With this formulation and assumption, it is apparent that problem C explicitly uses more information (contained in Y_x) to compute an estimate of θ.

[6] One could, instead, attempt to compute (or approximate) the variance of an imputed consumption value using the delta method. That is, letting $\theta = (\beta,\gamma)$ and the imputed consumption $c_{itp} = g_{itp}(\theta)$ for the individual and time specific function $g_{itp}(\cdot)$, $var(c_{itp}) \approx \nabla g_{itp}(\theta)^T var(\theta)\nabla g_{itp}(\theta)$.

[7] It is widely acknowledged that there is significant measurement error in CEX consumption measures. Additionally, a main emphasis of Blundell, Pistaferri, and Preston (2006) is that accounting for measurement error in total consumption is essential to obtain consistent estimates.

$$f_{itx} = D^{'}_{itx}\beta + \gamma c^{*}_{itx} + \sigma_f e_{itx} \quad (3)$$
$$c_{itx} = c^{*}_{itx} + \sigma_{cx} v_{itx} \quad (4)$$
$$f_{itp} = D^{'}_{itp}\beta + \gamma c^{*}_{itp} + \sigma_f e_{itp} \quad (5)$$

Here, variables with subscript x denote variables in the CEX, and p, variables in the PSID. f_{itx} and c_{itx}, are observable food and total nondurable consumption in the CEX, and f_{itp} is observable food consumption in the PSID. c^{*}_{itx} and c^{*}_{itp} are true (unobserved) nondurable consumption. $D^{'}_{itx}$ and $D^{'}_{itp}$ are observed covariates present in both data sets, and $(e_{itx}, v_{itx}, e_{itp})$ is a (jointly) normally distributed error term. Note the parameters are shared and not indexed by data source. For notational convenience group variables into parameters, θ, missing variables, X, and data, Y. We have N_{tx} individual observations in the CEX at time t and T_x total time periods in the CEX, with corresponding notation for the PSID. Thus, we define $N_x = \sum_{j=1}^{Tx} N_{jx}$ and $N_p = \sum_{j=1}^{Tp} N_{jp}$ as the total number of observations in the CEX and PSID.

D contains all of the individual/time specific characteristics as well as the aggregate variables. Following BPP we include: $constant$, age, age^2, $familysize$, $HSDropout$, $HSGrad$, $RegionNE$, $RegionMW$, $RegionSO$. $RaceWhite$, $Cohort2529$, $Cohort3034$, $Cohort3539$, $Cohort4044$, $Cohort4549$, $Cohort5054$, $Cohort5559$, $1Child$, $2Children$, $3+Children$, $PriceAlcohol$, $PriceFood$, $PriceFuel$, and $PriceTransportation$.

With equations 3, 4, and 5, we have cast the imputation problem into a traditional missing variables problem by treating c^{*}_{itp} as an unobserved latent variable. We have explicitly modeled measurement error by treating c^{*}_{itx} as an unobserved latent variable, related to measured total nondurable consumption with error.

3 Prior, Likelihood, and Posterior Distributions

Our goal is to draw from the joint posterior of parameters and unobservable consumption, given the information contained in *both* data sets, $p(X, \theta|Y) \propto p(Y|\theta, X)p(\theta, X)$. In order to do this, we first derive an analytical expression for the log likelihood generated by our statistical model and then specify priors.

In our notation, we distinguish between data, collected in Y, unobservables, collected in X, and parameters, collected in θ. Define $X_{tx} := \{c^{*}_{itx}\}_{i=1}^{N_{tx}}$, $X_{tp} := \{c^{*}_{itp}\}_{i=1}^{N_{tp}}$, $Y_{tx} := \{f_{itx}, D_{itx}, c_{itx}\}_{i=1}^{N_{tx}}$, $Y_{tp} := \{f_{itp}, D_{itp}\}_{i=1}^{N_{tp}}$, $Y_x := \{Y_t\}_{t=1}^{Tx}$, $Y_p := \{Y_t\}_{t=1}^{Tp}$, $X_x := \{X_t\}_{t=1}^{Tx}$, and $X_p := \{X_t\}_{t=1}^{Tp}$. Then, $X := (X_x, X_p)$, $Y := (Y_x, Y_p)$, and $\theta := (\beta, \sigma_f, \sigma_{cx}, \gamma)$.

3.1 Likelihood

The likelihood is

$$p(Y|\theta, X) = \prod_{t=1}^{T_x}\prod_{i=1}^{N_{tx}} p(f_{itx}, c_{itx}|\theta, X) \prod_{t=1}^{T_p}\prod_{i=1}^{Ntp} p(f_{itp}|\theta, X). \tag{6}$$

Given our assumption that $f_{itx}, c_{itx}, f_{itp}$ are independent given (θ, X), and normally distributed,

$$p(Y|\theta, X) = \left(\frac{1}{2\pi\sigma_f\sigma_{cx}}\right)^{T_x N_x} \prod_{t=1}^{T_x}\prod_{i=1}^{N_{tx}} p(\widehat{e}_{itx}|\theta, X) p(\widehat{v}_{itx}|\theta, X) \left(\frac{1}{\sqrt{2\pi}\sigma_f}\right)^{T_p N_p} \prod_{t=1}^{T_p}\prod_{i=1}^{N_{tp}} p(\widehat{e}_{itp}|\theta, X), \tag{7}$$

where, given our specification,

$$\begin{aligned}
\widehat{e}_{itx} &= \frac{1}{\sigma_f}(f_{itx} - (D'_{itx}\beta + \gamma c^*_{itx})), \\
\widehat{v}_{itx} &= \frac{1}{\sigma_{cx}}(c_{itx} - c^*_{itx}), \\
\widehat{e}_{itp} &= \frac{1}{\sigma_f}(f_{itp} - (D'_{itp}\beta + \gamma c^*_{itp})).
\end{aligned}$$

with

$$\begin{aligned}
p(\widehat{e}_{itx}|\theta, X) &= N(0, 1), \\
p(\widehat{v}_{itx}|\theta, X) &= N(0, 1), \\
p(\widehat{e}_{itp}|\theta, X) &= N(0, 1).
\end{aligned}$$

Thus, the log likelihood is

$$\begin{aligned}
\log(p(Y|\theta, X)) = &-(T_x N_x)\log(2\pi\sigma_f\sigma_{cx}) - (T_p N_p)\log\left(\sqrt{2\pi}\sigma_f\right) \\
&- \sum_{t=1}^{T_x}\sum_{i=1}^{N_{tx}} \frac{1}{2}(\widehat{e}^2_{itx} + \widehat{v}^2_{itx}) - \sum_{t=1}^{T_p}\sum_{i=1}^{N_{tp}} \frac{1}{2}\widehat{e}^2_{itp}.
\end{aligned} \tag{8}$$

3.2 Prior

In choosing priors, we want to center our estimates around economically relevant areas that are perceived to be likely based on previous research, but allow enough variability in the distribution to let the posterior significantly differ from the prior as the data informs the estimates. We specify our priors such that standard deviations are distributed according to Inverse Gamma distributions, parameters β and γ are distributed according to normal distributions, and each nondurable consumption is distributed according to a normal distribution. We detail specific values for the prior distributions in Section 5.2. We choose the Inverse Gamma distribution for standard deviation parameters to ensure positivity. We choose the normal distribution to reflect symmetric uncertainty in both directions from the mean, while ensuring a decent portion of the probability mass is in the *ex ante* economically reasonable region surrounding the chosen mean.

Now that we have defined the likelihood, prior, and, thus, posterior distributions, we proceed in Section 4 to discuss algorithms that sample from the posterior distribution.

4 Sampling

Given the data contained in the CEX and PSID, Y, we can construct a joint posterior density of parameters θ and unobservable states (total nondurable consumption) X: $p(\theta, X|Y)$. This density is a function defined in approximately 25,000 dimensions which we need to characterize succinctly. One way to characterize a distribution is to draw samples from it. The ability to draw from a distribution allows one to characterize any moment of that distribution, as well as expectations of any function of random variables that are distributed according to it. We will let s denote some random variable distributed according to a distribution, f, $s \sim f(\cdot)$ with density $f(s)$. For the inference problem at hand, we have $s = (\theta, X)$ and and $f(s) = p(\theta, X|Y)$. Our goal is to be able to produce a series of draws $\{s^m\}_{m=1}^M$, each (approximately) distributed according to our target $f(\cdot)$. Below we outline some common methods in the Markov chain Monte Carlo (MCMC) literature that have been proposed to achieve this goal—Gibbs Sampling and the the Metropolis-Hastings algorithm—and outline our choice of sampler for the problem at hand, the Metropolis-Adjusted Langevin Algorithm (MALA). MALA is a particular implementation of the Metropolis-Hastings algorithm. Below we discuss its relative merits compared to the more standard algorithms, as well as, potential extensions of the MALA algorithm. Essentially, sequences of draws generated by MALA have attractive properties when sampling from high-dimensional distributions.

4.1 Gibbs Sampling

In certain cases, it is possible to exactly compute and sample from $f(s_j|s_1, \ldots, s_{j-1}, s_{j+1}, \ldots, s_n)$ for each coordinate j in the state space. It is then possible to use the Gibbs sampling algorithm to build a Markov chain by iteratively sampling from $f(s_j|s_1, \ldots, s_{j-1}, s_{j+1}, \ldots, s_n)$, $j = 1, \ldots, n$, to produce draws $s \sim f(\cdot)$. In our Bayesian context, exact sampling is typically sensitive to the use of conjugate priors and hence is not robust to changes in the statistical model that defines $f(s)$. In this paper we wish to provide a general imputation framework, and, thus, we want to remain agnostic toward the "correct" demand system that links food and total consumption. That is, we want to avoid implementing an algorithm that is not robust to changes in the specification of the demand system. Beyond being sensitive to the specification of particular conjugate priors, Gibbs sampling can perform poorly when the components of s are correlated. For example, sampling an agent's total consumption on some date t independently of that individual's consumption at other dates may result in poor mixing of the Markov chain. Lastly, the performance of such an algorithm can degrade quickly in the dimension of s and there are examples where the mixing time increases exponentially with the dimension (e.g. the witch's hat of Matthews (1993)). For these reasons, we choose not to implement a Gibbs sampling algorithm for our problem.

4.2 Metropolis-Hastings

Rather than Gibbs sampling, we can use the Metropolis-Hastings algorithm to sample from $f(\cdot)$. The Metropolis-Hastings algorithm reverse engineers a Markov chain whose invariant distribution is the distribution from which one wants to sample. The algorithm requires a proposal distribution, $q(s,\cdot)$ with density $q(s,s')$, and proceeds as follows. At iteration m when the state is s, denoted by $s^m = s$, propose $s' \sim q(s,\cdot)$ and set $s^{m+1} = s'$ with probability

$$\alpha(s,s') = \min\left\{1, \frac{f(s')q(s',s)}{f(s)q(s,s')}\right\}$$

and $s^{m+1} = s$ with probability $(1-\alpha(s,s'))$. Under some regularity conditions, this algorithm produces a Markov chain whose ergodic distribution is $f(\cdot)$. The generality of this algorithm can be seen by noting that there are many proposal distributions $q(s,\cdot)$ that can be used to sample from $f(\cdot)$. Choice of $q(s,\cdot)$ along with the properties of the target density $f(s)$ will govern how efficiently this algorithm generates samples from $f(\cdot)$.

4.2.1 Random Walk Metropolis-Hastings

One choice of proposal density starts by searching for proposal densities that satisfy $q(s',s) = q(s,s')$. This criteria, for example, is satisfied by a Normal distribution centered at the current value s, $q(s,\cdot) = N(s,\Sigma)$. The Metropolis-Hastings algorithm with this particular proposal density is commonly referred to as Random Walk Metropolis-Hastings (RWMH). With $q(s',s) = q(s,s')$, the acceptance probability becomes

$$\alpha(s,s') = \min\left\{1, \frac{f(s')}{f(s)}\right\}$$

and draws are accepted with certainty when the target density at the proposed point rises relative to the target density evaluated at the current point, and are accepted with probability $\frac{f(s')}{f(s)}$ when this value decreases. This implementation of the Metropolis-Hastings algorithm has seen substantial use in the physics, statistics, and econometrics literature. The algorithm is tuned by the choice of the covariance matrix of the random walk proposal Σ, that when chosen optimally, should lead to an acceptance probability of approximately 0.234 (see Roberts, Gelman, and Gilks (1997)). Such an acceptance probability maximizes the rate at which the Markov chain generated by the RWMH algorithm explores its invariant distribution. Usually Σ is chosen to be a scaled version of the covariance matrix of the distribution of interest, the idea being that correlated components should be moved together. Roberts, Gelman, and Gilks (1997) show that for i.i.d. target densities the proposal variance should scale with n^{-1} and $O(n)$ steps are needed for the RWMH chain to explore its invariant distribution.

4.2.2 The Metropolis Adjusted Langevin Algorithm

Another choice for the proposal density of a Metropolis-Hastings algorithm is motivated by the Langevin SDE,

$$dS = \frac{1}{2}\nabla \log(f(S))dt + dW, \tag{9}$$

where dW is Brownian motion. In steady state, this SDE has $S \sim f(\cdot)$. A discretization of the Langevin SDE is

$$S_{t+1} = S_t + \frac{h^2}{2}\nabla \log(f(S_t)) + h\epsilon_{t+1}, \tag{10}$$

where the discretization size is $\Delta t = h^2$ and $\epsilon_{t+1} \sim N(0, I)$. Such a discretization would no longer have $f(\cdot)$ as its invariant distribution, but as suggested by Besag (1994), we can use this discretized SDE to generate proposals in an otherwise standard Metropolis-Hastings algorithm.[8] This algorithm, the Metropolis-Adjusted Langevin Algorithm (MALA), exploits the discretization of a Langevin SDE with the target density $f(s)$ and uses it as a proposal in a Metropolis-Hastings algorithm targeting the same density. MALA proceeds as follows: for a given h, at $s^m = s$, propose $s' = s + \frac{h^2}{2}\nabla \log(f(s)) + h\epsilon_{t+1}$ and set $s^{m+1} = s'$ with probability

$$\alpha(s, s') = \min\left\{1, \frac{f(s')q(s', s)}{f(s)q(s, s')}\right\}$$

and $s^{m+1} = s$ with probability $(1 - \alpha(s, s'))$. The MALA algorithm can be augmented by premultiplying the SDE by a constant matrix Λ as in Roberts and Stramer (2002),

$$dS = \frac{1}{2}\Lambda\nabla \log(f(S))dt + \sqrt{\Lambda}dW. \tag{11}$$

Then the discretized evolution equation is,

$$S_{t+1} = S_t + \frac{h^2}{2}\Lambda\nabla \log(f(S_t)) + h\sqrt{\Lambda}\epsilon_{t+1} \tag{12}$$

where $\sqrt{\Lambda}$ is the matrix square root of Λ. This preconditioning matrix Λ can be any positive definite matrix and is typically chosen to match the (target) posterior covariance, or a diagonal matrix that scales each dimension so that the algorithm makes proposals of relatively similar magnitude. Preconditioning by a constant matrix does not alter the invariant distribution of the SDE and can improve the sampling properties of the resulting MALA algorithm.

In our applications below, the density that we are trying to sample is the joint posterior of parameters and states (unobservable total nondurable consumption) given the data: $f(s) = p(\theta, X|Y)$. Our discretized Langevin SDE becomes the proposal distribution,

$$q((\theta, X), (\{\cdot, \cdot\})) = N((\theta, X) + \frac{h^2}{2}\Lambda\nabla \log(p(\theta, X|Y)), h^2\Lambda). \tag{13}$$

[8] Alternatively, we can view this as correcting the discretization error by a Metropolis accept/reject step.

For the preconditioning matrix, we use a diagonal matrix whose elements are approximations to the inverse Hessian. We would like to use the inverse of the Hessian, but due to the dimension of the space that we are trying to sample, it is not feasible to compute $-H^{-1}$. To implement this algorithm we need to be able to compute the gradient. See Appendix A for analytical expressions of these objects.

Statistical considerations lean toward our use of MALA rather than RWMH to sample our high-dimensional distribution. Specifically, Roberts and Rosenthal (1998) show that MALA has favorable properties compared with a Random Walk algorithm in high dimensions. They show the mixing time of the Markov chain for i.i.d. target densities is $O(n^{1/3})$ for MALA, while Roberts, Gelman, and Gilks (1997) show that it is $O(n)$ for RWMH. Alternatively viewed, Roberts and Rosenthal (1998) show that MALA has a larger optimal acceptance rate than RWMH, 0.574 versus 0.234, resulting in a more efficient algorithm. Generalizing these results to more complex target densities is an active area of research. Roberts and Rosenthal (2001) examine product densities in which the components have different scalings, finding that RWMH and MALA still scale with $O(n)$ and $O(n^{1/3})$ respectively. Breyer, Piccioni, and Scarlatti (2004) extend the MALA scaling results to the case of nonlinear regression. Bedard (2007) extends convergence results of Roberts, Gelman, and Gilks (1997) for RWMH to a more general setting and still recovers the $O(n)$ scaling. Beskos, Roberts, and Stuart (2009) further extend RWMH and MALA results in Roberts, Gelman, and Gilks (1997) and Roberts and Rosenthal (1998) to a class of non-product target densities. Mattingly, Pillai, and Stuart (2012) and Pillai, Stuart, and Thiéry (2012) extend the $O(n)$ and $O(n^{1/3})$ scaling results to a class of infinite-dimensional target densities.

Our takeaway from this literature is that, for the types of densities studied, MALA typically outperforms RWMH in high dimensions when both algorithms are started in equilibrium, $s^0 \sim f(\cdot)$. We are mindful of the limits to the types of target densities studied above and that densities implied by a particular economic model that relates food and total consumption may not satisfy all the assumptions of a given theorem. We are comforted by the fact that the literature consistently recommends the acceptance rate for MALA be tuned to approximately 0.574, as opposed to only 0.234 for a RWMH algorithm, and that mixing time of MALA scales with $O(n^{1/3})$ rather than $O(n)$. Using MALA is not a free lunch however, as Christensen, Roberts, and Rosenthal (2005) show that MALA can have unpredictable behavior when the algorithm is started far away from equilibrium. Further, as discussed in Roberts and Rosenthal (2001), the performance of MALA is relatively more sensitive to variation in the scale of different components of the target density than the analogous RWMH algorithm.

Computational complexity and the flexibility of the algorithm further weigh on the decision to use MALA. MALA involves storing and multiplying vectors of length n rather than matrices of size n^2 as is the case of RWMH with correlated proposals. There is a larger start up cost of MALA that involves computing the derivatives of the log posterior analytically. Furthermore, derivatives must be computed at each iteration, although for our problem, these are all, at most, $O(n)$ operations. In the case that the entire covariance matrix should be used in preconditioning MALA, Simpson, Turner, Strickland,

and Pettitt (2013) propose an approximation method based on Krylov subspaces that decreases the amount of floating point operations needed at each iteration of the algorithm. Economically, MALA is general enough that one can change the demand system for food consumption substantially or explicitly add a time dimension to the PSID data set, and the algorithm will remain largely intact. In sections below we will outline various extensions of our model in which the MALA algorithm can sample with only minor modifications to the code.

4.2.3 Further Extensions to MALA

We can augment the MALA algorithm with a truncated drift term, replacing $\nabla log(f(S_t))$ with $D(S_t) = \frac{D}{max\{D, |\nabla log(f(S_t))|\}}\nabla log(f(S_t))$ in the above equations. Further, we can precondition the SDE by a state-dependent matrix $\Lambda(S)$, so that the discretized SDE becomes,

$$S_{t+1} = S_t + \frac{h^2}{2}\Lambda(S_t)\nabla log(f(S_t)) + h\sqrt{\Lambda}(S_t)\epsilon_{t+1}.$$

Truncating the drift improves the ergodicity properties of the sampler, as demonstrated by Roberts and Tweedie (1996), while preconditioning adjusts the size of the proposed move by the local curvature in the appropriate dimension. The Metropolis algorithm that uses the above proposal mechanism has been called a simplified manifold MALA in Girolami and Calderhead (2011) or position specific preconditioned MALA in Herbst (2012).

As noted by Livingstone and Girolami (2014) and Xifara, Sherlock, Livingstone, Byrne, and Girolami (2014), a state-dependent preconditioning matrix alters slightly the invariant distribution of the SDE, so that S is no longer distributed according to $f(\cdot)$ in stationarity.[9] As described in Xifara, Sherlock, Livingstone, Byrne, and Girolami (2014), one can correct this by augmenting the drift term, so that

$$dS = \frac{1}{2}\Lambda(S)\nabla log(f(S))dt + \Upsilon(S)dt + \sqrt{\Lambda}(S_t)dW$$

where $\Upsilon_i(S) = \frac{1}{2}\sum_j \frac{\partial}{\partial S_j}\Lambda_{i,j}(S)$. Using a discretization of this SDE as a proposal density is the (full) Manifold Metropolis-Adjusted Langevin Algorithm (MMALA) of Girolami and Calderhead (2011). As noted in Liu (2001) and Girolami and Calderhead (2011), among others, the MALA algorithm is also a special case of Hybrid or Hamiltonian Monte Carlo.

5 Imputation Results

5.1 Simulating Data

Before proceeding to use a MALA sampler to impute total nondurable consumption using the PSID and CEX, we present evidence on the performance of the estimator using simulated data. The idea is to create artificial data sets that resemble the CEX and PSID, but that have known parameters and

[9] It should be emphasized that even though the SDE no longer has $f(s)$ as its invariant density, the Metropolis sampling algorithm that uses the discretized SDE to generate proposals is still a valid algorithm to draw from $f(\cdot)$.

latent states. To make our DGP realistic, we first run a BPP-style regression on real CEX data to estimate $(\widehat{\beta}, \widehat{\gamma}, \widehat{\sigma}_f)$. These will be the "true" parameters used in the food demand equation to generate data. With these known parameters we need $c^{*sim}_{itx}, c^{*sim}_{itp}, D_{itx}, D_{itp}$ in order to use the food demand equation to generate $f^{sim}_{itx}, f^{sim}_{itp}$.

Start to create the simulated dataset by keeping the observed D_{itx}, D_{itp}. To generate latent total nondurable consumption in the CEX, sample c^{*sim}_{itx} with replacement from measured $\{c_{itx}\}$ in the CEX dataset, ensuring the distribution of true consumption resembles measured consumption in the CEX.[10]

To generate latent total nondurable consumption in the PSID, first scale $\{c_{itx}\}$ in such a way that c^{*sim}_{itp} has the correct mean and variance if backed out of a BPP regression using $(\widehat{\beta}, \widehat{\gamma}, \widehat{\sigma}_f)$. To generate the simulated PSID nondurable consumption data, sample c^{*sim}_{itp} with replacement from the scaled distribution of $\{c_{itx}\}$, ensuring the distribution of true consumption is consistent with the statistical relationship implied by the BPP regression.

To simulate c^{sim}_{itx}, we need to take a stand on the amount of measurement error present in the data. For this exercise we set $\widehat{\sigma}_{cx} = 0.05 * \sigma_{c_{itx}}$, such that 5% of the standard deviation of measured total nondurable consumption is attributable to measurement error. With $\widehat{\sigma}_{cx}$ calculated, generate $c^{sim}_{itx} = c^{*sim}_{itx} + \widehat{\sigma}_{cx}\nu_{itj}$. Finally, use the statistical model to generate $f^{sim}_{itj} = D_{itj}\widehat{\beta} + \widehat{\gamma} c^{*sim}_{itj} + \widehat{\sigma}_f e_{itj}$, using draws from the normal distribution for e_{itj}.

5.2 Specifying Priors

Given our simulated data, we need to specify values for the prior distributions before estimating the model:

$$\begin{aligned} p(c^*_{itx}) &= N(\bar{c}_{itx}, \sigma^2_{c_{itx}}) & p(c^*_{itp}) &= N(1.08\bar{c}_{itx}, 1.25\sigma^2_{c_{itx}}) \\ p(\beta) &= N(0,1) & p(\gamma) &= N(0,1) \\ p(\sigma_f) &= IG(11,3) & p(\sigma_{cx}) &= IG(6,0.5). \end{aligned}$$

There are no strong reasons to pick exactly the standard normal distribution for β and γ. However, the principle is to have a weak prior with large variance such that the posterior is largely informed by the likelihood. For the standard deviation of the food demand equation, σ_f, we choose parameters of the inverse gamma distribution such that the prior mean is around $\widehat{\sigma}^{BPP}_f$, with a standard deviation of 0.1. For the standard deviation of the CEX nondurable consumption measurement equation, σ_{cx}, we choose parameters of the inverse gamma distribution such that the prior mean matches $0.05\widehat{\sigma}_{cx}$, reflecting our prior belief on the amount of measurement error in the CEX, with a standard deviation of 0.05.

We set $p(c^*_{itx})$ equal to a normal distribution with mean and variance equal to the mean and variance of the empirical distribution of measured total nondurable consumption in the CEX. We set $p(c^*_{itp})$ equal to a normal distribution with mean and variance a scaled up version of that for

[10]Note that this sampling removes correlation between D_{itx} and c^{*sim}_{itx}, which could potentially complicate estimation.

c^*_{itx}, reflecting higher average food expenditure in the PSID. The particular scaling factors are chosen to generate the mean and variance of the empirical distribution of total nondurable consumption in the PSID implied by the estimated BPP regression. Note that these are priors for an individual's consumption at a given time period. This prior suggests that each c^*_{itj} was likely drawn from a good guess of the empirical distribution of c^*_{itj} across all people and times.[11]

In analyzing the imputed values, it is important to analyze the sensitivity of the estimator to prior specification. One approach is to measure the difference between prior and posterior distributions, as any change is a measure of how informative the data are. As will be shown, in our application, the posteriors are very distinct from the priors.

5.3 Imputation Results Using Simulated Data

We can now use the artificial dataset, (c^{sim}_{itx}, f^{sim}_{itx}, f^{sim}_{itp}, D_{itx}, D_{itp}), to draw from the joint posterior distribution of parameters and latent states, (c^{*sim}_{itx}, c^{*sim}_{itp}, $\widehat{\beta}$, $\widehat{\sigma}_f$, $\widehat{\sigma_{cx}}$, $\widehat{\gamma}$). Since these values are known via simulation, we can test the accuracy and precision of our proposed Bayesian imputation algorithm to analyze the degree of imputation uncertainty surrounding the latent variables.

Although theoretically MALA has comparatively shorter mixing times relative to a RWMH algorithm, experience has shown that this sampler still has high integrated autocorrelation times in absolute terms for the problem at hand and thus a long chain is needed for accurate inference. We simulate a chain length of 25 million, thinning the chain to reduce correlation in draws by keeping every 5,000th draw. We then burn the first 1,000 draws, ultimately leaving 4,000 draws for inference. See Table 1 for statistics of the posterior distribution for parameters θ. One key result is that the estimated β parameters are almost all tightly centered very close to the true parameter values. The second key result in this table is that γ is well estimated. This parameter relates total nondurable consumption to food consumption, and thus has a very strong influence on the posterior of c^*_{itj}. The posterior of σ_f is very tight and centered at its true value, while the posterior for σ_{cx} is slightly upward biased. Furthermore, as β and γ had standard normal priors, it is immediately apparent from analyzing moments of the posterior that the data strongly informed the estimates. The prior for σ_f had a mean of 0.3 and standard deviation of 0.1, while the prior for σ_{cx} had a mean of 0.1 and a standard deviation of 0.05, revealing that the posterior is also very different from the prior in these dimensions.

Table 2 and Table 3 provide estimation results on the posterior distributions of c^*_{itx} and c^*_{itp}. For each latent observation of an individual's nondurable consumption at a point in time we have a sequence of draws from the posterior distribution. The first row presents statistics of the true simulated data. The second row of these tables presents statistics of the posterior distribution. To highlight the effects of abstracting from imputation uncertainty, in the third row we calculate the posterior mean for each c^*_{itj} as a "point estimate" (for each itj, we average across the draws of c^*_{itj} produced

[11] Note that in the estimation results presented in Table 2 and Table 3 the posteriors are substantially tighter than the prior, as can be seen by comparing StDev(Truth) with Mean(StDev(c^*_{itj})).

by the Markov chain) and then compare statistics of the resulting distribution of mean estimates of individual consumption to statistics of the true simulated data.[12] Table 2 documents that mean c^*_{itx} is distributed very similarly to the truth, with the mean, standard deviation and 95% interval of mean c^*_{itx} all being very close to corresponding statistics of the true simulated data. The fourth row shows that, on average, the standard deviation of the posterior of c^*_{itx} is quite small, given that mean c^*_{itx} is on average 9.398 and the standard deviation is on average 0.0375. Also, the posterior standard deviation does not vary much across individual/time observations, as the standard deviation of the posterior standard deviation is 0.0005. These results suggest highly accurate and precise estimates of c^*_{itx}, which is not surprising given that the data contain c_{itx}, which is a noisy measure of c^*_{itx}.

More importantly, Table 3 suggests quite accurate estimation results for imputed c^*_{itp}, even though the data contain no direct measure of this variable. While not lining up as exactly as the CEX estimates, the statistics of mean c^*_{itp} are also very close to the corresponding statistics of the simulated data, with the true mean equal to 10.1199 and the mean of the posterior mean c^*_{itp} equal to 10.1424. Not surprisingly, the standard deviation and 95% interval for PSID imputed consumption are not as close to the truth as for the CEX, but they are still quite close. The second major result seen in Table 3 is that the mean posterior standard deviation of c^*_{itp} is 0.2439—significantly larger than the 0.0375 seen in the CEX. This provides a measure of the uncertainty surrounding the imputed consumption estimate. Because of the increased imputation uncertainty, the point estimate given by the mean of the posterior distribution does a worse job at matching the variability of the true data, as seen in the standard deviation and tail-quantile statistics. This suggests that any research that would use a point estimate of imputed consumption as "data" and ignore that it is an estimated value could be significantly understating standard errors and making possibly worse mistakes.

Finally Table 4 and Table 5 measure the imputation error when using the mean of the posterior distribution as a point estimate of missing consumption. Define imputation error $\varepsilon_{itj} := \widehat{c}^*_{itj} - c^*_{itj}$, with $\widehat{c}^*_{itj}$ the mean of the posterior distribution of c^*_{itj}. Confirming previous results, mean imputation error in the CEX is extremely small, with 0.0003 mean error measuring consumption that averages 9.398. Reflecting the size of measurement error, the standard deviation of this imputation error is 0.0208. While Table 2 demonstrated that the distribution of estimated consumption matched the distribution of actual consumption in the CEX, Table 4 suggests that the estimation of each observation is highly accurate and precise. Table 5 presents the corresponding results for the PSID, again documenting that the mean imputation error in the PSID is quite small, with 0.0225 mean error measuring consumption that averages 10.1424. However, in the PSID, the standard deviation of the imputation error is significantly larger than in the CEX, at 0.2394. This significant variability in estimation error again suggests that the imputation results should not be treated as "data" without acknowledging estimation error.

[12] The mean of the posterior can heuristically be thought of as a point estimate for c^*_{itj} corresponding to a particular loss function.

5.4 Imputation Results Using PSID and CEX Data

After providing evidence that the Bayesian estimator performs quite well on simulated data, this section focuses on performing the imputation using real CEX and PSID data. For this exercise we limit the estimation to the same time period—1980–1992—and selection criteria as BPP to facilitate comparison across estimators. This results in approximately 8,600 person-time observations in the CEX and 15,500 person-time observations in the PSID.

We estimate the model using the same procedure outlined for the simulated data with the same priors. We simulate a chain length of 25 million draws. We thin the chain to reduce correlation in draws by keeping every 5,000th draw, burn the first 1,000 of the remaining draws, ultimately leaving 4,000 draws for inference.

Table 6 presents statistics of the posterior distribution for parameters θ. The mean of the posterior for β is largely in line with the output from a BPP regression. As in Blundell, Pistaferri, and Preston (2008), some of the elements of β have large confidence intervals while others have quite tight posteriors, but, importantly, essential parameters like σ_f, σ_{cx}, and especially γ are tightly estimated. Table 7 and 8 provide information on the distribution of the posterior mean and standard deviation of latent total nondurable consumption. As was the case in the simulated data exercise, the standard deviation of the consumption posterior in the CEX is quite small at 0.07, while it is larger for the PSID at 0.25. We view the uncertainty surrounding the PSID imputed values as large enough that it must be accounted for in subsequent use, but not so large as to invalidate the usefulness of the imputed data. We explore this further in the next section. However, we also note that there is clearly room for improved statistical models and the use of more data to reduce the substantial imputation uncertainty, as discussed in Section 7.

6 Applications

The goal of this section is to demonstrate how to use imputed consumption in economic analysis and also to demonstrate how ignoring imputation uncertainty can generate inaccurate results.[13] The data used are from the PSID and CEX. Sample selection and data treatment can significantly alter quantitative results, so these excercises are not designed to be authoritative measures of the calculated statistics. However, the methodology and qualitative results are valuable in demonstrating how to use the imputed data and the gains from using the full information from the posterior distribution in second stage analysis.

We document the change in consumption inequality over time by analyzing a time series of statistics from the consumption distribution, with an emphasis on the uncertainty surrounding these measures. In order to highlight the importance of imputation uncertainty, the mean, cross-sectional standard

[13] In some cases, the imputed distribution of consumption is directly the data of interest. In other cases, the consumption measures may be necessary regressors in a second stage analysis. As an alternative to multiple imputation, if feasible, one option is to explicitly stack the imputation and the second stage equations in to one system and to perform the estimation jointly.

deviation, and quantiles of the consumption distribution are calculated using two different methods. The first approach uses the full posterior distribution while the second uses the mean of the posterior distribution as a point estimate. To use the entire posterior, calculate each statistic of interest for each draw, generating a distribution of each statistic. Quantiles of the distribution of the statistic can be used to demonstrate how the uncertainty surrounding imputed values translates into uncertainty around the statistics of interest. Instead, to use the reduced information method, calculate the mean of the posterior distribution (for each latent state), and then calculate the statistics of interest using this mean as a point estimate of consumption. The approach using the full posterior distribution is related to the methodology proposed in the multiple imputation literature, where the pitfalls of using a single imputation have been well established.[14]

Figure 1 documents the evolution of the mean of the consumption distribution over time. The median of the mean consumption statistics when using the full posterior matches nearly identically the mean calculated using the point estimates. Using the full information of the posterior, the 10th percentile and 90th percentile are tightly around the median of the mean consumption statistic, suggesting this statistic is tightly estimated.

Figure 2 graphs the evolution of the cross-sectional standard deviation of consumption over time. This is a common measure of inequality. There is no consensus in the literature on whether or not consumption inequality has risen over time. The striking result is that using the point estimate to calculate the standard deviation of the consumption distribution significantly underestimates the true standard deviation. As can be seen in Figure 2 the level of inequality measured by the point estimate is well below the 10th percentile of the true distribution of the standard deviation statistic. However, the change over time seems to be the same across the two methods, suggesting no rise in consumption inequality over time.

Figure 3 graphs percentiles of the consumption distribution using the two methods. Using the first method, the 90th, 50th, and 10th percentiles of the consumption distribution in each year are calculated from the posterior distribution. Since each draw is a draw from the consumption distribution, simply calculate the percentile across all draws for all individuals in a given year. Using the second method, the 90th, 50th, and 10th percentiles of the consumption distribution for each year are calculated as the 90th, 50th, and 10th percentile of the point estimates. Again, the methods agree near the center of the distribution, as the median of the posterior is essentially the same as the median of the means of the posterior. However, using the point estimates significantly understates the variance in the consumption distribution, as the 10th and 90th percentiles of the full-information consumption distribution are outside of the 10th and 90th percentiles derived from the point estimates. To highlight

[14] Quoting from Rubin (1988) describing the use of multiply-imputed data sets, "The proper multiple imputations within each model are called repetitions and are combined ... to create one inference under each model. The inferences across models are not combined but are contrasted to reveal sensitivity of inference to assumptions about the reasons for the missing data. The critical issue then is how to analyze the repetitions within one model to yield a valid inference under the posited reasons for missing data. The key idea is that M repetitions yield M completed data sets, each of which can be analyzed by standard complete-data methods just as if it were the real data set. The M complete-data analyses based on the M repeated imputations are then combined to create one repeated-imputation inference."

the size of the discrepancy, for 1985 the 90th percentile of the means of consumption is \$39,340, while the 90th percentile of the consumption distribution is approximately \$43,477, an 11 percent difference. Thus, for any analysis that depends on the variability or tail properties of the imputed consumption distribution, accounting for imputation uncertainty by using multiple imputation methods is essential.

7 Extensions

In the above analysis, we highlighted our proposed imputation methodology by adopting a simple statistical model close to that of Blundell, Pistaferri, and Preston (2006). However, our sampling algorithm is designed to be flexible enough to be able to accommodate a wide array of demand systems. Below we outline a few promising extensions that might one day improve the quality of the imputed values. This is done primarily either by providing the estimator with more information from additional data sources or by improving the structure of the statistical model.[15] While we emphasize that the MALA algorithm as described above is still able to estimate each of these in kind, we note that the sampling algorithm itself may be improved along a few dimensions.

7.1 Interaction Terms

Our basline model assumes that the true (unobservable) consumption does not interact with any of the elements D_j in the food-demand equation. For ease of exposition, we will assume there is only one obserable that interacts with c^*_{itj}, which we'll call H_{itj}. Then, food demand is given by

$$f_{itj} = D'_{itj}\beta + \gamma c^*_{itj} + \alpha c^*_{itj} H_{itj} + \sigma_f e_{itj}$$

for each of the CEX and PSID datasets. The likelihood computation with such an addition is straightforward, with some minor adjustments needed to the gradient to implement MALA.

7.2 A richer measurement error structure

Our baseline includes iid across time, agent, and dataset, measurement errors. Recently, Aguiar and Bils (2013) have emphasized income-quintile specific errors in the food demand equation. This can be incorporated into the baseline model by expanding the individual characteristics matrix D_i as follows. Suppose that the level of observed food consumption, $e^{f^*_{itx}}$, is related to the level of true food consumption, $e^{f^*_{itx}}$, as

$$e^{f_{itx}} = e^{f^*_{itx} + \phi_{itx} + \varepsilon_{itx}}.$$

Here ϕ_{itx} is an income-level specific measurement error, motivated by the Aguiar and Bils (2013) finding that the rich may under report food consumption relative to the poor, while ε_{itx} is classical

[15] The extensions discussed all maintain the log demand system structure used by BPP. Another possible direction for improvement would be to base the statistical model on a richer demand system (such as those discussed in Pollak and Wales (1995)).

measurement error. We can plug this into our original food demand equation to get,

$$\begin{aligned} f_{itx} &= \phi_{itx} + D_{itx}^{'}\beta + \gamma c_{itx}^{*} + \sigma_f e_{itx} + \varepsilon_{itx} \\ &= \widetilde{D}_{itx}^{'}\widetilde{\beta} + \gamma c_{itx}^{*} + \widetilde{\sigma}_f \widetilde{e}_{itx} \end{aligned}$$

where $\widetilde{D}$ now has a set of dummy variables representing agent i's income quintile, with associated loadings embedded in $\widetilde{\beta}$. That the measurement error in food ε_{itx} is uncorrelated with the variables in $\widetilde{D}$ allows us to estimate a single food demand error $\widetilde{\sigma}_f \widetilde{e}_{itx}$. Further following Aguiar and Bils (2013), additional good-specific measurement error in food would be subsumed in our estimated constant in the food-demand equation.

7.3 Instruments

The Blundell, Pistaferri, and Preston (2006) specification uses instrumental variables in order to estimate the γ parameter. As described in Minka (1999) and Lopes and Polson (2014), one can add instruments in a Bayesian context as well. One way to add an instrument Z_t to the estimation routine is by adding another observation equation,

$$Z_{itj} = \delta_0 + \delta_1 c_{itj}^{*} + \sigma_z \epsilon_{itj}$$

This specification adds 3 parameters (that are shared across datasets), but brings in $T_x N_x + T_p N_p$ observables Z_{itj}. Adding the instruments in this way effectively models the total consumption with an errors-in-variables model. In a simple linear model such as ours, Minka (1999) describes how to marginalize the c_{itj}^{*} parameters to perform inference on θ. As explored by Lopes and Polson (2014), one must be careful with the specification of priors in such an instance.

7.4 Incorporating additional datasets

Another extension is motivated by Attanasio, Battistin, and Ichimura (2007), who combine the two CEX datasets: Diary and Survey. In addition to the CEX Survey data that we have been using throughout the paper, the CEX also collects data from consumer diaries. Consumers are asked to keep logs of all consumption expenditures, as opposed to answering questions at the end of the quarter about the previous quarters expenditures. The CEX Diary and CEX Survey both contain measures on all consumption categories. It is believed that the CEX Diary has less measurement error for daily small consumption categories, while the Survey is better at measuring large infrequent purchases. We can treat the CEX Diary measure and CEX Survey measure of the same consumption category as two noisy observations of the same variable. In such an instance, we will let xd and xs denote the Diary and Survey measures, respectively. Then, observable consumption in the CEX is

$$\begin{aligned} c_{itxd} &= c_{itx}^{*} + \sigma_{cxd} v_{itxd} \\ c_{itxs} &= c_{itx}^{*} + \sigma_{cxs} v_{itxs}. \end{aligned}$$

In addition to using more data from the CEX, there is more data in the PSID that could be useful for the imputation procedure. In 1999 the PSID added survey questions to start measuring consumption categories beyond just food. In 2005 the PSID expanded this set of questions. While this may reduce the need for imputation going forward, for many questions analyzing time trends it may be desirable to use a variable that is consistently available across all time periods, instead of using the imputed measure only when direct measures are not available. Thus, we can augment the estimator to use measured consumption in the PSID post 1999 to improve the imputed values for all time periods, in a similar manner to the treatment of measurement error in total nondurable consumption in the CEX.

7.5 A time dimension

In our baseline model, the unobserved total consumption in the PSID and CEX does not have a time dimension. We can easily add one, for example, by assuming that

$$c^*_{itj} = c^*_{i,t-1,j} + \sigma_\varepsilon \varepsilon_{itj}$$

with $\varepsilon_{itj} \sim N(0,1)$. Now, total consumption is intertemporally linked for each member of the survey. In the CEX, which is constructed by rolling panels of agents, there will only be 1 year to link together. The PSID has an important time dimension and presumably taking this into account will sharpen inference on c^*_{itp}. Such a specification adds one static parameter to the dimension of the state space: σ^2_ε.

8 Conclusion

This paper develops a Bayesian estimator to impute total nondurable consumption in the PSID using CEX data. We treat an individual agent's total nondurable consumption in each year as a latent variable and sample from the joint posterior distribution of latent states and parameters governing the food demand equation, given the data contained in the PSID and CEX simultaneously. This extremely high dimensional posterior distribution imposes sampling challenges. The main technical contribution of this paper is to develop a suitable sampler by exploiting derivatives of the posterior distribution using a Metropolis Adjusted Langevin Algorithm. The developed imputation algorithm is demonstrated to perform well on the baseline model and is also flexible, as it is adaptable to extensions that use more advanced statistical models and additional sources of data. The Bayesian estimator has advantages over existing methods in terms of efficiency and also provides essential information about imputation error and parameter uncertainty. Results using simulated data and applications using PSID and CEX data highlight the need to treat imputed values as estimates as opposed to measurements.

References

Aguiar, M., and M. Bils (2013): "Has Consumption Inequality Mirrored Income Inequality," *mimeo*.

Altonji, J. G. (1986): "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *The Journal of Political Economy*, 94(3), 176–215.

Attanasio, O., E. Battistin, and H. Ichimura (2007): "What Really Happened To Consumption Inequality in the US?," in *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, ed. by E. R. Berndt, and C. R. Hulten. University of Chicago Press.

Attanasio, O., and S. J. Davis (1996): "Relative Wage Movements and the Distribution of Consumption," *Journal of Political Economy*, 104(6), 1227–1262.

Bedard, M. (2007): "Weak Convergence of Metropolis Algorithms for Non-iid Target Distributions," *Annals of Applied Probability*, 17, 1222–1244.

Besag, J. E. (1994): "Comments on 'Representations of Knowledge in Complex Systems' by Ulf Grenander and Michael I. Miller," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56, 591–592.

Beskos, A., G. Roberts, and A. Stuart (2009): "Optimal Scalings for Local Metropolis-Hastings Chains on Non-product Targets in High Dimensions," *Annals of Applied Probability*, 19(3), 863–898.

Blundell, R., L. Pistaferri, and I. Preston (2006): "Imputing Consumption in the PSID Using Food Demand Estimates in the CEX," *Stanford University Working Paper*.

——— (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98:5, 1887–1921.

Breyer, L. A., M. Piccioni, and S. Scarlatti (2004): "Optimal Scaling of MALA for Nonlinear Regression," *The Annals of Applied Probability*, 14(3), 1479–1505.

Christensen, O. F., G. O. Roberts, and J. S. Rosenthal (2005): "Scaling Limits for the Transient Phase of Local Metropolis-Hastings Algorthims," *Journal of the Royal Statistical Society: Series B*, 67(2), 253–268.

Cochrane, J. H. (1991): "A Simple Test of Consumption Insurance," *Journal of Political Economy*, 99(5), 957–976.

Cooper, D. (2010): "Imputing Household Spending in the Panel Study of Income Dynamics: A Comparison of Approaches," *FRB Boston Working Paper*, 10(12).

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2014): *Bayesian Data Analysis*. Chapman & Hall, 3rd edn.

Geweke, J. (2005): *Contemporary Bayesian Econometrics and Statistics*. Wiley Series in Probability and Statistics.

Girolami, M., and B. Calderhead (2011): "Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods," *Journal of the Royal Statistical Society*, 73, 123–214.

Guvenen, F., and A. Smith (2014): "Inferring Labor Income Risk and Partial Insurance from Economic Choices," *Econometrica*, forthcoming.

Hall, R. E., and F. S. Mishkin (1982): "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50(2), 461–481.

Hayashi, F., J. G. Altonji, and L. Kotlikoff (1996): "Risk-Sharing Between and Within Families," *Econometrica*, 64(2), 261–294.

Herbst, E. (2012): "Gradient and Hessian-based MCMC for DSGE Models," *working paper*.

Krueger, D., and F. Perri (2006): "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *The Review of Economic Studies*, 73(1), 163–193.

Liu, J. S. (2001): *Monte Carlo Strategies in Scientific Computing*, Springer Series in Statistics. Springer.

Livingstone, S., and M. Girolami (2014): "Information-Geometric Markov Chain Monte Carlo Methods using Diffusions," *Entropy*, 16(6), 3074–3102.

Lopes, H. F., and N. G. Polson (2014): "Bayesian Instrumental Variables: Priors and Likelihoods," *Econometric Reviews*, 33, 100–121.

Matthews, P. (1993): "A Slowly Mixing Markov Chain with Implications for Gibbs Sampling," *Statistics and Probability Letters*, 17, 231–0236.

Mattingly, J. C., N. S. Pillai, and A. M. Stuart (2012): "Diffusion Limits of the Random Walk Metropolis Algorithm in High Dimensions," *The Annals of Applied Probability*, 22(3), 881–930.

Minka, T. P. (1999): "Linear Regression with Errors in Both Variables: A Proper Bayesian Approach," Discussion paper, MIT Media Lab.

Pillai, N. S., A. M. Stuart, and A. H. Thiéry (2012): "Optimal Scaling and Diffusion Limits for the Langevin Algorithm in High Dimensions," *The Annals of Applied Probability*, 22(6), 2320–2356.

Pollak, R. A., and T. J. Wales (1995): *Demand System Specification and Estimation*. Oxford University Press.

Roberts, G. O., A. Gelman, and W. R. Gilks (1997): "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms," *Annals of Applied Probability*, 7(1), 110–120.

Roberts, G. O., and J. S. Rosenthal (1998): "Optimal Scaling of Discrete Approximations to Langevin Diffusions," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(1), 255–268.

——— (2001): "Optimal Scaling for Various Metropolis-Hastings Algorithms," *Statistical Science*, 16(4), 351–367.

Roberts, G. O., and O. Stramer (2002): "Langevin Diffusions and Metropolis-Hastings Algorithms," *Methodology and Computing in Applied Probability*, 4, 337–357.

Roberts, G. O., and R. L. Tweedie (1996): "Exponential Convergence of Langevin Distributions and Their Discrete Approximations," *Bernoulli*, 2(4), 341–363.

Rubin, D. B. (1987): *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.

——— (1988): "An Overview of Multiple Imputation," in *Proceedings of the Survey Research Methods Section*, pp. 79–84. American Statistical Association.

——— (1996): "Multiple Imputation After 18+ Years," *Journal of the American Statistical Association*, 91(434), 473–489.

Simpson, D. P., I. W. Turner, C. M. Strickland, and A. N. Pettitt (2013): "Scalable Iterative Methods for Sampling from Massive Gaussian Random Vectors," *arxiv:1312.1476v1*.

Skinner, J. (1987): "A Superior Measure of Consumption from the Panel Study of Income Dynamics," *Economics Letters*, 23(2), 213–216.

Xifara, T., C. Sherlock, S. Livingstone, S. Byrne, and M. Girolami (2014): "Langevin Diffusions and the Metropolis-adjusted Langevin Algorithm," *Statistics and Probability Letters*, 91, 14–19.

Zeldes, S. P. (1989): "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy*, 97(2), 305–346.

Ziliak, J. (1998): "Does the Choice of Consumption Matter: An Application to the Permanent Income Hypothesis," *Journal of Monetary Economics*, 41, 201216.

A Derivatives of the Log Posterior

$$\frac{\partial log(p(Y|\theta,X))}{\partial c^*_{itx}} = -\widehat{e}_{itx}\frac{\partial \widehat{e}_{itx}}{\partial c^*_{itx}} - \widehat{v}_{itx}\frac{\partial \widehat{v}_{itx}}{\partial c^*_{itx}}$$

$$\frac{\partial log(p(Y|\theta,X))}{\partial c^*_{itp}} = -\widehat{e}_{itp}\frac{\partial \widehat{e}_{itp}}{\partial c^*_{itp}}$$

$$\frac{\partial log(p(Y|\theta,X))}{\partial \beta} = -\sum_t^{T_x}\sum_i^{N_{tx}} \widehat{e}_{itx}\frac{\partial \widehat{e}_{itx}}{\partial \beta} - \sum_t^{T_p}\sum_i^{N_{tp}} \widehat{e}_{itp}\frac{\partial \widehat{e}_{itp}}{\partial \beta}$$

$$\frac{\partial log(p(Y|\theta,X))}{\partial \sigma_f} = -\sum_t^{T_x}\sum_i^{N_{tx}} \widehat{e}_{itx}\frac{\partial \widehat{e}_{itx}}{\partial \sigma_f} - \frac{T_xN_x}{\sigma_f} - \sum_t^{T_x}\sum_i^{N_{tx}} \widehat{e}_{itp}\frac{\partial \widehat{e}_{itp}}{\partial \sigma_f} - \frac{T_pN_p}{\sigma_f}$$

$$\frac{\partial log(p(Y|\theta,X))}{\partial \sigma_{cx}} = -\sum_t^{T_x}\sum_i^{N_{tx}} \widehat{v}_{itx}\frac{\partial \widehat{v}_{itx}}{\partial \sigma_{cx}} - \frac{T_xN_x}{\sigma_{cx}}$$

$$\frac{\partial log(p(Y|\theta,X))}{\partial \gamma} = -\sum_t^{T_x}\sum_i^{N_{tx}} \widehat{e}_{itx}\frac{\partial \widehat{e}_{itx}}{\partial \gamma} - \sum_t^{T_p}\sum_i^{N_{tp}} \widehat{e}_{itp}\frac{\partial \widehat{e}_{itp}}{\partial \gamma}$$

With our specification:

$$\frac{\partial \widehat{e}_{itx}}{\partial c^*_{itx}} = \frac{\partial \widehat{e}_{itp}}{\partial c^*_{itp}} = \frac{-\gamma}{\sigma_f}$$

$$\frac{\partial \widehat{v}_{itx}}{\partial c^*_{itx}} = \frac{-1}{\sigma_{cx}}$$

$$\frac{\partial \widehat{e}_{itx}}{\partial \beta} = \frac{-D'_{itx}}{\sigma_f}; \quad \frac{\partial \widehat{e}_{itp}}{\partial \beta} = \frac{-D'_{itp}}{\sigma_f}$$

$$\frac{\partial \widehat{e}_{itx}}{\partial \sigma_f} = \frac{-\widehat{e}_{itx}}{\sigma_f}; \quad \frac{\partial \widehat{e}_{itp}}{\partial \sigma_f} = \frac{-\widehat{e}_{itp}}{\sigma_f}$$

$$\frac{\partial \widehat{v}_{itx}}{\partial \sigma_{cx}} = \frac{-\widehat{v}_{itx}}{\sigma_{cx}}$$

$$\frac{\partial \widehat{e}_{itx}}{\partial \gamma} = \frac{-c^*_{itx}}{\sigma_f}; \quad \frac{\partial \widehat{e}_{itp}}{\partial \gamma} = \frac{-c^*_{itp}}{\sigma_f}$$

Plugging this in, we get the gradient of the log likelihood as,

$$\begin{aligned}
\frac{\partial log(p(Y|\theta, X))}{\partial c^*_{itx}} &= \widehat{e}_{itx}\frac{\gamma}{\sigma_f} + \widehat{v}_{itx}\frac{1}{\sigma_{cx}} \\
\frac{\partial log(p(Y|\theta, X))}{\partial c^*_{itp}} &= \widehat{e}_{itp}\frac{\gamma}{\sigma_f} \\
\frac{\partial log(p(Y|\theta, X))}{\partial \beta} &= \sum_t^{T_x}\sum_i^{N_{tx}} \widehat{e}_{itx}\frac{D'_{itx}}{\sigma_f} + \sum_t^{T_p}\sum_i^{N_{tp}} \widehat{e}_{itp}\frac{D'_{itp}}{\sigma_f} \\
\frac{\partial log(p(Y|\theta, X))}{\partial \sigma_f} &= \sum_t^{T_x}\sum_i^{N_{tx}} \frac{\widehat{e}^2_{itx}}{\sigma_f} - \frac{T_x N_x}{\sigma_f} + \sum_t^{T_x}\sum_i^{N_{tx}} \frac{\widehat{e}^2_{itp}}{\sigma_f} - \frac{T_p N_p}{\sigma_f} \\
\frac{\partial log(p(Y|\theta, X))}{\partial \sigma_{cx}} &= \sum_t^{T_x}\sum_i^{N_{tx}} \frac{\widehat{v}^2_{itx}}{\sigma_{cx}} - \frac{T_x N_x}{\sigma_{cx}} \\
\frac{\partial log(p(Y|\theta, X))}{\partial \gamma} &= \sum_t^{T_x}\sum_i^{N_{tx}} \widehat{e}_{itx}\frac{c^*_{itx}}{\sigma_f} + \sum_t^{T_p}\sum_i^{N_{tp}} \widehat{e}_{itp}\frac{c^*_{itp}}{\sigma_f}
\end{aligned}$$

The derivatives of the log prior are,

$$\begin{aligned}
\frac{\partial log(p(c^*_{itx}))}{\partial c^*_{itx}} &= -\frac{(c^*_{itx} - \mu_{c^*_{itx}})}{\sigma^2_{c^*_{itx}}} \\
\frac{\partial log(p(c^*_{itp}))}{\partial c^*_{itp}} &= -\frac{(c^*_{itp} - \mu_{c^*_{itp}})}{\sigma^2_{c^*_{itp}}} \\
\frac{\partial log(p(\beta))}{\partial \beta} &= -\Sigma_\beta^{-1}(\beta - \mu_\beta) \\
\frac{\partial log(p(\sigma_f))}{\partial \sigma_f} &= \frac{\beta_{\sigma_f}}{\sigma_f^2} - \frac{(\alpha_{\sigma_f} + 1)}{\sigma_f} \\
\frac{\partial log(p(\sigma_{cx}))}{\partial \sigma_{cx}} &= \frac{\beta_{\sigma_{cx}}}{\sigma_{cx}^2} - \frac{(\alpha_{\sigma_{cx}} + 1)}{\sigma_{cx}} \\
\frac{\partial log(p(\gamma))}{\partial \gamma} &= -\frac{(\gamma - \mu_\gamma)}{\sigma^2_\gamma}
\end{aligned}$$

The diagonal elements of the Hessian of the log posterior are,

$$\frac{\partial^2 log(p(\theta, X|Y))}{\partial c_{itx}^{*2}} = -\frac{\gamma^2}{\sigma_f^2} - \frac{1}{\sigma_{cx}^2} - \frac{1}{\sigma_{c_{itx}^*}^2}$$

$$\frac{\partial^2 log(p(\theta, X|Y))}{\partial c_{itp}^{*2}} = -\frac{\gamma^2}{\sigma_f^2} - \frac{1}{\sigma_{c_{itp}^*}^2}$$

$$\frac{\partial^2 log(p(\theta, X|Y))}{\partial \beta \partial \beta'} = -\sum_{t}^{T_x}\sum_{i}^{N_{tx}} \frac{1}{\sigma_{fx}^2} D_{itx}^{'} D_{itx} - \sum_{t}^{T_p}\sum_{i}^{N_{tp}} \frac{1}{\sigma_f^2} D_{itp}^{'} D_{itp} - \Sigma^{-1}$$

$$\frac{\partial^2 log(p(\theta, X|Y))}{\partial \sigma_f^2} = -3\sum_{t}^{T_x}\sum_{i}^{N_{tx}} \frac{\widehat{e}_{itx}^2}{\sigma_f^2} + \frac{N_x T_x}{\sigma_f^2} - 3\sum_{t}^{T_p}\sum_{i}^{N_{tp}} \frac{\widehat{e}_{itp}^2}{\sigma_f^2} + \frac{N_p T_p}{\sigma_f^2} - 2\frac{\beta_{\sigma_f}}{\sigma_f^3} + \frac{(\alpha_{\sigma_f} + 1)}{\sigma_f^2}$$

$$\frac{\partial^2 log(p(\theta, X|Y))}{\partial \sigma_{cx}^2} = -3\sum_{t}^{T_x}\sum_{i}^{N_{tx}} \frac{\widehat{v}_{itx}^2}{\sigma_{cx}^2} + \frac{N_x T_x}{\sigma_{cx}^2} - 2\frac{\beta_{\sigma_{cx}}}{\sigma_{cx}^3} + \frac{(\alpha_{\sigma_{cx}} + 1)}{\sigma_{cx}^2}$$

$$\frac{\partial^2 log(p(\theta, X|Y))}{\partial \gamma^2} = -\sum_{t}^{T_x}\sum_{i}^{N_{tx}} \frac{c_{itx}^{*2}}{{\sigma_f}^2} - \sum_{t}^{T_p}\sum_{i}^{N_{tp}} \frac{c_{itp}^{*2}}{{\sigma_f}^2} - \frac{1}{\sigma_\gamma^2}$$

B Tables

Table 1: Estimation Results for θ on Simulated Data

Parameter	True	Mean	StDev	2.5%	97.5%
β_1	-0.0343	0.1447	0.0794	0.0069	0.3051
β_2	0.0228	0.0219	0.0068	0.0086	0.0355
β_3	0.0580	0.0522	0.0082	0.0363	0.0681
β_4	0.0996	0.0990	0.0120	0.0760	0.1227
β_5	0.0388	0.0308	0.0066	0.0179	0.0439
β_6	0.0176	0.0153	0.0048	0.0062	0.0246
β_7	1.5772	1.3453	0.3502	0.6506	1.9647
β_8	-1.5944	-1.2917	0.3857	-1.9838	-0.5438
β_9	0.0268	0.0264	0.0062	0.0141	0.0384
β_{10}	-0.0226	-0.0207	0.0057	-0.0324	-0.0095
β_{11}	-0.0384	-0.0388	0.0058	-0.0506	-0.0273
β_{12}	0.2611	0.2648	0.0288	0.2105	0.3222
β_{13}	0.0902	0.2067	0.1312	-0.0466	0.4708
β_{14}	0.4818	0.2600	0.1575	-0.0487	0.5676
β_{15}	-0.2073	-0.0963	0.0727	-0.2382	0.0476
β_{16}	-0.3930	-0.4282	0.1022	-0.6263	-0.2326
β_{17}	-0.0267	-0.0307	0.0149	-0.0595	0.0000
β_{18}	0.0026	0.0130	0.0134	-0.0129	0.0402
β_{19}	0.0059	0.0128	0.0130	-0.0122	0.0386
β_{20}	0.0030	0.0148	0.0132	-0.0109	0.0403
β_{21}	0.0040	0.0030	0.0128	-0.0215	0.0277
β_{22}	0.0026	0.0112	0.0118	-0.0113	0.0344
β_{23}	0.0128	0.0070	0.0116	-0.0160	0.0294
β_{24}	0.0507	0.0387	0.0075	0.0241	0.0530
σ_f	0.2340	0.2348	0.0019	0.2310	0.2386
σ_{cx}	0.0206	0.0374	0.0065	0.0259	0.0515
γ	0.8343	0.8200	0.0039	0.8122	0.8273

Table 2: Estimation Results for c_x^* on Simulated Data

Object	Mean	StDev	2.5%	97.5%
Truth	9.3980	0.4144	8.5646	10.2035
$\widehat{c}_{itx}^*$	9.3983	0.4136	8.5666	10.1967
Mean($\widehat{c}_{itx}^*$)	9.3983	0.4119	8.5748	10.1972
StDev($\widehat{c}_{itx}^*$)	0.0375	0.0005	0.0366	0.0384

Table 3: Estimation Results for c_p^* on Simulated Data

Object	Mean	StDev	2.5%	97.5%
Truth	10.1199	0.4566	9.2141	11.0123
$\widehat{c}_{itp}^*$	10.1424	0.4642	9.2202	11.0450
Mean($\widehat{c}_{itp}^*$)	10.1424	0.3950	9.3514	10.9057
StDev($\widehat{c}_{itp}^*$)	0.2439	0.0028	0.2384	0.2493

Table 4: Estimation Results for c_x^* on Simulated Data

Mean(ε_{itx})	StDev(ε_{itx})	Max(Abs(ε_{itx}))	mean(ε_{itx}^2)	StDev(ε_{itx}^2)
0.0003	0.0208	0.0752	0.0004	0.0006

Table 5: Estimation Results for c_p^* on Simulated Data

Mean(ε_{itp})	StDev(ε_{itp})	Max(Abs(ε_{itp}))	mean(ε_{itp}^2)	StDev(ε_{itp}^2)
0.0225	0.2394	1.0312	0.0578	0.0847

Table 6: Estimation Results for θ on CEX and PSID Data

Parameter	Mean	StDev	2.5%	97.5%
β_1	-0.1552	0.0828	-0.3036	0.0205
β_2	0.0304	0.0068	0.0169	0.0433
β_3	0.0571	0.0081	0.0409	0.0723
β_4	0.0697	0.0118	0.0467	0.0924
β_5	-0.0840	0.0069	-0.0976	-0.0706
β_6	-0.0495	0.0048	-0.0586	-0.0400
β_7	2.9222	0.3562	2.2178	3.5613
β_8	-2.7681	0.3924	-3.4906	-2.0028
β_9	0.0444	0.0062	0.0324	0.0567
β_{10}	-0.0529	0.0057	-0.0641	-0.0418
β_{11}	-0.0385	0.0059	-0.0501	-0.0271
β_{12}	0.5027	0.0289	0.4470	0.5596
β_{13}	-0.0897	0.1310	-0.3476	0.1681
β_{14}	0.1973	0.1561	-0.1077	0.5013
β_{15}	-0.2153	0.0720	-0.3552	-0.0771
β_{16}	-0.0886	0.1019	-0.2830	0.1151
β_{17}	-0.0071	0.0149	-0.0353	0.0230
β_{18}	0.0110	0.0134	-0.0145	0.0374
β_{19}	0.0187	0.0131	-0.0065	0.0448
β_{20}	0.0266	0.0134	0.0008	0.0530
β_{21}	0.0294	0.0129	0.0046	0.0543
β_{22}	0.0261	0.0118	0.0034	0.0496
β_{23}	0.0338	0.0115	0.0114	0.0571
β_{24}	0.0904	0.0077	0.0756	0.1056
σ_f	0.2317	0.0027	0.2264	0.2368
σ_{cx}	0.0725	0.0103	0.0512	0.0919
γ	0.8004	0.0047	0.7910	0.8095

Table 7: Estimation Results for c_x^* on CEX and PSID Data

Object	Mean	StDev	2.5%	97.5%
$\widehat{c}_{itx}^*$	9.3935	0.4066	8.5743	10.1808
Mean($\widehat{c}_{itx}^*$)	9.3935	0.4005	8.5847	10.1766
StDev($\widehat{c}_{itx}^*$)	0.0700	0.0009	0.0683	0.0718

Table 8: Estimation Results for c_p^* on CEX and PSID Data

Object	Mean	StDev	2.5%	97.5%
$\widehat{c}_{itp}^*$	10.1496	0.4582	9.2347	10.9984
Mean($\widehat{c}_{itp}^*$)	10.1496	0.3870	9.3710	10.8499
StDev($\widehat{c}_{itp}^*$)	0.2454	0.0028	0.2400	0.2509

C Figures

Figure 1: Mean of the Consumption Distribution

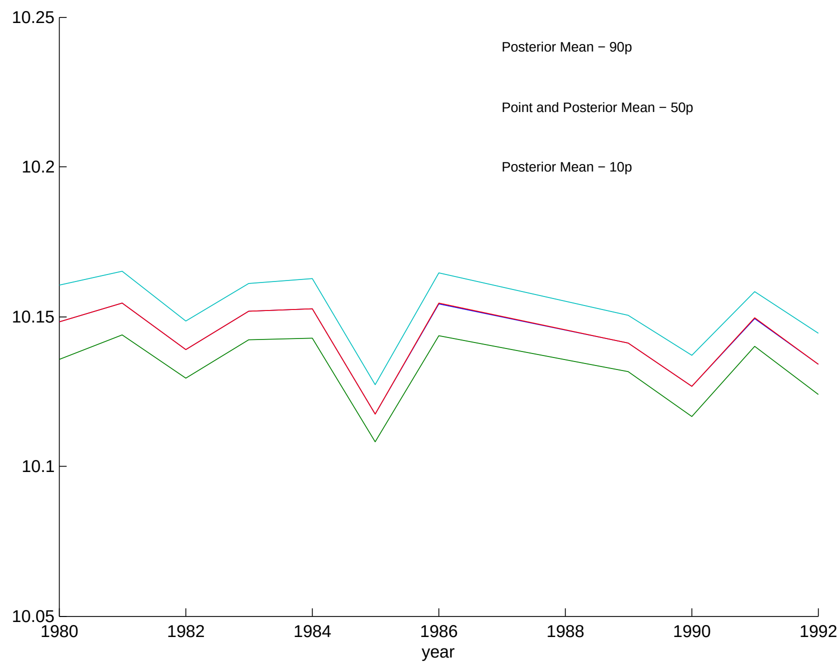

Figure 2: Cross-Sectional Standard Deviation of the Consumption Distribution

Figure 3: Quantiles of the Consumption Distribution