Imputing Rent in Consumption Measures, With an Application to Consumption Poverty in Canada 1997-2009

Sam B. Norris and Krishna Pendakur[1]

University of Toronto and Simon Fraser University

Abstract

We consider two econometric problems in the measurement of poverty, both relating to rent imputation. First, we account for quality differences correlated with selection into owner-occupied versus rental tenure. This correction increases estimated household consumption by 5 per cent over uncorrected estimates and decreases estimated poverty rates quite dramatically. Second, we propose that measurement error induced by the imputation be corrected by imputing a consumption distribution, rather than a consumption level, for each household. This correction increases estimated poverty rates slightly.

We use our methods to measure consumption poverty in Canada, and find that the imputation strategy used influences the patterns observed. For example, measured poverty among the elderly barely declines when one uses our methods, in contrast to the almost six percentage point reduction we find using traditional methods. In our assessment of the over-time evolution of consumption poverty, we find that substantial progress has been made on overall poverty and on child poverty, but that poverty among the elderly hardly changed.

JEL: C34, I32.

[1] We would like to thank two anonymous referees for thoughtful and useful feedback. In addition, we thank our friends and colleagues who helped us think about this work: David Green, Stefan Hoderlein, Arthur Lewbel, and Chris Muris. Finally, we acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada through its Strategic Research Grants (Homelessness) and Standard Research Grants programs.

1. Introduction

In the last ten years, the volume of empirical work on consumption poverty in Canada has grown, complementing the traditional choice of income poverty (Pendakur, 2001; Milligan, 2008; Brzozowski & Crossley, 2011). There are three main reasons for using consumption rather than income as an indicator of material well-being. First, because consumption is a choice variable, it is more closely connected with the lifetime wealth constraint faced by the household than is contemporaneous income. Second, consumption determines welfare more directly than does income. Indeed, in standard models, consumption—not income—is the argument of the utility function. Finally, recent empirical work has shown that survey data may measure consumption better than it measures income, especially at the tails of the distributions (Brzozowski & Crossley, 2011).

The analysis of consumption poverty requires household-level consumption data. For most households, shelter is the dominant consumption good. For renters, expenditures on shelter may capture shelter consumption. However, for owners, we must somehow impute shelter consumption. As noted in Brzozowski and Crossley (2011), such *rent imputation* is in practice very difficult, but cannot be avoided in the measurement of consumption because more than half of households are owner-occupied.

We develop two methodological improvements to use in rent imputation. First, we account for possible differences in quality correlated with selection into owner-occupied versus rented accommodation. Second, we account for measurement error in the imputation when we estimate poverty rates. Empirically, these innovations raise the estimated mean and variance of household consumption, and have a big effect on estimated poverty rates.

This paper uses household expenditure microdata from the Surveys of Household Spending (SHS) 1997-2009. The SHS includes information on household spending in many categories of goods and services for about 15,000 households in each year. In addition, the SHS contains information on household size and composition and other demographic factors, so that we can adjust for these in our calculations. Our household-level consumption measure (which can be compared to a poverty threshold to determine a poverty rate) is the total of household expenditure in 10 categories of goods and services.

Shelter is one of our expenditure categories, and it is observable for renters (their gross rent paid). The standard approach is to regress rent on household characteristics for renters, and to construct a predicted value for owner-occupiers that can be used as their *imputed rent*. We implement a selection-corrected version of this approach, which is similar in spirit to the Heckman correction used in wage equations in applied labour economics. Our imputed rent thus corrects for the fact that the unobserved quality of shelter chosen by renters may differ from that chosen by owners. Use of this correction increases the imputed rent of owners by almost 30 per cent. Since non-renters (mostly owners) make up about two thirds of the population, and since shelter is about a third of the budget, this increases estimated average consumption by about 5 per cent. Use of this strategy to impute rental flows to non-renters decreases measured poverty quite substantially.

The standard approach in consumption poverty measurement is to construct household-level consumption data, imputing rent where necessary, and compare the household-level consumption to pre-defined poverty threshold. If household consumption is below the threshold, the household would be called *poor*. We argue that if the researcher is imputing rent, and the rent imputation has measurement error, then it is more appropriate to assign each household a consumption *distribution* rather than a consumption *level*. The household-specific consumption distribution takes account of the fact that the rent imputation step in fact generates an implied rent distribution for each household. Then, instead of asking "is the household poor?", the researcher asks, "what is the probability that the household is poor?".

Considering household-level poverty probabilities rather than household-level binary poverty indicators increases the amount of measured poverty. This is because there are more households whose imputed consumption (imputed rent plus non-shelter consumption) distribution is centered just above the poverty line than just below the poverty line. Consequently, there may be more seemingly non-poor households that have a big probability of being poor than seemingly poor households with a big probability of being non-poor. Overall, we find a slight increase in the level of poverty when one uses the imputed consumption distribution.

We find that absolute consumption poverty decreased quite dramatically over the period 1997-2009. Using our preferred specification, the national consumption poverty rate dropped from 12.7% to 7.7% between 1997 and 2009. By way of comparison, the official after-tax low-income rate declined from 15.0% to 9.6%.

The rate of child consumption poverty was higher than the overall poverty rate for the entire study period, but it, too, dropped significantly, from 18.5% to 9.9% over the decade. In comparison, the official after-tax low-income rate declined from 17.4% to 9.5% over the period.

For children, consumption poverty decline by 8.6 percentage points, but official child poverty declined by only 7.9 percentage points.

The lesson is that different households have low consumption from those that have low income. Since consumption connects more closely with material well-being than does income, by using available consumption data we can get a more accurate picture of material deprivation than we can using income data.

The structure of the paper is as follows: we briefly review the literature and explain the model in Section 2, present the methodology in Section 3, and review the data and results in Section 4. Section 5 concludes.

2. The Literature

The theoretical rationale for measuring poverty using consumption data rather than income data in the assessment of material well-being is straightforward given a life-cycle consumption model. Income has much more variation than does consumption, but some of this variation is related to transitory income shocks rather than to changes in permanent income. Households smooth consumption with knowledge of their income history and predictions of their future incomes. So, a household with low income but high consumption is more likely to have high lifetime wealth. Conversely, a household with high income but low consumption may have received a positive income shock — a lot of overtime, for example— that they do not expect to continue. Consumption may be smoothed over variation in needs as well, for example, over child-rearing years. Thus, the distribution of consumption may be a better indicator of distribution of utility, or of lifetime wealth, than the distribution of income. Blundell and Preston (1997) show exact conditions under which this is true, and they are rather demanding, but consumption has an intuitive advantage over income in that it is consumption, not income, that enters most economic models of utility.

Consumption (expenditure) data may have the additional advantage of better capturing utility derived from income coming from illegal and informal arrangements. Recent work using American data has shown that consumption measures better capture government transfers and social insurance, especially in-kind transfers like food stamps and health insurance (Meyer & Sullivan, 2011)[2]. Deaton (1997) and, more recently, Brzozowski and Crossley (2011) raise another argument: sample surveys may do a better job of measuring the consumption of households than

[2] Typical data sources in Canada for both income and consumption link to tax records, and so would have 'equal access' to government transfers.

measuring the income of households. Meyer and Sullivan (2011) show that income data has a lower response rate, and that income appears to be more heavily under-reported than consumption.

The recent literature on consumption poverty in Canada is somewhat ambiguous with respect to the evolution of poverty rates over time. The ambiguity is driven by conceptual differences — do we use relative or absolute poverty measures? — and by measurement choices — how do we adjust for price variation and demographic variation across households?

Relative poverty lines are based on the outcomes of society as a whole. For example, in the 1980s, Canada's Low Income Cut Off (LICO) was set at the income level associated with allocating 70 per cent of household expenditure to necessities. Over time, households got richer and devoted a smaller proportion of expenditure to necessities. Now, the LICO is set at the income level associated with allocating 58 per cent of expenditure to necessities, a higher income level than that associated with the 70 per cent cutoff. The usual rationale for relative poverty lines is that they account for societal consumption expectations, but they have been criticized as a measure of inequality rather than deprivation (e.g.: Sen, 1982; Pendakur 2001; Sarlo, 2008).

In contrast, *absolute* poverty lines are based on the cost of attaining a certain standard of living, or the cost of purchasing a certain basket of goods. They therefore guarantee a certain utility level regardless of the distribution of utilities in the surrounding society, although they are usually constructed with some cognizance of societal expectations. For example, the familiar US$2 per day global poverty line is not meaningful in Canada.

Using an absolute poverty measure, Pendakur (2001) found that poverty rates declined dramatically between 1969 and 1992, and rose between 1992 and 1998. Sarlo (2008) used a similar concept, but differing measurement choices, and found that consumption poverty declined over the period 1996 to 2005.

Using relative poverty measures, Crossley & Curtis (2006) present evidence that child poverty stayed constant over the 1990s (this is consistent with the findings of Pendakur 2001). Milligan's (2008) work on elderly poverty shows slight increases in consumption poverty from the late 1990s to 2003. Unlike both Crossley & Curtis and Milligan, but like Pendakur (2001) and Sarlo (2008) (and indeed like the LICOs since 1992 when they were last rebased), this paper uses an absolute poverty line.

We use an absolute poverty line that generates a poverty rate equal to Statistics Canada's 2002 after-tax low-income rate for all persons (11.6%). Given our rent imputation, this implies a consumption poverty line of $13,740 for single persons in 2002. This strategy allows easy comparisons with Canada's official poverty statistics, and consequently we will focus more on differences in estimated poverty across groups and over time than in the level of poverty itself. In any case, we also found that the choice of poverty line has a small impact on trends and between-group differences, so we can safely choose poverty lines that are comparable to official ones without sacrificing the generality of our results.

Adjustments for price differences across time and space and for demographic differences across households are vary substantially across researchers. While Milligan (2008) and Sarlo (2008) use the national Consumer Price Index (CPI), this paper uses a model-based price deflator based on provincial-level commodity prices. In particular, like Pendakur (2001) and Crossley & Curtis (2006), we estimate a demand system to create a household-specific price deflator. In comparison with Pendakur (2001), we use provincial- rather than regional-level price variation. The use of provincial- rather than national-level prices is a strength of this paper, because ignoring spatial variation in prices may induce missing variables bias in the poverty estimation. For example, Pendakur (2001) found that ignoring spatial variation led to an overstatement of poverty rates in Quebec, where the price of shelter has been historically low.

Given total nominal expenditures (including imputed rent) for each household, we use a price index depending on provincial-level prices to construct *real expenditures*. The deflator used is that corresponding to the Exact Affine Stone Index (EASI) demand system (Lewbel and Pendakur, 2009). It is numerically very close to the venerable *Stone Index* (Stone, 1954), defined as the weighted geometric mean of prices, where weights are equal to expenditure shares for a given household.

Equivalence scales are used to adjust total expenditure to make it comparable across family types — if two families have the same equivalent expenditure, they are considered to have the same utility. Dividing real expenditure by an equivalence scale yields *real equivalent expenditures*. We use three common equivalence scales to account for variation in household sizes and compositions: (1) the equivalence scales implicit in Statistics Canada's low-income cut-offs; (2) the square root of household size; and (3) the OECD (modified) equivalence scale.

We compare household-level real equivalent expenditures to our poverty threshold to create a poverty indicator for each household. Our strategy to accommodate measurement error in the rent imputation results in estimating a probability of poverty for each household. Finally, we weight our household-level poverty indicators by the number of members in the household, so that poverty rates are for people rather than for households. Our reported poverty rates are either population averages of the poverty indicator or of the poverty probability.

Our methodological innovations have to do with the imputation of a rental flow for accommodation. Brzozowski and Crossley (2011) emphasize that since shelter accounts for a large chunk of total consumption, and since rental flows are not observed for owner-occupied households, any accounting of consumption inequality or poverty must take the imputation of rent seriously. Our two methodological innovations---the selection-corrected imputation and the use of an imputed rent distribution rather than an imputed rent value---are quite novel. The use of an imputed rent distribution in an investigation of any aspect of household-level consumption is apparently entirely novel. To our knowledge, the selection-correction methods have been used once in rent imputation. In their application to estimating the cost-of-living index, Arevalo and Ruiz-Castillo (2006) use a selection-correction in the imputation of rent for owner-occupied households. They find that it makes a modest difference in estimated inflation rates. In our application to estimating average consumption and consumption poverty, we find that it makes a very substantial difference.

3. Methodology

3.1 Rent Imputation

The majority of households live in owned accommodation, which means that their housing expenditures are not necessarily equal to the consumption flow they receive from accommodation. For example, a family that has paid off their mortgage would record only utilities and a small amount of maintenance under housing expenditures, which add up to significantly less than the consumption flow they receive from their home. This is a serious problem when we measure consumption poverty: if we used recorded housing expenditures to construct our consumption measure, as does Sarlo (2008), a household that has paid off their mortgage could well be considered poor.

This problem can be solved by rent imputation. A common approach in the literature is to regress rent on a vector of explanatory variables, and then impute a value for homeowners. The rent imputation is usually done by OLS (Pendakur, 2001; Crossley & Curtis, 2006). However, there are significant differences between renters and owners, and even controlling for explanatory variables, the owner is likely to live in higher-quality housing.

In this context, assigning the OLS predicted rent to owner-occupied households would underestimate their housing consumption flow, and consequently lead to an overestimate of the poverty rate for this group. Further, the OLS predicted value is equal to the *mean* of the conditional distribution of rent given the explanatory variables. Some of this conditional distribution is higher than the predicted value, and some lower. In the context of poverty estimation, these two types of measurement error do not typically cancel. In the next subsections, we discuss how to deal with both of these issues.

3.1.1 A Model for Rent with a Selection Correction

To allow for quality differences between rented and owned housing, we use a Heckman correction (Heckman, 1979). Although the technique has been used extensively in other contexts (labour economics in particular) and at least once in the imputation of rental flows for owner-occupiers (Arevalo and Ruiz-Castillo, 2006) and occasionally by government statistical agencies for calculating income (Frick et al., 2008), it has not been used at all in investigations of consumption inequality or poverty. Our selection correction approach is tailored to a rent imputation context where the statistical object of interest is a function of the distribution of consumption, such as consumption inequality or consumption poverty (or even average consumption). In particular, we want to allow for the very great degree of heteroskedasticity observed in real-world data.

In the discussion that follows, we will use the term 'renter' to describe a household for which observed rent may be taken as an accurate reflection of their consumption flow from shelter. We exclude from this category all households that are not rental-tenure and rental-tenure households that have reduced rent or pay part of their rent through in-kind transfers (e.g., gardening). Non-renters by this definition comprise about two-thirds of households, of which 52 percentage points are owner-occupied households. We will use the terms "renters" and "owners" somewhat sloppily to refer to renters and non-renters, respectively.

Our problem is the familiar selection-correction model due to Heckman (1979), modified to allow for heteroskedasticity in both equations. Letting r_i denote nominal rent expenditures and p_{rent} denote the price of rental accommodation, we write

$$r_i / p_{rent} = \mathbf{v}_{1i}\beta + u_{1i} \quad if \;\; t_i = 1 \qquad (1.1)$$

$$\begin{aligned} t_i &= 1 \quad if \;\; \mathbf{v}_{1i}\Gamma_1 + \mathbf{v}_{2i}\Gamma_2 + u_{2i} \geq 0 \\ t_i &= 0 \quad otherwise \end{aligned} \qquad (1.2)$$

and

$$\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_1^2(\mathbf{v}_{1i}) & \rho(\mathbf{v}_{1i})\sigma_1(\mathbf{v}_{1i})\sigma_2(\mathbf{v}_{1i}) \\ \rho(\mathbf{v}_{1i})\sigma_1(\mathbf{v}_{1i})\sigma_2(\mathbf{v}_{1i}) & \sigma_2^2(\mathbf{v}_{1i}) \end{pmatrix}\right) \tag{1.3}$$

and

$$\begin{aligned} \sigma_1(\mathbf{v}_{1i}) &= \mathbf{v}_{1i}\Delta_1, \\ \sigma_2(\mathbf{v}_{1i}) &= \exp(\mathbf{v}_{1i}\Delta_2), \\ \rho(\mathbf{v}_{1i}) &= \rho_0 + \mathbf{v}_{1i}\mathbf{C}\rho_{\mathbf{v}}. \end{aligned}$$

Here, equation (1.1) determines the real rent paid (or, equivalently, shelter quantity purchased) , r_i / p_{rent} , and equation (1.2) is the tenure equation determining who is a renter ($t_i = 1$) and who is not a renter ($t_i = 0$). Note that we observe r_i / p_{rent} only for renters. The row-vector $\mathbf{v}_{1i}$ is a set of explanatory variables determining rent for household *i* if household *i* is a rental-tenure household. The row-vector $\mathbf{v}_{2i}$ is an additional set of explanatory variables that determines the tenure of a household, and which does not affect rents for renters.

An important feature of this model is that both the discrete choice and the continuous part are heteroskedastic in the explanatory variables $\mathbf{v}_{1i}$. Both variances in the model are linear indices in the explanatory variables, with the discrete-choice variance as an exponential in that linear index (this is convenient for Stata's heteroskedastic probit routine). The correlation coefficient, $\rho(\mathbf{v}_{1i}) = \rho_0 + \mathbf{v}_{1i}\mathbf{C}\rho_{\mathbf{v}}$, is a linear index in $\mathbf{v}_{1i}\mathbf{C}$, a row-vector of the first *c* principal components of $\mathbf{v}_{1i}$ (the *c* columns of **C** give the weights on each variable in each principal component). Here, we reduce dimensionality of the correlation coefficient by making it linear in the first 3 principal components of $\mathbf{v}_{1i}$, rather than linear in $\mathbf{v}_{1i}$ itself. For tractability, we impose that the two disturbance terms are distributed jointly normal and independently across observations.[3]

3.1.2 Estimation of Selection-Corrected Model

[3] Joint normality may be a bit hard to swallow, since a normal variate can be negative for any finite mean (implying nonzero densities for negative imputed rents). Given the independence assumption and the existence of an infinite support continuous variable in $\mathbf{v}_{2i}$, one could specify the model with unknown distributions for u_{1i} and u_{2i} and still identify all the parameters of the model. Estimation could be by semiparametric sieve maximum likelihood. However, as discussed in the description of the data, we do not have such instruments in $\mathbf{v}_{2i}$, and so cannot invoke the "identification at infinity" arguments usually used to guarantee identification in semiparametric selection models.

This model may in principle be estimated by full-information maximum likelihood, but, in practice, the heteroskedastic covariance makes this computationally quite burdensome. We instead employ a 2-step limited-information maximum likelihood (LIML) approach as follows.

The **first step** is to estimate a heteroskedastic probit of equation (1.2) and compute the estimated inverse mills ratio $\hat{\lambda}_i$ for renters and for owners using predicted values as follows:

$$\hat{\lambda}_i = \frac{\phi(\hat{z}_i)}{\Phi(\hat{z}_i)} \;\; if \;\; t_i = 1, \qquad \hat{\lambda}_i = \frac{\phi(-\hat{z}_i)}{\Phi(-\hat{z}_i)} \;\; if \;\; t_i = 0, \qquad (1.4)$$
$$\hat{z}_i = \frac{\mathbf{v}_{1i}\hat{\Gamma}_1 + \mathbf{v}_{2i}\hat{\Gamma}_2}{\hat{\sigma}_2(\mathbf{v}_{1i})}$$

where ϕ and Φ refer to the probability and cumulative density functions, respectively, of the standard normal distribution. We note that the argument of ϕ and Φ for owners is the negative of that for renters.

The selection correction in the rent equation is determined as follows. The conditional expectation of u_2 given t_i is given by (see, e.g., Greene 2011, pg 836):

$$E\left[u_{2i}\right]|_{t_i=1} = \sigma_2\left(\mathbf{v}_{1i}\right)\lambda_i \qquad and \qquad E\left[u_{2i}\right]|_{t_i=0} = -\sigma_2\left(\mathbf{v}_{1i}\right)\lambda_i \qquad (1.5)$$

The conditional distribution of u_1 given u_2 is (see, e.g., Greene 2011, pg 1038):

$$u_{1i} \mid u_{2i} \sim N\left(\frac{\rho(\mathbf{v}_{1i})\sigma_1(\mathbf{v}_{1i})}{\sigma_2(\mathbf{v}_{1i})}u_{2i}, \sigma_1^2(\mathbf{v}_{1i})\left(1-\rho(\mathbf{v}_{1i})^2\right)\right). \qquad (1.6)$$

Since the mean of u_1 is linear in u_2, the conditional expectation of u_1 given $t_i=1$ is computed by substituting the conditional expectation of u_2 given $t_i=1$ into the distribution of u_1 given u_2, using the estimated inverse mills ratio $\hat{\lambda}_i$:

$$E\left[u_{1i}\right]|_{t_i=1} = \rho(\mathbf{v}_{1i})\sigma_1(\mathbf{v}_{1i})\lambda_i, \qquad and \qquad E\left[u_{1i}\right]|_{t_i=0} = -\rho(\mathbf{v}_{1i})\sigma_1(\mathbf{v}_{1i})\lambda_i \qquad (1.7)$$

Defining $\tilde{u}_{1i} = u_{1i} - E\left[u_{1i}\right]|_{t_i=1}$, this yields the following selection-corrected model for real rent:

$$r_i / p_{rent} = \mathbf{v}_{1i}\beta + \rho(\mathbf{v}_{1i})\sigma_1(\mathbf{v}_{1i})\lambda_i + \tilde{u}_{1i} \;\; if \;\; t_i = 1 \qquad (1.8)$$

In the canonical, homoskedastic LIML Heckman model, the coefficient on λ_i corresponds to a fixed number equal to $\rho(\mathbf{v}_{1i})\sigma_1(\mathbf{v}_{1i}) = \rho\sigma_1$, so one need add only the estimated inverse-mills ratio, $\hat{\lambda}_i$, as a regressor to the rent equation. In our case, the heteroskedasticity in both equations makes this more complex: the coefficient on λ_i corresponds to $\rho(\mathbf{v}_{1i})\sigma_1(\mathbf{v}_{1i})$.

The **second step** of our LIML procedure is a two-stage FGLS regression of equation (1.8). For the first stage, given the estimated inverse mills ratio, $\hat{\lambda}_i$, computed from the heteroskedastic probit regression, we regress real rent, r_i / p_{rent}, on $\mathbf{v}_{1i}$, $\hat{\lambda}_i$, $\mathbf{v}_{1i}\hat{\lambda}_i$, and the squared terms given by the elements of $\left(\mathbf{v}_{1i}\mathbf{C}\mathbf{v}_{1i}'\right)\cdot\hat{\lambda}_i$. The last 3 terms make this regression a consistent but inefficient estimator of equation (1.8). It is inefficient because it does not impose the restriction that the coefficient on $\hat{\lambda}_i$ is the product of two linear indices ($\sigma_1(\mathbf{v}_{1i})$ and $\rho(\mathbf{v}_{1i})$) *and* because it ignores the heteroskedasticity in u_{1i}. We then compute the residuals e_i, and regress their absolute values, $|e_i|$, on regressors $\mathbf{v}_{1i}$, generating predicted absolute residuals, $\hat{e}_i$.[4] From this regression, we compute predicted values $\hat{\sigma}_{1i} = \hat{\sigma}_1(\mathbf{v}_{1i})$ as follows:

$$\hat{\sigma}_{1i} = \frac{\hat{e}_i}{\sqrt{2/\pi}} \tag{1.9}$$

For the second stage, using the weights $\hat{\sigma}_{1i}^{-1}$ (the reciprocal of the estimated root-variance of u_{1i}), we run a weighted least squares regression of real rent, r_i / p_{rent}, on $\mathbf{v}_{1i}$, $\hat{\sigma}_{1i}\hat{\lambda}_i$ and $\mathbf{v}_{1i}\mathbf{C}\hat{\sigma}_{1i}\hat{\lambda}_i$. The coefficient on $\hat{\sigma}_1(\mathbf{v}_{1i})\hat{\lambda}_i$ is an estimate of ρ_0, and the coefficients on $\mathbf{v}_{1i}\mathbf{C}\hat{\sigma}_{1i}\hat{\lambda}_i$ are estimates of ρ_v. Note that the second stage uses the predicted variance of u_{1i} (that is, $\hat{\sigma}_{1i}$) in two ways: it provides the weights for the second stage, and it enters the second-stage regressor list.

3.1.3 Selection-Corrected Imputation of the Conditional Mean of Rent

Standard methods take imputed rent to be the conditional mean of rent given the estimated rent model. The estimated conditional expectation of nominal rent expenditures, $\hat{\mu}_i$, uses estimated values of all model parameters, and is given by:

[4] A random variable, y, equal to the absolute value of a normal variable, x, is said to follow a "half-normal" distribution. If x is mean-zero with variance σ^2, the half-normal random variable y has an expected value equal to $\sigma\sqrt{2/\pi}$. Thus, the predicted value from a regression of $|e_i|$ on regressors provides an estimate of $\sigma\sqrt{2/\pi}$, which may be divided by $\sqrt{2/\pi}$ and get an estimate of σ.

$$\hat{\mu}_i = E\left[r_i\right]|_{\mathbf{v}_{1i},\hat{\lambda}_i} = p_{rent}\left(\mathbf{v}_{1i}\hat{\beta} + \left(\hat{\rho}_0 + \hat{\rho}_v{}'\mathbf{C}\mathbf{v}_{1i}\right)\hat{\sigma}_{1i}\hat{\lambda}_i\right) \quad if \;\; t_i = 1$$
$$= p_{rent}\left(\mathbf{v}_{1i}\hat{\beta} - \left(\hat{\rho}_0 + \hat{\rho}_v{}'\mathbf{C}\mathbf{v}_{1i}\right)\hat{\sigma}_{1i}\hat{\lambda}_i\right) \quad if \;\; t_i = 0 \qquad (1.10)$$

We note that we bring the rent price over to the right hand side, so that we are imputing nominal rental flows rather than real ones. Since this imputation kills off variance due to u_1, imputing only for owners would not be evenhanded. Thus, for standard poverty measures we compute this imputation for all households, not just renters.

We then replace observed shelter expenditures with imputed rent in our computation of total nominal expenditure for all households. Assume that other categories of expenditure are measured without error, and let $\tilde{x}_i$ denote the total expenditure of household *i* on expenditures other than shelter. Imputed nominal total consumption, $\hat{x}_i$, is given by

$$\hat{x}_i = \tilde{x}_i + \hat{\mu}_i \qquad (1.11)$$

If a household has total nominal expenditures less than the nominal poverty threshold, we call it *poor*, otherwise not. We refer to this classification as *binary* poverty, and the associated poverty rate is equal to the average of this indicator variable across the population.

Binary poverty suffers from a methodological drawback. We know that our rent imputation is imprecise (because the conditional variance of real rent $\sigma_1(\mathbf{v}_{1i})$ is nonzero), so that households whose conditional mean rent puts them just above the poverty line might actually be poor, and vice versa for households just below the poverty line.

3.1.4 Selection-Corrected Imputation of the Conditional Distribution of Rent

For renters, rent is observed exactly, so in fact one need not impute at all. Indeed the only reason that researchers have imputed rent for renters is to put them 'on par' with owners, subjecting them to the same sort of measurement error that must be swallowed for owners. In this section, we propose a different way to get parity between owners and renters. Instead of killing off a similar amount of variance from each, we use observed rents (which include variance due to u_1) and add an appropriate amount of variance to the imputed rent of owners.

For owners, we create a measure of the *probability* of poverty for owners, based on the estimate, $\hat{\sigma}_{1i}$, of the variance $\sigma_1(\mathbf{v}_{1i})$ in the rent equation. Each household is assigned a probability of being poor, equal to the probability that the sum of their imputed rent and nonhousing expenditures is less than the poverty threshold. We refer to this as the *poverty probability* of each household. Then, our poverty rate is equal to the average of this poverty probability across the population.

We implement our estimation of the poverty probability for owners as follows. The predicted rent, $\hat{r}_i$, is now taken to be a distribution, rather than a fixed value:

$$\hat{r}_i \mid_{\mathbf{v}_{1i},\hat{\lambda}_i} \sim N\left(\hat{\mu}_i, \tilde{\sigma}_1^2(\mathbf{v}_{1i})\left(p_{rent}\right)^2\right) \quad if \;\; t_i = 0 \tag{1.12}$$

where

$$\tilde{\sigma}_1^2 = \frac{\hat{\rho}(\mathbf{v}_{1i})^2 \hat{\sigma}_{1i}^2}{\hat{\sigma}_2^2(\mathbf{v}_{1i})} \hat{\sigma}_{1i}^2 \left(1 + \hat{z}_i \hat{\lambda}_i - \hat{\lambda}_i^2\right) + \hat{\sigma}_{1i}^2 \left(1 - \hat{\rho}(\mathbf{v}_{1i})^2\right). \tag{1.13}$$

Here, $\hat{\mu}_i$ is the estimated expected value of rent expenditures defined in (1.9). We define $\tilde{\sigma}_1^2\left(\mathbf{v}_{1i}\right)$ as the estimate of the conditional variance of u_{1i} given that u_{2i} is in the range that makes household i choose $t_i=0$.[5] The conditional variance $\tilde{\sigma}_1^2\left(\mathbf{v}_{1i}\right)$ depends on the estimated inverse mills ratio, $\hat{\lambda}_i$, the probit predicted value, $\hat{z}_i$, the estimated rent variance, $\tilde{\sigma}_{1i}$, the estimated probit variance, $\hat{\sigma}_2\left(\mathbf{v}_{1i}\right)$, and the estimated correlation, $\hat{\rho}\left(\mathbf{v}_{1i}\right)$. Here, the conditional variance $\tilde{\sigma}_1^2\left(\mathbf{v}_{1i}\right)$ is multiplied by the squared rent price because r_i is nominal rent, but the estimating equation used real rent, r_i / p_{rent}, as the regressand.

The imputed nominal total consumption flow of each household is thus $\hat{x}_i = \tilde{x}_i + \hat{r}_i$, which is a normally distributed random variable:

$$\hat{x}_i \mid_{\tilde{x}_i,\mathbf{v}_{1i},\hat{\lambda}_i} \sim N\left(\tilde{x}_i + \hat{\mu}_i, \tilde{\sigma}_1^2(\mathbf{v}_{1i})\left(p_{rent}\right)^2\right) \quad if \;\; t_i = 0 \tag{1.14}$$

The probability of poverty for household i is then equal to the cumulative density of $\hat{x}_i$ below the nominal poverty threshold. If nominal poverty thresholds are available for every price regime and every household size/type, then this probability may be computed easily via the cdf of the normal distribution.

---

[5] The variance of u_{1i} given that u_{2i} is in the range that makes household i choose $t_i=0$ is a bit tedious to work out. It is equal to the conditional variance of u_{1i} given in equation (1.6), $\sigma_1^2(\mathbf{v}_{1i})\left(1-\rho(\mathbf{v}_{1i})^2\right)$, plus the variance of the truncated normal distribution for u_{2i} given $t_i=0$, $\sigma_1^2(\mathbf{v}_{1i})\left(1+\hat{z}_i\hat{\lambda}_i-\hat{\lambda}_i^2\right)$, multiplied by the square of the factor multiplying u_{2i} in the conditional distribution of u_{1i} given u_{2i}, $\rho(\mathbf{v}_{1i})\sigma_1(\mathbf{v}_{1i})/\sigma_2(\mathbf{v}_{1i})$.

In our analysis, rather than assuming nominal poverty thresholds for every price regime, we use price indices to convert the nominal expenditures of households to real expenditures comparable to those of a household facing a reference price vector. However, this means that the variation in imputed rent enters the numerator and denominator of the real expenditure measure. There is no convenient closed-form solution, so we estimate poverty probabilities via simulation.

The next several sections describe how we deal with price and demographic variation, and show how to simulate poverty probabilities.

3.2 The Equivalence Scale and Poverty Threshold

We use the 2002 After-Tax Low Income Cutoffs for residents of cities with populations of 500,000 or more to compute our equivalence scale in the reference price regime, that facing residents of Ontario in 2002 (more on adjusting for price variation in the next section). As shown below in Table 1, these equivalence scales do not exhibit some standard properties of equivalence scales, such as decreasing marginal cost of people in the household. They are, however, standard for poverty measurement in Canada. In our empirical work, we also consider two alternative base-period equivalence scales: the square-root of household size, and the OECD (adjusted series) equivalence scales.

We choose the poverty threshold so that our measured binary poverty rate with the selection-uncorrected conditional mean rent imputation (*straight binary poverty*) matches the officially reported 2002 After-Tax Low-Income rate for all persons of 11.6%. In our consumption poverty exercise, the poverty threshold that achieves this is $13,740, which corresponds to the threshold for a single adult in 2002. That is, if we use a poverty threshold of $13,740 we get a *straight binary poverty* rate of 11.6% in 2002. We think of this as benchmarking the *level* of the poverty threshold to the officially published poverty rates---if we use a higher threshold, we get higher estimated poverty rates than the officially reported measures, and if we use a lower one, we get lower estimated poverty rates.

By way of comparison, the after-tax low-income cutoff for a single person in a big city in 2002 was $16,102. The difference between our poverty threshold and the official LICO comes from two sources: consumption goods that are not included in our basket, and savings. Table 1 gives the After-Tax Low-Income Cutoffs, their implied equivalence scale, and our consumption poverty thresholds for households of different sizes facing the base price regime (that of residents of urban Ontario in 2002).

Table 1: After-Tax Low-Income Cutoffs, Equivalence Scales and Poverty Thresholds

Household Size	1	2	3	4	5	6	7+
After Tax LICO	16102	19598	24404	30445	34668	38448	42227
Equivalence Scale	1	1.22	1.52	1.89	2.15	2.39	2.62
Poverty Threshold	13740	16763	20885	25969	29541	32839	35999

We may compare a nominal consumption measure to the poverty thresholds above for residents of urban Ontario in 2002 (regardless of whether or not rent is imputed). For people facing other price regimes, we need to adjust their nominal consumption to create real consumption levels that are welfare-equivalent to those facing the reference price vector, and we may compare those real consumption levels to the consumption poverty thresholds in Table 1.

3.3 Dealing with Price Variation: Consumer Demand Estimation

A consumer demand system relates the share of expenditure commanded by a commodity $j=1,\ldots,J$, w^j, to the budget constraint faced by a household, a T-vector of observable characteristics, **z**, and a J-vector of unobservable characteristics, **e**, of that household. Let $\mathbf{z}=\mathbf{0}$ for a reference household type, defined as one comprised of a single childless adult. Let $\mathbf{1}_J\,'\mathbf{e}=0$. The budget constraint is characterised by a J-vector of prices, **p**, and a level of total expenditure, x. Let $\mathbf{p}=\mathbf{1}$ in a reference price regime, defined as that faced by residents of urban Ontario in 2002.

Let $x=C(\mathbf{p},u,\mathbf{z},\mathbf{e})$ give the cost to attain a utility level u for a household facing prices **p** with a observed characteristics **z** and unobserved characteristics **e**. Given the cost function, we can compute the amount of money needed to hold utility constant across any price change. The question is how to estimate $C(\mathbf{p},u,\mathbf{z},\mathbf{e})$.

A large literature considers the estimation of the cost function (for a survey, see Slottje 2009), and some recent progress has made the estimation of cost functions with unobserved preference heterogeneity (**e** in our case) quite easy to implement. In particular, Lewbel and Pendakur (2009) specify the following Exact Affine Stone Index (EASI) cost function:

$$\ln x = \ln C(\mathbf{p},u,\mathbf{z},\mathbf{e}) = u + \mathbf{d}'\mathbf{z} + \ln \mathbf{p}'\left[\mathbf{b}_0 + \sum_{r=1}^{5}\mathbf{b}_r u^r + \mathbf{D}\mathbf{z} + \frac{1}{2}\mathbf{A}\ln\mathbf{p} + \mathbf{e}\right] \qquad (1.15)$$

Here, prices affect cost through level effects ($\mathbf{b_0}$), interactions with utility ($\mathbf{b_r}$) and observed demographics ($\mathbf{D}$), and a quadratic term ($\mathbf{A}$). Observed demographics are additionally cost shifters through the parameters $\mathbf{d}$. Here, $\mathbf{b_0}$ and $\mathbf{b_r}$ are J-vectors, $\mathbf{D}$ is a J x T matrix and $\mathbf{A}$ is a J x J matrix. Homogeneity of cost (aka: no money illusion) implies that scaling $\mathbf{p}$ scales C by the same amount. This implies $\iota'\mathbf{b}_0 = 1, \iota'\mathbf{b}_r = 0\, r \neq 0, \iota'\mathbf{A} = \iota'\mathbf{D} = \mathbf{0}$, and $\iota'\mathbf{e} = 0$. Slutsky symmetry is satisfied if and only if $\mathbf{A}$ is symmetric. If these constraints hold, then a unique cost-of-living index may be constructed.

The parameters to be estimated are $\mathbf{b_r}$, $\mathbf{D}$ and $\mathbf{A}$. The vector $\mathbf{d}$ is set a priori rather than estimated---it is the equivalence scale given in Table 1.

Let $\mathbf{w}$ be the J-vector of expenditure shares, w^j. We may apply Shepherd's lemma to obtain the EASI expenditure share vector as:

$$\mathbf{w} = \mathbf{b}_0 + \sum_{r=1}^{5} \mathbf{b}_r u^r + \mathbf{Dz} + \mathbf{A}\ln\mathbf{p} + \mathbf{e} \qquad (1.16)$$

which is very similar to the term in square brackets in (1.15). This similarity allows us to substitute $ln\mathbf{p}'\mathbf{w}$ into the cost function (1.15) as:

$$\ln x = \ln C(\mathbf{p}, u, \mathbf{z}, \mathbf{e}) = u + \mathbf{d}'\mathbf{z} + \ln\mathbf{p}'\mathbf{w} - \ln\mathbf{p}'\frac{1}{2}\mathbf{A}\ln\mathbf{p} \qquad (1.17)$$

Since cost, C, equals expenditure, x, we may write utility, u, as a function of observables ($\mathbf{p}$,x,$\mathbf{z}$,$\mathbf{w}$) and parameters ($\mathbf{d}$,$\mathbf{A}$) as follows:

$$\ln y = u = \ln x - \mathbf{d}'\mathbf{z} - \ln\mathbf{p}'\mathbf{w} + \frac{1}{2}\ln\mathbf{p}'\mathbf{A}\ln\mathbf{p} \qquad (1.18)$$

Here, lny equals utility, and is a function just of parameters and observable variables. We call y as "real equivalent expenditure", because it is a money metric for utility which takes prices and demographic characteristics into account. That is, if two households have the same value of y (and therefore of lny), then they have the same utility level.

Real equivalent expenditures y has three components which relate it log-linearly to nominal expenditures x. The first component is $\mathbf{d'z}$, which is the log of the equivalence scale in the reference price regime. Pendakur (1999) shows how such an object can be estimated semi-parametrically, but in this context it is important to maintain comparability with other research. So, we assume the value of $\mathbf{d}$, rather than estimate it. Table 1 gives our values for $\mathbf{d'z}$, which are imposed in the estimation. In our empirical work, we also use two other equivalence scales: (1) the OECD-modified equivalence scales; and (2) the square-root of household size.

The remaining factors in *y* deal with price variation. The second component is $ln\mathbf{p}'\mathbf{w}$, which is the log of the Stone Index for the household, $\prod_{j=1}^{J}\left(\frac{p_i^j}{\bar{p}^j}\right)^{w_i^j}$, equal to the expenditure share-weighted geometric mean of price differences between prices and prices in the reference price regime. Stone (1954) proposed this as a 'natural' price index, and it is equal to a first-order approximation of the cost of living index.

The third component is $ln\mathbf{p}'\mathbf{A}ln\mathbf{p}/2$, which accounts for second-terms in the approximation. Given the EASI cost function above, all higher-order terms vanish. Lewbel and Pendakur (2009) find that this third component is quite small, relative to the Stone Index component, and we find that in our setting as well.

In the poverty measurement that we do below, we compare real equivalent expenditures *y* to the poverty threshold to evaluate whether or not a household is poor. Real equivalent expenditure *y* can be computed if we have information on prices, **p**, household demographics, **z**, expenditure shares, **w**, the vector **d** and the matrix **A**. Since **d** is assumed and **p**, **z** and **w** are observed for each household, all we need to estimate is the matrix **A**.

Substituting *y* for *u* in equation (1.16), the equation we estimate is

$$\mathbf{w} = \mathbf{b}_0 + \sum_{r=0}^{5} \mathbf{b}_r \left(\ln y\right)^r + \mathbf{Dz} + \mathbf{A}\ln\mathbf{p} + \mathbf{e} \qquad (1.19)$$

Given *y*, this is a linear regression on observables. That is, given *y*, one regresses each expenditure share w^j on a constant, a set of powers of *lny* (up to the fifth power in this case), the demographic variables **z** and the log-prices ln**p**. Then, the estimate of the matrix **A** is comprised of the estimated coefficients on ln**p** in each equation (for a low-tech description of the EASI model, see Pendakur 2009). The system is estimated for J-1 equations, and coefficients for the J^{th} equation are recovered from the symmetry and adding up restrictions.

There are four small problems with the equation-by-equation linear regression above. First, the matrix **A** should be symmetric. This may be addressed via linear restrictions in a linear Seemingly Unrelated Regression (SUR) setting. Second, log real expenditure *y* introduces endogeneity through the **p'w** term, so we use Three Stage Least Squares (3SLS) instead of SUR. The use of 3SLS allows us to correct for endogeneity via instruments based on ln**p,** *x*, and *z*.

Third, log real equivalent expenditure, *lny*, is linear in $ln\mathbf{p}'\mathbf{A}'ln\mathbf{p}$. Since the model includes, for example, coefficients on the square of *lny*, the expenditure share equations are (slightly) nonlinear in parameters. This may be solved via iterating the 3SLS estimation. At each iteration, we update the formula for *lny* with new estimated parameter values, and iterate until convergence. Iteration allows us to use linear methods to solve the full nonlinear problem that accounts for nonlinearity coming through *y*, as in Blundell and Robin (1999)[6] (see also Lewbel and Pendakur 2009 for details and Stata code).

The fourth issue is that since u_2, the unobserved heterogeneity parameter in the rental-selection equation, is correlated with u_1, the unobserved heterogeneity parameter in rent expenditures, we need to allow for correlations between u_2 and $\mathbf{e}$, the unobserved preference heterogeneity in the expenditure shares. (Demand estimation, like the rent equation in the rent imputation, is done on a sample of rental-tenure households.) The easiest way to do this is to include $\widehat{\sigma}_{1i}\widehat{\lambda}_i$ and the vector $\mathbf{Cv}_{1i}\widehat{\sigma}_{1i}\widehat{\lambda}_i$ as covariates. Since real rent, r_i/p_{prent}, is linear in these objects, the rent budget share, r_i/x_i, is linear in these objects multiplied by $\frac{p_{rent}}{x_i}$. Thus, we include the selection corrections $\frac{p_{rent}}{x_i}\widehat{\sigma}_{1i}\widehat{\lambda}_i$ and $\frac{p_{rent}}{x_i}\mathbf{Cv}_{1i}\widehat{\sigma}_{1i}\widehat{\lambda}_i$ as elements of $\mathbf{z}$. Empirically, although these selection terms strongly affect budget shares, their inclusion hardly affects the estimated value of the matrix $\mathbf{A}$.

4.4 The Estimation of Poverty Probabilities and Poverty Rates

Let $\bar{y}$ denote \$13,740, our real poverty threshold for a single adult living in urban Ontario in 2002. Imputed total consumption, $\hat{x}_i$, is nominal. To convert it to real equivalent consumption, $\hat{y}_i$, which can be compared to our real poverty threshold, we use the EASI expression for real equivalent expenditure (1.18) and exponentiate:

$$\hat{y}_i = \hat{x}_i \frac{\sqrt{\exp\left(\ln \mathbf{p}_i{}'\widehat{\mathbf{A}}\ln \mathbf{p}_i\right)}}{\exp\left(\mathbf{d}'\mathbf{z}_i\right)\exp\left(\ln \mathbf{p}_i{}'\mathbf{w}_i\right)} \qquad (1.20)$$

where $\widehat{\mathbf{A}}$ is the estimated value of $\mathbf{A}$ from the demand estimation and $\mathbf{p}_i$ is the price vector faced by household *i*. Our imputed rent measure $\hat{r}_i$ shows up in three places: it is a component of the leading term $\hat{x}_i$; it enters the denominator of each expenditure share in the expenditure share

[6] One might imagine that a nonlinear model with many equations could be difficult to estimate or subject to multiple solutions and/or solution instability. However, because this model is nonlinear only via *lny*, which depends only on parameters that are also show up outside *lny*, these issues do not arise. Blundell and Robin (1999) estimate a 22 equation model and find no evidence of multiple solutions (indeed, their identification theorem requires such uniqueness).

vector through $\hat{x}_i$; and it is the numerator of the expenditure share corresponding to shelter. For binary poverty measures imputed rent, $\hat{r}_i$, is a single value for each household. Given this value, if $\hat{y}_i < \bar{y}$, we call the household poor, and the estimated poverty rate is given by

$$\hat{P} = \frac{1}{N}\sum_{i=1}^{N} I(\hat{y}_i < \bar{y}) \tag{1.21}$$

In the case where $\hat{r}_i$ is a distribution for each household owner-occupier household (with $t_i=0$), then we instead consider the *poverty probability*, p_i, for each household. For renters, the household-level poverty probability distribution is degenerate, with a probability of 0 or 1 for each household. This is because for these households, rent is observed exactly and without measurement error, so the poverty probability is given by $p_i = I\left[\hat{y}_i < \bar{y}\right]$.

In contrast, for owner occupied households, rent is imputed with error, so for them the probability of poverty is non-degenerate. For these households, we estimate the probability that $\hat{y}_i < \bar{y}$ by simulating under the model as follows. For each observation, we take $b=1,\ldots,B$ draws u_{1i}^b and u_{2i}^b from the bivariate normal distribution (1.3). Then, we focus on the subset of draws of u_{2i}^b is such that the household would have $t_i=0$. For this subset of the B draws, we compute $\hat{x}_i^b$ and the budget share vector $\hat{\mathbf{w}}_i^b$ and, given these, real equivalent expenditures, $\hat{y}_i^b$. Then, the estimated poverty probability, p_i, for an owner-occupier household is average of the poverty indicator over the B draws of of u_{1i}^b and u_{2i}^b , given a draw of u_{2i} low enough to classify the household as an owner. Thus, we have

$$\hat{P} = \frac{1}{N}\sum_{i=1}^{N} p_i \tag{1.22}$$

where

$$
\begin{aligned}
p_i &= I\left[\hat{y}_i < \bar{y}\right] && \text{if } t_i = 1 \\
&= \frac{1}{\sum_{b=1}^{B} \hat{t}_{0i}^b} \sum_{b=1}^{B} I\left[\hat{y}_i^b < \bar{y}\right]\hat{t}_{0i}^b && \text{if } t_i = 0
\end{aligned}
$$

where

$$\hat{t}_{0i}^b = I\left[\mathbf{v}_{1i}\hat{\Gamma}_1 + \mathbf{v}_{2i}\hat{\Gamma}_2 + u_{2i}^b < 0\right].$$

4. The Data

4.1 Price Data

Price data are combined from three publicly available sources: the 2002-basket and 2006-basket commodity-level Consumer Price Indices 1997-2009 (for each province: commodity prices by year); and Statistics Canada's intercity price indices (for each commodity: prices by city and year). The commodity-level consumer price indices are normalized to 100 for each good and province in a base year, which means that they do not measure level differences in prices across provinces. More information is needed to create a price index that measures differences in price across time and province. We use the intercity indices, which consist of comparable prices (which average 100 in each year) for the largest city in each province over the period 2002-2009 (the only years these indices are available). We assume that each city is representative of its province, and use these indices to link across provinces in a given year. Hereafter, we refer to these as "inter-provincial commodity price links".

Prices obviously vary between urban and nonurban residents. Commodity prices are generally not available at this level of stratification (province/year/urban-nonurban). However, the 2006 Census collects data on rents, housing characteristics and city residence status. From this, we are able to compute for 2006 an inter-provincial price index for rent that is stratified by city resident versus non-city resident.[7] We use the assumption that the price differential between urban rents and nonurban rents is invariant over time for each province. Thus, we are able to create (for the price of rent only) a set of prices stratified at the level of province/year/urban-nonurban.

Newfoundland has no large cities, so housing price data is the same for the entire province, and PEI is excluded because data required in the rent imputation is masked. Combining the Statistics Canada price data and the Census housing data, there are 17 price regions in each year (8 provinces by urban/nonurban plus Newfoundland). For all commodities except rent, these prices are the same for urban and nonurban within a province/year; for rent, they differ.

[7] Unfortunately, for the SHS in 1997-1999 the city size flag is set at 30,000+, and for 2000-2009 the city size flag is set at 100,000+. This inconsistency means that in poverty estimates which use city-resident stratified rental prices, the series are not comparable over these two periods. However, poverty estimates implemented without the use of this stratification show the same patterns as those using the stratification.

Given the values of inter-provincial commodity price links in any one link year, one could take those as true measures of commodity prices in their respective provinces, and compute the commodity prices in other time periods via indexing by the consumer price index for that commodity and province. However, Statistics Canada uses more underlying data to compute the consumer price indices for each province and commodity than it does to compute the inter-provincial commodity price links. Thus, it is desirable to 'pool' the information across the different years of the inter-provincial commodity price links, so as to reduce the variance in the inter-provincial links.

We use a simple strategy to pool this information. Let $\bar{l}_{kt}$ be the inter-provincial commodity price link for a certain commodity in year *t* and province *k*. These average 100 across all provinces *k* in any given year. Let p_{kt} be the provincial consumer price index for a certain commodity in year *t* and province *k*. Define , $l_{kt} = \bar{l}_{kt} / \bar{l}_{0t}$ where $\bar{l}_{0t}$ is the price index in the base region, urban Ontario. We assume l_{kt} is measured with proportional error and related to the true value, L_{kt}, by $\ln l_{kt} = \ln L_{kt} + u_{kt}$, where u_{kt} is a mean-zero iid error term. The true value of our index of prices using a particular link year *s*, P_{kst}, is defined in terms of the true values of the links as

$$P_{kst} = \frac{\frac{p_{kt}}{p_{ks}} L_{ks}}{\frac{p_{0,2002}}{p_{0,s}} L_{0,s}} \tag{1.23}$$

Taking logs and substituting l_{kt} for L_{kt}, we get

$$\ln P_{kst} = \ln p_{kt} - \ln p_{ks} + \ln l_{ks} + u_{ks} - \ln p_{0,2002} + \ln p_{0,s} - \ln l_{0,s} - u_{0,s} \tag{1.24}$$

This should be invariant to the choice of link year, but since l_{kt} is a noisy measure of L_{kt}, our empirical estimate of those prices will be different across different choices of link years. We solve this problem by averaging over all possible choices of link years *s*. The prices we use in our empirical work are geometric means of P_{kst} over all seven possible choices (2002 to 2009). Since u_{kt} is additive noise relating $\ln l_{kt}$ and $\ln L_{kt}$, this averages out (and therefore reduces the variance of) the noise in the inter-provincial commodity price links.

4.2 Expenditure Data

Expenditure data are drawn from the annual Surveys of Household Spending (SHS) 1997-2009. We use expenditure and price data for all years, and for all provinces except Prince Edward

Island (dropped due to data masking) [8]. The SHS contains a rich set of demographic data. Included in the vector of demographic variables, **z**, are the following: the age of the household reference person less 42 and its square; a dummy indicating that the household's reference person is a female; year of the survey minus 2002 and its square; Environment Canada's heating and cooling degree-days for each year/province less the overall average of these quantities; a car non-ownership dummy; household type dummies for couple only, couple with children, couple with children and others, single parent, other with relatives only, and other; dummies for households of size 2, 3, 4, 5, 6 and larger; a indicator that the household receives more than 10% of its income as government transfer payments; a dummy for living in a smaller urban area (less than 100,000 residents), and a dummy for living in a rural area. The car non-ownership dummy assigns 1 to households who spend less than $50 on gasoline and 0 to all others. Given our demographic vector, the reference household type, for whom **z**=**0**, is a single-member household in 2002 comprised of a car-owning male aged 40 whose income was less than 10% from government transfers and whose heating/cooling days were average.

Since car ownership is a choice and therefore endogenous, this demand system is a conditional demand system (Browning, 1991). Conditional demand systems allow welfare comparisons holding z constant. Our model can be interpreted as one in which the income and gasoline price changes over the study period are not large enough to create changes in the extensive margin for car usage.

Expenditures are broken down into 10 categories: food purchased at home; food purchased in restaurants; housing; fuel for principal accommodation; electricity; clothing and footwear; health and personal care; recreation, education and reading; alcohol and tobacco; and transportation (excluding car and RV purchase). In our sample, these 10 categories account for 84% of total current consumption. Since we are considering consumption flows, care was taken to exclude durables.

For our purposes, a household is considered a *renter* if and only if they report: that they rent their accommodation; they spent more than $100 on rent in the year; they do not pay reduced rent; and they do not pay any of their rent as in-kind. To ensure that reported rents reflect market rents, we exclude subsidized renters and others whose reported rent is not informative.

[8] The city size flag is masked for all observations in Prince Edward Island, so the province is dropped from the entire model. The city size and urban/rural flags are also masked for a large majority of observations in Newfoundland and Manitoba in 1997-1999, so those observations are also dropped from the dataset..

We note that 22 per cent of households report either reduced/subsidized rent or payment via in-kind.

Instruments for the rent imputation include the log of the price of owned accommodation, the square and cube of the difference between the logged rent and owned accommodation prices, logged mean national mortgage rates for the previous three years, provincial unemployment rates, and dummies for married and single (excluded is separated/divorced).

The SHS is generally considered high quality. However, concerns have been raised about the quality of the 2006 data, especially for low-income households. As detailed in Brzozowski & Crossley (2011), the switch to computer-aided fieldwork may have created problems. Traditionally, Statistics Canada has used a "balance edit" to make sure that consumption and expenditure accounts are in line. If, during the interview, the discrepancy between consumption and income (including net flows into assets) is larger than 20%, the interviewer tries to correct the balance by asking the subject for clarifications and more information. If they cannot reduce the discrepancy to less than 20%, the information for the household is discarded in the processing stage. In 2006, the first year of computer-aided interviewing, there was no mechanism to perform a balance edit. During processing, Statistics Canada realized that 4,300 observations (29% of the total) were out of balance, as opposed to 546 the previous year. Brzozowski & Crossley (2011) argue convincingly that the balance edit principally corrects for under-reporting of income, but allow that the SHS interview process likely ameliorates some over-reporting of consumption. As our results show, poverty rates in 2006 are lower than in the surrounding years, suggesting that there may indeed be some expenditure over-reporting in 2006, particularly for low-consumption households.

Summary statistics for household expenditures and prices, as well as the demographic variables used in the rent imputation, can be found in Online Appendix Table 1.

4.3 Rent Imputation

We impute rent in two ways: (1) by OLS, and (2) using the selection correction techniques described above. The imputation based on OLS regression, which we call the *straight* imputation, uses estimates from regression of real rent, r_i/p_{rent}, on the vector of explanatory variables, $\mathbf{v}_{1i}$.

The explanatory variables $\mathbf{v}_{1i}$ include all variables used in demand analysis---these are demographics $\mathbf{z}_i$ and log-prices $ln\mathbf{p}_i$ and proxies for log nominal total expenditure. In addition, $\mathbf{v}_{1i}$ includes quality and quantity information about various aspects of the housing chosen: year built;

self-reported repairs needed; dwelling type; washer and dryer indicators; number of rooms bathrooms and bedrooms (dummied out). Since log nominal total expenditure is not available for owner-occupiers, we use the following proxies instead: log nominal non-shelter expenditures plus its second to fifth powers, a log nominal household income plus its second to fifth powers, and interactions up to the third order between log nominal non-shelter expenditures and log nominal household income.

It may seem odd that the rent imputation conditions on variables that are not included in the demand analysis. Indeed, the rent imputation conditions on choice variables, like the number of rooms, which induces endogeneity in coefficient estimates. It is therefore important to remember that in the rent imputation, we are not interested in the parameter estimates; rather we are interested in the prediction of rent for owner-occupiers. The best predictor in this setting is one that conditions on all available information, even endogenous choices.[9]

For regressions using the selection correction, our explanatory vector $\mathbf{v}_{1i}$ is the same as in the straight regression. The vector of selection instruments presumed to be associated with the rent/buy decision, $\mathbf{v}_{2i}$, are listed in Online Appendix Table 3. Mortgage rates and house purchase prices (relative to rental prices) are obviously relevant to this decision. In addition, we use marital status (3 categories, with 1 left-out) conditional on household size and composition (which are elements of $\mathbf{v}_{1i}$) as an instrument. The idea here is that, conditional on rent prices, mortgage rates and house prices give the relative cost of purchasing, and that, conditional on household size and structure, marital status relates to liquidity constraints faced by would-be purchasers.

The sample value of the Wald test statistic for the exclusion of these 7 instruments, $\mathbf{v}_{2i}$, in the first stage is 29.9 suggesting that these instruments are moderately informative (conditional on the other regressors) on the rent-buy decision. (The F-test is asymptotically equal to the Wald

[9] To be specific, let $\mathbf{v}_{1i} = \left[\ln \mathbf{p}_i \;\ln x_i \;\mathbf{z}_i \;\mathbf{s}_i\right]$ where $\ln \mathbf{p}_i, \ln x_i, \mathbf{z}_i$ are included in demand estimation and $\mathbf{s}_i$ are aspects of the *quantity* of housing purchased. Thus, $\mathbf{s}_i$ are choice variables for the consumer. Let rent and these choice variables be related by a SUR structure with multivariate normal errors: $\begin{pmatrix} r_i \\ \mathbf{s}_i \end{pmatrix} \sim N\left(\left(\ln \mathbf{p}_i \quad \ln x_i \quad \mathbf{z}_i\right)\mathbf{G}, \Sigma\right)$. In this case, the conditional distribution of rent given all observables, $r_i |_{\ln \mathbf{p}_i \ln x_i \mathbf{z}_i \mathbf{s}_i}$, is a normal variate whose mean is linear in $\mathbf{v}_{1i} = \ln \mathbf{p}_i \;\ln x_i \;\mathbf{z}_i \;\mathbf{s}_i$, and whose variance is smaller than that of the predictor which doesn't condition on $\mathbf{s}_i$. However, this predictor is based on regression coefficients which are endogenous---that is, the matrix of true coefficients $\mathbf{G}$ is not recovered in this prediction regression.

test statistic divided by its degrees of freedom, so, in this case, the F-test on the exclusion restriction is asymptotically equal to 4.3.)

Table 2 gives the means and standard deviations of observed rents, imputed rents and some objects involved in their calculation for all households. The bottom panel gives statistics for the sample of renters and owners whose nonshelter consumption is below the 25th percentile. The middle panel gives statistics for the remaining, relatively richer sample of renters (on the left) and owners (on the right). The former group is more relevant to the estimation of poverty. The correlation coefficient between u_{1i} and u_{2i} is equal to $\rho_0 + \rho_v'\mathbf{Cv}_{1i}$, which varies across observations, and is denoted *rho* in the table.

Table 2: Imputations

			Renters		Owners	
	Rent	**Model**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
All	Actual (real)		7781	3809	22	34
	Imputed	Straight	8161	1757	9218	1591
		Heckman	8148	1739	11832	2496
	Sigma	Straight	2814	1264	4849	1761
		Heckman	2723	1141	4488	1489
	Lambda*sigma	Heckman	1890	2708	8168	4472
	Rho	Heckman	-0.29	0.15	-0.36	0.07
High non-durable	Actual (real)		8723	3963	43	45
spending	Imputed	Straight	9044	1227	9492	1397
		Heckman	8988	1250	12238	2290
	Sigma	Straight	3237	1375	5039	1752
		Heckman	3130	1202	4656	1471
	Lambda*sigma	Heckman	2756	3242	8619	4467
	Rho	Heckman	-0.33	0.13	-0.37	0.05
Low non-durable	Actual (real)		6652	3278	9	9
spending	Imputed	Straight	7113	1714	7093	1387
		Heckman	7151	1712	8700	1666
	Sigma	Straight	2312	887	3379	952
		Heckman	2240	837	3188	849
	Lambda*sigma	Heckman	861	1270	4681	2563
	Rho	Heckman	-0.24	0.15	-0.30	0.10

The main thing to notice from the top panel of Table 2 is that the selection correction has a big effect on the imputed rent assigned to owners. Without a selection correction, owners have imputed rents only $1000 higher than renters. However, with a selection correction, owners have imputed rents $3700 higher than that of renters.

The second thing to notice from Table 2 is that the heteroskedasticity in the model matters for the imputation of rents to owners. Since in our preferred model (wherein we estimate poverty probabilities) we only impute for owners, this is an important point. The standard deviation of u_1 is somewhat lower for poorer subset of owners than for other owners, with an average value of 3188 for the poorer group in comparison to an average value of 4656 for the richer sample.

The value of the selection correction is also smaller for the poorer sample. The selection correction for owners is equal to $-\rho\sigma_{11}(\mathbf{v}_{1i})\lambda_i$. Since $\sigma_{11}(\mathbf{v}_{1i})\lambda_i$ is weakly positive and since the average estimate of ρ is -0.36, the selection correction increases the predicted value of log-rents for owners. However, it does so much less for poor owners than for all owners. Since the average value of $\sigma_{11}(\mathbf{v}_{1i})\lambda_i$ is about 4000 smaller for poorer owners than for richer owners (4681 in comparison to 8619), the selection correction increases imputed rents by about $1400 less for poor owners than for richer owners. Taken together, these results suggest that essentially all of the difference between the straight and selection-corrected rent imputation for owners, a difference of about $1600 for low consumption owners, is due to the selection terms, and not due to differences in prices attributed to the observed characteristics of dwellings.

In our investigation of poverty, these magnitudes matter. With a straight imputation, low consumption renters and owners would be assigned about the same rental flow. With the selection-corrected imputation, low consumption owners are assigned a rental flow about 20 per cent larger than low consumption renters, and thus are less likely to be identified as poor.

4.4 Estimated Substitution Effects and Price Indices

Table 3 gives the estimated **A** matrix recovered from demand estimation, with standard errors in *italics*. The main lesson from Table 3 is that the A matrix is reasonably precisely estimated, and that its coefficients are roughly reasonable. The units of the A matrix are budget shares. For example, the own-price element for food-out is -0.048. This means that if the price of food rises by 10 per cent (so that the log price goes up by 0.10), the food budget share would go down by 0.48 percentage points. The standard errors are in the range of 0.002 to 0.004, which is small enough to ensure that the estimates are not dominated by noise.

The **A** matrix is the matrix of compensated semi-elasticities of budget shares, and it is restricted to by symmetric in the estimation. This matrix is related to the Slutsky matrix for the household, $\mathbf{S}_i$, by $\mathbf{S}_i = k\left(\mathbf{A} + \mathbf{w}_i\mathbf{w}_i' - diag(\mathbf{w}_i)\right)$, where *k* is a scalar function of prices and expenditures. The diagonal of the Slutsky matrix can thus be negative even if the diagonal of **A** is positive. In our sample, only a small fraction of observations have a diagonal element of the Slutsky matrix that is statistically significantly positive (mainly in the recreation equation).

Table 3: Estimated A matrix

	Food in	Food out	Rent	Heat	Electricity	Clothing	Health	Recreation	Sins
Food in	0.035								
	0.012								
Food out	0.025	-0.048							
	0.008	*0.012*							
Rent	-0.051	0.001	0.216						
	0.004	*0.003*	*0.005*						
Heat	-0.005	-0.004	0.002	0.009					
	0.001	*0.001*	*0.001*	*0.000*					
Electricity	0.003	-0.006	-0.043	0.001	0.009				
	0.002	*0.002*	*0.001*	*0.000*	*0.001*				
Clothing	-0.004	0.039	-0.020	-0.002	0.020	0.026			
	0.008	*0.007*	*0.003*	*0.001*	*0.001*	*0.009*			
Health	0.080	-0.029	-0.031	-0.008	0.009	-0.024	0.074		
	0.007	*0.006*	*0.003*	*0.001*	*0.001*	*0.005*	*0.007*		
Recreation	-0.085	0.023	-0.042	0.013	-0.018	-0.022	-0.024	0.122	
	0.007	*0.006*	*0.004*	*0.001*	*0.002*	*0.005*	*0.005*	*0.009*	
Sins	-0.015	-0.002	-0.032	0.005	0.006	-0.006	-0.007	0.033	0.035
	0.005	*0.004*	*0.003*	*0.001*	*0.001*	*0.003*	*0.003*	*0.004*	*0.004*

Thus, although we did not impose negative semi-definiteness on the Slutsky matrix (concavity on utility), it is “mostly negative”.

Table 4 provides descriptive statistics for log expenditure, log price indices, log equivalence scales, and log real equivalent expenditure (ln REE), *y*. The first observation from Table 4 is that the variation of log real equivalent expenditures, *y*, is driven mainly by variation in *lnx*, and secondarily by variation in the log equivalence scale and the log Stone Index. The standard deviation of the remaining term in *y*, ***ln*p’A*ln*p/2**, is nearly two orders of magnitude smaller than that of *y* itself. Thus, we can think of all the complicated price index action as reducing to that of the Stone Index, which gives a lot of weight to changes in the prices of commodities to which households dedicate a lot of expenditure.

The second observation we take from Table 4 is that the selection correction increases our estimate of average log real equivalent consumption by about 0.05 log-points. This is driven by the direct effect of the imputation on nominal expenditures, *x*, rather than by the effect of the imputation on the price index (since the mean of the latter is nearly the same across the two imputations). That is, neglecting the selection correction in the rent imputation would lead to an underestimate of average consumption on the order of 5 per cent. An implication of this is that naively using consumption data and OLS rent imputation to estimate GDP will leave out about 5 per cent of the economy.

Table 4: Price Index, Equivalence Scale, and Real Equivalent Expenditure

	Variable	Mean	Std. Dev.	Min	Max
Straight	ln REE	10.03	0.42	-76.99	11.91
	ln expenditure	10.30	0.51	3.19	12.48
	Equivalence scale, d'z	-0.31	0.27	-0.96	0.00
	Stone index, lnp'w	-0.07	0.18	-19.20	77.77
	Substitution, lnp'Alnp	0.02	0.02	0.00	0.08
Selection-corrected	ln REE	10.08	0.40	-4.39	11.96
	ln expenditure	10.30	0.51	3.19	12.48
	Equivalence scale, d'z	-0.31	0.27	-0.96	0.00
	Stone index, lnp'w	-0.08	0.11	-14.89	7.59
	Substitution, lnp'Alnp	0.02	0.02	0.00	0.08

5. Estimated Poverty Rates

Table 5 presents calculated binary poverty rates and poverty probability rates with standard errors in *italics*[10] using both the straight and the selection-corrected imputation methods. We present estimates for the entire population of Canada, and for the sub-populations of children and the elderly. Children are defined as persons 17 and younger, and seniors as persons 65 and older. Our preferred specification is the rightmost column, which uses the Heckman corrected

[10] Reported standard errors are equal to *sqrt(P(1-P)/N)*, where *P* is the estimated poverty rate These standard errors account solely for the sampling variability induced by the fact that we have only a sample from the population of poverty indicators or poverty probabilities. They ignore sampling variability induced by the sampling variability in the cost-of-living adjustment (the parameters in the matrix **A**) and sampling variability induced by the sampling variation in the estimated parameters in the rent imputation model. These latter two factors may be accounted for via bootstrapping the entire procedure from start to finish. However, this is computationally quite expensive. We did bootstrap the poverty probability for all persons (3rd column from right), and this did not affect the first decimal place of any reported standard error in that column. So, we conclude that ignoring sampling variability in estimated parameters is tolerable in a sample of this size.

imputation and presents the average poverty probability. In the leftmost column, we present the published After-Tax Low-Income rates for comparison.

Figures 1-3 present this information graphically. Across all specifications, the trend in poverty has been downwards. The rightmost block gives our preferred specification, the probability of poverty using the selection-corrected rent imputation. Here, the overall poverty rate dropped from 12.7% to 7.7% between 1997 and 2009, a decline of 5 percentage points.

Our methodological advances in rent imputation matter. If one instead uses the binary poverty measure and a straight rent imputation, we see a measured poverty rate of 16.1% in 1997 and 7.6% in 2009. The binary straight measure shows poverty declining by about 8.5 percentage points over the period, a larger decline than is observed for our preferred measure.

Table 5: Consumption Poverty Rates in Canada, 1997-2009

	Official	Straight Imputation			Heckman Imputation					
		Binary			Binary			Probability		
Year	**All**	**All**	**Children**	**Seniors**	**All**	**Children**	**Seniors**	**All**	**Children**	**Seniors**
1997	15.0	16.1	24.2	12.5	10.7	16.9	5.1	12.7	18.5	9.5
		0.3	*0.4*	*0.3*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.2*
1998	13.7	15.3	22.1	13.6	9.9	14.6	5.4	12.4	17.1	11.2
		0.3	*0.4*	*0.3*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
1999	13.0	14.3	20.8	14.5	10.0	15.3	6.5	11.6	16.7	11.0
		0.3	*0.4*	*0.3*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2000	12.5	12.6	19.3	11.5	9.4	14.8	6.4	10.8	15.8	9.7
		0.3	*0.3*	*0.3*	*0.2*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2001	11.2	14.5	21.2	12.3	10.0	14.9	5.8	10.9	15.3	9.0
		0.3	*0.3*	*0.3*	*0.2*	*0.3*	*0.2*	*0.3*	*0.3*	*0.2*
2002	11.6	11.6	16.7	10.1	9.0	13.0	5.7	9.6	13.6	8.6
		0.3	*0.3*	*0.3*	*0.2*	*0.3*	*0.2*	*0.3*	*0.3*	*0.2*
2003	11.6	12.1	18.0	11.1	10.0	15.1	7.2	11.1	16.1	10.1
		0.3	*0.3*	*0.3*	*0.2*	*0.3*	*0.2*	*0.3*	*0.3*	*0.2*
2004	11.4	10.4	15.4	8.6	9.4	13.9	6.8	9.9	14.1	9.1
		0.3	*0.3*	*0.2*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.2*
2005	10.8	10.4	15.3	7.8	9.9	14.1	7.8	10.3	14.3	9.2
		0.3	*0.3*	*0.2*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.2*
2006	10.5	7.7	10.9	7.1	7.7	10.8	7.7	8.2	10.7	9.8
		0.2	*0.3*	*0.2*	*0.2*	*0.3*	*0.2*	*0.2*	*0.3*	*0.3*
2007	9.2	8.0	11.9	6.6	8.1	11.4	7.5	8.6	12.2	9.1
		0.2	*0.3*	*0.2*	*0.2*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2008	9.4	8.0	12.6	6.8	8.4	12.8	7.7	8.0	11.4	9.2
		0.3	*0.4*	*0.3*	*0.3*	*0.4*	*0.3*	*0.3*	*0.3*	*0.3*
2009	9.6	7.6	10.7	6.7	7.7	11.0	6.8	7.7	9.9	9.1
		0.3	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*	*0.3*

Correcting only the imputation, but keeping a binary poverty measure dramatically decreases the measured level of poverty: instead of a measured rate of 16.1% in 1997 shown for the straight binary measure, we see a measured poverty rate of 10.7%. If we take the selection corrected imputation as 'true', this means that nearly a third of those households classified as poor given a straight imputation are misclassified. The basic point is that owner-occupied households are richer than they appear given their observed dwelling characteristics.

These measures also differ in their over-time trend. The binary selection-corrected poverty measure (middle panel) drops from 10.7% in 1997 to 7.7% in 2009, a decline of 3 percentage points. This is a statistically significantly smaller drop in poverty than 8.5 percentage point decline seen in the binary straight imputation measure. The better rent imputation reduces the observed level of poverty, and makes the over-time decline smaller.

Turning now to the selection-corrected imputation of the probability of poverty (rightmost panel), we see that the estimated poverty rate is slightly higher in 1997. Looking only at selection-corrected imputations, the probability of poverty is 12.7% compared to an average for the binary indicator in 1997 of 10.7%. This difference declines over time and as poverty rates decrease, to erase the difference by the end of the study period.

The basic point here is rent imputations matter, and that given that we have to do them, accounting for the measurement error inherent to them also matters.. Although the selection-corrected rent imputation only identifies a small number of owners as having an expected value of consumption below the poverty line, it puts a much larger number of owners as having a significant amount of density of their consumption distribution below the poverty line. That is, there are a lot of owners whose expected value of consumption is in the neighbourhood of, but above, the poverty line. Since the rent imputation has measurement error, these owners should be treated as 'possibly poor', rather than 'not poor'.

The probability of poverty also has an over-time trend in poverty that is more decreasing than that shown by the binary measure. Between 1997 and 2009, the poverty probability dropped by 5.0 percentage points in comparison with a drop of 3.0 percentage points recorded for the selection-corrected binary poverty measure.

Consider now our subgroups defined by age. For children, consumption poverty is quite responsive to the type of imputation. Considering only binary poverty measures, the straight imputation yields a child poverty rate of 24.2% in 1997, but the selection-corrected binary imputation yields a much lower child poverty rate of 16.9. The key point here is that children are more likely to live in owner-occupied households than the population as a whole, and that since these owner occupied households are richer than their observed dwelling characteristics might indicate, these children are less poor than they appear.

Turning now to poverty among the elderly, we see a pattern noted in many other studies of poverty in Canada (see, e.g., Pendakur 2001): poverty is much less prevalent among the elderly than among children and non-elderly adults. In 1997, the straight binary measure suggests that the elderly have about half the poverty rate of children, and about a third of the rate when correcting for selection. For example, the selection-corrected binary poverty measure is less than 5.1 per cent for seniors in comparison with 16.9 per cent for children. However, the binary poverty rates declined much less for seniors than for children. The selection-corrected binary poverty rate for seniors actually increased by 1.7 percentage point between 1997 and 2009, in comparison with a drop of 5.9 percentage points for children and 3.0 percentage points for the population as a whole.

Turning to our preferred measure, the selection-corrected probability of poverty, we see much higher overall poverty rates for seniors than in the binary measure. This has two explanations. First, seniors are more likely to own a house than non-seniors--71% of households containing a senior are homeowners, compared to only 61% of households without seniors. Second, households with seniors are more likely to be in the vicinity of the poverty line. Over the study period, 6.6 per cent of households without seniors in them had a total expenditure within $1000 of the poverty line, compared to 11.9 per cent for households containing seniors. These two factors — the wider range of predicted equivalent expenditure caused by a larger share of homeowners, and a higher density of households near the poverty line — mean that the probability measure matters more. However, like what is seen with the binary measures, the probability of poverty among seniors did not decline as much as it did for other population groups. That measure declined from 9.5 per cent in 1997 to 9.1 per cent in 2009, a decline of only 0.4 percentage points.

The differences between selection-corrected poverty probability estimates (our preferred specification) and the straight results are driven by two factors that work in opposite directions. First, the Heckman correction pushes down measured poverty by better accounting for the consumption flow accruing to owners. In particular, although low-consumption owners have only slightly better observed dwelling characteristics than their renter counterparts, they have much better unobserved dwelling characteristics. This means that the straight imputation used in all previous studies of consumption poverty understate their consumption, and thus overstate their poverty.

Second, the probability measure pushes measured poverty slightly upwards. This can be attributed to the relative density of the distribution of imputed consumption on either side of the poverty line. The standard deviation of the second-step real rent regression, $\sigma_{11}(\mathbf{v}_{1i})$, for households near the poverty threshold is fairly large, about $2900. If the poverty threshold were at a flat part of the density function of imputed consumption (for example, near the mode of a unimodal distribution), there would be lots of households on both sides of the poverty line, so that mis-classification would largely cancel. In this case, the binary and probability measures would be similar. However, when the poverty threshold is at a point where the density is increasing (as it would be below the mode of a unimodal distribution), then there is more density to the right than to the left, which drives up the probability of poverty. As the poverty threshold in our empirical work is well below the modal (and median and mean) consumption level, accounting for measurement error in imputed consumption increases the estimated poverty rate for owners.

In Table 6, we consider several alternative specifications, to evaluate the effect of our various measurement choices. In the leftmost column, we present the baseline average probability of poverty for all persons from Table 5. In the next block, we consider other groups in the population, including single-parents. In the next block we use different equivalence scales. In the next block, we use different poverty lines. Finally, in the rightmost block, we consider 2 specifications that are "closer to the data". The first, uses the CPI instead of our price index, using the Heckman rent imputation and a binary poverty measure. The second measures binary poverty using a consumption basket that excludes shelter expenditures entirely, and uses the Stone Index alone to deflate expenditures (that is, we set $\ln\mathbf{p}'\mathbf{A}\ln\mathbf{p}/2=0$). In this last exercise, we use no econometrics at all: there is no demand system estimation, no imputation and no probability of poverty estimation. Since this final model measured a smaller set of expenditures, a poverty line generating a measured rate of 11.6% in 2002 (corresponding to the official rate After-Tax LICO poverty rate in 2002) was applied.

Table 6: Consumption Poverty Rates, Various Specifications

	Probability of poverty								Less model	
Year	**Baseline**	**Childless**	**2 parents**	**1 parent**	**Root-n**	**OECD**	**90% LICO**	**110% LICO**	**Heck CPI**	**No rent**
1997	12.7	8.8	10.8	32.3	16.0	16.5	8.6	17.5	14.7	14.6

	0.3	*0.2*	*0.3*	*1.3*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
1998	12.4	9.7	10.0	31.5	15.7	16.4	8.2	17.0	13.8	13.9
	0.3	*0.3*	*0.3*	*1.5*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
1999	11.6	8.6	9.9	24.1	15.0	15.1	7.9	16.8	14.1	13.6
	0.3	*0.2*	*0.3*	*1.3*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2000	10.8	7.8	9.4	21.2	13.8	14.3	6.8	15.3	14.3	11.9
	0.3	*0.2*	*0.2*	*1.2*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2001	10.9	8.2	9.7	23.2	14.1	14.0	7.2	15.7	17.2	14.5
	0.3	*0.2*	*0.2*	*1.2*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2002	9.6	7.6	7.6	19.6	12.7	12.9	6.2	14.1	13.5	11.6
	0.3	*0.2*	*0.2*	*1.2*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2003	11.1	8.2	9.6	21.3	14.0	14.1	7.2	15.4	14.8	12.6
	0.3	*0.2*	*0.2*	*1.2*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2004	9.9	7.7	9.2	19.9	12.8	13.2	6.4	14.3	13.9	11.3
	0.3	*0.2*	*0.2*	*1.2*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2005	10.3	8.0	9.5	21.1	12.7	13.0	6.7	14.1	14.1	11.3
	0.3	*0.2*	*0.3*	*1.3*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2006	8.2	7.0	6.9	14.3	10.9	11.0	4.8	12.1	12.6	9.5
	0.2	*0.2*	*0.2*	*1.1*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.2*
2007	8.6	6.9	7.4	15.5	11.0	11.4	5.3	12.5	12.8	10.1
	0.3	*0.2*	*0.2*	*1.2*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*
2008	8.0	6.5	7.3	16.0	10.5	11.1	5.2	12.1	13.5	9.5
	0.3	*0.3*	*0.3*	*1.5*	*0.3*	*0.3*	*0.2*	*0.3*	*0.4*	*0.3*
2009	7.7	6.5	6.6	15.0	10.5	10.4	4.7	12.1	12.8	9.9
	0.3	*0.3*	*0.3*	*1.4*	*0.3*	*0.3*	*0.2*	*0.3*	*0.3*	*0.3*

Turning first to the numbers for other population groups, the biggest feature here is the very high consumption poverty rates faced by single-parent families. In 1997, the consumption poverty rate for persons in single-parent households was 32.3%. By 2009, this proportion had declined to 15.0%. In contrast, people living in childless households were drastically less likely to be poor throughout the period, with a poverty rate ranging from 8.8% in 1997 to 6.5% in 2008.

The next block considers 2 alternative equivalence scales. The basic pattern here is that although both alternative equivalence scales show higher poverty than our baseline specification, they show the same pattern over time. We take this similarity in the over-time trend as suggesting that the choice of equivalence scale is not driving our conclusions.

The next block of Table 6 uses two alternative poverty thresholds from our baseline case. The first is a lower threshold, set at 90% of the value of our baseline threshold of $13,740 for a single adult in Ontario in 2002, and the second is set at 110% of that value. One can clearly see in this part of the table that the distribution of consumption is quite compressed. With the threshold at 90% of our baseline, poverty rates are nearly a full third lower, and with the threshold at 110%, over a third higher. We take this as a signal that more attention should be paid to the choice of the poverty threshold. The methods developed in this paper allow the research to extend a well-considered threshold to other household types and to other price regimes, but they do not solve the problem of what threshold to use.

We also consider two specifications that are closer to the data. The first does not use our price index, but rather uses the published value of the Consumer Price Index (CPI) to adjust for price variation. The CPI is used to adjust Canada's LICOs and official poverty statistics. As only national-level CPIs are used for this purpose, it does not account for interprovincial price variation. Here, we see that the same downward trend is evident, but that the level of measured poverty is quite a bit higher than when we account for provincial price variation. The reason is simple: households in cheaper parts of the country have less nominal consumption, which makes them look poor given a national-level price index, but not poor given a provincial-level price index.

Thus our major findings do not depend critically on either our specific rental imputation strategies, or our use of provincial-level commodity prices. However, it does show that accurate provincial level poverty statistics require use of provincial level prices.

The rightmost column is a specification that uses no econometrics whatsoever. Here, the Stone Index is used to account for interprovincial price variation, so no demand estimation is used. In addition, shelter is excluded from the consumption basket, so no rent imputation is used. This column shows a different pattern from that we found in Table 3. Here, the decline in the poverty rate is flatter than in our other scenarios, declining from 14.6% to 9.9% throughout the period. We note that this is not due to: (1) the use of consumption rather than income, because officially reported income poverty drops by a larger amount; (2) the use of provincial rather than national price variation, because CPI deflated consumption poverty tends to drop slower than provincially-deflated poverty; or (3) the use of a demand-based deflator, because the variance of the substitution term lnp'Alnp is so small.

One possible explanation is that there is unobserved heterogeneity in the price of shelter. For example, if the true price of shelter is high, then shelter expenditures will be high (because shelter is a necessity), which will make total expenditures high, but those shelter expenditures will crowd out other expenditures, making non-shelter expenditures low. However, if the observed price of shelter is not high, these high shelter expenditures will contribute to a high level of total consumption. In this case, the Stone-deflated non-shelter expenditures of the household would be low, but the Stone-deflated total consumption of the household would not be low.

Table 7: Poverty probability by province, all persons

Year	National	Alberta	BC	Manitoba	NB	Newfoundland	NS	Ontario	Quebec	Saskatchewan
1997	12.7	8.8	11.3	10.0	12.9	20.4	13.9	12.2	15.4	12.0
	0.3	*0.6*	*0.7*	*1.8*	*0.8*	*1.6*	*0.8*	*0.7*	*0.7*	*0.9*
1998	12.4	7.6	12.2	16.5	12.0	18.1	12.4	12.4	13.9	13.4
	0.3	*0.7*	*0.8*	*2.2*	*1.0*	*1.9*	*1.0*	*0.7*	*0.7*	*0.9*
1999	11.6	6.9	12.8	6.5	11.8	13.2	12.3	11.3	13.3	11.7
	0.3	*0.6*	*0.7*	*1.4*	*0.9*	*1.4*	*0.8*	*0.7*	*0.8*	*0.8*
2000	10.8	6.6	11.4	10.3	11.8	15.3	12.1	10.3	12.2	13.0
	0.3	*0.6*	*0.7*	*0.8*	*0.9*	*1.0*	*0.9*	*0.7*	*0.8*	*0.9*
2001	10.9	6.9	9.7	10.9	12.0	12.9	11.8	10.8	10.6	13.0
	0.3	*0.6*	*0.7*	*0.8*	*0.8*	*0.9*	*0.8*	*0.7*	*0.7*	*0.9*
2002	9.6	7.2	10.4	10.6	10.1	12.4	12.5	9.7	9.2	11.5
	0.3	*0.7*	*0.7*	*0.8*	*0.9*	*0.9*	*0.9*	*0.7*	*0.6*	*0.8*
2003	11.1	9.2	11.1	9.3	12.7	13.9	11.2	11.6	10.6	12.7
	0.3	*0.7*	*0.7*	*0.8*	*0.9*	*0.9*	*0.8*	*0.7*	*0.5*	*0.9*
2004	9.9	7.1	12.2	12.4	11.3	13.8	11.8	10.0	8.7	10.7
	0.3	*0.7*	*0.8*	*0.9*	*0.9*	*0.9*	*0.9*	*0.7*	*0.6*	*0.8*
2005	10.3	6.2	9.8	11.1	11.2	11.4	12.0	11.7	9.3	10.7
	0.3	*0.6*	*0.7*	*0.8*	*0.9*	*0.9*	*0.9*	*0.8*	*0.7*	*0.8*
2006	8.2	4.1	6.5	11.6	10.1	11.1	8.9	9.3	8.1	9.1
	0.2	*0.5*	*0.6*	*0.8*	*0.8*	*0.9*	*0.8*	*0.6*	*0.6*	*0.8*
2007	8.6	5.3	8.3	8.9	7.7	11.4	8.2	9.6	8.5	7.5
	0.3	*0.6*	*0.7*	*0.8*	*0.8*	*0.9*	*0.8*	*0.7*	*0.7*	*0.7*
2008	8.0	4.4	6.6	12.2	8.1	9.5	9.6	9.2	8.0	5.5
	0.3	*0.7*	*0.8*	*1.1*	*0.9*	*1.0*	*1.0*	*0.8*	*0.8*	*0.7*
2009	7.7	6.4	7.6	9.5	7.2	6.9	10.0	8.5	6.7	6.4
	0.3	*0.8*	*0.8*	*1.0*	*0.9*	*0.8*	*1.0*	*0.7*	*0.7*	*0.8*

To explain the over-time pattern that we observe, the variance of such unobserved shelter price heterogeneity would have to be rising over time. Moretti (2010) suggests that such a process might be associated with the housing bubble in the USA over the 1990s and 2000s. Using city-level price data for just 2 goods, shelter and non-shelter, he finds that this effect is quite large and affects estimates of overall inequality quite significantly. Consequently, we believe that creating better data on local-level price variation is essential to getting a real handle on consumption poverty and inequality. Further, the fact that the current best case is provincial-level price variation should lead one to take all poverty measures---be they based on consumption or income---with a few grains of salt.

Despite these reservations, we present provincial-level poverty rates in Table 7. Although poverty has decreased in every province, there has been substantial variation in poverty levels and trends across provinces. Newfoundland has shown the most dramatic progress as the economy boomed on the strength of oil and natural gas development, with the poverty rate dropping from 20.4% to 6.9%. Saskatchewan, which similarly benefited from higher oil prices and the development of its energy resources, saw its poverty rate fall from 12.0% to 6.4%. Resources were not the only factor in lowering poverty rates, however: the rate in Quebec fell from 15.4% to 6.7% despite having no oil and gas. Alberta's poverty rate fell only 2.4 percentage points to 6.4% despite massive oil and gas development, although it ended the study period with the lowest poverty rate in the country.

In British Columbia, poverty was lower than the national average at the beginning of the period with an estimated poverty probability of 11.3%. But, poverty in B.C. fell by less than the national average over the study period, dropping by only 3.7 percentage points, leaving the province with a rate almost identical to the national average by 2009. Manitoba showed the smallest decline in poverty: its rate fell 0.5 percentage points, to 9.5%.

6. Conclusion

We propose new methods for dealing with rent imputation in consumption data. We control for differences in unobserved quality correlated with the selection between owned and rented housing using a modified a Heckman correction. Further, we invent a model-based correction to poverty rates for the measurement error in the rent imputation. Instead of averaging a binary poverty indicator across observations in the population, we suggest instead averaging the probability of poverty across observations in the population.

We bring our methods to the data in the analysis of poverty in Canada. We describe patterns in the evolution of consumption poverty over the period 1997-2009, adding to previous work by Pendakur (2001), Crossley and Curtis (2006), Milligan (2008) and Sarlo (2008). This work uses provincial-level commodity price data to match expenditure data in the Surveys of Household Spending. We find that poverty decreased in Canada between 1997 and 2009. The overall poverty rate using our preferred measure---the average probability of poverty for all households---declined from 12.7 per cent in 1997 to 7.7 per cent in 2009. Reductions in child poverty were more pronounced, dropping from 18.5 per cent to 9.9 per cent over the same period.

References

Blundell, R. & Robin, J.M., 1999. Estimation in Large and Dissaggregated Demand Systems: An Estimator for Conditionally Linear Systems. *Journal of Applied Econometrics*, 14(3), pp.209–232.

Browning, M. & Meghir, C., 1991. The Effects of Male and Female Labor Supply on Commodity Demands. *Econometrica*, 59(4), p.925.

Brzozowski, M. & Crossley, T.F., 2011. Viewpoint: Measuring the well-being of the poor with income or consumption: a Canadian perspective. *Canadian Journal of Economics/Revue canadienne d'économique*, 44(1), pp.88–106.

Crossley, T.F. & Curtis, L.J., 2006. Child poverty in Canada. *Review of Income and Wealth*, 52(2), pp.237–260.

Deaton, A., 1997. *The analysis of household surveys: a microeconometric approach to development policy*, Johns Hopkins University Press.

Frick, M; Grabka, M; Smeeding, T & Tsaklogou, P; 2008. Distributional effects of imputed rents in seven European countries. *Accurate Income Measurement for the Assessment of Public Policies.*

Greene, W.H., 2011. *Econometric analysis*, Prentice Hall Upper Saddle River, NJ.

Heckman, J.J., 1979. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pp.153–161.

Lewbel, A. & Pendakur, K., 2009. Tricks with Hicks: The EASI demand system. *The American Economic Review*, 99(3), pp.827–863.

Meyer, B.D. & Sullivan, J.X., 2011. Viewpoint: Further results on measuring the well-being of the poor using income and consumption. *Canadian Journal of Economics/Revue canadienne d'économique*, 44(1), pp.52–87.

Milligan, K., 2008. The evolution of elderly poverty in Canada. *Canadian Public Policy*, 34(4), pp.79–94.

Pendakur, K., 2001. Consumption poverty in Canada, 1969 to 1998. *Canadian Public Policy/Analyse de politiques*, 27(2), pp.125–149.

Pendakur, K., 1999. Semiparametric estimates and tests of base-independent equivalence scales. *Journal of Econometrics*, 88(1), pp.1–40.

Sarlo, C.A., 2001. *Measuring poverty in Canada*, Fraser Institute.

Sarlo, C.A., 2008. *What is poverty? Providing clarity for Canada*, Fraser Institute.

Sen, A.K. & Press, O.U., 1982. *Poverty and famines: an essay on entitlement and deprivation*, Clarendon Press Oxford.

Stone, R., 1954. Linear expenditure systems and demand analysis: an application to the pattern of British demand. *The Economic Journal*, 64(255), pp.511–527.

Figure 1: Poverty, all persons

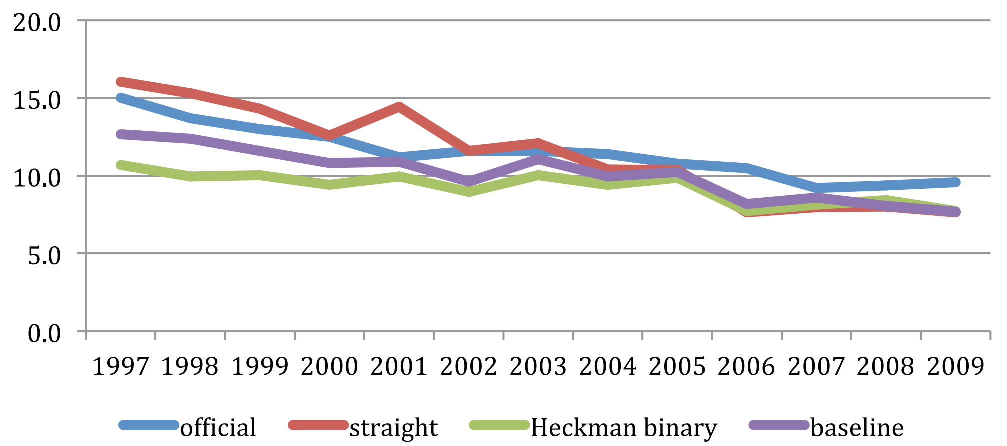

Figure 2: poverty, children

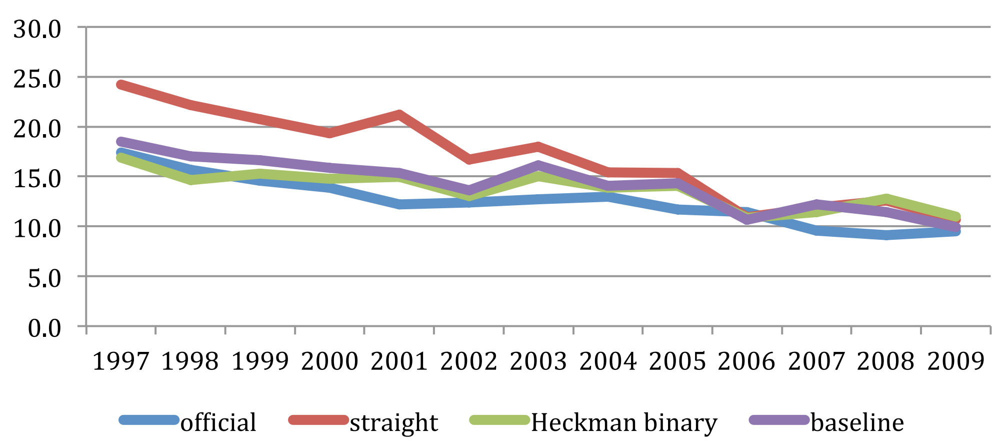

Figure 3: poverty, seniors