# Lost in translation: What do Engel curves tell us about the cost of living?*

Ingvild Almås
IIES Stockhom University
Norwegian School of Economics

Timothy K.M. Beatty
University of California, Davis

Thomas F. Crossley
University of Essex
Institute for Fiscal Studies

May 25, 2016

**Abstract**

The Hamilton method for estimating CPI bias is simple and intuitively appealing and has been widely adopted. We show that the Hamilton method is internally inconsistent and conflates CPI bias with variation in cost of living across income levels. We demonstrate a method by which these two components can be disentangled. When applied to Hamilton's data our method gives larger estimates of CPI bias. Our approach can recover changes in cost of living for any income level and we exploit this to study growth in the median income for the United States.

*Crossley acknowledges support from the ESRC-funded Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies (CPP, reference RES-544-28-5001). We would like to thank Orazio Attansio, Garry Barrett, Richard Blundell, Chris Carroll, Jeff LaFrance, Valerie Lechene, Ian Preston, David Stromberg, Guglielmo Weber, participants in a number of seminars and especially Bruce Hamilton for helpful comments. All remaining errors are our own.

# 1 Introduction

Measuring economic progress over time and assessing cross-country differences in living standards are central concerns of researchers, international institutions, policy makers and the broader public. Such comparisons require accurate measures of the cost of living with which to deflate nominal measures of economic activity or economic resources. Price indices and PPP calculations, such as the Consumer Price Index (CPI) or the purchasing power parities implicit in the Penn World Tables, are commonly employed to this end. However, there are longstanding concerns regarding the many biases associated with traditional approaches to price index construction. These include outlet bias, substitution bias, and biases associated with failing to deal fully with quality changes and the introduction of new goods. See Diewert (1987, 1993, 1995), Reinsdorf (1993), Boskin et al. (1998), Hausman (2003) and Hausman and Leibtag (2009) for biases in the CPI, and Neary (2004), Hill (2004), Deaton and Heston (2010) and Almås (2012) for corresponding concerns with PPP calculations.

In this paper we reconsider a popular approach to this problem: the "Hamilton Method" (Hamilton, 1998, 2001a). Stagnant "real incomes" (i.e. income deflated by CPI) in the United States during the 1970s and 80s were of considerable concern to economists and policy makers. Nakamura (1995) noted that aggregate food budget shares declined throughout the period. That the food budget share falls with income (Engel's "law" (Engel, 1857, 1895)) is among the best known empirical regularities in economics. This suggested that perhaps incomes were growing, but standard methods had failed to capture this growth. An appealing explanation was that the CPI overestimated increases in the cost of living over time, so that real income growth had been underestimated. Hamilton formalized this intuition, proposing a way of using micro data to estimate the correction to the CPI required for food budget shares to be stable over time. Among the attractions of this method are that it is simple and intuitively appealing, it has modest data requirements and it offers an omnibus approach to assessing the multitude of potential biases in previously available indices like the CPI, rather than a piecemeal approach of assessing one bias at a time. Consequently, the Hamilton method has been widely applied over the last fifteen years. Hamilton (2001a,b), Costa (2001), Beatty and Larsen (2005), Brzozowski (2006), Larsen (2007), Gibson et al. (2008), Xu and Zeng (2009), Barrett and Brzozowski (2010), Gibson and Scobie (2010), Chung et al. (2010), Clerc and Coudin (2011), Olivia and Gibson (2012), de Carvalho Filho and Chamon (2012) and Nakamura et al. (2016) study temporal price change indifferent countries. It has also recently been applied to spatial comparisons across countries. Chattopadhyay (2010), Coondoo et al. (2011) and Almås (2012) apply the method to spatial comparisons across countries or regions. The method has has been cited, for example, as important complementary evidence of upward CPI bias in reassessments of the time path of poverty in the U.S. (Meyer and Sullivan,

2009).

We begin by noting that the Hamilton method is internally inconsistent. On the one hand, the method estimates a single price index and interprets that price index as the cost of living index for all households. This interpretation implies that the cost of living index for an individual household does not depend on the utility (and hence income) of the household. This can only be true if preferences are homothetic. On the other hand, the method *requires* that preferences are not homothetic. Homotheticity implies that budget shares are independent of income. In that case there is no translation (or scaling of nominal income) that aligns the Engel curves, and the method can not be implemented. So, if preferences are such that the method can be implemented, there is not a single cost of living index to identify, and if preferences are such that there is a single cost of living index to identify, the method would not work.

As food shares do fall with income, preferences cannot be homothetic. Richer and poorer households have different consumption baskets, and hence experience different changes in the cost of living. The Hamilton method and the CPI each estimate changes in the cost of living at specific points in the income distribution. However, they need not and in general will not, estimate changes in the cost of living at the same point in the income distribution. Thus the Hamilton method conflates CPI bias with variation in inflation across income levels.

We then develop a consistent method in the spirit of Hamilton. The first step is to apply a method due to Muellbauer (1976)(see also Deaton (1998)) to determine the point in the income distribution where changes in the cost of living are best approximated by the CPI. We next determine, theoretically, *whose* cost of living is being measured by the Hamilton method and show how this can be recovered empirically. This allows to separate genuine bias in the CPI from differences in the cost of living at the income levels corresponding to the CPI and to the Hamilton method. Applying our method to Hamilton's data leads to larger estimates of the degree of CPI bias.

We also show, that with no further assumptions, the method recovers changes in the cost of living at any point in the income distribution. We exploit this to study growth in median income in the United States.

The paper is organized as follows. Section 2 reviews the Hamilton method and demonstrates the fundamental inconsistency. Section 3 determines theoretically whose cost of living is being measured by the Hamilton method. Section 4 shows how to recover this income level empirically, and separate genuine CPI bias the consequences of nonhomothetic preferences. Section 5 applies our method to Hamilton's data and Section 6 concludes.

# 2 The Hamilton Method

## 2.1 Overview

The reasoning underpinning the Hamilton approach seems straightforward. If preferences are stable over time and the CPI accurately measures costs of living then we would expect households at the same level of CPI-deflated income to allocate the same share of their budgets to food. Controlling for other possible influences on spending patterns, the food budget share at a given level of deflated income should be constant through time. If households at the same levels of deflated-income appear to be spending different shares of their budget on food over time, and this cannot be explained by other factors, it suggests that true real income is changing, even when measured real income is not. The Hamilton method ascribes this to mismeasurement of change in the cost of living.[1]

Most applications of the Hamilton method assume Almost Ideal Demand preferences (Deaton and Muellbauer, 1980) which implies Engel curves of the Working-Leser form. The food Engel curve is given by:

$$w^f_{h,r,t} = \alpha^f + \gamma^{ff} \ln\left(\frac{p^f_{r,t}}{p^n_{r,t}}\right) + \beta^f \ln\left(\frac{x_{h,t}}{P_{r,t}}\right), \tag{1}$$

where $h$ indexes households[2], $r$ indexes regions, $t$ indexes time, $x_{h,r,t}$ is nominal income (or total expenditure), $\frac{p^f_{r,t}}{p^n_{r,t}}$ is the price of food ($f$) relative to nonfood ($n$) and $P_{r,t}$ is a price index.[3]

In practice, with U.S. data, *levels* of food and nonfood prices are not observed but their inflation rates are. These inflation rates are assumed to be measured with error. So we can write

$$p^k_{r,t} = p^k_{r,0}(1 + \Pi^k_{r,t})(1 + E^k_{r,t}) \tag{2}$$

where $\Pi^k_{r,t}$ is the observed inflation rate in the price of good $k$ to period $t$ in region $r$ and $E^k_{r,t}$ captures the difference between the true and observed inflation rates. Similarly, the overall price index $P_{r,t}$ can be written as a function of the base period value of the price index, the observed CPI and a term $E_{r,t}$ that captures the difference between the growth of the CPI and the true growth in $P_{r,t}$:

$$P_{r,t} = P_{r,0} CPI_{r,t}(1 + E_{r,t}). \tag{3}$$

[1] A different critique of this approach emphasized by Blow (2003) and Logan (2009) is that it is a *residual-based* method. Any time varying determinants of demand that are not accounted for will be attributed to CPI bias. See also Deaton and Dupriez (2011) and Ravallion (2015).

[2] We use household as the unit of observation throughout because micro level expenditure data typically come at the household level.

[3] This is a two-good demand system for food and nonfood as in Hamilton (2001a) or Costa (2001). Costa (2001) also considers Recreation / Non Recreation but her main results focus on food/nonfood. The two good system implies either within group homotheticity or fixed within group relative prices.

Then:

$$
\begin{aligned}
w^f_{h,r,t} &= \alpha^f + \gamma^{ff} \ln\left(\frac{p^f_{r,0}(1+\Pi^f_{r,t})(1+E^f_{r,t})}{p^n_{r,0}(1+\Pi^n_{r,t})(1+E^N_{r,t})}\right) + \beta^f \ln\left(\frac{x_{h,r,t}}{P_{r,0}\mathrm{CPI}_{r,t}(1+E_{r,t})}\right) \qquad (4)\\
&= \alpha^f + \gamma^{ff} \ln\left(\frac{1+\Pi^f_{r,t}}{1+\Pi^n_{r,t}}\right) + \beta^f \ln\left(\frac{x_{h,r,t}}{\mathrm{CPI}_{r,t}}\right)\\
&\quad +\gamma^{ff}[\ln(1+E^f_{r,t}) - \ln(1+E^n_{r,t})] - \beta^f \ln(1+E_{r,t}) + \gamma^{ff} \ln\left(\frac{p^f_{r,0}}{p^n_{r,0}}\right) - \beta^f \ln P_{r,0}.
\end{aligned}
$$

The last term $\left[\gamma^{ff} \ln\left(\frac{p^f_{r,0}}{p^n_{r,0}}\right) - \beta^f \ln P_{r,0}\right]$ varies only by region and so can be captured by region dummies. Hamilton makes two further assumptions. First, he assumes that $E^f_{r,t}$, $E^n_{r,t}$ and $E_{r,t}$ vary with time only. Second, in Hamilton's reported estimates, and in all the literature that follows, it is assumed that the difference between the true and observed inflation rates are identical for food and nonfood, i.e. $\ln(1+E^f_t) - \ln(1+E^n_t) = 0$.[4] Thus:

$$
\begin{aligned}
w^f_{h,r,t} &= \alpha^f + \gamma^{ff} \ln\left(\frac{1+\Pi^f_{r,t}}{1+\Pi^n_{r,t}}\right) + \beta^f \ln\left(\frac{x_{h,r,t}}{\mathrm{CPI}_{r,t}}\right) \qquad (5)\\
&\quad +\gamma^{ff}[\ln(1+E^f_t) - \ln(1+E^n_t)] - \beta^f \ln(1+E_t) + \sum_r \delta^r D^r\\
&= \alpha^f + \gamma^{ff} \ln\left(\frac{1+\Pi^f_{r,t}}{1+\Pi^n_{r,t}}\right) + \beta^f \ln\left(\frac{x_{h,r,t}}{\mathrm{CPI}_{r,t}}\right) + \sum_t \delta^t D^t + \sum_r \delta^r D^r \qquad (6)
\end{aligned}
$$

where $D^r$ are region dummies, $D^t$ are time dummies, $\delta^r$ and $\delta^t$ are coefficients, and in particular, $\delta^t = -\beta^f \ln(1+E_t)$. This equation can be taken to data.

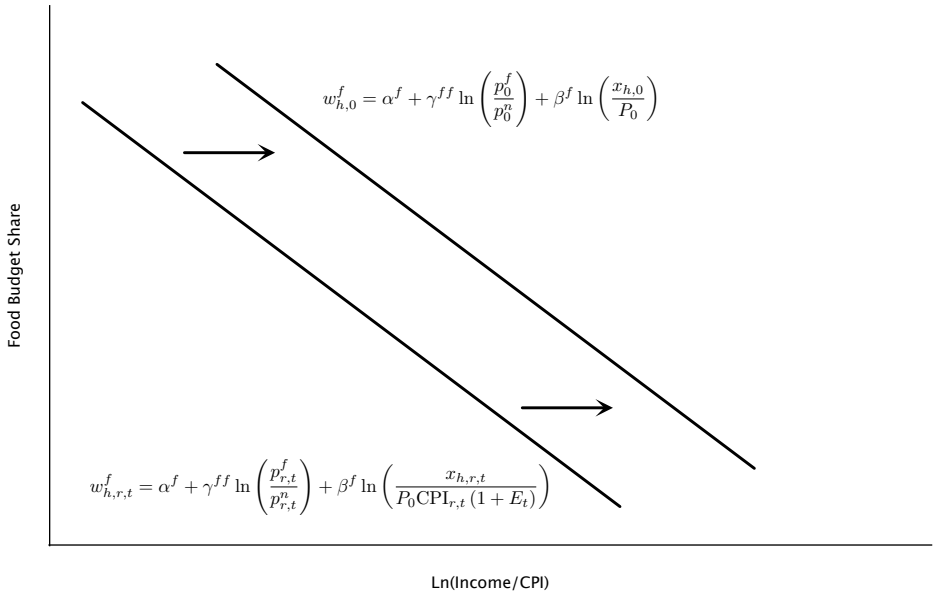
Intuitively, $E_t$ is chosen to minimize the distance between the Engel curves in the base period and period $t$. Equivalently, the Hamilton method estimates the price index, $P_{r,t} = P_{r,0}\mathrm{CPI}_{r,t}(1+E_t)$ as the translation to nominal income that aligns the food Engel curves. This is summarized in Figure 1. The correction $E_t$ shifts the region/period $r,t$ Engel curve back onto the region/period 0 Engel curve, as indicated by the arrows.

Hamilton incorrectly identifies $P_{r,t}$ as the true cost of living index.[5] $E_t$ is then interpreted as the bias in the CPI relative to growth in the cost of living. In the next section, we show that while $P_{r,t}$ is a price index, if relative prices vary, it is the true cost of living index at, at most, one point in the income distribution.

[4] In his development of the method, Hamilton allows that $E^f_t$ and $E^n_t$ may differ. He suggests that his method will understate the bias in the overall CPI as observed food inflation likely understates true food inflation by less than observed nonfood inflation understates true nonfood inflation.

[5] Page 622, just above Equation 2.

Figure 1: The Hamilton Approach




This means that $E_t$ can *not* be interpreted as the bias in the CPI relative to changes in the cost of living.

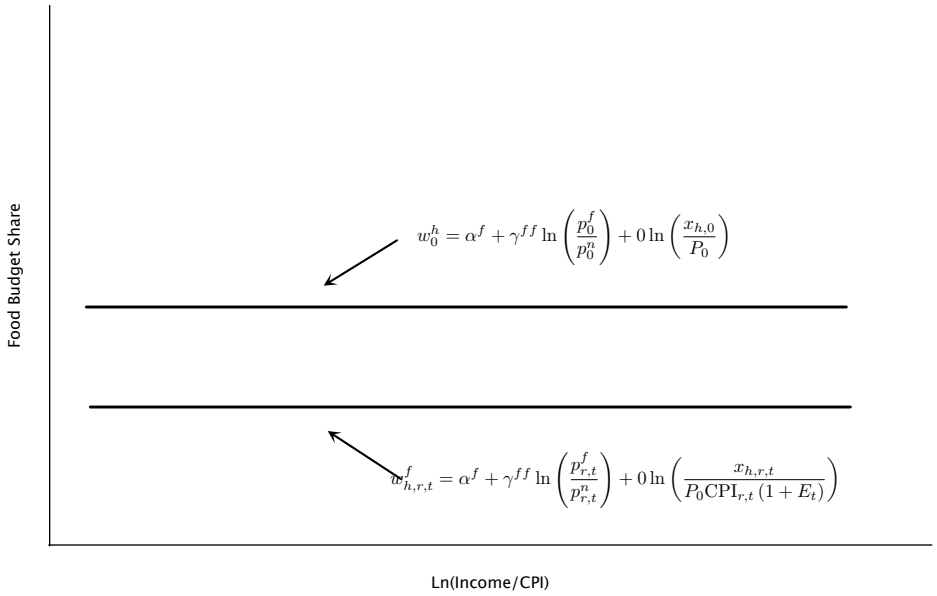
## 2.2 Lost in Translation

The Hamilton method is internally inconsistent. The method estimates a single price index and interprets that price index as the cost of living index for all households. This interpretation implies that the cost of living index for an individual household does not depend on the utility – and hence income – of the household. This can only be true if preferences are homothetic. But if preferences are homothetic, the Hamilton method cannot be implemented. Homotheticity implies that budget shares are flat with regard to income, as illustrated in Figure 2 with the 0 coefficient on the logarithm of deflated income. In this case there is no horizontal translation (or scaling of nominal income in region/period $r, t$) that aligns the two Engel curves, and the method cannot identify the cost of living index. To see this another way, note that the Hamilton method estimates $\ln(1 + E_t)$ by $\frac{-\delta^t}{\beta^f}$; if preferences are homothetic, this involves dividing by zero.

The evidence that preferences are not homothetic is overwhelming: that food shares fall with income is Engel's law and has been confirmed by a multitude of empirical studies of consumer demand . Thus the Hamilton method can be implemented, but the true cost of living differs across households; poor and rich households experience different changes in the cost of living because they consume different bundles (for example, the poor spend a greater share of their budget on food).[6]

[6] These differences might be small if the relative price of food was relative stable. But relative prices can and do change over

Figure 2: The Hamilton Approach with Homothetic Preferences




Of course, the nonhomotheticity of preferences also implies that the CPI, which is single price index, estimates the cost of living at one point in the income distribution – an observation made long ago by Prais (1959) and Nicholson (1975). Therefore the implication of preferences being nonhomothetic is that the CPI and $P_t$ each estimate a cost of living index for a specific level of income, but these two income levels need not be the same. We refer to the virtual household at the level of income at which the CPI estimates the cost of living as the *CPI household*. Correspondingly, we refer to the virtual household at the level of income at which $P_t$ estimates the cost of living as the *Hamilton household*.

Therefore the Hamilton estimate of "bias", $E_t$, contains two elements:

1. CPI bias: the CPI differs from the true cost of living index of the CPI household due to substitution, new goods, quality changes, etc.

2. Nonhomotheticity: The CPI household and Hamilton household have different levels of income, and so experience different changes in their true cost of living.

Previous analyses using Hamilton's method to assess CPI bias have conflated these two things. The rest of the paper develops and implements a method that does not suffer from this flaw.

time. Moreover, if they did not, one would not need a price index to adjust nominal quantities. Any single price would do.

# 3 Whose cost of living?

## 3.1 The CPI household

To disentangle genuine CPI bias from the effect of nonhomotheticity, we first need to determine the locations of the CPI household and the Hamilton household in the income distribution.

For the CPI there is a well-established method of doing this due to Muellbauer (1976). The CPI household has expenditure shares which match the expenditure weights in the CPI. The expenditure weights in the CPI are the *aggregate* expenditure shares (that is the share of each good in total household spending). The aggregate expenditure share of good $k$ is in turn a weighted mean of household expenditure shares for good $k$, where each household is weighted by its share of total income.

$$W_{r,t}^k = \sum_h \frac{x_{h,r,t}}{\sum x_{h,r,t}} w_{h,r,t}^k. \tag{7}$$

Given this, and the Working-Leser functional form for household-level shares assumed above, we can write the aggregate share for food at time $t$, as:

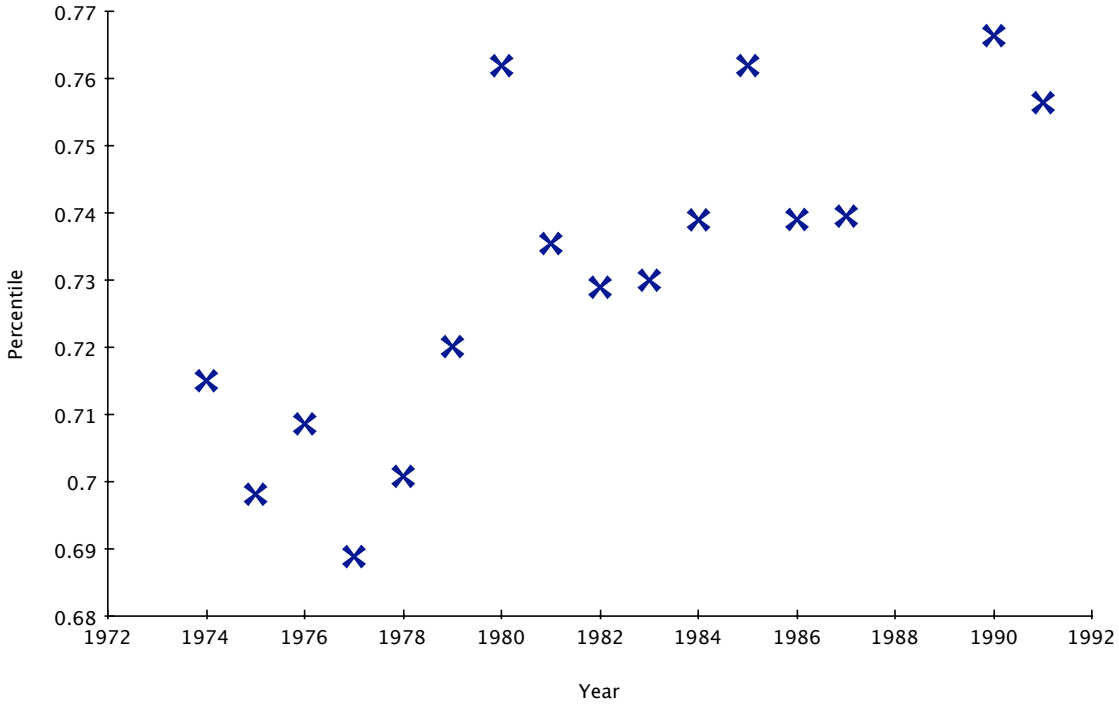
$$\begin{aligned} W_{r,t}^f &= \sum_h \frac{x_{h,r,t}}{\sum x_{h,r,t}} \left[ \alpha_{r,t}^f + \gamma^{ff} \ln \left( \frac{p_{r,t}^f}{p_{r,t}^n} \right) + \beta_{r,t}^f \ln \left( \frac{x_{h,r,t}}{P_{r,t}} \right) \right] \\ &= \alpha_{r,t}^f + \gamma^{ff} \ln \left( \frac{p_{r,t}^f}{p_{r,t}^n} \right) - \beta^f \ln P_{r,t} + \beta_{r,t}^f \sum_h \frac{x_{h,r,t} \ln \left( x_{h,r,t} \right)}{\sum x_{h,r,t}} \end{aligned} \tag{8}$$

Thus the CPI household has income equal to $\sum_h \frac{x_{h,r,t} \ln(x_{h,r,t})}{\sum x_{h,r,t}}$.

We compute where the income of the CPI household lies in the household income distribution for each year from 1974 to 1991 using the same PSID data as in Hamilton (2001a). These calculations are illustrated in Figure 3. Over this period, the aggregate shares used in the CPI correspond most closely to a household somewhere between the 68th and 77th percentile of the household income distribution.[7] The CPI is a plutocratic index and so measures changes in the cost of living for a fairly affluent household. As noted above, as a measure of this household's true cost of living, the CPI may be biased because of substitution, new goods, quality changes and other biases discussed in the literature.

[7] Deaton (1998) similarly finds that the U.S. CPI is usually around the 75th percentile of the household income distribution.

Figure 3: Income percentile of CPI household by Year




## 3.2 The Hamilton Household

Turning to the Hamilton household, we begin with the general definition (in logarithms) of the true cost of living index for a household $h$, in region $r$, and period $t$. We define reference utility for household $h$, $u_{h,0}$, to be the utility of that household in the base region ($r = 0$) and base period ($t = 0$), where it faces the price vector $\mathbf{p}_0 (= \mathbf{p}_{0,0})$:

$$\ln COLI_{h,r,t} = \ln C(u_{h,0}, \mathbf{p_{r,t}}) - \ln C(u_{h,0}, \mathbf{p_0}). \tag{9}$$

$C\left(u_{h,0}, \mathbf{p_{r,t}}\right)$ is the money cost of obtaining utility level $u_{h,0}$ at prices $\mathbf{p_{r,t}} = [p^f_{r,t}, p^n_{r,t}]$. This measures the resources necessary to maintain base-period utility at the new price vector $\mathbf{p_{r,t}}$ for this household, and the logarithm of the COLI measures the proportional change in required resources.

For AIDS preferences,

$$\begin{aligned} \ln C\left(u_{h,0}, \mathbf{p_{t,r}}\right) &= \ln a\left(\mathbf{p_{r,t}}\right) + b\left(\mathbf{p_{r,t}}\right) u_{h,0} \\ \ln a(\mathbf{p_{r,t}}) &= \alpha_0 + \sum_{k=f,n} \alpha_k \ln p^k_{r,t} + \sum_{k=f,n} \sum_{l=f,n} \gamma_{k,l} \ln p^k_{r,t} \ln p^l_{r,t} \\ b(\mathbf{p}_{r,t}) &= \prod_{k=f,n} (p^k_{r,t})^{\beta^k}, \end{aligned} \tag{10}$$

where $\sum_{k=f,n} \alpha_k = 1$ and $\sum_{k=f,n} \gamma_{k,l} = \sum_{l=f,n} \gamma_{k,l} = \sum_{k=f,n} \beta^k = 0$. Thus, given Hamilton's assumption

of AIDS preferences, the true COLI is log-linear in utility, and not independent of utility (or income) as Hamilton asserts.

The food share is:

$$w^f_{h,r,t} = \alpha^f + \sum_{k=f,n} \gamma^k \ln p^k_{r,t} + \beta^f \ln\left(\frac{x_{h,r,t}}{a(\mathbf{p_{r,t}})}\right) \tag{11}$$

$$\left[= \alpha^f + \sum_{k=f,n} \gamma^k \ln p^k_{r,t} + \beta^f \ln\left(\frac{x_{h,r,t}}{P_{r,t}}\right) \text{ in Hamilton's notation}\right]$$

Note in particular that $\ln a(\mathbf{p_{r,t}}) = \ln P_{r,t}$, which is the price index corresponding to the Hamilton household.

In the base period and region, prices are 1 (and log prices are zero), so that, $\ln a(\mathbf{p_0}) = \alpha_0$ and $b(\mathbf{p_0}) = 1$. As cost in period/region 0 must be equal to observed income in period/region 0, inverting gives utility in the base period and region as a function of observed income ($x_{h,0}$) :

$$u_{h,0} = \ln x_{h,0} - \alpha_0 \tag{12}$$

The cost of obtaining base period/region utility facing prices $\mathbf{p_{r,t}}$ is then,

$$\ln C(\mathbf{p_{r,t}}, u_{h,0}) = \ln P_{r,t} + b\left(\mathbf{p_{r,t}}\right)\left(\ln x_{h,0} - \alpha_0\right) \tag{13}$$

Equation 13 demonstrates that the Hamilton household has income level $\ln x_{h,0} = \alpha_0$. $P_{r,t}$ is the true cost of living index only at this level of income. The value of $\alpha_0$ is not identified by the Hamilton method. Apriori, there is no reason to think that $\alpha_0$ would also be the income level of the CPI household.

### 3.3 Decomposing the Hamilton Estimate

Solving Equation (13) for $\ln P_{r,t}$ and taking changes give:

$$\triangle \ln P_{r,t} = \triangle \ln C(\mathbf{p_{r,t}}, u_{h,0}) - \Delta b\left(\mathbf{p_{r,t}}\right)\left(\ln x_{h,0} - \alpha_0\right) \tag{14}$$

Also recall from Section 2 that $E_t$ is defined by $P_{r,t} = P_{r,0} CPI_{r,t}(1 + E_t)$. Taking logs, and rearranging this expression yields:

$$\Delta \ln P_{r,t} = \ln CPI_{r,t} + \ln(1 + E_t). \tag{15}$$

Combining Equations (14) and (15) and rearranging then gives:

$$\ln(1 + E_t) = [\triangle \ln C(\mathbf{p_{r,t}}, u_{h,0}) - \ln CPI_{r,t}] - \Delta b\,(\mathbf{p_{r,t}})\,(\ln x_{h,0} - \alpha_0)\,. \tag{16}$$

Evaluating this expression at the income level of the CPI household ($x_{h,0} = x^{CPI}$) gives a natural decomposition of the Hamilton correction:

$$-\ln(1 + E_t) = \ln CPI_{r,t} - \Delta \ln P_{r,t} = \left[\ln CPI_{r,t} - \triangle \ln C(\mathbf{p_{r,t}}, u(x^{CPI}, \mathbf{p}_{r,t}))\right] + \left[\Delta b(\mathbf{p_{r,t}})(\ln x^{CPI} - \alpha_0)\right]. \tag{17}$$

Hamilton's measure, $-\ln(1+E_t)$, captures the difference between the CPI ($\ln CPI_{r,t}$) and the price index ($\Delta \ln P_{r,t}$), and this measure can be further decomposed into actual CPI bias, which is given by the first term $\left[\ln CPI_{r,t} - \triangle \ln C(\mathbf{p_{r,t}}, u(x^{CPI}, \mathbf{p}_{r,t}))\right]$, and the part due to nonhomotheticity, which is given by the second term $\left[\Delta b(\mathbf{p_{r,t}})(\ln x^{CPI} - \alpha_0)\right]$.[8]

This is illustrated in Figure 4. The four panels illustrate four possible cases. In each panel, the logarithm of base period nominal income is on the horizontal axis and the CPI, changes in $P_t$ and changes in the true cost of living (all in logarithms) are measured on the vertical axis . The change in the logarithm of the true cost of living from period/region 0 to period/region $t, r$ ($\Delta \ln C$) is given by the diagonal line. It varies with the household's income in the base period. The slope of this diagonal line is:

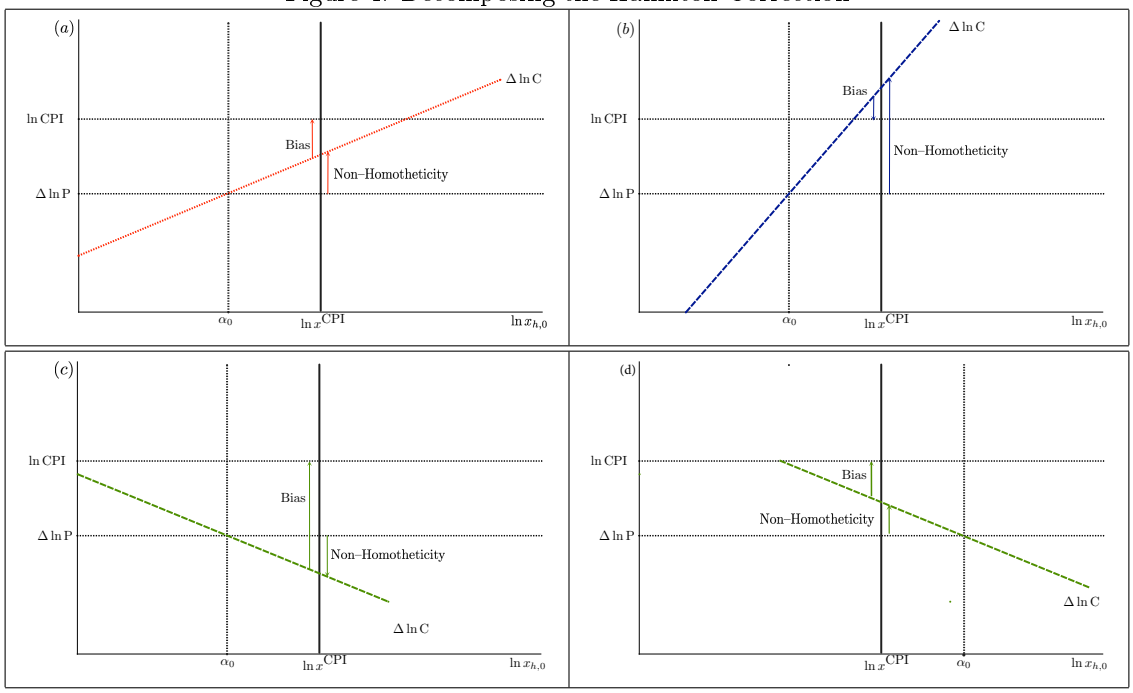
$$\Delta b(\mathbf{p_{r,t}}) = \left(\frac{p^f_{r,t}}{p^n_{r,t}}\right)^{\beta^f} - \left(\frac{p^f_0}{p^n_0}\right)^{\beta^f} = \left(\frac{p^f_{r,0}}{p^n_{r,0}}\right)^{\beta^f} \left[\left(\frac{(1 + \Pi^f_{r,t})}{(1 + \Pi^n_{r,t})}\right)^{\beta^f} - 1\right]. \tag{18}$$

Food is a necessity, so $\beta_f < 0$ . Thus, if the relative price of food is lower in region/period $r, t$ than in region/period 0, then $\Delta b(\mathbf{p_{r,t}}) > 0$ and the diagonal line slopes up (as in panels (a) and (b)). If the relative price of food is higher in period/region $r, t$ than in period/region 0 then $\Delta b(\mathbf{p_{r,t}}) < \mathbf{0}$ and the diagonal line slopes down (as in panel (c)). Whatever the slope, the diagonal line for $\Delta \ln C$ always passes through the intersection of the vertical line at $\ln x_{h,0} = \alpha_0$ and the horizontal line at $\triangle \ln P_{r,t}$. This is because the Hamilton method estimates the change in the cost of living at $\ln x_{h,0} = \alpha_0$.

The composition above can be seen at the vertical line at $\ln x_{h,0} = \ln x^{CPI}$. The difference between $\Delta \ln C$ and $\Delta \ln P$ is the part due to nonhomotheticity ($\Delta b(\mathbf{p_{r,t}})(\ln x^{CPI} - \alpha_0)$). In panel (a), this is positive, but

[8] Note that Hamilton's assumption that $E$ varies only with time can only be approximately correct as the two components of $E$ vary with both time and region.

Figure 4: Decomposing the Hamilton Correction




less than $\Delta \ln P_{r,t} - \ln CPI_{r,t}$. The remaining gap between $\Delta \ln C$ and $\ln CPI$ is then CPI bias (the extent to which the CPI miss-measures the cost of living increase for the CPI household), and this is also positive. In Panel (b), the difference between $\Delta \ln C$ and $\Delta \ln P$ (due to nonhomotheticity) is positive and *exceeds* $\Delta \ln P_{r,t} - \ln CPI_{r,t}$. This implies that the actual CPI bias must be negative, as indicated in the figure. In Panel (c), the difference between $\Delta \ln C$ and $\Delta \ln P$ (due to nonhomotheticity) is negative, so that actual CPI bias is greater than $\Delta \ln P_{r,t} - \ln CPI_{r,t}$. Note that this does not depend only on the slope, as can be seen by comparing Panel (c) and Panel (d). Panel (d) shows an expenditure $\Delta \ln C$ function the same slope as in Panel (c) but with $\alpha_0$ above $\ln x^{CPI}$. In this case the difference between $\Delta \ln C$ and $\Delta \ln P$ (nonhomotheticity) is positive but but less than $\Delta \ln P_{r,t} - \ln CPI_{r,t}$, so that CPI bias is positive but less than what the Hamilton Method would estimate. Of course, with the same slope, if $\alpha_0$were sufficiently large (far to the right) the effect of nonhomotheticity at $\ln x^{CPI}$ would exceed $\Delta \ln P_{r,t} - \ln CPI_{r,t}$, and CPI bias would be negative.

To quantify true CPI bias we need to empirically implement this decomposition of $\Delta \ln P_{r,t} - \ln CPI_{r,t} = \ln(1 + E_t)$ into the part due to nonhomotheticity and true CPI bias. This means calculating

$$\Delta b(\mathbf{p_{r,t}})(\ln x^{CPI} - \alpha_0) = \left(\frac{p^f_{r,0}}{p^n_{r,0}}\right)^{\beta^f} \left[\left(\frac{(1+\Pi^f_{r,t})}{(1+\Pi^n_{r,t})}\right)^{\beta^f} - 1\right](\ln x^{CPI} - \alpha_0) \tag{19}$$

where$\beta^f$ is the slope of the Engel curve and easily estimated. The calculation of $\ln x^{CPI}$ was discussed

above. Thus for a given region, $r$, expression 19 contains two unobservable quantities: $\alpha_0$ and the base period price ratio $\frac{p^f_{r,0}}{p^n_{r,0}}$. Note that in the base region $\frac{p^f_{r,0}}{p^n_{r,0}} = 1$. But the choice of base period is an arbitrary normalization of prices. We can re-normalize prices so that, essentially, each region is the base region in turn. This requires an adjustment to $\alpha_0$ (because $\alpha_0$ is the logarithm of the cost of base utility when prices are 1), but this is straight forward.[9] This means that if $\alpha_0$ could be identified, then the this decomposition of $\Delta \ln P_{r,t} - \ln CPI_{r,t} = \ln(1 + E_t)$ into the part due nonhomotheticity and true CPI bias could be quantified. We take up identification of $\alpha_0$ in the next section.

# 4 Identification

## 4.1 The Problem

The parameter $\alpha_0$ would in principle be identified if price levels were perfectly observed, but is not identified given the available data. To see this, first rewrite the two good demand system.

$$w_{h,r,t} = \alpha^f + \gamma^{ff}(\ln p^f_{r,t} - \ln p^n_{r,t}) + \beta^f \ln x_{h,r,t} - \beta^f \ln P_{r,t} \tag{20}$$

$$\ln P_{r,t} = \alpha_0 + \alpha^f \ln p^f_{r,t} + (1 - \alpha^f) \ln p^n_{r,t} + \frac{\gamma_{ff}}{2}(-2 \ln p^f_{r,t} \ln p^n_{r,t} + (\ln p^f_{r,t})^2 + (\ln p^n_{r,t})^2) \tag{21}$$

using equation (20) to eliminate $P_t$ from the food share equation (19) and express the share as a function of nominal income and additional price terms. Grouping terms gives:

$$w_{h,r,t} = \left\{\alpha^f - \beta^f \alpha_0\right\} + \left\{\gamma^{ff} - \beta^f \alpha^f\right\} \ln p^f_{r,t} - \left\{\gamma^{ff} + \beta^f (1 - \alpha^f)\right\} \ln p^n_{r,t} + \beta^f \ln x_{h,r,t} \tag{22}$$

$$-\frac{\beta^f \gamma^{ff}}{2}\left\{-2 \ln p^f_{r,t} \ln p^n_{r,t} + (\ln p^f_{r,t})^2 + (\ln p^n_{r,t})^2\right\}$$

Note that $\alpha_0$ is *not* identified from income variation alone but is in principle, identified from the non-linearity in price responses in this expanded form of the share equation.[10]

[9] Appendix A describes the calculations

[10] First, $\beta^f$ is identified by the nominal log income term and $\gamma^{ff}$ is identified by the square of the log prices term. Given $\beta^f$ and $\gamma^{ff}$ , $\alpha^f$ is identified from the log relative price term, and then $\alpha_0$ is identified from the constant (since we already know $\alpha^f$ and $\beta^f$ ). Such an estimate of $\alpha_0$ is likely to be imprecise because its a nonlinear function of estimated (reduced form) parameters in the share equation, and identification depends on the coefficient on the quadratic term in prices. These in turn would not be precisely estimated unless there was a great deal of relative price variation. For this reason, demand modelers often fix $\alpha_0$ even though it is in principle identified. See for example Banks et al. (1997) who set $\alpha_0$ just below smallest observed value of log income in the base year. A normalization does not help us here, as the resulting decomposition of the difference between the CPI and the Hamilton estimate would then be completely arbitrary.

However the Hamilton method is motivated by the observation that prices are *not* perfectly observed, and the goal is to estimate changes in the cost of living using limited data. Recall from Section 2.1 that there are two unobservables: base period price levels, and the error in observed inflation rates. The Hamilton methods proceeds by capturing these unobservables with, respectively, region and time dummies. However in the expanded form of the share equation (Equation (22)) the quadratic terms in price levels imply interactions between the unobservables. To captures these would require a full of set of interactions between time and region dummies. However, this would leave no variation in observed inflation rates with which to estimate the parameters of the equation. Consequently $\alpha_0$ is not identified by this approach and data.

Too see this, recall from Section 2.1 that Hamilton assumes the following error structure:

$$\ln p_{r,t}^k = \ln(1+\Pi_{r,t}^k) + \ln p_{r,0}^k + \ln(1+E_t^k) \tag{23}$$

where $\Pi_{r,t}^k$ is reported good specific inflation. Substituting this structure into Equation (22) above gives:

$$\begin{aligned} w_{h,r,t} \quad = \quad & \left\{\alpha^f - \beta^f \alpha_0\right\} + \left\{\gamma^{ff} - \beta^f \alpha^f\right\} \left[\ln(1+\Pi_{r,t}^f) + \ln p_{r,0}^f + \ln(1+E_t^f)\right] \\ & -\left\{\gamma^{ff} + \beta^f(1-\alpha^f)\right\} \left[\ln(1+\Pi_{r,t}^n) + \ln p_{r,0}^n + \ln(1+E_t^n)\right] + \beta^f \ln X_{h,r,t} \qquad (24) \\ & -\frac{\beta^f \gamma^{ff}}{2}\{-2\left[\ln(1+\Pi_{r,t}^f) + \ln p_{r,0}^f + \ln(1+E_t^f)\right]\left[\ln(1+\Pi_{r,t}^n) + \ln p_{r,0}^n + \ln(1+E_t^n)\right] \qquad (25) \\ & + \left[\ln(1+\Pi_{r,t}^f) + \ln p_{r,0}^f + \ln(1+E_t^f)\right]^2 + \left[\ln(1+\Pi_{r,t}^n) + \ln p_{r,0}^n + \ln(1+E_t^n)\right]^2\} \end{aligned}$$

As in Section 2.1, base period price differences ($\ln p_{r,0}^f$ and $\ln p_{r,0}^n$) can be captured with region dummies, and the time-specific errors captured by year dummies ($\ln(1+E_t^k)$). However in the two last lines of Equation (25), the time-specific errors interact with unobserved regional base period prices so that time-by-region interaction dummies are required.

## 4.2 Solutions

There are three ways to deal with non-identification of a parameter: bounding the parameter, bringing additional data to bear, and adding additional structure to the model. We pursue all three possibilities.

### 4.2.1 Bounds

Note that budget shares are naturally bounded between zero and one. At base period prices, the food budget share for household $h$ is $w_{h,r,0}^f = \alpha^f + \beta^f u_{h,r,0} = \alpha^f - \beta^f \alpha_0 + \beta^f \ln x_{h,r,0}$. Thus to ensure that budget shares are bounded between zero and one at zero utility requires $0 < \alpha^f < 1$ and to ensure that budget shares are bounded between zero and one at all observed incomes (again at base period prices)

requires $0 < \alpha^f - \beta^f \alpha_0 + \beta^f \ln x_{h,0} < 1 \ \forall x_{h,0} \in \left[x_0^{min}, x_0^{max}\right]$ where $x_0^{min}$ and $x_0^{max}$ are the minimum and maximum nominal incomes observed in the data (in the base year). Combining these conditions gives $\ln x_0^{\min} + \frac{1}{\beta^f} < \alpha_0 < \ln x_0^{\max} - \frac{1}{\beta^f}$. Food is a necessity so $\beta^f$ is negative and these are bounds are larger than the support of household income in the base period, and so not particularly informative. We nevertheless compute them for the Hamilton data and report them in our application below.

#### 4.2.2 Better data

Since Hamilton conducted his study, information of regional price levels in the U.S. has become available through the important work of Bettina Aten (Aten, 2008). With these data, $\ln p_{r,0}^f$ and $\ln p_{r,0}^n$ are observed eliminating the need for regional dummies. Define the measured log price level as $\ln \tilde{p}_{r,t}^k = \ln(1+\Pi_{r,t}^k) + \ln p_{r,0}^k$. so that from Equation (2) we have $\ln p_{r,t}^k = \ln \tilde{p}_{r,t}^k + \ln(1 + E_t^k)$. With this, and maintaining the Hamilton assumption that $\ln(1 + E_t^f) - \ln(1 + E_t^n) = 0$, we can derive an estimable equation of the form:

$$\begin{aligned} w_{h,r,t} &= \left\{\alpha^f - \beta^f \alpha_0\right\} + \beta^f \ln x_{h,r,t} + \left\{\gamma^{ff} - \beta^f \alpha^f\right\} \ln p_{r,t}^{*f} - \left\{\gamma^{ff} + \beta^f (1 - \alpha^f)\right\} \ln p_{r,t}^{*n} \\ &- \frac{\beta^f \gamma^{ff}}{2} \left\{ -2 \ln p_{r,t}^{*f} \ln p_{r,t}^{*n} + \left[\ln p_{r,t}^{*f}\right]^2 + \left[\ln p_{r,t}^{*n}\right]^2 \right\} \\ &- \beta^f \ln(1 + E_t^n) \end{aligned} \tag{26}$$

The terms in the first two lines of this array are observables. The term in the third line of this array can be captured just by time dummies. $\beta^f$ is identified by the nominal log income term and $\gamma^{ff}$ is identified by the second order terms in log prices. Given $\beta^f$ and $\gamma^{ff}$ , $\alpha^f$ is identified from the log relative price terms, and then $\alpha_0$ is identified from the constant (since we already know $\alpha^f$ and $\beta^f$). We have implemented this on Hamilton's data augmented with Aten's regional price levels. Unfortunately, the resulting estimate of $\alpha_0$ has a confidence interval which is as large as the theoretical bounds described above. This is because identification rests on base-period regional variation in relative prices, which in this context is not large. In other contexts, regional price level variation may be more helpful.

#### 4.2.3 Preference Restriction

Finally, identification can be achieved in this setting by imposing restrictions on preferences. An obvious candidate is to set $\gamma^{ff} = 0$. These preferences rule out relative price effects on the budget shares if we condition on $P_{r,t}$:

$$w^f_{h,r,t} = \alpha^f + \beta^f \ln\left(\frac{x_{h,r,t}}{P_{r,t}}\right) \tag{27}$$

Some authors without access to regional price variation have estimated exactly this model when implementing the Hamilton method (see e.g., Beatty and Larsen (2005)), and researchers – including Hamilton – estimating the more general formulation often cannot reject the restriction that $\gamma^{ff} = 0$ (see e.g., Hamilton (2001a) where the coefficient estimate is small and insignificant or Costa (2001) where the effect is negative in one period and positive in another.))

Combining this restriction with the error structure in Equation (2), the expanded form of the share equation (Equation (22)), and Hamilton assumption that $\ln(1+E^f_t) - \ln(1+E^n_t) = 0$, we get:

$$\begin{aligned} w^f_{h,r,t} &= \left\{\alpha^f - \beta^f\alpha_0\right\} + \beta^f \ln x_{h,r,t} + \beta^f\alpha^f \ln\left(\frac{1+\Pi^n_{r,t}}{1+\Pi^f_{r,t}}\right) - \beta^f \ln(1+\Pi^n_{r,}) - \beta^f \ln(1+E^n_t) \\ &\quad - \left[\beta^f\alpha^f \ln p^f_{r,0} + \beta^f(1-\alpha^f)\ln p^n_{r,0}\right] \end{aligned} \tag{28}$$

$$w^f_{h,r,t} = \left\{\alpha^f - \beta^f\alpha_0\right\} + \beta^f \ln\left(\frac{x_{h,r,t}}{1+\Pi^n_{r,t}}\right) + \beta^f\alpha^f \ln\left(\frac{1+\Pi^n_{r,t}}{1+\Pi^f_{r,t}}\right) - \sum_t \tilde{\delta}^t D^t + \sum_r \tilde{\delta}^r D^r \tag{29}$$

where the coefficients on the time dummies are $\tilde{\delta}^t = \beta^f \ln(1+E^n_t)$ and the coefficients on the region dummies are $\tilde{\delta}^r = -\left[\beta^f\alpha^f \ln p^f_{r,0} + \beta^f(1-\alpha^f)\ln p^n_{r,0}\right]$.

To implement the full decomposition described in Section 3.3, we first estimate Equation (6) (the specification that Hamilton and subsequent literature estimate) but impose the restriction that $\gamma^{ff} = 0$,

$$\begin{aligned} w^f_{h,r,t} &= \alpha^f + \beta^f \ln\left(\frac{x_{h,r,t}}{\text{CPI}_{r,t}}\right) - \beta^f \ln(1+E_t) + \sum_r \delta^r D^r \\ &= \alpha^f + \beta^f \ln\left(\frac{x_{h,r,t}}{\text{CPI}_{r,t}}\right) + \sum_t \delta^t D^t + \sum_r \delta^r D^r. \end{aligned} \tag{30}$$

We recover an estimate of $E_t$ from $\delta^t = -\beta^f \ln(1+E_t)$. Then, estimate Equation (29) to get an estimate of $\alpha_0$ and following Section 3.3, calculate part of $E_t$ that is due to non-homotheticity as

$$\Delta b(\mathbf{p_{t,r}})(\ln x^{CPI} - \alpha_0) = \left(\frac{p^f_{r,0}}{p^n_{r,0}}\right)^{\beta^f} \left[\left(\frac{(1+\Pi^f_{r,t})}{(1+\Pi^n_{r,t})}\right)^{\beta^f} - 1\right](\ln x^{CPI} - \alpha_0) \tag{31}$$

In our application, in the next Section, we implement this procedure on the Hamilton data.

# 5 Application

In this section we apply the same PSID data as Hamilton (2001a) to estimate the log income level of the Hamilton household as well as calculating the bounds on that level. This then allows us to calculate the change in cost of living for the CPI household, and hence CPI bias, as described in the previous section.

Figure (5) displays our estimate of the log income level of the Hamilton household (that is, $\alpha_0$), and the CPI household, against the full distribution of log income.[11] We find that the Hamilton household is much poorer than the the CPI household, and indeed, has a log income level below the minimum log income in the the sample. This means that this household will be much more sensitive to price changes in food than the CPI household.

Figure (6) adds the theoretical bounds, and shows that the range of log income levels that the Hamilton household could representis very wide. That the log income level of the Hamilton household can be outside the observed log income distribution (as we find for Hamilton's data) and indeed potentially very far below or above the observed log income distribution (as the bounds illustrate), has important implications. In response to an very early version of this paper (Beatty and Crossley, 2012), Nakamura et al. (2016) propose to check the robustness of Engel Curve-based inflation estimates (using the Hamilton Method) by calculating exact price indices for different parts of the observed income distribution. They employ Divisia indices from Feenstra and Reinsdorf (2000) which are exact for the Almost Ideal Demand System (for one particular path of prices). These indices require only data on initial and final period expenditure shares and prices for the relevant group. Nakamura et al. (2016) argue that while inflation rates so calculated are different for different income groups in China, the differences are not large. This is a very elegant way of assessing slope of the change in cost of living (the line labeled $\Delta \ln C$ in Figure 4). But, the problem is that this proceedure is only informative about the slope of this line, and not its location. If the log income level of the Hamilton household is very far below, or very far above, the log income level of the CPI household, then even a modest slope can lead to significant differences between the Hamilton household and the CPI

[11] Here we use the 1985 New York income distribution and reference price vector, using the method for cycling through reference price vectors described in Appendix A

household in the change in cost of living. This means that calculations such as those reported by Nakamura, Steinsson and Liu cannot rule out nonhomotheticity as an important component of the difference between Engel Curve-based inflation estimates and other estimates. The role of non-homotheticiy can only be assessed with information about *both* the slope of the change in cost of living and the location (in log-income space) of the Hamilton household (again see figure 4). The slope of the change in cost of living depends on movements in relative prices and the slope of the food Engel curve and it is fairly straightforward to calculate it directly or to approximate it using methods such as those employed by Nakamura et al. (2016). Information about location (in log-income space) of the Hamilton household is more difficult to obtain, but is equally important to assessing the implications of nonhomotheticity.

Figure 7 displays, in the first column, estimates of national CPI bias using the Hamilton method on Hamilton's data. In the second column we report estimates of CPI bias using our method which corrects for the effects of nonhomotheticity. In these data the log income level of the Hamilton Household is far below the log income level of the CPI household. In this time period, food prices rose more quickly than non-food prices. Together these facts imply the true change in cost-of-living that was larger for the Hamilton Household than the for the CPI Household. This in turn implies that the difference between the Hamilton Household and the CPI understates the bias in the CPI *as a measure of the change in cost-of-living experienced by the CPI household.* The scenario that turns out to be relevant for these data and this period is captured by Panel (c) in Figure 4. Our corrected estimates of CPI bias are therefore larger than Hamilton's: by the end of the period the estimate of cumulative CPI bias is 10 percentage points (or 21 percent) higher.

A limitation of these calculations is that $\alpha_0$ (the location of the Hamilton household in log-income space) is not precisely estimated (in contrast, the slope of the change in cost of living is quite precisely estimated). Figure 8 illustrates the implications of the estimated confidence interval for $\alpha_0$. The top-left panel corresponds to the point estimate of $\alpha_0$ (and hence to the numbers in Figure 7). The bottom-left panel displays the lower bound of the confidence interval for $\alpha_0$, which places the Hamilton household even farther away from the CPI household in log-income space. The upper bound of the confidence interval for $\alpha_0$ places the Hamilton household just below the CPI household in log-income space, and the consequences of this are displayed in the bottom-right panel of Figure 8. In each panel, the red line tracks the effect of nonhomotheticity: the difference between the true change in cost-of-living experienced by the Hamilton and CPI households. The CPI household is more affluent than the Hamilton household, and food prices rose more quickly than nonfood prices over this period, so the red line is everywhere below zero, though at the the upper bound of the confidence interval $\alpha_0$, where the Hamilton household is just below CPI household in log-income space, the effect is very small. The blue line tracks the Hamilton estimate of CPI bias which, as described above, is in fact the sum of genuine CPI bias and the effect of nonhomotheticity. The green line tracks genuine

CPI bias, having adjusted the Hamilton estimate for non-homotheticity using the method proposed in this paper. As effect of nonhomotheticity, captured by the red line, is everywhere negative, the green line must lie everywhere above the blue line. At the point estimate for $\alpha_0$, the green line is significantly above the blue line (with a gap of 10 percentage points at the end of the period). At the lower bound of the confidence interval for $\alpha_0$, the gap is much larger, suggesting that, in failing to account for nonhomotheticity, Hamilton very significantly underestimated the degree of bias in the CPI. At the upper bound of the confidence interval the gap our estimate is only slightly above Hamilton's. Again, the location of the Hamilton household in log-income space ($\alpha_0$) is difficult to estimate precisely. However, there is sufficient information in these data to say bias in the CPI in this period was at least as large as Hamilton's estimates, and may have been significantly larger.

These calculations are based on the idea that if we have Hamilton estimate of the change in cost-of-living for the Hamilton household ($\Delta \ln P$), the location of the Hamilton household in log-income space ($\alpha_0$), and the slope of the cost-of-living function, we can calculate the change in the cost-of-living at the log-income level of the CPI household. But of course the calculation can be made for *any* income level. So we've actually shown that the assumptions typically made to use the Hamilton method are sufficient to recover changes in the cost of living right across the income distribution. This means we can study how inflation differs between rich and poor, or how real income has evolved at different points in the income distribution. As an illustration, we consider the development of the median income over time. Concern with stagnant real incomes was the original motivation for Nakamura (1995), and the lack of growth in median income continues to receive considerable attention in the literature.

The results are striking. Figure (9) shows the trend in median income estimated by the different methods. First, the CPI deflated median income (in blue) shows a rather modest increase from about \$17,000 to about \$19,000 from 1974 to 1991. The CPI however, is the wrong deflator both because the CPI best captures a household that is somewhat more affluent than the median income household, and because the CPI suffers from price index problems such as substitution bias, new goods and quality change. The Hamilton corrected CPI shows a much larger increase to about \$26,000. This is the red line in Figure (9). However, this is also the wrong deflator because it captures the true change in cost-of-living for a very poor household and that household experienced a larger change in cost-of-living the the median income household. Finally, our method, in green, calculates the true change in cost-of-living for the median income household and uses this to deflate. This shows an even large real increase to about \$27,000. The implied annual growth rates are 0.7%, 2.8% and 3.1%.

Figure 5: Hamilton and CPI households

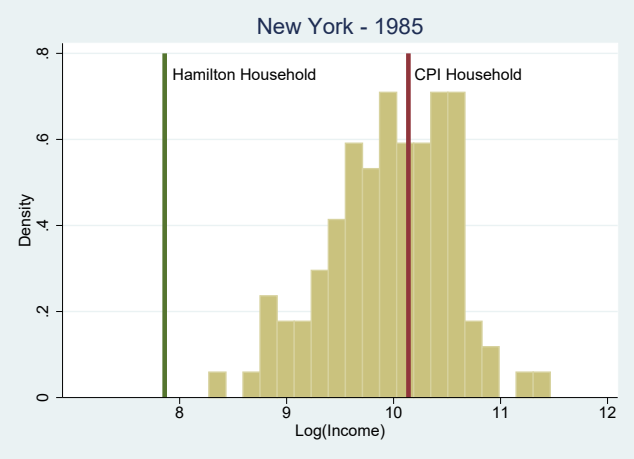



Figure 6: Bounds

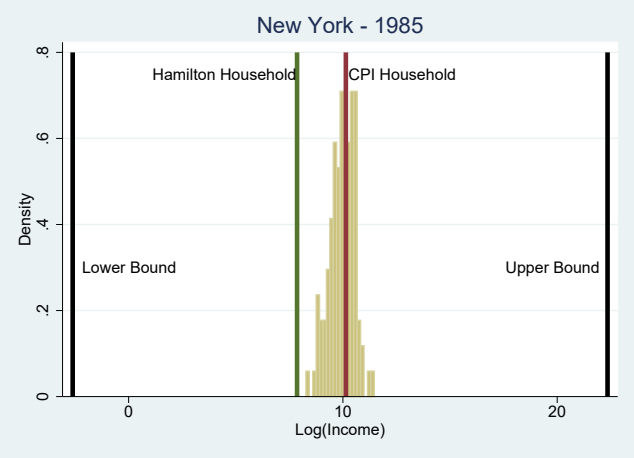



Figure 7: Results

| Year | Hamilton | Revised |
|---|---|---|
| 1975 | 5.4 | 5.5 |
| 1976 | 15.0 | 16.5 |
| 1977 | 12.6 | 14.2 |
| 1978 | 16.6 | 17.4 |
| 1979 | 22.4 | 23.8 |
| 1980 | 31.3 | 35.4 |
| 1981 | 28.5 | 33.7 |
| 1982 | 33.0 | 39.8 |
| 1983 | 37.7 | 46.7 |
| 1984 | 48.9 | 59.8 |
| 1985 | 42.4 | 53.8 |
| 1986 | 42.3 | 52.7 |
| 1987 | 44.9 | 55.6 |
| 1990 | 49.0 | 58.5 |
| 1991 | 47.9 | 57.9 |

Figure 8: Precision

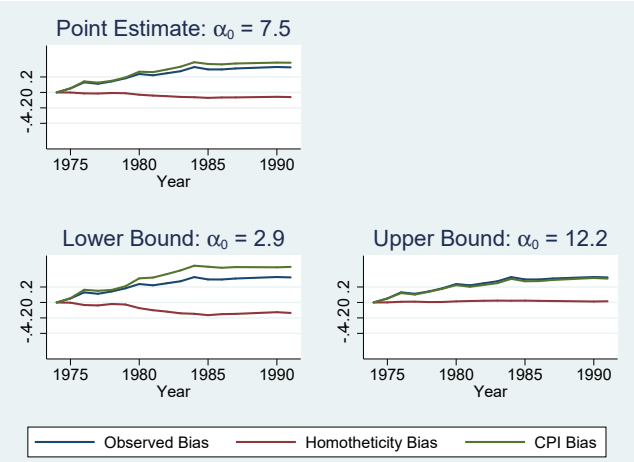
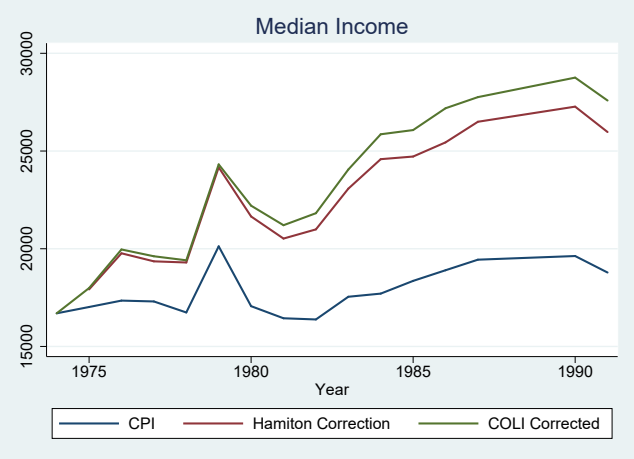



Figure 9: Median Income Growth



# 6 Summary and Conclusion

Hamilton method as originally proposed and much used confounds genuine CPI bias with differences in consumption baskets across the income distribution. We have shown that it is possible to disentangle these two things, and to make the method internally consistent. For the data and period that Hamilton studies, this leads to a larger estimate of CPI bias: and additional upward bias of 10 percentage points over the period from 1974 to 1991. We have also shown that the usual assumptions of this method are actually sufficient to recover changes in the cost of living at *any* income level. Our estimate of changes in the cost of living at median income over this period imply significant median income growth from 1974 to 1991, where deflating instead by the CPI implies none.

## References


Ingvild Almås. International Income Inequality: Measuring PPP Bias by Estimating Engel Curves for Food. *The American Economic Review*, 102(2):1093–1117, 2012.

Bettina H. Aten. Estimates of State and Metropolitan Price Parities for Consumption Goods and Services in the United States, 2005. *Washington DC: Bureau of Economic Analysis*, 2008.

James Banks, Richard Blundell, and Arthur Lewbel. Quadratic Engel Curves and Consumer Demand. *The Review of Economics and Statistics*, 79(4):527–539, 1997.

Garry F. Barrett and Matthew Brzozowski. Using Engel Curves to Estimate the Bias in the Australian CPI. *Economic Record*, 86(272):1–14, 2010.

Timothy K. M. Beatty and Thomas F. Crossley. Lost in Translation: What Do Engel Curves Tell Us About the Cost of Living. Technical report, 2012.

Timothy K.M. Beatty and Erling Røed Larsen. Using Engel Curves to Estimate Bias in the Canadian CPI as a Cost of Living Index. *Canadian Journal of Economics*, 38(2):482–499, 2005.

Laura Blow. Explaining trends in household spending. Technical Report 03/06, IFS Working Papers, Institute for Fiscal Studies (IFS), 2003.

Michael J. Boskin, Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale W. Jorgenson. Consumer Prices, the Consumer Price Index, and the Cost of Living. *The Journal of Economic Perspectives*, 12(1): 3–26, 1998.

Matthew Brzozowski. Does One Size Fit All? The CPI and Canadian Seniors. *Canadian Public Policy*, 32 (4):387–411, 2006.

Somnath Chattopadhyay. District level poverty estimation: a spatial approach. *Economics Bulletin*, 30(4): 2962–2977, 2010.

Chul Chung, John Gibson, and Bonggeun Kim. CPI Mismeasurements and Their Impacts on Economic Management in Korea. *Asian Economic Papers*, 9(1):1–15, 2010.

Marie-Émilie Clerc and Élise Coudin. Does the CPI Reflect Changes in the Cost of Living in France? What the Analysis of Engel Curves Tells Us. *Economie et Statistique*, 2011.

Dipankor Coondoo, Amita Majumder, and Somnath Chattopadhyay. Estimating Spatial Consumer Price Indices Through Engel Curve Analysis. *Review of Income and Wealth*, 57(1):138–155, 2011.

Dora Costa. Estimating Real Income in the United States from 1888 to 1994: Correcting CPI Bias Using Engel Curves. *Journal of Political Economy*, 109(6):1288–1310, 2001.

Irineu de Carvalho Filho and Marcos Chamon. The myth of post-reform income stagnation: Evidence from Brazil and Mexico. *Journal of Development Economics*, 97(2):368–386, 2012.

Angus Deaton. Getting Prices Right: What Should be Done? *The Journal of Economic Perspectives*, 12 (1):37–46, 1998.

Angus Deaton and Olivier Dupriez. Purchasing Power Parity Exchange Rates for the Global Poor. *American Economic Journal: Applied Economics*, 3(2):137–66, April 2011.

Angus Deaton and Alan Heston. Understanding PPPs and PPP-Based National Accounts. *American Economic Journal: Macroeconomics*, 2(4):1–35, October 2010.

Angus Deaton and John Muellbauer. An Almost Ideal Demand System. *The American Economic Review*, 70(3):312–326, 1980.

W. Erwin Diewert. *The New Palgrave: A Dictionary of Economics*, volume 2, chapter Index Numbers, pages 767–780. The Macmillian Press, London, 1987.

W. Erwin Diewert. *Essays in index number theory*, chapter The early history of price index research, pages 33–65. North Holland, 1993.

W. Erwin Diewert. Axiomatic and Economic Approaches to Elementary Price Indexes. Working Paper 5104, National Bureau of Economic Research, May 1995.

Ernst Engel. Die Productions- und Consumtionsverhaltnisse des Königreichs Sachsen. *Zietscrift des Statischen Bureaus des Koniglich Sachischen Ministeriums des Innern*, 3, 1857.

Ernst Engel. *Die Lebenskosten belgischer Arbeiter-Familien früher und jetzt.* C. Heinrich, 1895.

Robert C. Feenstra and Marshall B. Reinsdorf. An exact price index for the almost ideal demand system. *Economics Letters*, 66(2):159–162, 2000.

John Gibson and Grant Scobie. Using Engel Curves to estimate CPI bias in a small, open, inflation-targeting economy. *Applied Financial Economics*, 20(17):1327–1335, 2010.

John Gibson, Steven Stillman, and Trinh Le. CPI Bias and Real Living Standards in Russia During the Transition. *Journal of Development Economics*, 87(1):140–160, 2008.

Bruce W. Hamilton. The True Cost of Living: 1974-1991. Working paper 395, John Hopkins University, Department of Economics, 1998.

Bruce W. Hamilton. Using Engel's Law to Estimate CPI Bias. *The American Economic Review*, 91(3): 619–630, 2001a.

Bruce W. Hamilton. Black-White Differences in Inflation: 1974-1991. *Journal of Urban Economics*, 50(1): 77–96, 2001b.

Jerry Hausman. Sources of Bias and Solutions to Bias in the Consumer Price Index. *The Journal of Economic Perspectives*, 17(1):23–44, 2003.

Jerry Hausman and Ephraim Leibtag. *Price Index Concept and Measurement*, chapter CPI Bias from Supercenters: Does the BLS Know that Wal-Mart Exists?, pages 203–231. University of Chicago Press, 2009.

Robert J. Hill. Constructing Price Indexes across Space and Time: The Case of the European Union. *The American Economic Review*, 94(5):1379–1410, 2004.

Erling Røed Larsen. Does the CPI Mirror the Cost of Living? Engel's Law Suggests Not in Norway. *Scandinavian Journal of Economics*, 109(1):177–195, 2007.

Trevon D. Logan. Are Engel Curve Estimates of CPI Bias Biased? *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 42(3):97–110, 2009.

Bruce D. Meyer and James X. Sullivan. Five Decades of Consumption and Income Poverty. Working Paper 14827, National Bureau of Economic Research, March 2009.

John Muellbauer. Community Preferences and the Representative Consumer. *Econometrica*, 44(5):979–999, 1976.

Emi Nakamura, Jón Steinsson, and Miao Liu. Are Chinese Growth and Inflation Too Smooth? Evidence from Engel Curves. *American Economic Journal (forthcoming)*, 2016.

Leonard I. Nakamura. Is U.S. economic performance really that bad? Working Papers 95-21, Federal Reserve Bank of Philadelphia, 1995.

Peter Neary. Rationalizing the Penn World Table: True Multilateral Indices for International Comparisons of Real Income. *The American Economic Review*, 94(5):1411–1428, 2004.

J. L. Nicholson. Whose Cost of Living? *Journal of the Royal Statistical Society. Series A (General)*, 138(4): 540–542, 1975.

Susan Olivia and John Gibson. Using Engel curves to measure CPI bias for Indonesia. Working Paper 13, Monash Business School, Department of Econometrics and Business Statistics, 2012.

S. J. Prais. Whose Cost of Living? *The Review of Economic Studies*, 26(2):126–134, 1959.

Martin Ravallion. Toward Better Global Poverty Measures. Working Paper 417, Center for Global Development, 2015.

Marshall Reinsdorf. *Price Measurements and Their Uses*, chapter The Effect of Outlet Price Differentials on the U.S. Consumer Price Index, pages 227 – 258. National Bureau of Economic Research, 1993.

Yonghong Xu and Wuyi Zeng. Estimation of CPI Bias with Chinese City Statistical Data. *Statistical Research*, 4:002, 2009.

# Appendix A: Renormalizing Prices

The calculation of the nonhomotheticity bias is done for the reference region. However, we can choose any of the R regions to make the reference price vector. To calculate the CPI bias at every region specific price vector, we do not have to re-estimate R times because we can work out quickly from the one set of estimates what the estimates would be if we left out a different region dummy.

If we leave out a region dummy, all that changes is the constant. This in turn changes our estimate of $\alpha_0$. This makes sense because $\alpha_0$ is the log cost of zero utility when prices are 1. If we redefine which region's prices are one (equivalently: redefine quantities) then we change $\alpha_0$.

Our "long form" is

$w^f_{h,r,t} = \{\alpha^f - \beta^f \alpha^1_0\} + \beta^f \ln\left(\frac{x_{h,r,t}}{1+\Pi^n_{r,t}}\right) + \beta^f \alpha^f \ln\left(\frac{1+\Pi^n_{r,t}}{1+\Pi^f_{r,t}}\right) - \sum_{t=2}^{T} \tilde{\delta}^t D^t + \sum_{r=2}^{R} \tilde{\delta}^r D^r$

with reduced form

$w^f_{h,r,t} = A^1 + Z_{r,t}\theta + \sum_{r=2}^{R} \tilde{\delta}^r D^r$

where $A^1$ is the constant and $Z_{r,t}\theta$ contains the income, relative price and time terms. Note that we are now indexing the the constant and $\alpha^r_0$ by the normalization (by which region is the ommitted region, or equivalently, which region is the reference region and has prices set to 1 in the base period).

Note that $\alpha^1_0 = -\frac{A^1 - \alpha^r}{\beta^f}$ .

The intercept if we left out region r instead would just be

$A^r = A^1 + \tilde{\delta}^r$

So $\alpha^r_0 = -\frac{A^r - \alpha^r}{\beta^f} = -\frac{A^1 + \tilde{\delta}^r - \alpha^r}{\beta^f} = -\frac{\alpha^f - \beta^f \alpha^1_0 + \tilde{\delta}^r - \alpha^r}{\beta^f} = \alpha^1_0 - \frac{\tilde{\delta}^r}{\beta^f}$.

So the proceedure would be as follows:

1. Estimate the long form as before leaving out the Region 1 dummy and use the ratio of inflation rates to do the bias/non-homotheticity calculation for that region. $\Delta b(\mathbf{p_{1,t}}))(\ln x^{CPI} - \alpha^1_0) = \left[\left(\frac{(1+\Pi^f_{1,t})}{(1+\Pi^n_{1,t})}\right)^{\beta^f} - 1\right](\ln x^{CPI} - \alpha^1_0)$.

2. Then do the calculation for each of other regions, again using just the inflation rates (so that each time we are renormalizing prices to be one in that region in the base period) but noting that renormalization of prices means we have to adjust $\alpha_0$ for each region:

$$\begin{aligned} \Delta b(\mathbf{p_{r,t}}))(\ln x^{CPI} - \alpha^r_0) &= \left[\left(\frac{(1+\Pi^f_{r,t})}{(1+\Pi^n_{r,t})}\right)^{\beta^f} - 1\right](\ln x^{CPI} - \alpha^r_0) \\ &= \left[\left(\frac{(1+\Pi^f_{r,t})}{(1+\Pi^n_{r,t})}\right)^{\beta^f} - 1\right](\ln x^{CPI} - \alpha^1_0 + \frac{\tilde{\delta}^r}{\beta^f}.) \end{aligned}$$

3. We then average over the bias from the different regions. Note that structurally

$$\delta^r = \beta^f \alpha^f \ln p^f_{r,0} + \beta^f (1 - \alpha^f) \ln p^n_{r,0}$$

so that:

$$\alpha^r_0 = \alpha^1_0 - (\alpha^f \ln p^f_{r,0} + \beta^f (1 - \alpha^f) \ln p^n_{r,0})$$

4. Finally, so see why this makes sense, note that a our identifying assumtipn is that $\gamma_{k,l} = 0$ so that

$$\ln a(\mathbf{p_{r,t}}) = \alpha_0 + \sum_{k=f,n} \alpha_k \ln p^k_{r,t}$$

$\alpha_0$ is the value of the price index when prices are one. If we renormalize prices we have to adjust $\alpha_0$ by $\sum_{k=f,n} \alpha_k \ln p^k_{r,t}$.