ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Almås, Ingvild; Beatty, Timothy K.M.; Crossley, Thomas F.

Working Paper

Lost in Translation: What do Engel Curves Tell us about the Cost of Living?

CESifo Working Paper, No. 6886

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Almås, Ingvild; Beatty, Timothy K.M.; Crossley, Thomas F. (2018) : Lost in Translation: What do Engel Curves Tell us about the Cost of Living?, CESifo Working Paper, No. 6886, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/176905

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

CESifo WORKING PAPERS

6886
2018

January 2018

Lost in Translation: What do Engel Curves Tell us about the Cost of Living?

Ingvild Almås, Timothy K.M. Beatty, Thomas F. Crossley

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6886*
Category 12: Empirical and Theoretical Methods

Lost in Translation: What Do Engel Curves Tell Us About the Cost of Living?

Abstract

The Hamilton method for estimating CPI bias is simple, intuitive, and has been widely adopted. We show that the method confiates CPI bias with variation in cost-of-living across income levels. Assuming a single price index across the income distribution is inconsistent with the downward sloping Engel curves that are necessary to implement the method. We develop and implement the Translated Engel curve (TEC) method that disentangles genuine CPI bias from differences caused by comparing changes in the cost of living across different income levels - non-homotheticity. The TEC method gives substantially different estimates of CPI bias prior to major reforms to the CPI in 1999 (post-Boskin), but both methods suggest very little CPI bias thereafter.

JEL-Codes: C430, C820, D120, D310, E310.

Keywords: Engel curves, current price index, cost of living.

Ingvild Almås
Institute for International Economic Studies
Stockholm University
Sweden – 10691 Stockholm
ingvild.almas@iies.su.se

Timothy K.M. Beatty
University of California, Davis
Agricultural and Resource Economics
USA - Davis, CA 95616
tbeatty@ucdavis.edu

Thomas F. Crossley
University of Essex
Department of Economics
United Kingdom - Colchester CO4 3SQ
tcross@essex.ac.uk

January 26, 2018
Almås acknowledges valuable support from Vetenskapsrådet (the Swedish Research Council), ESOP, University of Oslo, The Choice Lab, Norwegian School of Economics, and the "young research talents'" program of the Norwegian Research Council. Crossley gratefully acknowledges support for this research from the Economic and Social Research Council (ESRC) through the Research Centre on Micro-Social Change (MiSoC) at the University of Essex (grant number ES/L009153/1) and Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies (grant number ES/M010147/1.) We would like to thank Orazio Attansio, Garry Barrett, Richard Blundell, Chris Carroll, Angus Deaton, Erwin Diewert, Jeff LaFrance, Valerie Lechene, Ian Preston, David Strömberg, Guglielmo Weber, participants in a number of seminars and especially Bruce Hamilton for helpful comments, and Peter Levell for assistance with the CE data. 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, policy makers, and international institutions. 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) and the purchasing power parities implicit in the Penn World Tables, are commonly used 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); Bils and Klenow (2001); Greenlees and McClelland (2011); Reinsdorf (1998, 1993); Boskin and Jorgenson (1997); Boskin et al. (1998); Triplett (2001); Hausman (2003); Lebow and Rudd (2003); Schultze (2003); Nordhaus (1998); Pollak (1998); Hausman and Leibtag (2009) for biases in the CPI, Broda and Weinstein (2010) for discussion of measurement issues in the underlying BLS price series, and Neary (2004); Hill (2004); Deaton and Heston (2010a); Almås (2012) for corresponding concerns with PPP calculations.

In this paper we reconsider a popular approach to this problem: the "Hamilton Method" proposed by Bruce Hamilton (Hamilton, 1998, 2001a). Stagnant deflated incomes 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, 1857a, 1895a)) 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 deflated 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 Engel curves 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, see 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); Nakamura et al. (2016); Sacerdote (2017). A corresponding approach (referred to as the "Engel method") has also recently been applied to spatial comparisons across and within countries, see Almås (2012); Almås and Johnsen (2012); Chattopadhyay (2010); Coondoo et al. (2011). In a series of papers discussing the time path of poverty the U.S., Bruce Meyer and James Sullivan refer to the Hamilton method as important evidence of upward bias in the CPI (Meyer and Sullivan, 2009, 2011, 2012), and the World Bank has recently applied the Hamilton method in 16 African countries to assess CPI bias and reassess progress on poverty reduction (Dabalen et al., 2016).

In this paper, we begin by showing that the Hamilton method is internally inconsistent. We then develop a consistent method in the spirit of Hamilton which we refer to as the Translated Engel curve (TEC) method. We apply the TEC method to Hamilton's data, and then to more recent U.S. data, for the period 1990 to 2014.

The inconsistency in the original Hamilton method arises because, 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 real income, and so are also independent of any bias in, or corrections to, the deflator. In other words, 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 will 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. Put simply, the Hamilton method conflates CPI bias with variation in inflation across income levels.

The TEC method we develop overcomes this problem by determining (theoretically and empirically) *whose* cost of living is being measured by the Hamilton method and then making a theoretically consistent adjustment to obtain the cost of living for any point in the income distribution.

The TEC method can used to assess bias in the CPI. To do this we 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. The TEC method then gives us the change in cost of living at this point, from which we can calculate CPI bias. This allows us to separate genuine CPI bias from the differences in inflation at the different points in the income distribution. For the period Hamilton studied, the TEC method reveals significant upward bias in the CPI, but the estimated bias is smaller than that obtained by Hamilton's method.

Applying the TEC method to more recent U.S. data yields two key findings. First, for the 1990s, we find that the Hamilton method overstates genuine CPI bias by almost 40%. Second, our estimates imply very little CPI bias for the period 2000-2014. We speculate that this is because of improvements to the CPI undertaken by the BLS in 1999, in the wake of the Boskin Commission report. The TEC method and the Hamilton method show very similar patterns in this latter period because relative price changes in this period do not systematically favour the poor or the rich.

We also use the TEC method and U.S. data to calculate illustrative trends in deflated consumption for different quantiles of the consumption distribution. Again we see that in the period before post-Boskin reforms were implemented, the CPI, the Hamilton, and the TEC method show substantially different trends for deflated consumption, whereas in the

period after, the difference between the three is much smaller and not systematic.

Our emphasis on taking account of nonhomothetic preferences when evaluating price indices and calculating deflated incomes, is related to a recent literature on how non-homothetic preferences matter for the analysis of growth, income, and development: Bems and Di Giovanni (2017) study how shifts in income can drive expenditure switching between domestic and imported goods by looking at Latvian scanner data in the period of the financial crisis; Boppart (2014) develops a growth model with non-homothetic preferences and apply it to study and decompose the U.S. structural change into an income and substitution effect; Handbury (2013) studies spatial price differences in the U.S. by using Nielsen household-level purchase data of food products allowing for income- and city-specific price indexes; and Moretti (2013) studies the wage gap between high skilled and low skilled in the U.S. and reports that when deflating nominal wages using a location-specific CPI, the difference between the wage of college graduates and high school graduates is lower in deflated terms than in nominal terms and has grown less.

The paper proceeds as follows. Section 2 reviews the Hamilton method and demonstrates the inconsistency in it. Section 3 develops the TEC method theoretically. Section 4 discusses empirical implementation. Section 5 presents two empirical applications of the TEC method: one using Hamilton's original data and one using more recent data from the Consumer Expenditure survey. Section 6 concludes.

2 The Hamilton Method

2.1 Overview

The reasoning underpinning the Hamilton method 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 deflated income is changing, even when measured deflated 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 non-food prices are not observed but their inflation rates are. These inflation rates are assumed to be measured with error. Following Hamilton, 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/non-food. The two good system implies either within group homotheticity or fixed within group relative prices.

Then:

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

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 non-food, i.e. $\ln(1+E^f_t) - \ln(1+E^n_t) = 0$.[4] Thus:

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

where D^r are region dummies, D^t are time dummies, δ^r and δ^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 (after accounting for relative price effects and the effects of changing demographics). This is summarized in Figure 1. Figure 1 shows the region/period r,t Engel curve and the base period (0) Engel curve, but both at base period relative prices $(\frac{p^f_0}{p^n_0})$. The correction E_t shifts the region/period r,t Engel curve back onto the region/period 0 Engel curve, as indicated

[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 non-food inflation understates true non-food inflation.

Figure 1: The Hamilton Method

Note: The figure illustrates a shift in an Engel curve induced by a correction of a bias to the CPI. Without the correction for the bias correction E_t the Engel curve for period/region r, t would lie left of the Engel curve for period/region 0. The correction E_t shifts the Engel curve back onto the region/period 0 Engel curve.

by the arrows. If, as expected, the CPI is upward biased, from 3 we should have $(1+E_t) < 1$ and so $E_t < 0$. However, Hamilton reports $-E_t$ as cumulative bias, to give a positive number (and we shall follow this convention below.)

Hamilton incorrectly interprets $P_{r,t}$ 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. $P_{r,t}$ is a price index, but as we show in the next section, if relative prices vary it is the true cost of living index at, at most, one point in the income distribution. This means that E_t can *not* , in general, be interpreted as the bias in the CPI relative to changes in the cost of living.

[5] Hamilton (2001a), page 622, just above Equation 2.

2.2 Lost in Translation

The inconsistency in the Hamilton method is as follows. 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 generally true if preferences are homothetic.[6] 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, food shares contain no information about deflated income (or biases in deflated income). After adjusting for relative prices, the Engel curves lie on top of each other, and there is no horizontal translation (or scaling of nominal income in region/period r, t) that brings them closer together. 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, $\beta^f = 0$ and this calculation 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 (Engel, 1857b, 1895b; Working, 1943; Leser, 1963; Blundell, 1988). 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).

Of course, the non-homotheticity of preferences also implies that the CPI, which is a 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 non-homothetic 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

[6] With non-homothetic preferences, changes in the cost of living can be common across the income distribution in the specific instance of unchanging relative prices. But in general, relative prices do move. Moreover, if relative prices did not move, the problem of measuring cost of living differences across time and space would be simple: one need only find one well-measured price.

Figure 2: The Hamilton Method with Homothetic Preferences

Note : The figure illustrates Engel curves for homothetic preferences.

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 method's "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. Non-homotheticity: 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 the Hamilton 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.

3 Whose cost of living?

3.1 The CPI household

To disentangle genuine CPI bias from the effect of non-homotheticity, 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

Figure 3: Income Distribution Percentile of the CPI household

Note : The figure displays the percentile location of the CPI household in the U.S. income distribution for different years. The CPI household is defined as a virtual household with income $\sum_h \frac{x_{h,r,t} \ln x_{h,r,t}}{\sum_h x_{h,r,t}}$.

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.

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, at period t. We define reference

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

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 $\mathbf{p}_0 (= \mathbf{p}_{0,0})$. In logarithms, the true cost of living index is then given by:[8]

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

where $C\left(u_{h,0}, \mathbf{p_{r,t}}\right)$ 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 implies.

The food share is:

$$\begin{aligned} w^f_{h,r,t} &= \alpha^f + \sum_{k=f,n} \gamma^k \ln p^k_{r,t} + \beta^f \ln\left(\frac{x_{h,r,t}}{a(\mathbf{p_{r,t}})}\right), \\ &\left[= \alpha^f + \sum_{k=f,n} \gamma^k \ln p^k_{r,t} + \beta^f \ln\left(\frac{x_{h,r,t}}{P_{r,t}}\right) \text{in Hamilton's notation}\right]. \end{aligned} \tag{11}$$

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, $\ln a(\mathbf{p_0}) = \alpha_0$

[8] $u_{h,0}$ is not the only possible choice of reference utility level, but is the natural choice for comparison with price indices, like the CPI, that use base period quantities as weights. We discuss this further below.

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 ($x_{h,0}$)

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

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

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

Equation (13) shows 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 α_0 is not identified by the Hamilton method. A priori, there is no reason to think that α_0 would also be the log income level of the CPI household.

3.3 Decomposing the Hamilton Estimate

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

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

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 changes (recalling that $CPI_{0,0} = 1$ and $E_0 = 0$), 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:

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

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 non-homotheticity, which is given by the second term $\left[\Delta b(\mathbf{p_{r,t}})(\ln x^{CPI} - \alpha_0)\right]$.[9]

This decomposition 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 decomposition 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 non-homotheticity ($\Delta b(\mathbf{p_{r,t}})(\ln x^{CPI} -$

[9] 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.

$\alpha_0)$). In panel (a), this is positive, but 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 mismeasures the true 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 non-homotheticity) 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 non-homotheticity) 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 α_0 above $\ln x^{CPI}$. In this case the difference between $\Delta \ln C$ and $\Delta \ln P$ (non-homotheticity) is positive 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 α_0 were sufficiently large (far to the right) the effect of non-homotheticity 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 non-homotheticity 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 β^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: α_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 α_0 (because α_0 is the logarithm of the cost of base utility when prices are 1), but this is straight forward.[10] This means that if α_0 could be identified, then the decomposition of $\Delta \ln P_{r,t} - \ln CPI_{r,t} = \ln(1 + E_t)$ into the part due non-homotheticity and true CPI bias

[10] Appendix A describes the calculations.

Figure 4: Four possible cases

Note : The figure displays 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 t, r to period/region 0 (0,0) is given by the diagonal line. In panels (a) and (b) the relative price of food is lower in region/period t, r than in region/period 0 and the diagonal line slopes up. In panel (a) the Hamilton method reports a positive CPI bias and the slope of the diagonal line is such that the Hamilton method overestimates the CPI bias (the TEC bias is lower than what the Hamilton method reports). In panel (b) the Hamilton method also reports a positive bias and the slope is such that the TEC bias is negative. In panel (c) and (d) the relative price of food is higher in region/period t, r than in region/period 0 and the diagonal line slopes down. The Hamilton method reports a negative CPI bias, in panel (c) the Hamilton method underestimates the negative bias in CPI whereas in panel (d) the Hamilton method overestimates the negative CPI bias.

could be quantified. We take up identification of α_0 in the next section, but first consider the implications of non-homotheticity for volume measures based on Engel curve shifts.

3.4 Volume measures

While assessing bias in the CPI is important in its own right, the underlying motivation in this literature, going back to Nakamura (1995), is to construct volume measures (deflated income or deflated expenditure, for example). A price index $\Psi_{r,t}$ satisfies "weak factor reversal" if there exists an associated quantity index, $Q_{r,t}$, such that the nominal uplift in income, $\frac{x_{r,t}}{x_0}$, is the product of those price index and quantity indices:

$$\frac{x_{r,t}}{x_0} = \Psi_{r,t} Q_{r,t}. \tag{20}$$

Rearranging gives:

$$\frac{x_{r,t}}{\Psi_{r,t}} = x_0 Q_{r,t}, \tag{21}$$

where either side of this expression is deflated income, measured in the monetary units of period 0.

From Equation (9) we have taken the correct price index to be the Konüs cost of living index with reference utility level u_0.

$$\Psi_{r,t} = \frac{C(p_{r,t}, u_0)}{C(p_0, u_0)}. \tag{22}$$

The quantity index which is complimentary to this Konüs price index (in the sense of satisfying Equation (20)) is the Allen quantity index with reference price vector $p_{r,t}$:

$$Q_t = \frac{C(p_{r,t}, u_{r,t})}{C(p_{r,t}, u_0)}. \tag{23}$$

In our calculations below we use the left-hand side of Equation (21) to calculate deflated income, just as deflated income is often calculated by dividing nominal income by the CPI, but the Allen quantity index with reference price vector $p_{r,t}$ (multiplied by x_0) is numerically

identical.

In logarithms, deflated income is then $\ln x_t - \Delta \ln C(p_{r,t}, u_{h,0,r})$. Using Equation (13) we have:

$$\ln x_{h,r,t} - \Delta \ln C(p_{r,t}, u_{h,0,r}) = \ln x_{h,r,t} - \Delta \ln P_{r,t} - \Delta b(p_{r,t})(\ln x_{h,0} - \alpha_0). \tag{24}$$

Then Equation (3) gives:

$$\ln x_{h,r,t} - \Delta \ln C(p_{r,t}, u_{h,0,r}) = \ln x_{h,r,t} - \{\ln CPI_{r,t} + \ln(1 + E_t)\} - \Delta b(p_{r,t})(\ln x_{h,0} - \alpha_0). \tag{25}$$

Recall that $\frac{-\delta^t}{\beta^f} = \ln(1+E_t)$. Substituting this and Equation (18) and grouping terms gives:

$$\begin{aligned} \ln x_{h,r,t} - \Delta \ln C(p_{r,t}, u_{h,0,r}) = \{\ln x_{h,r,t} - \ln CPI_{r,t}\} + \left\{ \frac{\delta_t}{\beta^f} \right\} \\ - \left\{ \left(\frac{p^f_{r,0}}{p^n_{r,0}} \right)^{\beta^f} \left[\left(\frac{1 + \Pi^r_{r,t}}{1 + \Pi^n_{r,t}} \right)^{\beta^f} - 1 \right] (\ln x_{h,r,0} - \alpha_0) \right\}. \end{aligned} \tag{26}$$

Note that the first term in curly brackets on the right-hand side of Equation (26) is the standard measure of deflated income: nominal income adjusted by the CPI. The addition of the second term in curly brackets, $\left\{ \frac{\delta_t}{\beta^f} \right\}$, gives the deflated income by the Hamilton method. The third term is the TEC correction for the non-homotheticity of preferences. As Engel-curve based methods require non-homotheticity to implement, this gives a theoretically coherent measure of deflated incomes based on movements in Engel curves.[11] We implement and compare all three measures in our second application below.

[11] Note that we have taken the Konüs cost of living index with reference utility level u_0 to the be the appropriate price index. This leads Equation (26) to depend on $\ln x_{h,r,0}$. With repeated cross-sectional data sets, $\ln x_{h,r,t}$ and $\ln x_{h,r,0}$ are not observed for the same household. Thus Equation (26) can only be implemented with panel data, or at a group level. We take the latter approach in our second application below. An alternative approach would be to take the Konüs price index with with reference utility $u_{t,r}$ as the correct price index (so that an Allen index with reference price vector p_0 is the implied volume measure). Through manipulations similar to those given above, this would lead to an estimator of an extended money metric which the literature has sometimes called "adjusted expenditure" (Pendakur, 2002; Donaldson, 1992; King, 1983).

$$\ln x_{h,r,t} - \Delta \ln C(p_{r,t}, u_{h,t,r}) = \ln C(p_0, U(p_{r,t}, x_{h,r,t})). \tag{27}$$

We eschew that approach in this paper because the Konüs cost of living index with reference utility level u_0 seems the natural comparator for price indices, such as the CPI, that employ base-period quantities as weights. The Konüs cost of living index with reference utility level $u_{t,r}$ would seem a more natural comparator for a Paasche price index.

Finally, before turning to implementation, it is important to consider the interpretation of deflated incomes calculated by Equation (26), and in particular concerns raised by Angus Deaton. Deaton has strongly criticized the Hamilton method, arguing that it is subject to the Pollak and Wales critique (Deaton, 2016, 2010b). To clarify, suppose that the cost of a reference level of utility depends not only on prices but also on characteristics of households or the environment, z, so that the cost function is $C(p, z, u)$. An equivalence scale compares the cost of some reference level utility (for example u_0) across values of z, holding prices constant at some reference price level (for example, p_0):

$$S_0^1 = \frac{C(u_0, p_0, z^1)}{C(u_0, p_0, z_0)}. \tag{28}$$

In an important paper, Pollak and Wales (1979) demonstrate that S_0^1 is not identified from demand data, because monotonic transformations of the utility function (the inverse of the cost function) alter the cost-ratio (Equation(28)) without altering demands.

When the cost of a reference level of utility depends not only prices but also on characteristics of households or the environment, we have to be explicit about how we define a cost-of-living index. If we define it strictly to be the cost-ratio between two price vectors *holding all possible z constant*,

$$\Psi_{r,t} = \frac{C(p_{r,t}, z_0, u_0)}{C(p_0, z_0, u_0)}, \tag{29}$$

this object is identified by demand data and can be recovered by the TEC method, as shown above. The Pollak and Wales critique does not apply. However, the Pollack and Wales critique does apply to a mixed cost-ratio, in which both p and z are changing. The cost-ratio

$$\frac{C(p_{r,t}, z_1, u_0)}{C(p_0, z_0, u_0)}, \tag{30}$$

is not identified by demand data. Of course, across time and space, as prices change, other things also change. For example, climate change will affect welfare and well-being, perhaps particularly in less developed countries. So the adjustments we would really like to make

are of the type given by Equation (30). Deaton's reference to Pollak and Wales reminds us that we cannot, at least not without different kinds of data or very strong assumptions. The implication is that we should be very cautious about interpreting the deflated incomes developed in this subsection as welfare measures. It is exactly for this reason that we employ the term deflated income, rather than the more common "real" income. Of course, this caution applies equally whether we deflate income by the CPI, by a price index obtained from the Hamilton method or the TEC method, or by some other price index derived from demands.

4 Identification

4.1 The Problem

The TEC method requires knowledge of the parameter α_0. This parameter would in principle be identified if price levels were perfectly observed, but is not identified given data like that used by Hamilton. To see this, first rewrite the two good demand system:

$$w_{h,r,t} = \alpha^f + \gamma^{ff}(\ln p^f_{r,t} - \ln p^n_{r,t}) + \beta^f \ln x_{h,r,t} - \beta^f \ln P_{r,t}, \tag{31}$$

$$\ln P_{r,t} = \alpha_0 + \alpha^f \ln p^f_{r,t} + (1-\alpha^f)\ln p^n_{r,t} + \frac{\gamma_{ff}}{2}(-2\ln p^f_{r,t}\ln p^n_{r,t} + (\ln p^f_{r,t})^2 + (\ln p^n_{r,t})^2). \tag{32}$$

Using Equation (21) to eliminate P_t from the food share Equation (20), expressing the share as a function of nominal income and additional price variables, and then grouping terms yields:

$$\begin{aligned} w_{h,r,t} = {} & \left\{\alpha^f - \beta^f\alpha_0\right\} + \left\{\gamma^{ff} - \beta^f\alpha^f\right\}\ln p^f_{r,t} - \left\{\gamma^{ff} + \beta^f(1-\alpha^f)\right\}\ln p^n_{r,t} + \beta^f \ln x_{h,r,t} \\ & - \frac{\beta^f\gamma^{ff}}{2}\left\{-2\ln p^f_{r,t}\ln p^n_{r,t} + (\ln p^f_{r,t})^2 + (\ln p^n_{r,t})^2\right\}. \end{aligned} \tag{33}$$

Note that α_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.[12]

[12] First, β^f is identified by the nominal log income term and γ^{ff} is identified by the square of the log price

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 method proceeds by capturing these unobservables with, respectively, region and time dummies. However, in the expanded form of the share equation (Equation (33)) 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 α_0 is not identified by this approach.

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

$$\ln p^k_{r,t} = \ln(1+\Pi^k_{r,t}) + \ln p^k_{r,0} + \ln(1+E^k_t), \tag{34}$$

where $\Pi^k_{r,t}$ is reported good specific inflation. Substituting this structure into Equation (33) above gives:

$$\begin{aligned} w_{h,r,t} = {} & \left\{\alpha^f - \beta^f\alpha_0\right\} + \left\{\gamma^{ff} - \beta^f\alpha^f\right\}\left[\ln(1+\Pi^f_{r,t}) + \ln p^f_{r,0} + \ln(1+E^f_t)\right] \\ & - \left\{\gamma^{ff} + \beta^f(1-\alpha^f)\right\}\left[\ln(1+\Pi^n_{r,t}) + \ln p^n_{r,0} + \ln(1+E^n_t)\right] + \beta^f \ln X_{h,r,t} \\ & - \frac{\beta^f\gamma^{ff}}{2}\{-2\left[\ln(1+\Pi^f_{r,t}) + \ln p^f_{r,0} + \ln(1+E^f_t)\right]\left[\ln(1+\Pi^n_{r,t}) + \ln p^n_{r,0} + \ln(1+E^n_t)\right] \\ & + \left[\ln(1+\Pi^f_{r,t}) + \ln p^f_{r,0} + \ln(1+E^f_t)\right]^2 + \left[\ln(1+\Pi^n_{r,t}) + \ln p^n_{r,0} + \ln(1+E^n_t)\right]^2\}. \end{aligned} \tag{35}$$

As in Section 2.1, base period price differences ($\ln p^f_{r,0}$ and $\ln p^n_{r,0}$) can be captured with region dummies, and the time-specific errors captured by year dummies ($\ln(1+E^k_t)$). However, in the two last lines of Equation (35), the time-specific errors interact with unobserved

term. Given β^f and γ^{ff} , α^f is identified from the log relative price term, and then α_0 is identified from the constant (since we already know α^f and β^f). Such an estimate of α_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 α_0 even though it is in principle identified. See for example Banks et al. (1997) who set α_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 deflator would then be completely arbitrary.

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 of these 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 $w^f_{h,r,0} = \alpha^f + \beta^f u_{h,r,0} = \alpha^f - \beta^f \alpha_0 + \beta^f \ln x_{h,r,0}$. Thus to ensure that budget shares are bounded between zero and one at base 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 β^f is negative and these 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 an empirical 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, $\ln p^f_{r,0}$ and $\ln p^n_{r,0}$ are observed eliminating the need for regional dummies. Define the measured log price level as $\ln \tilde{p}^k_{r,t} = \ln(1 + \Pi^k_{r,t}) + \ln p^k_{r,0}$ so that from Equation (2) we have $\ln p^k_{r,t} = \ln \tilde{p}^k_{r,t} + \ln(1 + E^k_t)$. With this, and maintaining the Hamilton assumption that $\ln(1 + E^f_t) - \ln(1 + E^n_t) = 0$, we can derive an estimable equation of the form:

$$
\begin{aligned}
w_{h,r,t} = & \left\{\alpha^f - \beta^f \alpha_0\right\} + \beta^f \ln x_{h,r,t} + \left\{\gamma^{ff} - \beta^f \alpha^f\right\} \ln p^{*f}_{r,t} - \left\{\gamma^{ff} + \beta^f (1-\alpha^f)\right\} \ln p^{*n}_{r,t} \\
& - \frac{\beta^f \gamma^{ff}}{2} \left\{ -2 \ln p^{*f}_{r,t} \ln p^{*n}_{r,t} + \left[\ln p^{*f}_{r,t}\right]^2 + \left[\ln p^{*n}_{r,t}\right]^2 \right\} \\
& - \beta^f \ln(1 + E^n_t).
\end{aligned} \tag{36}
$$

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. β^f is identified by the nominal log income term and γ^{ff} is identified by the second order terms in log prices. Given β^f and γ^{ff}, α^f is identified from the log relative price terms, and then α_0 is identified from the constant (since we already know α^f and β^f). We have implemented this on Hamilton's data augmented with Aten's regional price levels. Unfortunately, the resulting estimate of α_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 A 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 $P_{r,t}$:

$$
w^f_{h,r,t} = \alpha^f + \beta^f \ln \left(\frac{x_{h,r,t}}{P_{r,t}} \right). \tag{37}
$$

Some authors without access to regional price variation have estimated exactly this model when implementing the Hamilton method (see Beatty and Larsen (2005) amongst others), and researchers – including Hamilton – estimating the more general formulation often cannot reject the restriction that $\gamma^{ff} = 0$ (see 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 (33)), and Hamilton's assumption that $\ln(1+E_t^f)-\ln(1+E_t^n) = 0$, we get:

$$w_{h,r,t}^f = \left\{\alpha^f - \beta^f\alpha_0\right\} + \beta^f \ln x_{h,r,t} + \beta^f\alpha^f \ln\left(\frac{1+\Pi_{r,t}^n}{1+\Pi_{r,t}^f}\right) - \beta^f \ln(1+\Pi_{r,}^n) - \beta^f \ln(1+E_t^n)$$
$$- \left[\beta^f\alpha^f \ln p_{r,0}^f + \beta^f(1-\alpha^f)\ln p_{r,0}^n\right], \tag{38}$$

$$w_{h,r,t}^f = \left\{\alpha^f - \beta^f\alpha_0\right\} + \beta^f \ln\left(\frac{x_{h,r,t}}{1+\Pi_{r,t}^n}\right) + \beta^f\alpha^f \ln\left(\frac{1+\Pi_{r,t}^n}{1+\Pi_{r,t}^f}\right) - \sum_t \tilde{\delta}^t D^t + \sum_r \tilde{\delta}^r D^{r,} \tag{39}$$

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

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 $\gamma^{ff} = 0$,

$$w_{h,r,t}^f = \alpha^f + \beta^f \ln\left(\frac{x_{h,r,t}}{\mathrm{CPI}_{r,t}}\right) - \beta^f \ln(1+E_t) + \sum_r \delta^r D^r, \tag{40}$$
$$= \alpha^f + \beta^f \ln\left(\frac{x_{h,r,t}}{\mathrm{CPI}_{r,t}}\right) + \sum_t \delta^t D^t + \sum_r \delta^r D^r.$$

We recover an estimate of E_t from $\delta^t = -\beta^f \ln(1+E_t)$. Then, we estimate Equation (39) to get an estimate of α_0 and following Section 3.3, we calculate part of E_t that is due to non-homotheticity as

$$\Delta b(\mathbf{p_{t,r}})(\ln x^{CPI} - \alpha_0) = \left(\frac{p_{r,0}^f}{p_{r,0}^n}\right)^{\beta^f} \left[\left(\frac{(1+\Pi_{r,t}^f)}{(1+\Pi_{r,t}^n)}\right)^{\beta^f} - 1\right](\ln x^{CPI} - \alpha_0). \tag{41}$$

We now implement this procedure on the Hamilton data.

5 Applications

5.1 PSID 1974-1991 (Hamilton's data)

We first apply the TEC method to 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 estimates, with the preference restriction described above, of the log income level of the Hamilton household (that is, α_0), and the CPI household, against the full distribution of log income. The figure pools data across states. We can treat each state in turn as the base region and calculate α_0. The two vertical lines in the figure give the highest and lowest values of α_0 across states. We find that the Hamilton household is much poorer than the CPI household, and indeed, in every state has a log income level far below the minimum log income in the full sample. In practice, 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 developed in Section 4.2.1, and shows that the range of log income levels that the Hamilton household could represent is 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 a 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 (computed 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 over different income groups in China, the differences are small. This is a very elegant way of assessing slope of the change in cost of living (the line labeled $\Delta \ln C$ in

Figure 5: Hamilton and CPI households

Note : The figure shows the distribution of income in the PSID sample, the income of the CPI household, and the income of the Hamilton household (lowest and highest states).

Figure 4). However, this procedure is only informative about the *slope* of this line, and not its *location*. If the log income level of the Hamilton household is far below, or 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 household in the change in cost of living. This means that calculations such as those reported by Nakamura, Steinsson and Liu cannot rule out non-homotheticity as an important component of the difference between Engel Curve-based inflation estimates and other estimates. The role of non-homotheticity 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 non-homotheticity.

Figure 6: Bounds

Note : The figures show the bounds on α_0, the income of the Hamilton household (for the medium state), and the income of the CPI household.

Figure 7: Relative Prices & Non-homotheticity (1974–1991)

Note : The Figure charts the evolution of the relative price of food to non-food goods, over the period studied by Hamilton (solid line). The difference in the change in log cost-of-living for the CPI household and the change in the log cost of living of the Hamilton Household is given by the dashed line.

The solid line in Figure 7 charts the evolution of the relative price of food (to non-food goods) over the period studied by Hamilton. In this time period, food prices fell relative to non-food prices, so that rich households experienced larger cost-of-living increases than poor households. Together with the fact that the log income level of the Hamilton household is far below the log income level of the CPI household, this implies that the true change in cost-of-living was smaller for the Hamilton household than for the CPI household. The excess in the change in cost-of-living for the CPI household over the change in the cost of living of the Hamilton household is given by the dashed line in Figure 7. This difference in true cost-of-living changes in turn implies that the difference between the change in cost-of-living for the Hamilton household and the CPI overstates the bias in the CPI *as a measure of the change in cost-of-living experienced by the CPI household.* Hence, the scenario that turns out to be relevant for these data and this period is captured by Panel (a) in Figure 4.

Our TEC estimates of CPI bias are presented in right-hand side column of Table 1 . Our replication of Hamilton's original estimates are given to the left of these for comparison.[13] As we remove the difference between the change in cost-of-living for the CPI household and the change in the cost of living of the Hamilton household, the TEC estimates of cumulative CPI bias are smaller than Hamilton's. By the end of the period the TEC method reveals a cumulative CPI bias of 5 percentage points lower than Hamilton's: Nevertheless, after correcting for non-homotheticity, we continue to find a significant upward bias in the CPI in this period, with cumulative bias over the 15 years of 23 percentage points.

5.2 CE Data 1990-2014

In a second empirical application we consider the years 1990-2014. This period is interesting not only because it brings us up to date, but also because after Boskin et al. (1996) the BLS made a number of improvements to the CPI (see also Gordon (2006)). These included the use of a geometric means formula to account for lower level substitution, expanding the use of hedonic models to account for quality change, and introducing a procedure to

[13] The appropriate comparison is to his 25-SMSA sample estimates. Our numbers differ very slightly, due to subsequent revisions to the PSID

Table 1: Comparing CPI Bias computed according to Hamilton and TEC Methods

Year	Hamilton	TEC
1974	-	-
1975	0.050	0.049
1976	0.122	0.112
1977	0.106	0.096
1978	0.133	0.131
1979	0.167	0.166
1980	0.212	0.196
1981	0.199	0.177
1982	0.220	0.189
1983	0.240	0.204
1984	0.280	0.238
1985	0.257	0.250
1986	0.257	0.209
1987	0.266	0.220
1990	0.280	0.238
1991	0.277	0.230

Note: The table shows the cumulative bias for the Hamilton and TEC methods.The cumulative bias for the Hamilton method is given by $E_t = \exp[-\frac{\delta^t}{\beta f}] - 1$ and the cumulative bias for the TEC method is given by $B_t = \exp[-\frac{\delta^t}{\beta f} + \Delta b\left(\mathbf{p_{r,t}}\right)\left(\ln x_{h,0}^{CPI} - \alpha_0\right)] - 1$.

introduce new goods to the index more quickly (Johnson et al., 2006). It is important to assess the effect of these improvements and the TEC method provides a line of evidence on this question.

For these years, we switch from the PSID to the Consumer Expenditure (CE) Survey.[14] CE data are collected by the Census Bureau for the Bureau of Labor Statistics. The CE data contain household-level information on an aggregate of non-durable consumption, food consumption, and income (to be used as an instrument).[15] We also observe employment status and detailed demographics that can be used as controls. We follow Hamilton (2001) and use only households that are identified as "white" and with both adults older than 21. The area identifier for households is state of residency in addition to an identifier for whether or not the household lives in a SMSA. We have CPI data at SMSA level and assign a CPI to

[14] More recent waves of the PSID are not well suited to this exercise as the PSID is now biennial and income and expenditure questions refer to different years

[15] More precisely the CE contains data at the level of the consumer-unit. See the Appendices for more detail.

each household in the following manner. We drop households that do not live in a SMSA. If a state has no SMSA for which we have a CPI, we drop it. If we have one SMSA CPI for a state, we assign it to the state. If we have more than one SMSA CPI for a state we use the average. As some SMSA span more than one state, an SMSA is assigned (fully or as part of an average) to every state it spans.

Moreover, the CE is a natural dataset with which to estimate Engel curves. It allows us to use (non-durable) consumption, rather than net income, as the budget measure to which we relate the food share. This is conceptually desirable and means that we do not need to be concerned that Engel curves are shifting because of changes in inter-temporal allocation (saving). It also allows us to deal with possible measurement error in the budget measure by instrumenting for consumption with income, as is commonly done in empirical demand estimation.[16]

Figure 8 shows the estimated position of the Hamilton household relative to the consumption distribution for 1990. The figure again pools data across states, and the two vertical lines give the highest and lowest values of α_0 – the Hamilton household – across states. As in the earlier time period, the Hamilton household is very poor, and their consumption lies well below the observed consumption distribution.

However, the years from 1990 to 2014 differ from the period studied by Hamilton in the movement of the relative food price. In contrast to the steady decline seen between 1974 and 1991, which favored the poor, 1990 to 2014 saw an initial decline in the relative food price followed by a steep increase in the relative food price. This is shown in Figure 9. One implication is that from 2006 the poor experienced a faster increase in the cost-of-living than the rich, and the (very poor) Hamilton household had a higher cost-of-living increase than the CPI household (or, indeed, any household in the empirical consumption distribution). This in turn means that the TEC correction for non-homotheticity changes sign after 2006, and the original Hamilton method *understates*, rather than overstates, the degree of CPI bias in recent years.

We break the post-1990 years into two periods, 1990-1998 and 2000-2014, in order to

[16]The results are based on OLS estimates to maximize comparability with the PSID estimates, but IV estimates are similar.

Figure 8: Estimated log-income of the Hamilton household, 1990

Note : The figure shows the distribution of income in the CE sample, the income of the CPI household, and the income of the Hamilton household (lowest and highest states).

Figure 9: Relative Prices (1990-2014)

Note : The figure shows the development in the relative price of food (over non-food) in the years 1990-2014.

asses the major reforms to CPI undertaken after the Boskin report were introduced in 1999. Figure 10 displays our results for cumulative CPI bias using both the Hamilton method (dotted lines) and the TEC method (solid lines). There are three panels, all with the same vertical scale. For comparison purposes, our PSID results for the earlier Hamilton period are reproduced in the far left-hand panel. The middle panel displays our CE results for the 1990-1998 period (that is, after the Hamilton period but before the post-Boskin reforms to the CPI), and the farthest-right panel gives CE results for the 2000-2014 period, after the post-Boskin reforms.

In the period 1990-1998, we continue to find evidence of important bias in the CPI. As noted above, movements in relative food price through this period were largely to the benefit of the poor. So, as in the original Hamilton period, the Hamilton household experienced a smaller increase in the cost-of-living than the CPI household. The TEC method suggests cumulative bias in the CPI of 13 percentage points over these 8 years (compared to 28 percentage points over the 15 years of the Hamilton period). Note that there is a larger difference between the TEC method and the Hamilton method for these years and data. Our correction for non-homotheticity indicates that the original Hamilton method overstates genuine CPI bias by almost 40% by the end of this period.

Turning to the period 2000-2014, because the relative price of food falls and then rises in period, our non-homotheticity correction changes sign part way through the period, and the two sets of estimates do not diverge significantly. But more importantly, neither set of estimates suggests much CPI bias in this period, as is evident in the far-right panel. The TEC method estimates a cumulative upward bias in CPI of 6 percentage points over this 15 years period. This is in sharp contrast to the two earlier, pre-Boskin periods in the middle and left-hand panels. This suggests that the improvements made to the CPI in the late 1990s may have been effective. At a minimum, food Engel curves no longer seem to be shifting much over time.

Finally, as noted Section 3.4, the underlying motivation in much of this literature is to construct volume measures of deflated income or consumption measures. We conclude this application by using Equation (26) to calculate and compare alternative measures of

Figure 10: Cumulative CPI Bias Across Three Periods: Hamilton and TEC methods

Note : The figure displays the cumulative bias for the Hamilton method and the TEC method. The cumulative bias for the Hamilton method is given by $E_t = \exp[-\frac{\delta^t}{\beta^f}] - 1$ and the cumulative bias for the TEC method is given by $B_t = \exp[-\frac{\delta^t}{\beta^f} + \Delta b\left(\mathbf{p_{r,t}}\right)\left(\ln x_{h,0}^{CPI} - \alpha_0\right)] - 1$.

deflated consumption for the periods 1990-1998 and 1999-2014.

These calculations of genuine CPI bias are based on the idea that if we have the right deflator for the Hamilton household ($\Delta \ln P$), the location of the Hamilton household in log-income (or log-consumption) space (α_0), and the slope of the cost-of-living function, we can calculate the true change in the cost-of-living at the log-income or (log-consumption) level of the CPI household. But of course the calculation can be made for *any* income (or consumption) level. Put differently, we have 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 (or consumption) distribution. This means we can study how inflation differs between rich and poor, or apply Equation (26) to study how deflated income (or consumption) has evolved at different points in the relevant distribution. Here we consider three households with the mean consumption of the first, third and fifth quintile of the consumption distribution. We then calculate (using Equation (26)) alternative paths of deflated consumption for these households from 1990 to 1998, and from 1999 to 2014. To compare growth, we normalize each path to 100 in either 1990 or 1999.

The results are displayed in Figure 11. The top row shows results for 1990-1998 with panels for the lowest, middle and highest quintiles of the consumption distribution as we move from left to right. The bottom row shows results for 1999-2014, again with panels for the lowest, middle and highest quintiles of the consumption distribution arranged from left to right. In each panel, non-durable consumption deflated by the CPI is given by the darker solid line. Non-durable consumption deflated by the CPI corrected for the bias as estimated by the Hamilton method is given by the short-dashed line. Finally, non-durable consumption deflated by the TEC measured cost of living is given with by the long-dashed line.

Starting with the 1990-1998 period, consumption deflated by the CPI fell substantially for the lowest quintile. By the same measure, the middle and top quintiles experienced falls in non-durable consumption through the middle of the decade but recovered their initial position by the end of the decade. Using alternative deflators leads to a quite different story. If consumption is deflated by the CPI corrected according to the Hamilton method, the

lowest quintile experienced modest economic progress through the 1990s, while the middle and top quintiles appear to have experienced strong non-durable consumption growth. The TEC method reveals that part of the correction that the Hamilton method makes to the CPI is the spurious result of differences in the consumption basket of the fairly affluent CPI household and extremely poor Hamilton household. When we use the TEC method to calculate the correct cost-of-living change for each household, we see that the economic progress of each quintile is substantially understated by consumption deflated by the CPI, but substantially *overstated* by consumption deflated by the CPI corrected according to the Hamilton method. The middle and top quintiles of the consumption distribution have experienced more economic progress than the bottom quintile, but that progress was more modest than the Hamilton method would suggest. At the same time, the bottom quintile did not experience the consumption losses that the uncorrected CPI would suggest.

In contrast to the 1990s, during the 1999-2014 period, shown in the bottom row of Figure 11, the three measures of deflated-consumption are quite similar. This a consequence of two things. First, the disappearance of bias in the CPI as a measure of the cost-of-living increase in the CPI household, as noted above. Second, over this period the inflation experienced by the CPI-household is not very different to that experienced by poorer households over this period, because the relative price of food both rose and fell. Together, these facts mean that for the 1999-2014, the CPI did a reasonable job of capturing changes in the cost of living for all quintiles.[17]

6 Summary and Conclusions

The Hamilton method as proposed and often used confounds genuine CPI bias with differences in consumption baskets across the income distribution. We have argued that, as the Hamilton method *requires* non-homotheticity to implement, non-homotheticity must be

[17] In this analysis our focus is on the denominator of the deflated consumption measure, and hence across the three measures (deflated by CPI, Hamilton and TEC.) Comparisons across quintiles and time will also, of course, reflect the numerator (nominal consumption), and here there are important issues regarding the changing representativness of the CE data across the income distribution and over time. Those issues are beyond the scope of this paper but have been much discusssed elsehwere. See Aguiar and Bils (2015) and the further references therein.

Figure 11: Deflated Consumption indexes (1990-1998 & 1999-2014)

Note : The figure shows the trends for deflated incomes using the CPI, the Hamilton and the TEC deflators, respectively. It shows trends for the lowest, the middle and the highest quintile of the income distribution, and for the early and later period (before and after post-Boskin reforms) separately.

accounted for in interpreting the movement in Engel curves over time or space. We have demonstrated how to do this by developing a method that disentangles genuine price index bias from differences in consumption baskets across the income distribution. In this way, the Hamilton method can be made internally consistent.

For the data and period that Hamilton studies, the TEC method leads to smaller, but still important, estimates of cumulative bias in the CPI. When we extend the analysis to more recent periods, we have two interesting findings. First, between 1990 and 1998, we continue to find evidence of CPI bias. But in this period, the difference between the TEC deflator and the Hamilton deflator is larger. Our correction for non-homotheticity indicates that the original Hamilton method overstates genuine CPI bias by almost 40% by the end of this period. Second, we find little evidence of CPI bias for the period 2000-2014. We speculate that this is due to improvements to the CPI undertaken by the BLS in 1999, after the Boskin Commission report.

These findings have important implications for the study of growth in deflated household income or consumption. For example, Meyer and Sullivan (2009, 2011) argue that official poverty statistics in the U.S. understate the progress that has been made on eliminating poverty over time. While their calculations differ from the official statistics in a number of ways, one key factor is that they reduce growth in the CPI by one percent per year, citing the complementary evidence from the Boskin Commission and from Hamilton (2001a). Our analysis suggest this is a reasonable thing to do for the period originally studied by Hamilton. But in the 1990s the TEC method suggests a growth in deflated consumption that, while significantly above what the CPI would imply, is significantly below what the original Hamilton method would imply. This is true right across the income distribution. And for the post-2000 period, we find no systematic evidence of the CPI bias, so that there is no evidence from Engel curves against using income deflated by the CPI to measure household resources.

References

Mark Aguiar and Mark Bils. Has consumption inequality mirrored income inequality? *American Economic Review*, 105(9):2725–2756, September 2015.

Ingvild Almås and Åshild Auglænd Johnsen. The cost of living in china: Implications for inequality and poverty. 2012.

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

Bettina H. Aten. Estimates of State and Metropolitan Price Parities for Consumption Goods and Services in the United States, 2005. Working paper, Bureau of Economic Analysis, October 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. Working paper, 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.

Rudolfs Bems and Julian Di Giovanni. Income-induced Expenditure Switching. *American Economic Review*, 106(12):3898–3931, December 2017.

Mark Bils and Peter J. Klenow. Quantifying Quality Growth. *American Economic Review*, 91(4):1006–1030, September 2001.

Laura Blow. Explaining Trends in Household Spending. Working paper 03/06, Institute for Fiscal Studies, 2003.

Richard Blundell. Consumer behaviour: theory and empirical evidence–a survey. *The economic journal*, 98(389):16–65, 1988.

Timo Boppart. Structural Change and the Kaldor Facts in a Growth Model with Relative Price Effects and Non-Gorman Preferences. *Econometrica*, 82(6), 2014.

Michael J. Boskin and Dale W. Jorgenson. Implications of Overstating Inflation for Indexing Government Programs and Understanding Economic Progress. *American Economic Review*, 87(2):89–93, May 1997.

Michael J. Boskin, Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale W. Jorgenson. Toward a More Accurate Measure of the Cost of Living: Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index. Report, December 1996.

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. *Journal of Economic Perspectives*, 12(1):3–26, 1998.

Christian Broda and David E. Weinstein. Product Creation and Destruction: Evidence and Price Implications. *American Economic Review*, 100(3):691–723, May 2010.

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.

Andrew Dabalen, Isis Gaddis, and Nga Nguyet Nguyen. CPI Bias and its Implications for Poverty Reduction in Africa. *World Bank Policy Research Working Paper*, (7907), 2016.

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? *Journal of Economic Perspectives*, 12(1):37–46, 1998.

Angus Deaton. Price Indexes, Inequality, and the Measurement of World Poverty. *American Economic Review*, 100(1):5–34, March 2010b.

Angus Deaton. Measuring and understanding behavior, welfare, and poverty. *American Economic Review*, 106(6):1221–43, 2016.

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 2010a.

Angus Deaton and John Muellbauer. An Almost Ideal Demand System. *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.

David Donaldson. On the Aggregation of Money Measures of Well-being in Applied Welfare Economics. *Journal of Agricultural and Resource Economics*, 17(1):88–102, 1992.

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

Ernst Engel. Die productions-und consumtionsverhältnisse des königreichs sachsen. *Zeitschrift des Statistischen Bureaus des Königlich Sächsischen Ministeriums des Innern*, 8:1–54, 1857b.

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

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

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.

Robert J Gordon. The boskin commission report: A retrospective one decade later. Technical report, National Bureau of Economic Research, 2006.

John S. Greenlees and Robert McClelland. Does Quality Adjustment Matter for Technologically Stable Products? An Application to the CPI for Food. *American Economic Review*, 101(3):200–205, 2011.

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. *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.

Jessie Handbury. Are Poor Cities Cheap for Everyone? Non-homotheticity and the Cost of Living Across U.S. Cities. Working paper 71, University of Pennsylvania - The Wharton School, June 2013.

Jerry Hausman. Sources of Bias and Solutions to Bias in the Consumer Price Index. *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. *American Economic Review*, 94(5):1379–1410, 2004.

David S. Johnson, Stephn B. Reed, and Kenneth J. Stewart. Price measurement in the United States: a decade after the Boskin Report. *Monthly Labor Review*, 129(1):10–19, January 2006.

Mervyn A. King. Welfare Analysis of Tax Reforms Using Household Data. *Journal of Public Economics*, 21(2):183–214, 1983.

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

David E. Lebow and Jeremy B. Rudd. Measurement Error in the Consumer Price Index: Where Do We Stand? *Journal of Economic Literature*, 41(1):159–201, 2003.

Conrad Emanuel Victor Leser. Forms of engel functions. *Econometrica: Journal of the Econometric Society*, pages 694–703, 1963.

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.

Bruce D. Meyer and James X. Sullivan. Viewpoint: Further Results on Measuring the Well-being of the Poor Using Income and Consumption. *Canadian Journal of Economics*, 44 (1):52–87, February 2011.

Bruce D. Meyer and James X. Sullivan. Identifying the Disadvantaged: Official Poverty, Consumption Poverty, and the New Supplemental Poverty Measure. *Journal of Economic Perspectives*, 26(3):111–136, Summer 2012.

Enrico Moretti. Real Wage Inequality. *American Economic Journal: Applied Economics*, 5 (1):65–103, 2013.

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: Macroeconomics*, 8 (3):113–144, 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. *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.

William D. Nordhaus. Quality Change in Price Indexes. *Journal of Economic Perspectives*, 12(1):59–68, 1998.

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.

Krishna Pendakur. Taking Prices Seriously in the Measurement of Inequality. *Journal of Public Economics*, 86(1):47–69, 2002.

Robert A. Pollak. The Consumer Price Index: A Research Agenda and Three Proposals. *Journal of Economic Perspectives*, 12(1):69–78, 1998.

Robert A. Pollak and Terence J. Wales. Welfare Comparisons and Equivalence Scales. *American Economic Review*, 69(2):216–221, 1979.

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 B. 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.

Marshall B. Reinsdorf. Formula Bias and Within-stratum Substitution Bias in the U.S. CPI. *The Review of Economics and Statistics*, 80(2):175–187, 1998.

Bruce I. Sacerdote. Fifty Years Of Growth In American Consumption, Income, and Wages. Working Paper 23292, National Bureau of Economic Research, March 2017.

Charles L. Schultze. The Consumer Price Index: Conceptual Issues and Practical Suggestions. *Journal of Economic Perspectives*, 17(1):3–22, 2003.

Jack E. Triplett. Should the Cost-of-living Index Provide the Conceptual Framework for A Consumer Price Index? *The Economic Journal*, pages 311–334, 2001.

Holbrook Working. Statistical laws of family expenditure. *Journal of the American Statistical Association*, 38(221):43–56, 1943.

Yonghong Xu and Wuyi Zeng. Estimation of CPI Bias with Chinese City Statistical Data. *Statistical Research*, 4:2, 2009.

Online Appendices for Lost in Translation

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

January 29, 2018

Appendix A: Re-normalizing Prices

The calculation of the non-homotheticity 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 the TEC estimate of α_0. This makes sense because α_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 α_0.

Our "long form" is

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

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 constant and α^r_0 by the normalization (by which region is the omitted 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 procedure 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 re-normalizing prices to be one in that region in the base period) but noting that renormalization of prices means we have to adjust α_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, to see why this makes sense, note that the identifying assumption is that $\gamma_{k,l} = 0$ so that

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

where α_0 is the value of the price index when prices are one. If we renormalize prices we have to adjust α_0 by $\sum_{k=f,n} \alpha_k \ln p^k_{r,t}$.

Appendix B: Data and Estimates

PSID 1974-1991 (Hamilton's data)

In our first application we apply the TEC method to the Panel Study of Income Dynamics (PSID) data studied in Hamilton (2001). While Hamilton's original data set is no longer available, there is sufficient information in Hamilton (2001) to recreate the his data from raw PSID files. However, revisions to the underlying PSID data with subsequent releases mean that there are very minor differences between our data set and Hamilton's original data, particularly with regards to geography.

In creating our PSID data set for analysis we follow Hamilton's sample selection rules. In particular, we set aside the poverty sample in the PSID and then select white, two-adult families with any number of children (including zero), with both adults aged 21 or older. We delete families that report a change in household composition since the previous year, and those receiving food stamps or AFDC. We also delete families with food shares greater than 80% or less than 2%, those with missing tax information, and those with net income less than $150 or with top-coded income. Note that in the PSID income refers to the previous year, so for each family-year observation we take the one-year-ahead value of the income variable.

We use the CPI Urban Consumers (Old Series) for all MSAs used by Hamilton. From the Bureau of Labor Statistics (BLS), we obtained the per city All Items CPI (SA0) and the per city Food-At-Home price index (SA111). Note that the BLS does not report All Items Less Food (SA0L1) at the MSA level. To create this series, we use the relative importance files provided by BLS to back this series out from the All Item and Food series. Note that the relative importance files are missing for 1978, 1979, and 1980 and so we interpolate the weights by city for these years.

In estimating Engel curves, we again follow Hamilton exactly. We take the share of food at home in net income as our dependent variable, net income as our income measure, and we include the ratio of the Food-At-Home CPI to the All Items Less Food CPI as our relative price. Additional controls are: age of the husband, age of the wife, education of the husband, education of the wife, hours worked by the husband, hours worked by the wife, the number of children, share of income spent on food away from home, and finally measures of positive and negative income growth (equal to the change in year over year income $y_t - y_{t-1}$, times indicator variables for positive and negative changes). Appendix Tables 1 and 2 correspond to Tables 2 and 3 in Hamilton (2001). Note here that we are not imposing the preference restriction that use to identify α_0 in the TEC method, so as to match Hamilton precisely. These estimates underlay cumulative bias estimates using the Hamilton method presented in Table 1 of the paper.

Table 1: PSID Main Regression Results

Variable	Coefficient	Std. Error
Constant	0.9789	0.0138
Age of Husband	0.0005	0.0001
Age of Wife	0.0005	0.0001
Number of Children	0.0219	0.0006
County Unemployment	-0.0001	0.0003
Husband's Annual Hours	0.0037	0.0008
Wife's Annual Hours	-0.0026	0.0007
Husband's Education	-0.0011	0.0003
Wife's Education	-0.0002	0.0003
ln(Income)	-0.0919	0.0015
ln(Relative Food Price)	0.0181	0.0201
Income Growth +	7.01e-07	1.34e-07
Income Growth -	-8.40e-07	1.02e-07
Food Share at Resaturant	0.1400	0.0199
R2	0.5545	
N	7937	

Table 2: PSID Year and SMSA Dummies

Year	Coefficient (Std. Err.)	Cumulative bias estimate	SMSA	Coefficient(Std. Err.)
1975	-0.0047 (0.0033)	0.050	New York	0.0270 (0.0038)
1976	-0.0119 (0.0034)	0.122	Miami	0.0180 (0.0050)
1977	-0.0103 (0.0035)	0.106	Los Angeles	0.0118 (0.0050)
1978	-0.0131 (0.0032)	0.133	Buffalo	0.0131 (0.0091)
1979	-0.0168 (0.0033)	0.167	San Francisco	0.0096 (0.0044)
1980	-0.0218 (0.0036)	0.212	Portland, OR	0.0083 (0.0068)
1981	-0.0203 (0.0039)	0.199	Chicago	0.0069 (0.0040)
1982	-0.0227 (0.0043)	0.220	Cincinnati	-0.0178 (0.0051)
1983	-0.0251 (0.0048)	0.240	Houston	0.0055 (0.0053)
1984	-0.0302 (0.0049)	0.280	San Diego	0.0010 (0.0068)
1985	-0.0273 (0.0052)	0.257	Washington, DC	0.0056 (0.0047)
1986	-0.0273 (0.0049)	0.257	Philadelphia	0.0077 (0.0047)
1987	-0.0284 (0.0048)	0.266	Milwaukee	0.0081 (0.0077)
1988			Detroit	0.0033 (0.0040)
1989			Boston	0.0043 (0.0038)
1990	-0.0302 (0.0044)	0.280	Baltimore	0.00717 (0.0048)
1991	-0.0298 (0.0045)	0.277	Denver	-0.0030 (0.0056)
			Pittsburgh	0.0005 (0.0044)
			Kansas City	-0.0019 (0.0062)
			Seattle	-0.0058 (0.0045)
			St. Louis	-0.0078 (0.0044)
			Cleveland	0.0061 (0.0050)
			Dallas	-0.0102 (0.0051)
			Minneapolis/St. Paul	-0.0130 (0.0041)
			Atlanta	Base

CE Data 1990-2014

In our second empirical application we consider the years 1990-2014, and for these years we switch from the PSID to the Consumer Expenditure (CE) Survey.

The CE data are collected by the Census Bureau for the BLS. The CE comprises two distinct surveys (with different samples): an annual diary survey and a quarterly interview survey. For a general descriptions of the data see Bee et al (2015) and Bureau of Labor Statistics (2011). We use data from the interview survey, which Bee et al (2015) argue is the more reliable source, particularly for distributional analyses.

The CE interview survey is a rotating panel and each consumer unit (roughly, a family) participates in up to five quarterly interviews. However, the first interview is a bounding interview and data from this interview are not publicly available. We use data from the 2nd quarterly interview for each consumer unit. Focusing on 2nd quarter avoids issues of attrition.

From the CE data we use consumer-unit level data on food spending, a measure of aggregate non-durable consumption spending, and income. Neither the interview survey nor the diary survey collects spending on all items, so our measure of aggregate non-durable consumption spending omits several small items (see Bee et all (2015) for further details). We also use data on employment status and detailed demographics as regression controls.

For the CE application in this paper we follow the selection rules that Hamilton employed for PSID closely, though some of those rules are either not relevant or not possible to apply to the CE. As with the PSID we select two adult households (with any number of children) that are identified as "white" and with both adults older than 21. We also once again delete families with food shares greater than 80% or less than 2% and those with net income less than $150.

We also follow, as far as is possible, his specification of the food Engel curve. As our CE dataset does not have a panel structure, we are unable to include the negative and positive income growth controls that Hamilton includes (and we include) for the PSID.

The area identifier for consumer units is the state of residency. In addition, the data contain an identifier for whether or not the household lives in a SMSA. As CPI data are available for SMSAs, we base our estimation on the households that live in SMASs, and drop consumer units that do not live in a SMSA.

If a state has no SMSA for which we have a CPI for all years, we drop that state. If we have one SMSA CPI for a state, we assign it to that state. If we have more than one SMSA CPI for a state we use the average. As some SMSAs span more than one state, an SMSA is assigned (fully or as part of an average) to every state it spans.

Appendix Tables 3 and 4 present the full regression results for our replication of Hamilton on this CE

Table 3: CE Main Regression Results

Variable	Coefficient	Std. Error
Constant	1.250	0.010
Age of head	0.000	0.000
Sex of spouse	0.000	0.000
Number of Children	0.024	0.001
Typical hours of head per week	0.002	0.001
Wife's Annual Hours	-0.006	0.001
Wife High school	-0.016	0.002
Wife more than high school	-0.020	0.002
Husband high school	-0.006	0.002
Husband more than high school	-0.008	0.002
ln(Income)	-0.113	0.001
Food Share at Restaurant	-0.204	0.007
R2	0.406	
N	33554	

sample. Note that here we are imposing the preference restriction discussed in the paper, though results without that restriction are very similar. These estimates result from estimating the Engel curve by OLS, but IV estimates are again quite similar. The full list of the states that we are able to include is given in Appendix Table 4.

Table 4: CE Year and State Dummies

Year	Coefficient (Std. Err.)	Cumulative bias estimate	State	Coefficient(Std. Err.)
1991	-0.011 (0.004)	0.091	California	-0.045 (0.004)
1992	-0.009 (0.004)	0.075	Colorado	-0.061 (0.005)
1993	-0.018 (0.004)	0.145	Connecticut	-0.058 (0.005)
1994	-0.020 (0.004)	0.159	Florida	-0.051 (0.004)
1995	-0.013 (0.004)	0.107	Georgia	-0.056 (0.004)
1996	-0.023 (0.004)	0.188	Hawaii	-0.038 (0.014)
1997	-0.016 (0.004)	0.127	Illinois	-0.061 (0.004)
1998	-0.021 (0.004)	0.168	Indiana	-0.085 (0.004)
1999	-0.018 (0.004)	0.151	Kansas	-0.077 (0.007)
2000	-0.017 (0.004)	0.143	Maryland	-0.045 (0.005)
2001	-0.024 (0.004)	0.194	Massachusetts	-0.053 (0.004)
2002	-0.019 (0.004)	0.152	Michigan	-0.073 (0.004)
2003	-0.016 (0.004)	0.130	Minnesota	-0.079 (0.005)
2004	-0.015 (0.004)	0.125	Missouri	-0.068 (0.005)
2005	-0.021 (0.003)	0.171	New Hampshire	-0.073 (0.008)
2006	-0.031 (0.004)	0.242	New Jersey	-0.049 (0.004)
2007	-0.026 (0.004)	0.209	New York	-0.041 (0.004)
2008	-0.020 (0.004)	0.160	Ohio	-0.076 (0.004)
2009	-0.016 (0.004)	0.136	Oregon	-0.074 (0.005)
2010	-0.024 (0.004)	0.193	Pennsylvania	-0.079 (0.004)
2011	-0.022 (0.004)	0.174	Texas	-0.075 (0.004)
2012	-0.024 (0.004)	0.193	Virginia	-0.009 (0.004)
2013	-0.015 (0.004)	0.122	Washington	-0.073 (0.005)
2014	-0.021 (0.004)	0.169	Wisconsin	-0.081 (0.004)
			Arizona	Base

References

[1] Bee, Adam, Bruce D. Meyer, and James X. Sullivan (2012). "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?", *National Bureau of Economic Research*, 18308.

[2] Bureau of Labor Statistics (2011), "Appendix A: Description of the Consumer Expenditure Survey", *Consumer Expenditure Survey Anthology*, *US Bureau of Labor Statistics Report*, 1030.

[3] Hamilton, B. W. (2001). "Using Engel's Law to estimate CPI bias", *The American Economic Review*, 91(3), 619-630.