COST-OF-LIVING INDEX OF AN ESTIMATED GENERALIZED CONSTANT-ELASTICITY-OF-SUBSTITUTION UTILITY FUNCTION THAT ACCOUNTS FULLY FOR CHANGES IN PREFERENCES FOR AND QUALITIES OF GOODS*

Peter A. Zadrozny

Bureau Labor Statistics
Division of Price Index Number Research
2 Massachusetts Ave., NE, Room 3105,
Washington, DC 20212, USA
zadrozny.peter@bls.gov
peterzadrozny@gmail.com

May 20, 2022

Key words:
continuous-time modelling with discrete-time data,
computations with multi-step (homotopy) perturbation,
probabilistic Bayesian estimation

JEL codes: C51, C63, D12

*This work represents the author's views and does not necessarily represent any positions of the Bureau of Labor Statistics or its staff. The author is also affiliated as Research Fellow with the Center for Financial Studies (CFS), Goethe University, Frankfurt, Germany, and with the Center for Economic Studies and Ifo Institute for Economic Research (CESifo), Munich, Germany. The paper benefitted from conversations with Gregory Kurtzon and Randal Verbrugge.

ABSTRACT

The paper develops and illustrates a new method for computing a cost of living index (COLI) based on an estimated newly proposed generalized constant elasticity of substitution utility function (GCESUF) that has the advantages of (i) generality, (ii) practicality, (iii) accuracy, (iv) integratedness, and (v) comprehensiveness.

(i) The method is general because the GCESUF allows a wide range of elasticity responses to changed prices and expenditure levels, in particular, isn't necessarily homothetic so that expenditure shares can vary with expenditure levels.

(ii) The method has two broad implementation steps: (a) estimating a GCESUF and (b) using the estimated GCESUF to compute a GCES/COLI. The method is practical because both steps (a)-(b) can be numerically implemented easily, reliably, and quickly for any number of goods (and services) by evaluating analytically-derived scalar-level expressions.

(iii) By definition, a COLI holds if and only if the utility of a representative consumer is constant over the sample of data being considered. The method is accurate because it can implement computations (b) to any desired degree of numerical accuracy simply by passing through an already programmed double loop a pre-determined number of times. That is, there's no convergence involved with an open-ended number of loop passes. Apparently, no one else has proposed or illustrated any practical method for accurately computing a COLI for a general UF and a large number of goods

(iv) The method is integrated because all aspects of estimation (a) and computations (b) are formally derived based on microeconomic foundations, mathematics, and statistics (econometrics). There are no "ad hoc" or "intuitive" steps or jumps in (a)-(b). The method includes an integrated version of standard adjustment of a price index for changes in qualities of goods based on "hedonic" regression.

(v) The method is comprehensive because it simultaneously accounts for the effects on GCES/COLI of changes in consumer's preferences for goods and for producer's changes in qualities of goods.

The GCES/COLI method is illustrated with an application to monthly data on 8 aggregate categories of goods in the U.S. CPIU from January 1990 to December 2008. The resulting GCES/COLI lies about halfway between uniformly higher and lower Laspeyres and Törnqvist price indexes computed from scratch with the same data. GCES/COLI is also comprehensive by being to handle turnover of goods without essential change, which is explained but not illustrated in the paper because it's inappropriate for the 8 aggregate categories of goods.

1. Introduction.

1.1. General introduction.

By definition a cost of living index (COLI) is a price index (PI) that holds the utility of a representative consumer constant over a sample of data. Although being a COLI has been a touchstone of a PI and PIs have been claimed to be COLIs, the present paper appears to be the first one to propose a practical method for accurately computing a COLI on a large scale. Because consumers' preferences for goods (and services) and qualities of goods change over time, an accurate COLI computation should account for these changes, which the present method does using an estimated time-varying generalized constant-elasticity-of-substitution utility function (GCESUF). The COLI of the estimated GCESUF (GCES/COLI) is illustrated in the paper using U.S. monthly price and quantity data from the Bureau of Labor Statistics (BLS) on 8 aggregate categories of goods (and services) from January 1990 to December 2008.

The GCESUF is specified in terms of constant "substitution parameters" in matrix B, time-varying "preferences" in vector A_t, and time-varying "qualities" in vector R_t. B defines curvature of indifference curves (surfaces for more than two goods), contributes to the determination of slopes of indifference curves, and is, therefore, the major determinant of substitutions among consumed goods when their relative prices change. Because A_t and R_t contribute to the determination of slopes of indifference curves, as such they also contribute to the determination of substitutions among goods. Because "parameters" usually refers to constants, time-varying A_t and R_t are called "preferences" and "qualities" and not "preference parameters" and "quality parameters".

Ideally B should also be time varying but is treated here as a constant, because a constant B can be estimated consistently and efficiently (with minimum variance) by linear regression, but a time-

varying B can be estimated only using a more specialized, nonstandard, and nonlinear estimation method. First-order conditions (FOC) of a consumer's optimization problem for GCESUF are considered demand equations for goods, are the basis for linearly estimating elements of B (or nonlinearly estimating "deeper" parameters underlying B), preferences, and qualities, which and are, then, used to compute GCES/COLI by applying multi-step perturbation (MSP) to the estimated FOC and the data.

Chen and Zadrozny (2009) developed and applied MSP for a closely related productivity analysis. MSP has a preset approximation order k and a preset number of steps h and, therefore, doesn't involve an open-ended number of steps that need to converge. Related 1st- and 2nd-order solution methods were developed and applied for related consumption analyses by Vartia (1983), Breslaw & Smith (1995), and Dumagan & Mount (1997) and analyzed for accuracy by Sun & Xie (2013). Versions of MSP are used in physics, where they're called homotopy perturbation (He, 2006).

Laspeyres (1871), Paasche (1874), and Fisher (1922) PIs are commonly used PIs, none of which are generally accurate as COLIs, because they are based on observed quantities of goods and not on computed quantities of goods that hold utility of an underlying representative consumer constant over a sample of data. The probability is zero that utility of any utility function underlying a PI computed with observed quantities of goods is constant over a sample of data. Therefore, the probability is zero that a common PI computed with observed quantities of goods over a sample is accurate as a COLI.

The Divisia (1928) PI is methodologically close to GCES/COLI by also being derived using marginal first-order conditions of a consumer's optimization problem for any time-varying and differentiable utility function. Therefore, the Divisia PI implicitly accounts for changes in preferences for goods, but not for changes in qualities of goods. A key difference between GCES/COLI and Divisia PI is that GCES/COLI is a COLI because it's based on computed quantities of goods that hold utility constant over a sample of data, whereas the Divisia PI is

based on observed quantities of goods for which utility changes with probability one over a sample of data.

Because it's formulated in continuous time, the Divisia PI can't be computed directly with discrete-time data but is commonly approximated for discrete-time data by the Törnqvist (1936) PI. However, the Törnqvist PI was determined intuitively, was not derived from the Divisia PI, so that its accuracy as approximation of a Törnqvist PI is unclear without further analysis. Approximation error aside, like the other aforementioned PIs, the Törnqvist PI generally also isn't accurate as a COLI because it's computed with observed quantities of goods. Quantities of goods for which utility is constant over a sample of data can't be computed for the Törnqvist PI, because it has no known unique underlying utility function. The Törnqvist PI can be derived from a particular utility function such as Cobb-Douglas, but a unique utility function can't be derived from a Törnqvist PI.

1.2. Review of more recent literature.

GCESUF generalizes common utility functions such as Cobb-Douglas (CDUF; Cobb & Douglas, 1928) and Constant-Elasticity-of-Substitution (CESUF; Arrow et al., 1961) and is comparable to a translog utility function (TLUF; Christensen et al., 1975). CDUF implies unit price and expenditure-level elasticities, so that expenditure shares don't vary with prices or expenditure levels; CESUF generalizes this to any constant price elasticities, so that expenditure shares vary with prices but not with expenditure levels; TLUF implies a wider range of price and expenditure-level elasticities and, like GCES/COLI, allows non-unit expenditure-level elasticities. As discussed below, TLUF and GCESUF can be related to each other and, therefore, should have very similar elasticities, although no attempt is made here to verify this. Both GCESUF and TLUF can be considered utility functions in their own right or linear-quadratic approximations of more general utility functions.

Diewert's (1976) notion of a "superlative" PI has been a main criterion for over 45 years for deciding whether or not a PI is good or acceptable. A PI is defined as superlative if and only if it's an "exact PI" for a 2nd-order Taylor approximation of a true homothetic utility function. GCES/COLI has the following four advantages over a PI whose only justification is that it's superlative.

1. Saying that a PI is superlative is a general qualitative statement without specific numerical qualification. By contrast, GCES/COLI is a quantitative method that produces a PI whose accuracy as a COLI can be checked numerically.

2. Diewert's proof relies on an underlying and approximate utility function being homothetic, which is unrealistic, because a homothetic utility function implies that expenditure shares don't vary with expenditure levels, which is rejected by almost all data. GCESUF allows nonhomotheticity and, therefore, allows expenditure shares to vary with expenditure levels, as in the application in section 4.

3. A 2nd-order Taylor approximation of an underlying utility function is accurate only locally in some limited region around the center point of the approximation. If data range widely, then, a PI can be accurate superlatively over a range of data only by having multiple, different, approximating utility functions which seems to contradict the notion of being superlative. By contrast, as proved in appendix B, GCES/COLI can be made as globally accurate as desired over any range of data for the same GCESUF simply by computing with enough MSP steps.

4. Like the aforementioned common PIs, a superlative PI would be accurate as a COLI only if it were based on computed quantities of goods for which utility is held constant over a sample of data, but this has nothing to do with being superlative.

Like GCES/COLI, Gabor-Toth & Vermeulen (GTV; 2018) and Redding & Weinstein (RW; 2020) developed PIs from explicit solutions of optimal consumption problems for estimated utility functions that account for changes in preferences for goods, but, unlike GCES/COLI, don't also

account for changes in qualities of goods, GCES/COLI has two other advantages over GTV and RW's PIs.

First, GTV and RW used tiered (composite) CDUF and CESUF that, although apparently sufficiently general to adequately fit the data they used, inherit homotheticity from underlying CDUF and CESUF and, therefore, unrealistically imply unit expenditure-level elasticities for their PIs. Although their time-varying preferences account for time-varying expenditure shares, they do so spuriously to the extent that expenditure shares vary because the UF is nonhomothetic.

Second, because GTV and RW's PIs are computed using observed quantities of goods, their PIs can't generally be accurate as COLIs. In principle, GTV and RW could have computed quantities of goods that hold utility constant over a sample of data, but, doing this in the usual way in which they proceeded with closed-form utility functions would be practically impossible, because it would require solving a nonlinear problem for the n quantities of goods for each sample period. Doing this for the n = 8 goods in the application in section 4 would be difficult enough, much less for the hundreds of goods in GTV and RW or the thousands of goods in the U.S. consumer price index of urban consumers (CPIU).

GCES/COLI avoids this practical impossibility by solving an n+1-dimensional linear-equation system h times every sample period. GCES/COLI does this not by using a closed-form utility function but by using a closed-form marginal utility function that determines the differential of the utility function, setting the differential to zero, and computing GCES/COLI by numerically integrating the first-order conditions of a consumer's optimal consumption problem in differential form to periods of the discrete-time data. If, as in the application in section 4, matrix B of substitution parameters is diagonal, then, the n+1-dimensional linear-equation system has a known analytical solution of the form of equations (4.1)-(4.3), so that GCES/COLI can be computed easily, quickly, and accurately by evaluating scalar-level expressions, one for each good, one good at a

time. Thus, GCES/COLI practically solves the problem of accurately computing a COLI for any number of goods.

1.3. Continuation of the paper.

The paper continues as follows. Section 2 defines GCESUF and states a representative consumer's optimal consumption problem, including its first- and second-order conditions (FOC, SOC). Section 3 explains how the FOC are used to estimate parameters, preferences, and qualities of a GCESUF with discrete-time data and, then, how the estimates are used to compute GCES/COLI with MSP. Section 4 illustrates section 3 with U.S. monthly data on 8 aggregate categories of goods in the CPIU. Section 5 summarizes the paper and discusses four possible extensions: estimating a full (nondiagonal) matrix B, estimating a time-varying matrix B, handling turnover of goods, and computing a monetary-financial services index in the manner of GCES/COLI. Appendix A proves that inequality (2.13) is sufficient for SOC of a consumer's optimal consumption problem to hold. Appendix B reviews definitions, representations, and rules of matrix differentiation. Appendix C uses the differentiation rules to state the details of MSP computation of GCES/COLI. Appendix D proves that MSP has global accuracy of order h^{-k} for MSP order k and number of steps h.

2. Optimal consumption problem.

2.1. Definitions and notation.

Consider a sample of data spanning the continuous-time interval $\tau \in [1, T+1) = \{\tau | 1 \leq \tau < T+1\}$. The sample is divided into T periods of equal unit length as $[1, T+1) = \bigcup_{t=1}^{T} [t, t+1)$, so that $[t, t+1) = \{\tau | t \leq \tau < t+1\}$ is period t. A variable such as c is denoted at continuous moment τ by $c(\tau)$ and at the beginning of discrete period t by $c_t = c(t)$.

Let $dc(\tau) = (\partial c(\tau)/\partial \tau)d\tau$ denote the first differential of $c(\tau)$ with respect to continuous time τ, where $\partial c(\tau)/\partial \tau$ denotes the first-partial derivative of $c(\tau)$ with respect to τ. Let $\Delta_1 c_t = c_{t+1} - c_t$ denote the discrete-time 1-period forward difference of c_t, so that $\Delta_1 c_t = \int_{\tau=t}^{t+1} dc(\tau)$ by the fundamental theorem of calculus.

In particular, let $c(\tau) = \ln(C(\tau))$ denote the natural logarithm (log) of the cost of expenditures $C(\tau) = \sum_{i=1}^{n} P_i(\tau)Q_i(\tau)$ on n consumed goods (and services) at moment τ, where $P(\tau) = (P_1(\tau), \ldots, P_n(\tau))^T$ and $p(\tau) = (p_1(\tau), \ldots, p_n(\tau))^T$ denote $n \times 1$ vectors of observed prices of goods in original units and logs ($p_i(\tau) = \ln(P_i(\tau))$), where superscript T denotes vector or matrix transposition, and $Q(\tau) = (Q_1(\tau), \ldots, Q_n(\tau))^T$ and $q(\tau) = (q_1(\tau), \ldots, q_n(\tau))^T$ denote $n \times 1$ vectors of observed quantities of goods in original units and logs. Let $A(\tau) = (A_1(\tau), \ldots, A_n(\tau))^T$ and $a(\tau) = (a_1(\tau), \ldots, a_n(\tau))^T$ denote $n \times 1$ vectors of unobserved preferences for goods in original units and logs, let $R(\tau) = (R_1(\tau), \ldots, R_n(\tau))^T$ and $r(\tau) = (r_1(\tau), \ldots, r_n(\tau))^T$ denote $n \times 1$ vectors of unobserved qualities of goods in original units and logs, and let $S(\tau) = (S_1(\tau), \ldots, S_n(\tau))^T$ denote the $n \times 1$ vector of shares of expenditures on goods, $S_i(\tau) = P_i(\tau)Q_i(\tau)/C(\tau)$. Quantity $Q_i(\tau)$ denotes a count of units of good i, like the number of automobiles, quality $R_i(\tau)$ denotes a multiplicative quality-adjusting factor per unit of $Q_i(\tau)$, so that $V(\tau) = (V_1(\tau), \ldots, V_n(\tau))^T = (R_1(\tau)Q_1(\tau), \ldots, R_n(\tau)Q_n(\tau))^T$ denotes the $n \times 1$ vector of quality-adjusted goods in original units. All variables in original units, observed $C(\tau)$, $P_i(\tau)$, $Q_i(\tau)$, and $S_i(\tau)$ and unobserved $A_i(\tau)$ and $R_i(\tau)$, are positive real numbers. The general notational rule is that upper-case Latin letters denote variables in original units and corresponding lower-case Latin letters denote their natural logarithms.

2.2. First-order conditions.

Let $U(V(\tau))$ denote utility as a differentiable function of quality-adjusted goods. At every moment $\tau \in [1, T+1)$ in periods $t = 1,\ldots,T$ and for given utility function, $U(V(\tau))$, qualities of goods, $R(\tau)$, level of utility, $\bar{U}$, and prices of goods, $P(\tau)$, a representative consumer purchases and consumes quantities of goods, $Q(\tau)$, at minimum cost of expenditures, $C(\tau) = \sum_{i=1}^{n} P_i(\tau)Q_i(\tau)$, such that utility is at level $U(V(\tau)) = \bar{U}$.

The optimal consumption problem has $n+1$ first-order conditions (FOC): n marginal FOC (MFOC) and one constraint FOC (CFOC). The MFOC are

$$L(\tau)\cdot\partial U(V(\tau))/\partial Q_i(\tau) = P_i(\tau), \tag{2.1}$$

for $i = 1,\ldots,n$, where $L(\tau)$ denotes a Lagrange multiplier. The MFOC are derived in terms of quality-unadjusted $Q_i(\tau)$, because they are the consumer's decision variables that affect cost $C(\tau)$. The CFOC is

$$U(V(\tau)) = \bar{U}. \tag{2.2}$$

The term "utility function" is generally a misnomer here and is used for convenience, because GCESUF is specified in terms of marginal utilities that integrate to a utility function if and only if 2nd-order cross-partial derivatives (further derivatives of marginal utilities) are equal (Ford, 1955, pp. 137-141). The conditions are called "integrability conditions" and hold for GCESUF only in special cases such as when matrix B of substitution parameters is diagonal. For GCES/COLI derivation and computation it makes no difference whether the differential of utility is integrable or not, because GCES/COLI depends entirely on marginal utilities.

If a GCESU differential isn't integrable, then, integrals over different connected paths produce different values of changes in utility and a utility function doesn't exist. If a GCESU differential is integrable, then, every integral over a connected path from a first point in goods space to second point in goods space produces the same change in utility regardless of the path taken, so that a utility function can be defined and exists.

We define GCESUF marginal utility of good i in terms of quality-adjusted quantities $V_i(\tau) = R_i(\tau)Q_i(\tau)$ by

$$(2.3) \qquad \partial U(V(\tau))/\partial V_i(\tau) = A_i(\tau) \prod_{j=1}^{n} (R_j(\tau)Q_j(\tau))^{-\beta_{ij}},$$

for $i = 1,\ldots,n$. Marginal utility of good i increases with increased preference $A_i(\tau)$ and decreases with increased quality $R_i(\tau)$, because $\beta_{ii} > 0$.

The difference between $A_i(\tau)$ and $R_i(\tau)$ can be unclear. For example, a garment may be preferred or not (have high or low $A_i(\tau)$) whether it objectively has high quality or not (has high or low $R_i(\tau)$). Such ambiguity is ameliorated in section 3.2 by introducing correlation coefficient $\hat{\rho}_{it}$ between preferences for and qualities of a good. For example, if a garment is highly preferred and considered to have high quality, whether objectively it has high quality or not, then, we may expect $\hat{\rho}_{it}$ to have a positive and high value. In any case, the present probabilistic method for computing GCES/COLI implicitly takes any value of $\hat{\rho}_{it}$ into account without explicitly using it.

Off-diagonal elements β_{ij} of B measure by how much marginal utility of good i is affected by consumption of good j. Upon renumbering of goods, if necessary, "related" ($\beta_{ij} \neq 0$, $i \neq j$) and "unrelated" ($\beta_{ij} = 0$, $i \neq j$) goods make B block diagonal. In the application in section 4, attempts to estimate a more general than diagonal B failed, which makes sense for the large categories of goods. For example, we don't

expect utility from consumption of medical services to be affected by consumption of apparel.

To proceed, we need the marginal utility of good i with respect to $Q_i(\tau)$. The chain rule of differentiation, $\partial U(V(\tau))/\partial Q_i(\tau) = \partial U(V(\tau))/\partial V_i(\tau) \cdot \partial V_i(\tau)/\partial Q_i(\tau) = \partial U(V(\tau))/\partial V_i(\tau) \cdot R_i(\tau)$ and definition (2.3) imply that

$$(2.4) \qquad \partial U(V(\tau))/\partial Q_i(\tau) = F_i(\tau)\prod_{j=1}^{n} Q_j(\tau)^{-\beta_{ij}},$$

for $i = 1,\ldots,n$, where $F_i(\tau) = A_i(\tau)\prod_{j=1}^{n} R_j(\tau)^{\delta_{ij}-\beta_{ij}}$ and δ_{ij} denotes the Kronecker delta ($\delta_{ii} = 1$; $\delta_{ij} = 0$, for $i \neq j$). In vector-log form,

$$(2.5) \qquad f(\tau) = a(\tau) + (I_n - B)r(\tau),$$

where $f(\tau) = (f_1(\tau),\ldots,f_n(\tau))^T$, $f_i(\tau) = \ln(F_i(\tau))$, and I_n denotes the nxn identity matrix.

We also need marginal utilities in terms of logs $q_i(\tau)$ and $r_i(\tau)$. Multiplying equation (2.4) by $Q_i(\tau)$ implies that

$$(2.6) \qquad \partial U(V(\tau))/\partial q_i(\tau) = F_i(\tau)\prod_{j=1}^{n} Q_j(\tau)^{\delta_{ij}-\beta_{ij}} = \mu_i(\tau),$$

for $i = 1,\ldots,n$. Because $Q_i(\tau)$ and $R_i(\tau)$ enter $U(V(\tau))$ symmetrically as $Q_i(\tau)R_i(\tau)$,

$$(2.7) \qquad \partial U(V(\tau))/\partial r_i(\tau) = \partial U(\tau)/\partial q_i(\tau) = \mu_i(\tau).$$

GCESUF relates to the translog utility function (TLUF; Christensen et al., 1975): GCESUF is defined by $\ln(\partial U(V(\tau)/\partial Q(\tau)) = f(\tau) - Bq(\tau)$, where $\partial Q(\tau) = (\partial Q_1(\tau),\ldots,\partial Q_n(\tau))^T$, and TLUF may be defined correspondingly by

$\partial \ln(U(V(\tau))/\partial \ln Q(\tau) = f(\tau) - Bq(\tau)$, where $\partial \ln Q(\tau) = (\partial \ln Q_1(\tau), \ldots, \partial \ln Q_n(\tau))^T$, so that, whereas GCESUF is defined linearly in terms of logs of marginal utilities with respect to goods in original units, TLUF is defined linearly in terms of marginal logs of utilities with respect to logs of goods.

With restrictions on B, GCESUF approximates or equals standard utility functions. For example, if B is diagonal with large positive diagonal elements, then, GCESUF is approximately a fixed-proportions Leontief utility function; if $B = \{\beta_{ij}\} = \{\delta_{ij} - \gamma_i\}$, where $0 < \gamma_i < 1$ and $\sum_{i=1}^n \gamma_i = 1$, then, GCESUF is a CDUF; if B is scalar (diagonal with equal positive diagonal elements), then, GCESUF is a standard CESUF; if B is diagonal with unequal positive diagonal elements, then, GCESUF is the utility function in the present application.

2.3. First-order conditions in differential form.

Inserting marginal utilities (2.4) into MFOCs (2.1) and taking logs gives

$$e\lambda(\tau) + f(\tau) - Bq(\tau) = p(\tau), \tag{2.8}$$

where $e = (1, \ldots, 1)^T = nx1$ and $\lambda(\tau) = \ln(L(\tau))$. Differentiating MFOC (2.8) totally with respect to τ gives

$$ed\lambda(\tau) + df(\tau) - Bdq(\tau) = dp(\tau). \tag{2.9}$$

Differentiating CFOC (2.2) totally with respect τ and inserting marginal utilities (2.6)-(2.7) into the result gives

$$\sum_{i=1}^n \mu_i(\tau)(dr_i(\tau) + dq_i(\tau)) = 0. \tag{2.10}$$

Because MFOCs (2.9) imply that $\mu_i(\tau) = S_i(\tau)$ in CFOC (2.10), MFOCs (2.9) and CFOC (2.10) combine as the differential-equation system

$$(2.11)\qquad \begin{bmatrix} -\beta_{11} & \cdots & -\beta_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ -\beta_{n1} & \cdots & -\beta_{nn} & 1 \\ S_1(\tau) & \cdots & S_n(\tau) & 0 \end{bmatrix} \begin{bmatrix} dq_1(\tau) \\ \vdots \\ dq_n(\tau) \\ d\lambda(\tau) \end{bmatrix} = \begin{bmatrix} dp_1(\tau) - da_1(\tau) - \sum_{j=1}^{n} (\delta_{1j} - \beta_{1j}) dr_j(\tau) \\ \vdots \\ dp_n(\tau) - da_n(\tau) - \sum_{j=1}^{n} (\delta_{nj} - \beta_{nj}) dr_j(\tau) \\ -\sum_{j=1}^{n} S_j(\tau) dr_j(\tau) \end{bmatrix},$$

such that "endogenous" variables, $dq_i(\tau)$ and $d\lambda(\tau)$, are determined by solving system (2.11) using MSP, for given values of "exogenous" variables, $dp_i(\tau)$, $da_i(\tau)$, and $dr_i(\tau)$.

The first-n scalar MFOC equations in system (2.11) are linear in variables, because their coefficients β_{ij}, 0, and 1 are constant, i.e., have the same values for all values of τ; the last-(n+1)th scalar CFOC equation in system (2.11) is nonlinear, because $S_i(\tau)$ depend nonlinearly on τ. The first-n scalar MFOC equations in MSP-integrated discrete-time system (2.11) are the key to estimating B and $\{a_{it}, r_{it}\}_{t=1}^{T}$ consistently and efficiently and computing $\{\text{GCES/COLI}_t\}_{t=1}^{T}$ with discrete-time data.

2.3. Second-order conditions.

Second-order conditions (SOC) must hold in order for a solution of the consumer's optimal consumption problem at the FOC to be locally unique. Having the SOC hold at the FOC is also necessary for GCES/COLI to be computed reliably and accurately using MSP. The necessary and sufficient SOC of a representative consumer's optimal consumption problem came to economists from Mann (1943) via Samuelson (1947) and impose alternating sign restrictions on the principal minors of system matrix

$$(2.12) \qquad \bar{B}(\tau) = \begin{bmatrix} -B & e \\ S(\tau)^T & 0 \end{bmatrix}$$

of system (2.11).

As a slightly more restrictive but much easier to enforce SOC, we assume that

(2.13) $\bar{B}(\tau)$ is negative definite ($\bar{B}(\tau) < 0$).

If $\bar{B}(\tau)$ is negative definite, then, B is positive definite (PD; $B > 0$). A real square matrix is PD if and only if its eigenvalues are real and positive. Usually, real square PD matrices are symmetric, but not necessarily according to this definition.

Appendix A proves that

(2.14) If $\underline{\lambda}(B) > \sqrt{1/n + 2 + n}$, then, $\bar{B}(\tau)$ is negative definite,

where $\underline{\lambda}(B)$ denotes the smallest (presumably, real and positive) eigenvalue of B. Because the diagonal elements of a diagonal matrix are its eigenvalues, table 2 shows that $\underline{\lambda}(\hat{B}) = 8.04 > 3.18 = \sqrt{1/8 + 2 + 8}$, so that condition (2.14) easily holds in the application in section 4.

3. Estimation of GCESUF and computation of GCES/COLI.

3.1. Estimation of substitution parameters in B.

We now discuss three possible ways of estimating substitution parameters in B: unrestricted ordinary-least-squares estimation (ULSE), restricted LSE (RLSE), and maximum likelihood estimation

(MLE). We consider the three estimation methods because they have the following tradeoffs.

ULSE is easy to implement and quick to execute, but may produce estimates that don't satisfy the SOC that are necessary for MSP computation of GCES/COLI to be reliable and accurate. RLSE can be as easy to implement and quick to execute as ULSE, but its restrictions may be considered arbitrary, hence, its estimates may be considered biased.

MLE has theoretical advantages and computational disadvantages. MLE can handle a wide range of restrictions on parameters in a statistically coherent way. Depending on restrictions, MLE can estimate parameters either sequentially and more easily, one scalar estimating equation at a time, or only simultaneously and with more difficulty, using all estimating equations simultaneously. In practice, MLE iterations may converge too slowly and are stopped or may diverge and stop by themselves due to overflowed computations. These difficulties with and failures of MLE tend to occur when "too many" parameters are poorly identified. However, MLE was used successfully in the application in section 4 to estimate the 8 parameters of diagonal B.

3.1.1. ULSE of substitution parameters in B.

The standard hedonic regression equation for adjusting observed prices of goods for their observed qualities is

$$(3.1) \qquad p_t = \Phi z_t + \zeta_t,$$

for $t = 1,\ldots,T$, where Φ denotes an nxm matrix of constant coefficients, z_t denotes an mx1 vector of observed qualities or supply-determined variables in log form that don't necessarily directly measure qualities but indicate them, and ζ_t denotes an nx1

vector of unobserved disturbances that are assumed to be uncorrelated with z_t, so that z_t is exogenous in equation (3.1).

In usual hedonic adjustment, estimated $\hat{\Phi}$ is obtained by applying ULSE to each scalar equation in vector equation (3.1) and is then used to adjust prices of goods for their qualities as $\tilde{p}_t = p_t - \hat{\Phi} z_t$ (Coats, 1939; Griliches, 1961), where a hat (^) denotes a value estimated using all data, $\{p_t, q_t, z_t\}_{t=1}^{T}$, and the absence of a hat denotes an observed value or an unobserved true value.

The more pertinent quality-accounting equation for GCES/COLI is

$$r_t = \Phi z_t + \zeta_t, \tag{3.2}$$

for t = 1,…,T. Although Φ can't be estimated directly using equation (3.2) because r_t is unobserved, Φ can be estimated indirectly using

$$q_t = \Pi p_t + \Xi z_t + \eta_t, \tag{3.3}$$

obtained by replacing r_t in the discrete-time form of MFOC (2.8) with the right side of equation (3.2) and rewriting the result, so that $\Pi = -B^{-1}$, $\Xi = (B^{-1} - I_n)\Phi$, $\eta_t = B^{-1}(e\lambda_t + a_t) + (B^{-1} - I_n)\zeta_t$ = nx1 vector of unobserved disturbances.

We assume that $(p_s^T, z_s^T)^T$ and η_t are uncorrelated in equation (3.3) within and across all periods,

$$E(p_s^T, z_s^T)^T \eta_t^T = 0_{(n+m)xn}, \tag{3.4}$$

for s,t = 1,…,T, where E denotes unconditional expectation, so that p_t and z_t are exogenous in equation (3.3) and the estimation methods discussed here yield consistent estimates $\hat{\Pi}$, $\hat{\Xi}$, $\hat{B} = -\hat{\Pi}^{-1}$, and $\hat{\Phi} = (\hat{B}^{-1} - I_n)^{-1}\hat{\Xi}$.

GCES/COLI is more accurate to the extent that z_t accounts for more of changes in qualities of goods. In fact, the discussion at the end of section 3.2 indicates that, if z_t accounts for all changes in qualities of all goods in every period (which never happens in practice), and the estimated GCESUF is an accurate estimate of the true UF of the representative consumer, then, the computed GCES/COLI can be considered an accurate estimate of the true COLI of the representative consumer.

Condition (3.4) in effect says that p_t and Φz_t are determined by suppliers (firms) and a_t, λ_t, and ζ_t are determined independently by demanders (consumers). Whereas the usual hedonic adjustment doesn't account for the effects of unobserved and unaccounted for ζ_t on a PI, the GCES/COLI method accounts probabilistically for the effects of ζ_t on GCES/COLI.

If condition (3.4) is questioned, then, a simultaneous-equations supply-demand equilibrium-analysis can be added. Scalar equations in vector equation (3.3) are demand-for-goods equations and supply-of-goods equations would need to be introduced. Gabor-Toth & Vermeulen (2018) and Redding & Weinstein (2020) estimated such demand-supply equilibria of prices and quantities for disaggregated categories of goods. Such demand-supply equilibria aren't estimated in the application in section 4, because they wouldn't be convincing for the large and heterogeneous categories of goods in the application.

If Π and Ξ are estimated with equation (3.3) and implied estimated $\hat{B} = \hat{\Pi}^{-1}$ satisfies the SOC of GCES/COLI, then, estimated $\hat{\Pi}$ and $\hat{\Xi}$ are statistically efficient (asymptotically unbiased with minimum variance), because $\hat{\Pi}$ satisfies all GCES/COLI restrictions on $\hat{B}$ and there aren't any GCES/COLI restrictions on $\hat{\Phi}$, hence, on $\hat{\Xi}$. Differentiated CFOC (2.10) imposes no restrictions on B, because MFOC (2.8) imply that $\mu_{it} = S_{it}$ in CFOC (2.10) for any differentiable utility function and any data.

We assume that disturbance vector η_t in demand equation (3.3) is distributed normally, identically, and independently, with zero mean vector and constant positive-definite covariance matrix Σ_η (> 0),

(3.5) $\eta_t \sim \text{NIID}(0, \Sigma_\eta)$,

for t = 1,…,T. Distributional assumptions like (3.5) are necessary to implement MLE and imply the usual desired asymptotic properties of parameter estimates (Theil, 1971).

In the application in section 4, graphs 1-4 show that prices and quantities have different trends and seasonalities that are mostly removed by double monthly-and-annual differencing, $\Delta_{12}\Delta_1$, where Δ_1 denotes 1-month forward differencing and Δ_{12} denotes 12-month forward differencing. Therefore, in the application, B was estimated with equation (3.3) in the differenced form

(3.6) $\Delta_{12}\Delta_1 q_t = \Pi\Delta_{12}\Delta_1 p_t + \Delta_{12}\Delta_1 \eta_t$,

with quality-accounting term $\Xi\Delta_{12}\Delta_1 z_t$ omitted.

To check that $\Delta_{12}\Delta_1$ differencing rendered residual vector $\Delta_{12}\Delta_1\hat{\eta}_t$ sufficiently IID, equation (3.6) was estimated undifferenced with η_t specified as a vector moving average (VMA) process with B and the VMA parameters estimated simultaneously by MLE. Because this resulted in effectively the same estimated $\hat{B}$, the reestimation wasn't considered further and isn't reported.

Differenced equation (3.6) is consistent with MFOC (2.8) if B is diagonal or if every scalar equation in vector equation (3.3) is differenced in the same way, which is the case in the application in section 4 for both of these reasons. Although in the application B was estimated with equation (3.6) using differenced and standardized data, GCES/COLI was computed with original, undifferenced, and

unstandardized data, so that computed GCES/COLI includes trends and seasonalities removed by the differencing.

3.1.2. RLSE of substitution parameters in B.

To compute GCES/COLI reliably and accurately, $\hat{B}$ must satisfy the SOC, but might not unless $\hat{\Pi}$ is restricted. In order for $\hat{B}$ to satisfy sufficient-for-SOC inequality (2.14), its eigenvalues must be real, positive, and sufficiently large. Because the eigenvalues of $\hat{B}$ are reciprocals of the eigenvalues of $\hat{\Pi}$, $\hat{B}$ satisfies inequality (2.14) if and only if the eigenvalues of $\hat{\Pi}$ are real, positive, and not too large.

To ensure that the eigenvalues of $\hat{\Pi}$ are real, $\hat{\Pi}$ can be restricted to be exactly or nearly symmetric. Compared with exact restrictions, weaker approximate restrictions lessen any perceived bias of estimates due to the restrictions being considered arbitrary. Approximate linear restrictions on coefficients can be imposed using Theil & Goldberger's (1961) mixed estimation (TGME), which amounts to adding "pseudo" data to actual data and estimating Π and Ξ exactly as in ULSE. The pseudo data reflect probabilistically specified restrictions according to specified "tightness" parameters, so that TGME is effectively a Bayesian estimation (Theil, 1971, pp. 346-351).

If exactly or nearly symmetric $\hat{\Pi}$ has all real but some negative eigenvalues, then, $\hat{\Pi}$ can also be "regularized" as in ridge regression (Hoerl & Kennard, 1970a,b) to have all positive eigenvalues. This may require some trial and error, which may cause the final estimated $\hat{\Pi}$ to be considered somewhat arbitrary. If not too many elements of B and Φ need to estimated, then, overall it may be easier, quicker, and statistically more coherent to estimate using MLE.

3.1.3. MLE of substitution parameters in B.

Assumptions (3.4)-(3.5) imply that MLE minimizes $|\hat{\Sigma}_{\eta}| = |\sum_{t=1}^{T} \hat{\eta}_t \hat{\eta}_t^T/T|$ with respect to elements of or parameters underlying B, restricted so that SOC hold, and elements of Φ, where $\hat{\eta}_t$ denotes the residual in estimated equation (3.3). If, as in the application in section 4, differenced data are used, then, MLE minimizes $|\hat{\Sigma}_{\Delta_{12}\Delta_1\eta}| = |\sum_{t=1}^{T} \Delta_{12}\Delta_1\hat{\eta}_t \cdot \Delta_{12}\Delta_1\hat{\eta}_t^T/T|$. In the application, SOC were enforced by parameterizing diagonal B as $\beta_{ii} = \rho + \theta_i^2$, for i = 1,...,8 and $\rho = 10^{-8}$, and estimating θ_i by MLE.

3.2. Estimation of preferences and qualities.

We assume that $\{f_t\}_{t=1}^{T}$ are identified by

(3.7) $\quad e^T f_t = 0,$

for t = 1,…,T, where $e = (1,\ldots,1)^T = nx1$. Restrictions (3.7) only prevent monotonic transformations of utility and leave indifference curves and optimal consumption of goods unchanged. Gabor-Toth & Vermeulen (2018) and Redding & Weinstein (2020) imposed the same restrictions for the same reason.

Suppose that $\{q_t, \lambda_t\}_{t=1}^{T}$ are at optimal values. If $\{f_t\}_{t=1}^{T}$ increase to $\{\ln(2)e + f_t\}_{t=1}^{T}$, then, optimal consumption of goods is unchanged and the only change in the solution of the optimal consumption problem is that $\{\lambda_t\}_{t=1}^{T}$ decline to $\{\ln(.5) + \lambda_t\}_{t=1}^{T}$. Condition (3.7) precludes such common changes of $\{a_t, r_t\}_{t=1}^{T}$ and also precludes changes in utility for "no apparent reasons".

Consider MFOC (2.11) in estimated discrete-time form, premultiply it by $(1/n)e^T$, apply identifying condition (3.7), and obtain estimated

$$(3.8) \qquad \hat{\lambda}_t = (1/n)e^T(\hat{B}q_t + p_t).$$

Replace $\hat{\lambda}_t$ in estimated discrete-time MFOC with the right side of equation (3.8) and obtain estimated

$$(3.9) \qquad \hat{f}_t = M(\hat{B}q_t + p_t),$$

for $t = 1,\ldots,T$, where $M = I_n - (1/n)ee^T$. Premultiplying an n-row vector or matrix by M transforms it into deviations from its column averages.

As in the application in section 4, assume that $\hat{B} = \mathrm{diag}(\hat{\beta}_{11},\ldots,\hat{\beta}_{nn})$ is diagonal with $\hat{\beta}_{ii} > 1$. Then, estimated discrete-time row-wise equation (2.5) is

$$(3.10) \qquad \hat{f}_{it} = \hat{a}_{it} + (1-\hat{\beta}_{ii})\hat{r}_{it},$$

for $i = 1,\ldots,n$ and $t = 1,\ldots,T$.

If $da(\tau)$ and $dr(\tau)$ entered FOC (2.9)-(2.10) in the same way through $df(\tau)$, then, $\{GCES/COLI_t\}_{t=1}^T$ could be computed without knowing the separate values of $\{\hat{a}_t\}_{t=1}^T$ and $\{\hat{r}_t\}_{t=1}^T$. However, because this isn't the case, because CFOC (2.10) depends only on $dr(\tau)$, $\{GCES/COLI_t\}_{t=1}^T$ can be computed only by knowing the separate values of $\{\hat{a}_t\}_{t=1}^T$ and $\{\hat{r}_t\}_{t=1}^T$.

We now describe a probabilistic method for estimating the separate values of $\{\hat{a}_t\}_{t=1}^T$ and $\{\hat{r}_t\}_{t=1}^T$. Row-wise estimated equation (3.2) is

$$(3.11) \qquad \hat{r}_{it} = \hat{\phi}_i^T z_t + \hat{\zeta}_{it},$$

for i = 1,…,n and t = 1,…,T, where $\hat{\phi}_i^T$ denotes the ith row of $\hat{\Phi}$. We consider $\hat{\phi}_i^T z_t$ as known (abstracting from sampling variability of $\hat{\phi}_i$) and attribute all or most uncertainty about accuracy of $\hat{a}_{it}$ and $\hat{r}_{it}$ as arising from $\hat{\zeta}_{it}$. The probabilistic method first determines $\hat{\zeta}_{it}$, then, determines $\hat{r}_{it}$ using equation (3.11) and, then, determines $\hat{a}_{it}$ using equation (3.10). The reverse order of determination can be used if $\hat{\beta}_{ii} \neq 1$ (more generally, if $\hat{B} \neq I_n$); otherwise, $\hat{r}_{it}$ can't be determined from $\hat{a}_{it}$.

Combined equations (3.10)-(3.11) imply that

$$\tilde{f}_{it} = \hat{a}_{it} + (1-\hat{\beta}_{ii})\hat{\zeta}_{it}, \tag{3.12}$$

where $\tilde{f}_{it} = \hat{f}_{it} - (1-\hat{\beta}_{ii})\hat{\phi}_i^T z_t$. Conditional on all data, $\{p_t, q_t, z_t\}_{t=1}^T$, the estimates $\{\hat{f}_{it}\}_{i=1t=1}^{n\ T}$, $\{\hat{\beta}_i, \hat{\phi}_i\}_{i=1t=1}^{n\ T}$, and $\{\tilde{f}_{it}\}_{i=1t=1}^{n\ T}$ are known. Therefore, conditional on all data, the expectation of the square of equation (3.12) is

$$\tilde{f}_{it}^2 = \sigma_{\hat{a}it}^2 + 2\hat{\rho}_{it}(1-\hat{\beta}_{ii})\sigma_{\hat{a}it}\sigma_{\hat{\zeta}it} + (1-\hat{\beta}_{ii})^2\,\sigma_{\hat{\zeta}it}^2, \tag{3.13}$$

where $\sigma_{\hat{a}it}^2 = \bar{E}\hat{a}_{it}^2$, $\sigma_{\hat{\zeta}it}^2 = \bar{E}\hat{\zeta}_{it}^2$, $\hat{\rho}_{it} = \bar{E}(\hat{a}_{it}\hat{\zeta}_{it})/\sigma_{\hat{a}it}\sigma_{\hat{\zeta}it}$, and $\bar{E}$ denotes expectation conditional on all data (not unconditional expectation, E, without the overbar). The Cauchy-Schwarz inequality implies that $\hat{\rho}_{it} \in [-1,1]$.

Square roots of second moments about zero, $\sigma_{\hat{a}it}$ and $\sigma_{\hat{\zeta}it}$, must be nonnegative real numbers, which equation (3.13) implies occurs if and only if, respectively,

$$\sigma_{\hat{\zeta}it}^2 \in [0, \tilde{f}_{it}^2/(1-\beta_{ii})^2], \quad \sigma_{\hat{a}it}^2 \in [0, \tilde{f}_{it}^2]. \tag{3.14}$$

Because squared means and variances sum to second moments about zero,

$$(3.15) \qquad \sigma^2_{\hat{a}it} = \mu^2_{\hat{a}it} + \upsilon^2_{\hat{a}it}, \quad \sigma^2_{\hat{\zeta}it} = \mu^2_{\hat{\zeta}it} + \upsilon^2_{\hat{\zeta}it},$$

where $\mu_{\hat{a}it} = \bar{E}\hat{a}_{it}$, $\upsilon^2_{\hat{a}it} = \bar{E}(\hat{a}_{it} - \mu_{\hat{a}it})^2$, $\mu_{\hat{\zeta}it} = \bar{E}\hat{\zeta}_{it}$, and $\upsilon^2_{\hat{\zeta}it} = \bar{E}(\hat{\zeta}_{it} - \mu_{\hat{\zeta}it})^2$. Thus, the data via $\tilde{f}_{it}$ and $\hat{\beta}_{ii}$ restrict means and variances of $\hat{a}_{it}$ and $\hat{\zeta}_{it}$.

Whereas the values of $\tilde{f}_{it}$ and $\hat{\beta}_{ii}$ are estimated using data, we proceed as if nothing is known about the separate values of $\hat{a}_{it}$ and $\hat{\zeta}_{it}$. Reflecting this lack of knowledge, we assume that $\upsilon^2_{\hat{\zeta}it}$ is at its maximum, so that equations (3.14)-(3.15) imply that $\mu^2_{\hat{\zeta}it} = 0$ and $\upsilon^2_{\hat{\zeta}it} = \tilde{f}^2_{it}/(1-\hat{\beta}_{ii})^2$.

We further assume that $\hat{\zeta}_{it}$ is distributed uniformly over the finite interval $[-\hat{\gamma}_{it}, \hat{\gamma}_{it}]$ or $\hat{\zeta}_{it} \sim UD[-\hat{\gamma}_{it}, \hat{\gamma}_{it}]$. An elementary integration shows that $\upsilon^2_{\hat{\zeta}it} = \tilde{f}^2_{it}/3(1-\hat{\beta}_{ii})^2$, which implies that $\hat{\gamma}_{it} = \sqrt{3}\,|\tilde{f}_{it}/(1-\hat{\beta}_{ii})|$. Assuming that $\hat{\zeta}_{it} \sim UD[-\hat{\gamma}_{it}, \hat{\gamma}_{it}]$ corresponds to assuming an uninformed prior distribution in Bayesian estimation. We could instead assume that $\hat{\zeta}_{it}$ is distributed unboundedly, so that $\hat{\zeta}_{it} \in [-\hat{\gamma}_{it}, \hat{\gamma}_{it}]$ with some probability less than one.

To compute the <u>fully probabilistic estimate</u> of GCES/COLI, first, for some chosen $\bar{J} \geq 1$, randomly (pseudo randomly in practice) draw $nT\bar{J}$ values from $UD[-\hat{\gamma}_{it}, \hat{\gamma}_{it}]$ and assign them to $\{\hat{\zeta}^{(j)}_{it}\}^{n}_{i=1}{}^{T}_{t=1}{}^{\bar{J}}_{j=1}$; then, use the obtained $\{\hat{\zeta}^{(j)}_{it}\}^{n}_{i=1}{}^{T}_{t=1}{}^{\bar{J}}_{j=1}$ and equations (3.10)-(3.11) to compute $\bar{J}$ values of $\{\hat{a}_t, \hat{r}_t\}^{T}_{t=1}$ denoted $\{\hat{a}^{(j)}_t, \hat{r}^{(j)}_t\}^{T}_{t=1}{}^{\bar{J}}_{j=1}$; then, use the computed values

of $\{\hat{a}_t^{(j)}, \hat{r}_t^{(j)}\}_{t=1}^{T}{}_{j=1}^{\bar{J}}$ to compute $\bar{J}$ values of $\{GCES/COLI_t\}_{t=1}^{T}$ denoted $\{GCES/COLI_t^{(j)}\}_{t=1}^{T}{}_{j=1}^{\bar{J}}$, which comprise a $\bar{J}$-value and T-dimensional distribution. Then, the fully probabilistic estimate of $\{GCES/COLI_t\}_{t=1}^{T}$ is the arithmetic average of $\{GCES/COLI_t^{(j)}\}_{t=1}^{T}$ over $j = 1,\ldots,\bar{J}$,

$$(3.16) \qquad \{\bar{C}_t\}_{t=1}^{T} = \{(\textstyle\sum_{j=1}^{\bar{J}} GCES/COLI_t^{(j)})/\bar{J}\}_{t=1}^{T}.$$

If $nT\bar{J}$ is large enough so that the time required to compute $\{\bar{C}_t\}_{t=1}^{T}$ is considered too long, then, the following approximately probabilistic estimate of $\{\bar{C}_t\}_{t=1}^{T}$ can be computed more quickly. Define four cases of variations of $\{\hat{a}_{it}, \hat{\xi}_{it}\}_{i=1}^{n}{}_{t=1}^{T}$:

$$(3.17) \quad \{C_t^{(a,-|\gamma|)}\}_{t=1}^{T} = \{GCES/COLI_t\}_{t=1}^{T} \text{ for } \{\hat{a}_{it}, \hat{\xi}_{it}\}_{i=1}^{n}{}_{t=1}^{T} = \{\hat{a}_{it}, -|\hat{\gamma}_{it}|\}_{i=1}^{n}{}_{t=1}^{T},$$

$$(3.18) \quad \{C_t^{(a,0)}\}_{t=1}^{T} = \{GCES/COLI_t\}_{t=1}^{T} \text{ for } \{\hat{a}_{it}, \hat{\xi}_{it}\}_{i=1}^{n}{}_{t=1}^{T} = \{\gamma_{it}, 0\}_{i=1}^{n}{}_{t=1}^{T},$$

$$(3.19) \quad \{C_t^{(a,|\gamma|)}\}_{t=1}^{T} = \{GCES/COLI_t\}_{t=1}^{T} \text{ for } \{\hat{a}_{it}, \hat{\xi}_{it}\}_{i=1}^{n}{}_{t=1}^{T} = \{\hat{a}_{it}, |\hat{\gamma}_{it}|\}_{i=1}^{n}{}_{t=1}^{T},$$

$$(3.20) \quad \{C_t^{(0,0)}\}_{t=1}^{T} = \{GCES/COLI_t\}_{t=1}^{T} \text{ for } \{\hat{a}_{it}, \hat{\xi}_{it}\}_{i=1}^{n}{}_{t=1}^{T} = \{0, 0\}_{i=1}^{n}{}_{t=1}^{T}.$$

In equations (3.17)-(3.19, $\hat{a}_{it}$ is given by equation (3.12), respectively, for $\hat{\xi}_{it} = -|\hat{\gamma}_{it}|$, $\hat{\xi}_{it} = 0$, and $\hat{\xi}_{it} = |\hat{\gamma}_{it}|$, for all i and t; in equation (3.20), both $\hat{a}_{it}$ and $\hat{\xi}_{it}$ are zero for all i and t.

Although $\{C_t^{(0,0)}\}_{t=1}^{T}$ contradicts data, because $\{\tilde{f}_t\}_{t=1}^{T} = \{0_{nx1}\}_{t=1}^{T}$ never holds in practice, $\{C_t^{(0,0)}\}_{t=1}^{T}$ is useful because it shows how $\{GCES/COLI_t\}_{t=1}^{T}$ changes when accounting for changes in preferences for

and qualities of goods is shut down. The application in section 4 reports $\{\tilde{\tilde{C}}_t\}_{t=1}^T$ for $\tilde{\tilde{J}} = 3$ and $\{C_t^{(0,0)}\}_{t=1}^T$ and discusses their differences.

Choose some $\tilde{\tilde{J}} \geq \max(3,n)$. For $i = 1,\ldots,n$ and $t = 1,\ldots,T$, partition $[-\hat{\gamma}_{it}, \hat{\gamma}_{it}]$ into $j = 1,\ldots,\tilde{\tilde{J}}$ equally-spaced points, including endpoints $\pm\hat{\gamma}_{it}$ and midpoint 0. Assign each of the $\tilde{\tilde{J}}$ points to $\{\hat{\zeta}_{it}^{(j)}\}_{i=1}^n{}_{t=1}^T{}_{j=1}^{\tilde{\tilde{J}}}$, necessarily with repetition because $nT\tilde{\tilde{J}} \geq \tilde{\tilde{J}}$. The assignments can be pseudo random or nonrandom in any order but preferably with minimal repetition.

Then, the <u>approximate</u> <u>probabilistic</u> <u>estimate</u> of $\{GCES/COLI_t\}_{t=1}^T$ denoted by $\{\tilde{\tilde{C}}_t\}_{t=1}^T$ is the arithmetic average of the three cases (3.17)-(3.19) with $\{\hat{\zeta}_{it}^{(j)}\}_{i=1}^n{}_{t=1}^T{}_{j=1}^{\tilde{\tilde{J}}}$ determined in each case as explained above.

We now derive an approximate upper bound for differences between true and estimated $\ln(GCES/COLI_t)$ due to differences between true ζ_t and estimated $\hat{\zeta}_t$. So far $c_t = \ln(C_t)$ for $C_t = \sum_{i=1}^n P_{it}Q_{it}$ and any observed $\{Q_{it}\}_{i=1}^n$, but now, more specifically, let $c_t = \ln(GCES/COLI_t)$. Also, let $\{\hat{c}_t\}_{t=1}^T$ = either $\{\bar{c}_t\}_{t=1}^T$ or $\{\tilde{\tilde{c}}_t\}_{t=1}^T$, where $\bar{c}_t = \ln(\bar{C}_t)$ and $\tilde{\tilde{c}}_t = \ln(\tilde{\tilde{C}}_t)$ and let Δ (not Δ_1 and Δ_{12} defined in section 3.1) denote differences between true and estimated (computed) $\ln(GCES/COLI_t)$, so that $\Delta c_t = c_t^* - \hat{c}_t$ and $\Delta\zeta_t = \zeta_t^* - \hat{\zeta}_t$, where * denotes a true value. The ζ_t are implicitly in log form. Computing $\hat{c}_t$ means computing the mapping from data to $\{\hat{c}_t\}_{t=1}^T$, which includes the 3-part composite submapping $\hat{c}_t = F(G(H(\hat{\zeta}_t)))$, where $\hat{c}_t = F(\tilde{q}_t)$, $\tilde{q}_t = G(\tilde{f}_t)$, and $\tilde{f}_t = H(\hat{\zeta}_t)$. The tilded $\{\tilde{q}_t\}_{t=1}^T$ denote computed logs of quantities of goods according to MSP computation of GCES/COLI.

Because the mappings $F(\tilde{q}_t)$, $G(\tilde{f}_t)$, and $H(\hat{\zeta}_t)$ are twice differentiable, $\Delta c_t = \Delta F(G(H(\hat{\zeta}_t)))$ has the first-order Taylor approximation (Apostol, 1974, pp. 113-114), given by

$$(3.21) \qquad \Delta c_t \cong \nabla F(\tilde{q}_t)\nabla G(\tilde{f}_t)\nabla H(\hat{\zeta}_t)\Delta\zeta_t ,$$

where $\nabla F(\tilde{q}_t) = (\tilde{S}_{1t},\ldots,\tilde{S}_{nt})$ is the 1xn gradient vector of first-partial derivatives of $\ln(C_t) = \ln(\sum_{i=1}^{n} P_{it}Q_{it})$ with respect to elements of $\tilde{q}_t = (\tilde{q}_{1t},\ldots,\tilde{q}_{nt})^T$, tilded $\tilde{S}_{it}$ denote expenditure shares computed with observed p_{it} and computed $\tilde{q}_{it}$, $\nabla G(\tilde{f}_t) = \begin{bmatrix} \partial\tilde{q}_{1t}/\partial\tilde{f}_{1t} & \cdots & \partial\tilde{q}_{1t}/\partial\tilde{f}_{nt} \\ \vdots & & \vdots \\ \partial\tilde{q}_{nt}/\partial\tilde{f}_{1t} & \cdots & \partial\tilde{q}_{nt}/\partial\tilde{f}_{nt} \end{bmatrix}$ is the nxn Jacobian matrix of first-partial derivatives of $G(\tilde{f}_t)$ with respect to elements of $\tilde{f}_t$, and $\nabla H(\hat{\zeta}_t) = I_n - \hat{B}$ is the nxn Jacobian matrix of first-partial derivatives of $H(\hat{\zeta}_t)$ with respect to the elements of $\hat{\xi}_t = (\hat{\xi}_{1t},\ldots,\hat{\xi}_{nt})^T$.

The matrix-product-norm rule applied to equation (3.21) implies that $\|\nabla F(\tilde{q}_t)\nabla G(\tilde{f}_t)\nabla H(\hat{\zeta}_t)\Delta\zeta_t\| \leq \|\nabla F(\tilde{q}_t)\|\cdot\|\nabla G(\tilde{f}_t)\|\cdot\|\nabla H(\hat{\zeta}_t)\Delta\zeta_t\|$, where $\|X\| = \max\{|x_{ij}|\}_{i=1}^{m}{}_{j=1}^{n}\}$ denotes the ∞-norm of an mxn vector or matrix X (Golub & Van Loan, 1983, pp. 12-13). For $\nabla h_i(\hat{\zeta}_t)\Delta\zeta_t \in [-\sqrt{3}|\tilde{f}_{it}|, \sqrt{3}|\tilde{f}_{it}|]$, where $\nabla H(\hat{\zeta}_t) = (\nabla h_1(\hat{\xi}_{1t}),\ldots,\nabla h_n(\hat{\xi}_{nt}))$, equation (3.21) implies that

$$(3.22) \qquad |\Delta\hat{c}_t| \leq 2\sqrt{3}\cdot\max\{\tilde{S}_{it}\}_{i=1}^{n}\cdot\max\{|\partial\tilde{q}_{it}/\partial\tilde{f}_{jt}|\}_{i=1}^{n}{}_{j=1}^{n}$$

$$\cdot\max\{|\delta_{ij} - \hat{\beta}_{ij}|\}_{i=1}^{n}{}_{j=1}^{n}\cdot\max\{|\tilde{f}_{it}|\}_{i=1}^{n} .$$

Although inequality (3.22) has the advantage of being easy to derive and is robust because it's derived using a minimum of properties of $\hat{c}_t$, it's also the "most pessimistic" upper bound of accuracy of $\hat{c}_t$.

In the application in section 4, sample averages of shares of expenditures on goods are $(\bar{S}_1,...,\bar{S}_8)$ = (.103, .114, .122, .137, .127, .155, .114, .128), so that we consider $\max\{\tilde{S}_{it}\}_{i=1}^{8}$ = .155. Table 2 implies that $\max\{|\delta_{ij} - \hat{\beta}_{ij}|\}_{i=1}^{8}{}_{j=1}^{8}$ = 185.0 and approximation $\|\tilde{f}_t\| \cong \sqrt{1-\|R^2\|}$ implies that $\max\{|\tilde{f}_{it}|\}_{i=1}^{n}$ = .837, where $R^2 = (R_1^2,\ldots,R_8^2)$. This last approximation possibly under- or over-states $\|\tilde{f}_t\|$ because it includes the effect of variations in λ_t on R^2. With $\partial q_{it}/\partial \tilde{f}_{jt}$ given by elasticity equations (4.1)-(4.2), table 3 implies that $\max\{|\partial q_{it}/\partial \tilde{f}_{jt}|\}_{i=1}^{8}{}_{j=1}^{8}$ = .123. Thus, in the application, inequality (3.22) implies that $|\Delta\hat{c}_t| \leq 10.2$. Because in the application there was no convincing z_t to use for hedonic adjustment, $|\Delta\hat{c}_t| \leq 10.2$ reflects the largest possible $\Delta\zeta_{it} \sim UD[-\hat{\gamma}_{it}, \hat{\gamma}_{it}]$ with no hedonic adjustment.

Inequality (3.22) is one of other possible measures of GCES/COLI accuracy. In an application, if inequality (3.22) implies that $|\Delta\hat{c}_t| \leq \varepsilon$, where ε is sufficiently small to indicate accuracy of $\hat{c}_t$, then, it makes no difference how ζ_t is estimated, how much or whether hedonic adjustment is done, so that any hedonic adjustment isn't worth doing. If ε is considered large, such as ε = 10.2, then, no conclusions can be drawn from inequality (3.22).

Inequality (3.22) could possibly be tightened up by using a different matrix norm or by exploiting the sizes of $\bar{J}$ or $\tilde{\bar{J}}$, the sampling variability of $\hat{B}$, or other statistical properties of GCES/COLI. For example, in the application, estimated diagonal $\hat{B}$ has some elements with large estimated standard errors. If $\|\hat{B}\|$ were instead evaluated for only the 6x6 submatrix in table 3 with the 6 smallest diagonal elements, then, inequality (3.22) would decline to

$|\Delta\hat{c}_t| \leq 1.71$, a still large upper bound, but one order of magnitude lower. Although hedonic adjustment has been advocated since Coats (1939), a general assessment of the contribution of hedonic adjustment the accuracy of a price index, like inequality (3.22), has apparently not been made before and warrants further study.

Different frequency components of $\{\tilde{f}_t\}_{t=1}^T$ could also be assigned separately to $\{\hat{a}_t\}_{t=1}^T$ and $\{\hat{r}_t\}_{t=1}^T$, for example, according to trend, seasonal, and residual noise variations. For example, seasonal variations of $\{\tilde{f}_t\}_{t=1}^T$ could be assigned to $\{\hat{a}_t\}_{t=1}^T$ by first Fourier transforming $\{\tilde{f}_t\}_{t=1}^T$ to obtain its spectral density, then, assigning the seasonal bands of the spectral density of $\{\tilde{f}_t\}_{t=1}^T$ to the spectral density of $\{\hat{a}_t\}_{t=1}^T$, and, then, inverse-Fourier transforming the spectral density of $\{\hat{a}_t\}_{t=1}^T$ to obtain $\{\hat{a}_t\}_{t=1}^T$. If the remaining trend and noise variations of $\{\tilde{f}_t\}_{t=1}^T$ can't be similarly and convincingly assigned to $\{\hat{a}_t\}_{t=1}^T$ and $\{\hat{r}_t\}_{t=1}^T$, then, they can be handled as discussed above.

The fully-probabilistic and approximately-probabilistic estimates $\{\bar{C}_t\}_{t=1}^T$ and $\{\tilde{\bar{C}}_t\}_{t=1}^T$ of true $\{GCES/COLI_t\}_{t=1}^T$ discussed above could be modified and extended in various ways. The GCES/COLI method discussed above and illustrated in section 4 is intended to be only a start of an integrated and comprehensive method for computing an estimated COLI.

3.3. MSP computation of GCES/COLI.

For given estimates $\hat{B}$ and $\{\hat{a}_t, \hat{r}_t\}_{t=1}^T$, MSP computes $\{GCES/COLI_t\}_{t=1}^T$ by numerically integrating differential-equation system (2.11) for each sample period t = 1,…,T. Appendix C states a reduced version of the MSP order k = 4 computations that were used in in Chen & Zadrozny (2003) for a related production analysis. The reduction involves

reducing the 4th-order polynomial accounting of intra-period variations of exogenous variables to the first order.

[Put table 1 here]

Table 1 states predicted accuracies ε of MSP computations for different orders k, different numbers of steps h, and upper bound $K = 1$ in inequality (B.6) in appendix B. In the table, achieving accuracy ε for a particular k, requires h = smallest integer $\geq \varepsilon^{-1/k}$. For example, for $K = 1$, MSP of order $k = 4$ with $h = 10$ steps is predicted to compute GCES/COLI to within 4-decimal-digit accuracy and MSP of order $k = 3$ with $h = 465$ steps is predicted to compute GCES/COLI to within 8-decimal-digit accuracy. Because computations in finite-precision arithmetic always accumulate some rounding errors, the errors in table 1 understate actual errors. Using related computational methods, Vartia (1983) considered $k = 1$ and large h and Dumagan & Mount (1997) and Breslaw & Smith (1995) considered $k = 2$ and large h. Sun & Xie (2013) analyzed the accuracies in these papers.

Initial GCES/COLI$_1$ is computed in the application with initial observed quantities of goods, $\{Q_{i1}\}_{i=1}^{n}$, because initial observed quantities are assumed to be optimally allocated by a representative consumer. In the application, GCESUF exists as a function, because the differential of GCSEUF is integrable, because B is diagonal. If GCESUF exists as a function, then, the COLI-constant level of utility over a sample of data can be computed according to initial utility based on observed quantities of goods. However, attempting to solve the COLI problem numerically using a known level of utility is generally impossible, because it's an n-dimensional nonlinear problem. GCES/COLI avoids having to solve this numerical problem. Instead of holding utility at the same level as initially over a sample of data, GCES/COLI computations keep utility from changing from the initial level over the sample. The level of utility is immaterial for computing GCES/COLI. Moreover, like any PI, GCES/COLI is normalized to one for the initial period. Before the normalization,

$$(3.19)\qquad \text{GCES/COLI}_1 = \sum_{i=1}^{n} P_{i1}Q_{i1}, \quad \{\text{GCES/COLI}_t\}_{t=2}^{T} = \{\sum_{i=1}^{n} P_{it}\tilde{Q}_{it}\}_{t=2}^{T},$$

where $\{\tilde{Q}_{it}\}_{i=1\ t=2}^{n\ \ T}$ are computed as explained in appendix C.

Symmetric $\hat{B}$ simplifies conditions for GCESUF integrability and, thereby, simplifies computing GCES/COLI strictly correctly. If GCESUF is integrable, then, any connected path between two points of observed prices and estimated preferences and qualities results in the same correctly computed GCES/COLI. If GCESUF isn't integrable, then, MSP computations should strictly be based on the best estimated connected paths between points of prices, preferences, and qualities, which are obtained by estimating a continuous-time process of observed prices and estimated preferences and qualities (Zadrozny, 1988) and, then, using the estimated process to estimate connected paths between discrete points of the prices, preferences, and qualities.

4. Application to 8 aggregate categories of goods.

In this section, the paper applies GCES/COLI to U.S. monthly data on Törnqvist PIs (TORN/PI) and expenditures for 8 aggregate categories of goods and services (Apparel, Education, Food, Other Goods, Housing, Medical, Recreation, and Transportation) in the CPIU from January 1990 to December 2008. The nominal-dollar TORN/PIs of the categories are considered to be prices of the categories and the nominal-dollar expenditures divided by the nominal-dollar TORN/PIs are considered to be their real quantities.

Graphs 1-4 display the TORN/PIs and their implied quantities in standardized original units, in standardized differences of logs of original units ($\Delta_{12}\Delta_1 \ln$), and their autocorrelations. The graphs show trends, seasonal variations, and irregular variations of the data. The graphs show that the $\Delta_{12}\Delta_1 \ln$ transformation renders prices and quantities IID except for possibly significant autocorrelations at monthly lags 1 and near 12. However, when the autocorrelations were

accounted for more carefully by re-estimating B along with parameters of a VMA process for η_t, the re-estimated $\hat{B}$ was negligibly different and, therefore, wasn't considered further.

[Put graphs 1-4 here]

In the application, diagonal B was estimated by applying MLE to equation (3.3) with exogenous term Ξz_t omitted, using only price and quantity data in standardized $\Delta_{12}\Delta_1 \ln$ form, because it would be difficult to find convincing observed exogenous quality-related z_t for the 8 aggregate categories of goods. An attempt to estimate GCESUF for a full B matrix failed because the MLE computations stalled before converging to a proper maximum. Table 2 reports p values of Ljung-Box Q statistics of autocorrelations of residuals of individual equations at monthly lags 1-36, R^2 coefficients of determination of individual equations, and, estimated diagonal $\hat{B}$.

[Put table 2 here]

Table 2 shows that different categories of goods have different $\hat{\beta}_{ii}$, so that the estimated GCESUF is nonhomothetic and expenditure shares vary with expenditure levels. The p values of Ljung-Box statistics in table 2 indicate that residual autocorrelations are insignificant, so that assumption (3.5) that the disturbance vector η_t ~ IID is acceptable. R^2 values in table 2 indicate that price variations account from about 30% to 70% of variations in quantities of goods.

Elasticity responses of optimally consumed goods to varying prices and expenditure levels are determined for diagonal B as follows. Consider the dual of the consumer's primal optimal consumption problem in section 2: for a diagonal-B GCESUF, for given preferences for goods, qualities of goods, prices of goods and cost of expenditures on goods, a consumer purchases and consumes goods so as to maximize

utility. The dual problem has the same MFOC (2.1) as the primal problem, but has the expenditure line as the CFOC.

The primal and dual problems have the same left-side system matrix $\bar{B}(\tau)$ as system (2.11). Preferences and qualities are now ignored, so that their logs are assumed to be constant at zero, $\{a_t\}_{t=1}^T = \{r_t\}_{t=1}^T = \{0_{nx1}\}_{t=1}^T$, and cost shares are set to sample averages, $\{\bar{S}_i\}_{i=1}^n = \{\sum_{t=1}^T S_{it}/T\}_{i=1}^n$. The CFOC of the primal problem is constant utility; the CFOC of the dual problem is the expenditure line at a constant level of expenditures. The CFOC of the dual problem is put into primal system (2.11) simply replacing the bottom element on the right side with dc_t. Then, using Laplace expansion of a determinant and Cramer's rule, the resulting system (2.11) is solved for dq_t in terms of dp_t and dc_t and, in scalar partial-derivative (elasticity) form, the solution is

$$(4.1) \qquad \partial q_{it}/\partial p_{it} = -\sigma_i(1 - \omega_i) < 0,$$

$$(4.2) \qquad \partial q_{it}/\partial p_{jt} = \sigma_i(1 - \omega_j) > 0,$$

$$(4.3) \qquad \partial q_{it}/\partial c_t = \sigma_i/\bar{\sigma} > 0,$$

for i,j = 1,…,n and i ≠ j, where $\sigma_i = 1/\beta_{ii} > 0$, $\bar{\sigma} = \sum_{k=1}^n \bar{S}_k\sigma_k > 0$, $0 < \bar{S}_k < 1$, $\sum_{k=1}^n \bar{S}_k = 1$, and $0 < \omega_i = \bar{S}_i(\sigma_i/\bar{\sigma}) < 1$. Inserting $\{\bar{S}_i\}_{i=1}^8$ computed with data and $\{\hat{\beta}_{ii}\}_{i=1}^8$ in table 3 into equations (4.1)-(4.3) gives the own-price, cross-price, and expenditure-level elasticities in table 3.

[Put table 3 here]

The own-price elasticities (4.1) in table 3 may be compared with Braithwait's (1980) own-price elasticities obtained using annual data, also on aggregate categories of consumer expenditures, although different ones. Braithwait obtained an average estimated own-price elasticity of -.400, compared with the average own-price elasticity of -.074 in table 3. Braithwait's larger absolute value makes sense, because consumers have more flexibility over a year than over a month to change their expenditure patterns. Because table 3 shows different estimated $\hat{\beta}_{ii}$ for different categories of goods, the estimated GCESUF is nonhomothetic and, in its last column, table 3 shows the different nonunit expenditure-level elasticities (4.3) for the different categories of goods.

$\{GCES/COLI_t\}_{t=1}^{T}$ was computed for t = 1 (January 1990), …, T = 228 (December 2008). For the same period, Laspeyres and Törnqvist PIs were computed as

$$(4.4) \qquad LASP/PI_t = \sum_{i=1}^{n} S_{it}(P_{it+1}/P_{it}),$$

$$(4.5) \qquad TORN/PI_t = \prod_{i=1}^{n} (P_{it+1}/P_{it})^{(S_{it+1} + S_{it})/2},$$

where $S_{it} = P_{it}Q_{it}/\sum_{j=1}^{n} P_{jt}Q_{jt}$.

Whereas $\{GCES/COLI_t\}_{t=1}^{T}$ are COLIs because, except for initial quantities of goods, $\{Q_{i1}\}_{i=1}^{n}$, they're based on computed quantities of goods, $\{\tilde{Q}_{it}\}_{i=1\ t=2}^{n\ T}$, $\{LASP/PI_t\}_{t=1}^{T}$ and $\{TORN/PI_t\}_{t=1}^{T}$ generally aren't COLIs because they're based entirely on observed quantities, $\{Q_{it}\}_{i=1\ t=1}^{n\ T}$. $\{GCES/COLI_t\}_{t=1}^{T}$ explicitly accounts for changes in preferences for and qualities of goods. $TORN/PI_t$, like Divisia PI that it approximates, implicitly accounts for changes in preferences for goods, but not for changes in qualities of goods. However, there's no way to check this for $TORN/PI_t$, because it has no known underlying utility function.

LASP/PI_t and TORN/PI_t are usually defined in terms of backward price relatives, P_{it}/P_{it-1}, but are here defined in terms of forward price relatives, $P_{i,t+1}/P_{i,t}$, to be consistent with forward differences, Δ_1 and Δ_{12}. CPIU and chained CPIU (CCPIU) produced and released to the public by BLS are based on similar equations but with expenditure shares timed differently. See Klick (2018). The three indexes here all have expenditure shares in "current" periods t and so are "fully chained".

The indexes are displayed in graphs 5 in differences between them in logarithmic form from January 1990 to December 2008. The indexes are graphed normalized, meaning divided by initial values, so that initial values are one. Graphs 5 depict $\ln(\text{LASP}/PI_t) - \ln(\text{GCES}/COLI_t)$ in the left graph and $\ln(\text{GCES}/COLI_t) - \ln(\text{TORN}/PI_t)$ in the right graph. Presumably, LASP/PI_t > GCES/$COLI_t$ in all sample periods because LASP/PI_t reflects no substitutions of goods induced by relative price changes, whereas GCES/$COLI_t$ reflects average own-price substitution elasticities of -.074. Presumably, GCES/$COLI_t$ > TORN/PI_t in all sample periods, because TORN/PI_t reflects larger price substitution elasticities than GCES/$COLI_t$. Because a TORN/PI_t can be derived from a CDUF, it's often assumed to reflect unit price elasticities, but a TORN/PI_t doesn't necessarily derive from a UF, like a CDUF, with unit price elasticities. GCES/$COLI_t$ is seen in graph 5 to lie about halfway between LASP/PI_t and TORN/PI_t in all periods.

[Put graphs 5 here]

Graphs 6 depict estimated log-form (but not demeaned or standardized) preferences-plus-qualities, $\{\hat{f}_{it}\}_{i=1\ t=1}^{8\ \ 228}$, implied by estimated diagonal $\hat{B}$ in table 3 and shows generally nonzero means, approximately linear trends, and seasonal variations, the latter especially prominent for education and recreation.

[Put Figure 6 here]

5. Conclusion.

5.1. Summary of the paper.

The paper develops a new, general, practical, accurate, integrated, and comprehensive method for computing a cost of living index (COLI) based on an estimated generalized CES utility function (GCESUF). The GCES/COLI fully accounts for changes in preferences for and qualities of consumed goods (and services) in the computed GCES/COLI. The changes in preferences and qualities are integrated into the development and computation of GCES/COLI. Unlike in usual hedonic adjustment of prices of goods for changes in their qualities accounted for by hedonic regression of prices on z_t, where z_t is a vector of observed variables that account for qualities of goods. GCES/COLI accounts for changes in qualities of goods by estimating a demand equation derived from first-order conditions (FOC) of the representative consumers' optimization problem that includes z_t as an explanatory variable and accounts for unobserved quality changes, ζ_t, probabilistically.

The paper illustrates GCES/COLI by estimating GCESUF by maximum likelihood for U.S. monthly data from the Bureau of Labor Statistics on prices and expenditures on 8 aggregate categories of goods in the consumer price index of urban consumers (CPIU) from January 1990 to December 2008. For the period, the resulting GCES/COLI lies about halfway between higher Laspeyres and lower Törnqvist PIs computed from scratch with the same data. The purpose of the application is only to illustrate GCES/COLI and isn't to compare it with CPIU and chained CPIU (CCPIU) produced and released to the public by BLS. See Klick (2018) for a discussion of recent CPIU and CCPIU.

5.2. Four possible extensions.

5.2.1. Estimating a full matrix B.

Iterative computations of MLE of substitution-parameter matrix B can fail by not converging when B is nondiagonal or much larger than the 8-dimensional B in the application in section 4. A properly set up Bayesian estimation shouldn't fail to estimate B of any size, at worst may be slow to compute, because Bayesian estimation doesn't need to optimize an estimation criterion function and satisfy its second-order conditions, but only needs to compute a numerical histogram of the posterior distribution of parameters and to compute its statistics of central tendency and dispersion. It would be interesting to compute GCES/COLI for more than 8 goods based on a full or fuller estimated B and compare it with standard formulaic PIs.

5.2.2. Estimating a time-varying matrix B.

Elements of time-varying B_t could be specified as generated by independent random walks,

$$\text{vech}(W_t) = \text{vech}(W_{t-1}) + \xi_t, \tag{5.1}$$

where W_t = lower-triangular Cholesky factor of symmetric B_t^{-1}, $\text{vech}(W_t)$ = column vectorization of the nonzero lower-triangular elements of W_t, $\xi_t \sim \text{NIID}(0, \Sigma_\xi)$, and Σ_ξ = diagonal with positive diagonal elements.

Specification (5.1) could be estimated either using MLE or a Bayesian method, in terms of diagonal elements of Σ_ξ. With constant B, GCES/COLI accounts for time-varying slopes of indifference curves only with estimated time-varying preferences and qualities, $\hat{a}_t$ and $\hat{r}_t$. With estimated time-varying $\hat{B}_t = (\hat{W}_t \hat{W}_t^T)^{-1}$, GCES/COLI would account with time-

varying W_t both for time-varying slopes and for time-varying curvatures of indifference curves, which should result in more accurate GCES/COLI.

5.2.3. Handling turnover of goods.

Turnover of goods means either "old" goods are removed from an index, "new" goods are added to an index, or old goods are replaced in an index by new versions. Supplying firms decide when and at what minimal levels of sales goods are withdrawn from and introduced to consumers. Because the minimal levels are always nonzero, GCES/COLI can handle turnover without essential change. Turnover isn't considered in the application in section 4 because it would be artificial to omit some aggregate category of expenditures from GCES/COLI from an initial part of the sample of data and, then, include it in the remaining part of the sample or vice versa.

Old good removed: If old good n is removed from GCES/COLI at the beginning of period t+1, then, the dimensions of B and of vectors and matrices in GCES/COLI computations decrease by one but otherwise the computations don't change. Values of elements of B of remaining goods can be kept or somehow determined but can't be meaningfully estimated in the new situation until enough new data have accumulated.

New good added: If new good n+1 is added to GCES/COLI at the beginning of period t+1, then, the dimensions of B and of vectors and matrices in GCES/COLI computations increase by one but otherwise the computations don't change. Values of elements of B of previously available goods can be kept or somehow determined along with elements of the new good but can't be meaningfully estimated in the new situation until enough new data have accumulated.

Old good replaced: If old good n is replaced in GCES/COLI by a new version, then, nothing changes in GCES/COLI computations. The previous value of B can be kept or somehow determined, but can't be meaningfully estimated in the new situation until enough new data have accumulated.

5.2.4. Computing a monetary-financial services index.

It would be interesting to compute a monetary-financial services quantity index based not on the Törnqvist index as Barnett (2011) does (although he calls the Törnqvist index a Divisia index), but on a quantity index corresponding to GCES/COLI and, thereby, account in the index for changes in preferences for and qualities of different monetary and financial assets as in GGES/COLI, for any reasons such as changes in perceived risks, changes in regulations, changes in technology, etc.

Table 1: Predicted accuracies of MSP computations.

Accuracy order ε	Polynomial order k	Step size h^{-1}	Number of steps h
Semi-single decimal-digit precision of order 10^{-4}	1	1.00×10^{-4}	10^{4}
	2	1.00×10^{-2}	10^{2}
	3	4.55×10^{-2}	22
	4	1.00×10^{-1}	10
	5	1.43×10^{-1}	7
	6	2.00×10^{-1}	5
Single decimal-digit precision of order 10^{-8}	1	1.00×10^{-8}	10^{8}
	2	1.00×10^{-4}	10^{4}
	3	2.15×10^{-3}	465
	4	1.00×10^{-2}	100
	5	2.50×10^{-2}	40
	6	4.55×10^{-2}	22
Double decimal-digit precision of order 10^{-16}	1	1.00×10^{-16}	10^{16}
	2	1.00×10^{-8}	10^{8}
	3	4.64×10^{-6}	215,444
	4	1.00×10^{-4}	10^{4}
	5	6.31×10^{-4}	1585
	6	2.15×10^{-3}	465

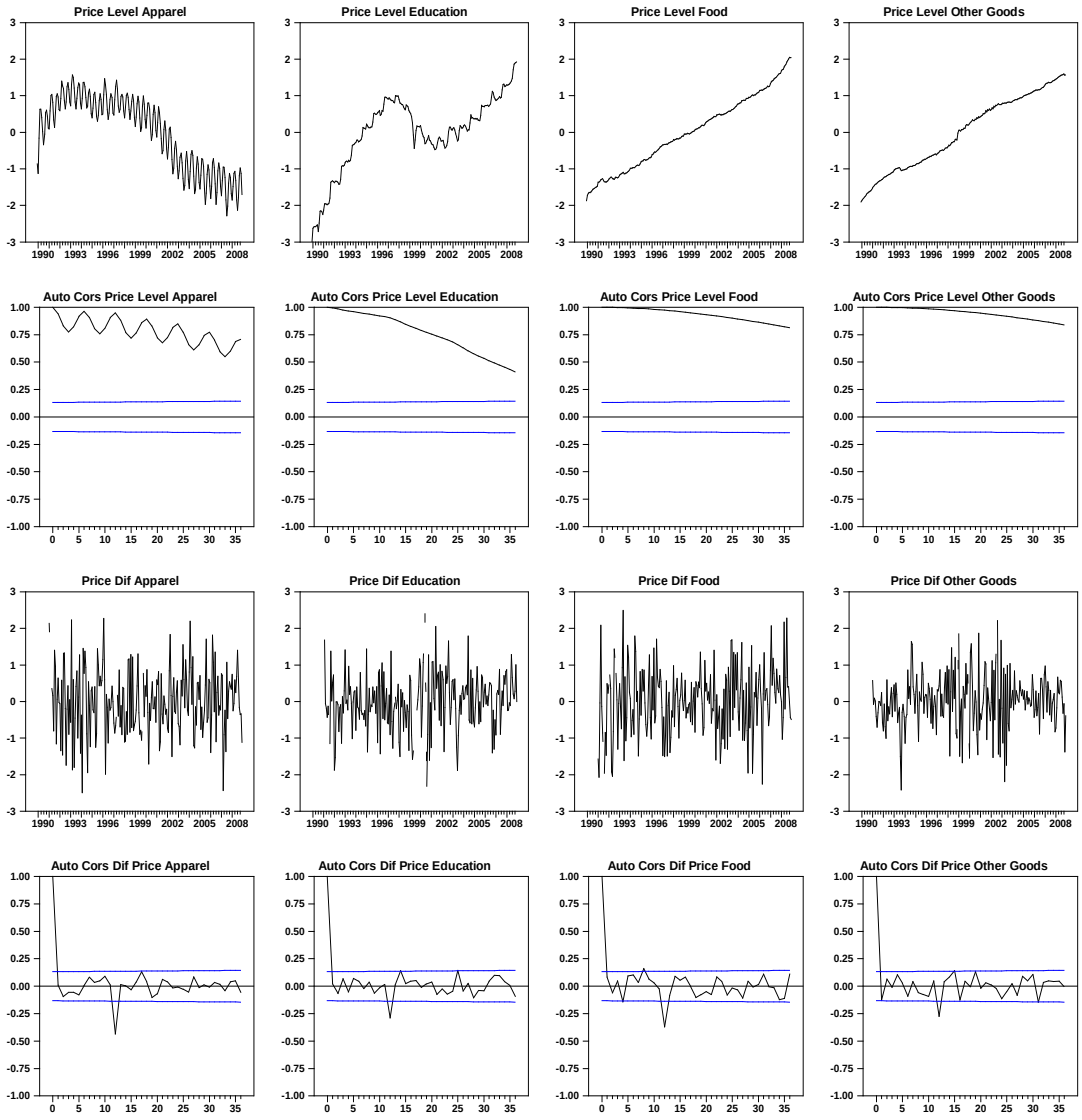
Table 2: MLE statistics of estimated diagonal B.

Aggregate expenditure categories	p values of Q statistic	R_i^2	Estimated $\hat{\beta}_{ii}$
Apparel	.267	.324	37.8
Education	.787	.312	8.04
Food	.745	.480	9.62
Other Goods	.239	.583	32.1
Housing	.181	.301	13.7
Medical	.135	.494	151.
Recreation	.487	.548	186.
Transportation	.066	.699	31.9

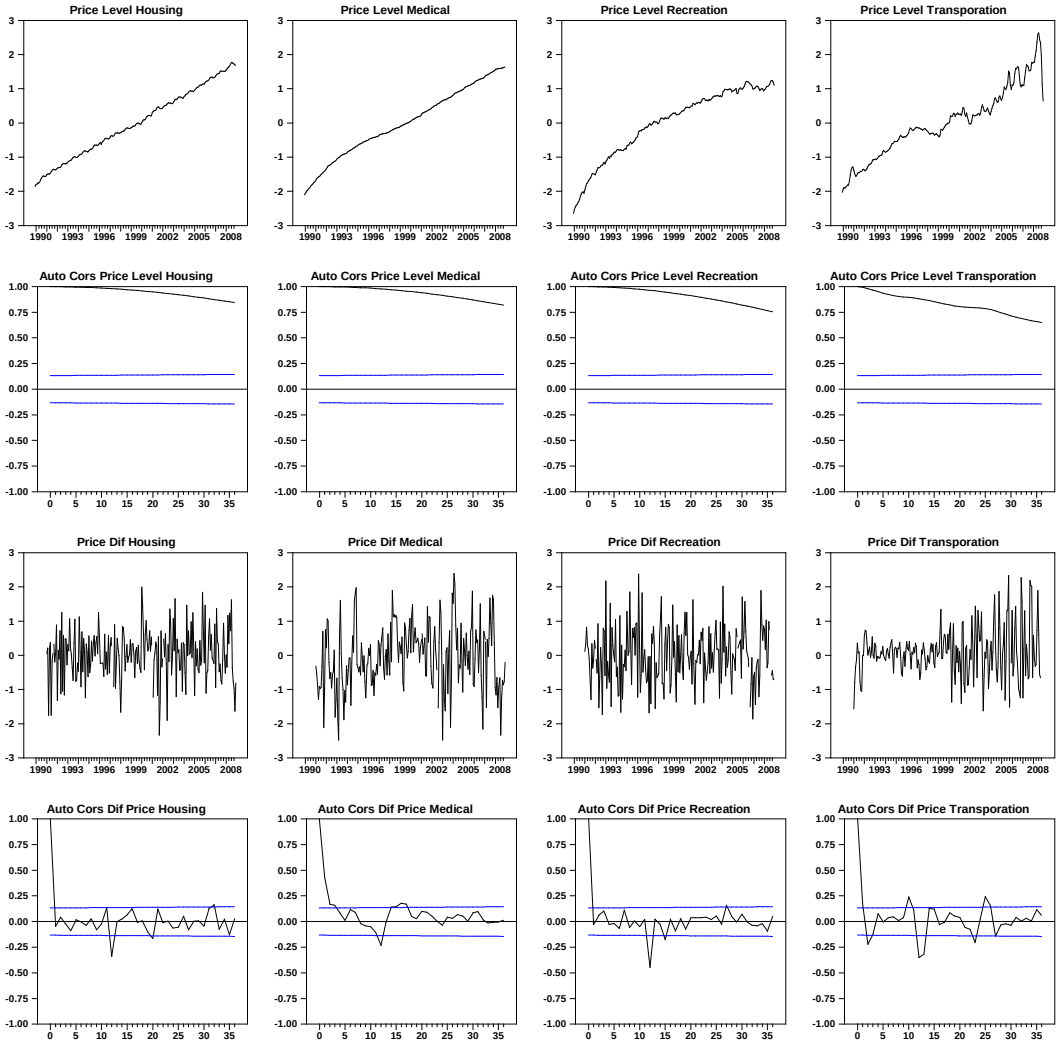
Table 3: Price and expenditure elasticities (4.1) - (4.3) implied by estimated diagonal B in table 2.

q_i	p_a	p_e	p_f	p_g	p_h	p_m	p_r	p_t	c
	Appa	Educ	Food	Good	Hous	Medic	Recr	Tran	Expe
Apparel	-.123	.012	.013	.015	.015	.019	.006	.018	1.76
Education	.013	-.059	.006	.007	.007	.009	.003	.008	.746
Food	.015	.006	-.062	.008	.008	.010	.003	.010	.801
Other Goods	.021	.009	.009	-.073	.011	.015	.004	.013	.990
Housing	.018	.008	.008	.010	-.071	.013	.004	.012	.941
Medical	.029	.012	.013	.016	.016	-.086	.006	.019	1.24
Recreation	.001	.003	.003	.004	.003	.004	-.030	.004	.367
Transportation	.022	.009	.010	.012	.012	.016	.005	-.084	1.14

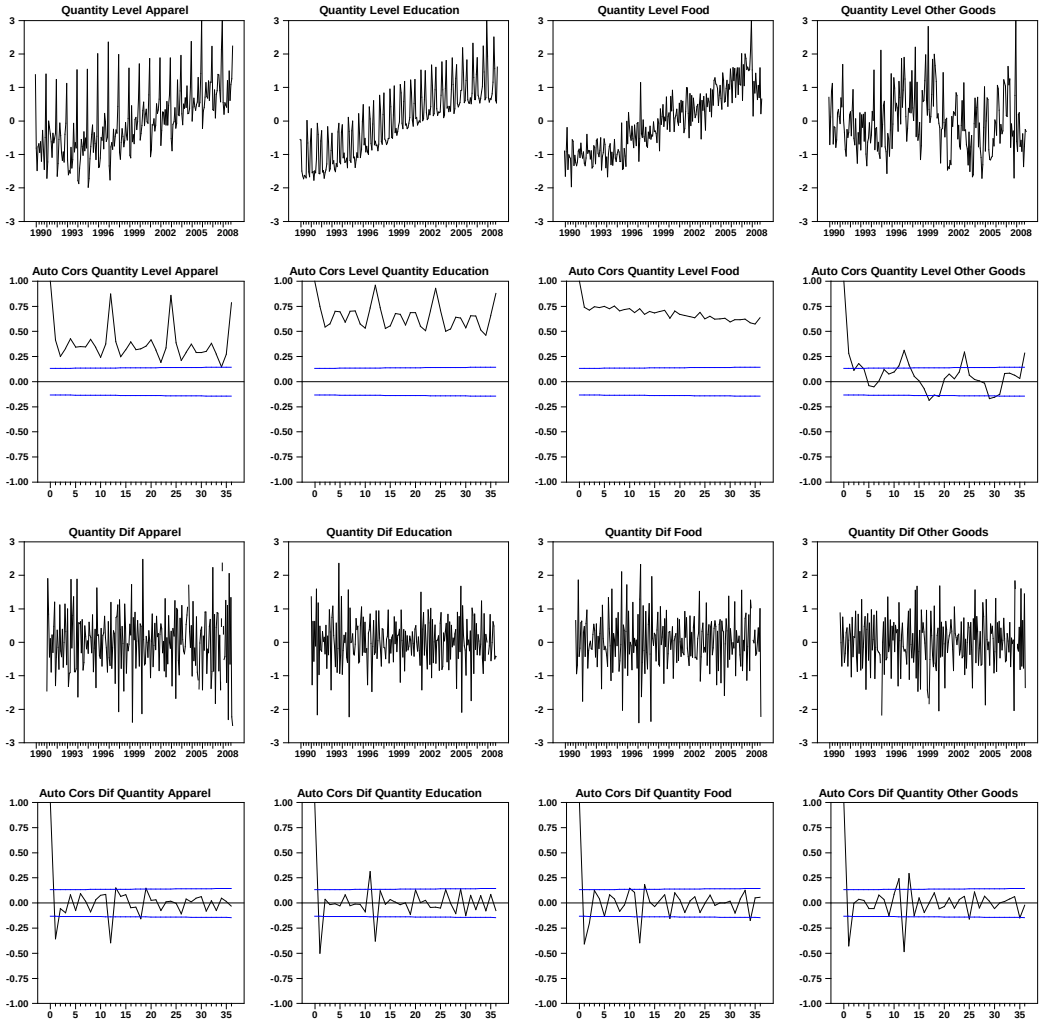
Graphs 1: Standardized P_{it} and $\Delta_{12}\Delta_1 \ln(P_{it})$ of Apparel, Education, Food, and Other Goods and their autocorrelations, 1/1990,…,12/2008.



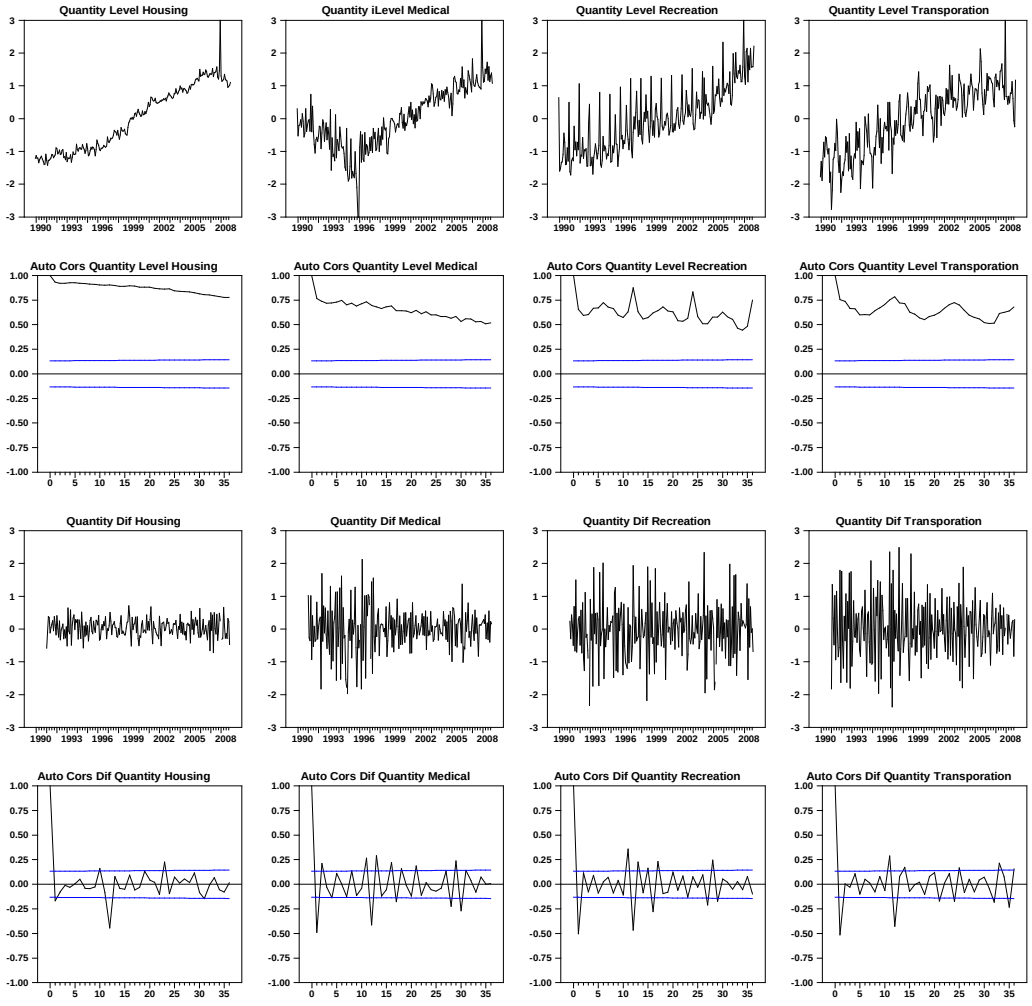
Graphs 2: Standardized P_{it} and $\Delta_{12}\Delta_1 \ln(P_{it})$ of Housing, Medical, Recreation, and Transportation and their autocorrelations, 1/1990,…,12/2008.



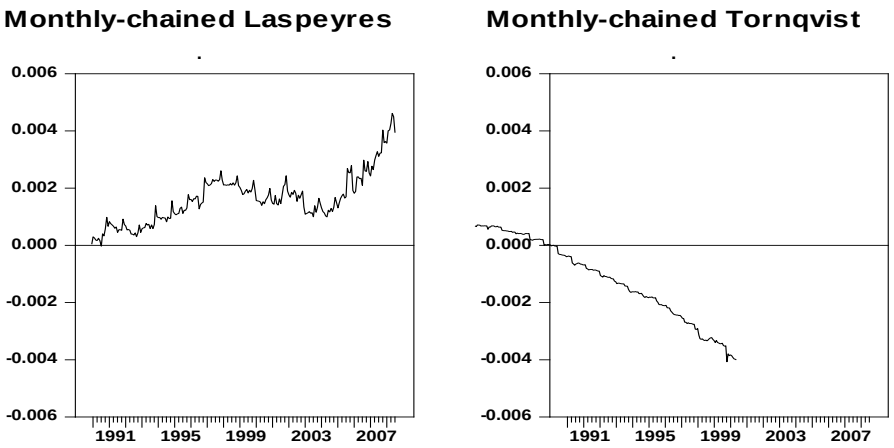
Graphs 3: Standardized Q_{it} and $\Delta_{12}\Delta_1 \ln(Q_{it})$ of Apparel, Education, Food, and Other Goods and their autocorrelations, 1/1990,…,12/2008.



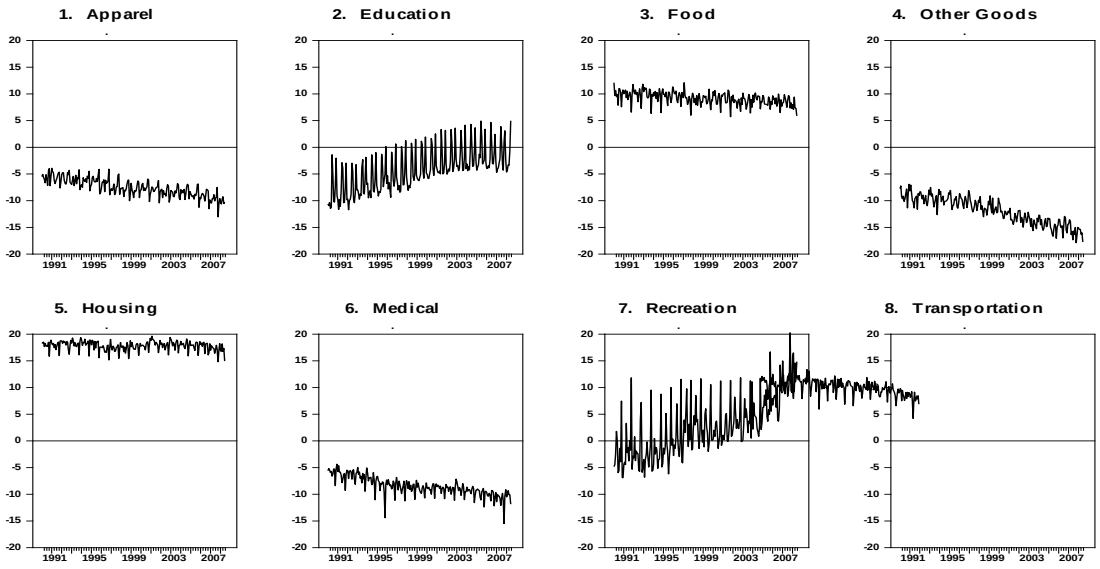
Graphs 4: Standardized Q_{it} and $\Delta_{12}\Delta_1 \ln(Q_{it})$ of Housing, Medical, Recreation, and Transportation and their autocorrelations, 1/1990,…,12/2008.



Graphs 5: $\ln(LASP/PI_t)-\ln(GCES/COLI_t)$ (left) and $\ln(TORN/PI_t)-\ln(GCES/COLI_t)$ (right), 1/1990,…,12/2008.



Graphs 6: Estimated log-form preferences-plus-qualities, $\hat{f}_{1t},\ldots,\hat{f}_{8t}$, 1/1990,…,12/2008.



APPENDICES A-D.

Appendix A: Proof that inequality (2.14) implies SOC.

Sufficient condition (2.13) for the second-order condition (SOC) of the consumer's optimization problem, stated for simplicity without time argument τ, is

$$\text{(A.1)} \qquad \bar{B} = \begin{bmatrix} -B & e \\ S^T & 0 \end{bmatrix} < 0 \text{ (negative definite)},$$

where $B = nxn$, $e = (1,\ldots,1)^T = nx1$, $S = (S_1,\ldots,S_n)^T$, $S_i > 0$, and $\sum_{i=1}^{n} S_i = 1$. Condition (A.1) hold if and only if

$$\text{(A.2)} \qquad -x^TBx + x^Tey + yS^Tx = -x^TBx + y(S + e)^Tx < 0,$$

for any $x = nx1$ and scalar y both not zero. If $B > 0$ (positive definite), then, inequality (A.2) holds if and only if $y(S + e)^Tx$ is positive and not too large. To determine what "not too large" means, first, consider x and y normalized as

$$\text{(A.3)} \qquad x^Tx + y^2 = 1,$$

with no loss of generality because (A.2) is homogeneous in x and y.

We consider inequality (A.2) as $y(S + e)^Tx < x^TBx$ and determine an upper bound for

$$\text{(A.4)} \qquad f = \left|y(S + e)^Tx\right| = \sqrt{\gamma}\left|\alpha^Tx\right|,$$

where $\gamma = y^2 \in (0,1)$ and $\alpha = S + e$. We maximize f sequentially with respect to x, γ, and S. For given γ and S, f is maximized when x points in the same direction as α, so that $x = c\alpha$ and, because $x^Tx = 1 - \gamma$, $c^2\alpha^T\alpha = 1 - \gamma$, $c = \sqrt{(1-\gamma)/(\alpha^T\alpha)}$, and maximized $f = \sqrt{\gamma(1-\gamma)(\alpha^T\alpha)}$. Now, f is further maximized when $\gamma = 1/2$, so that

(A.5) $$f = \sqrt{(\alpha^T\alpha)}/2 = \sqrt{(S + e)^T(S + e)}/2.$$

Finally, maximizing $(S+e)^T(S+e)$ with respect to S_i implies $S_i = 1/n$, for $i = 1,\ldots,n$, or $S = e/n$, so that

(A.6) $$f = \left[\sqrt{S^TS + 2S^Te + e^Te}\right]/2 = \left[\sqrt{(1/n) + 2 + n}\right]/2.$$

For the $n = 8$ in the application, (A.6) implies

(A.7) $$f = \left[\sqrt{81/8}\right]/2 = 1.591.$$

Now, x^TBx is minimized for $x^Tx = 1-\gamma = 1/2$ when x is the eigenvector of the smallest positive eigenvalue of B, which in Table 2 is 8.04. Therefore, in the application, the minimum x^TBx, for $x^Tx = 1/2$, is $8.04/2 = 4.02 > 1.591$, so that inequality (A.2) holds and sufficient condition (A.1) for SOC holds in the application.

Appendix B: Review of matrix differentiation.

B.1. Definitions of matrix derivatives.

Let $A(x) \in \mathbf{D}^k$: $\mathbf{R}^n \rightarrow \mathbf{R}^{p\times q}$ be a K-times differentiable p×q matrix function of the n×1 vector x. A(x) could be a function of the matrix X

$\in \mathbf{R}^{k \times m}$, such that $x = \mathrm{vec}(X)$, where $\mathrm{vec}(\cdot)$ is the columnwise vectorization of a matrix. We consider derivatives of elements of A with respect to elements of x in three forms: ∂ partial derivative form, d differential form, and ∇ Jacobian form.

For $k = 1, \ldots, K$ and $i_1, \ldots, i_k \in \{1, \ldots, n\}$, we define $\partial^k_{i_1 \cdots i_k} A \in \mathbf{R}^{p \times q}$ by

$$\text{(B.1)} \qquad \partial^k_{i_1 \cdots i_k} A = \begin{bmatrix} \dfrac{\partial^k A_{11}}{\partial x_{i_1} \cdots \partial x_{i_k}} & \cdots & \dfrac{\partial^k A_{1q}}{\partial x_{i_1} \cdots \partial x_{i_k}} \\ \vdots & & \vdots \\ \dfrac{\partial^k A_{p1}}{\partial x_{i_i} \cdots \partial x_{i_k}} & \cdots & \dfrac{\partial^k A_{pq}}{\partial x_{i_1} \cdots \partial x_{i_k}} \end{bmatrix},$$

as the <u>partial derivative form</u> of k-th order partial derivatives of the elements of A with respect to $x_{i_1}, \ldots, x_{i_k}$.

The <u>differential form</u> associated with (B.1) is

$$\text{(B.2)} \qquad d^k A = \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n \partial_{i_1 \cdots i_k} A \cdot dx_{i_1} \cdots dx_{i_k},$$

where the dx_i's are small (infinitesimal, i.e., nonzero but approaching zero) increments to the elements of $x = (x_1, \ldots, x_n)^T$.

The <u>Jacobian form</u> associated with (B.1) and (B.2) can now be built up recursively, starting with k = 1. We call the matrix representation of kth derivatives of a vector function or a vectorization of a matrix function a "k-Jacobian," which generalizes common terminology: for a scalar-valued function, the 1-Jacobian is a gradient vector of first-partial derivatives and the 2-Jacobian is the Hessian matrix of second-partial derivatives.

We use the following rule for vectorizing matrix products,

$$\text{(B.3)} \qquad \mathrm{vec}(ABC) = [C^T \otimes A]\mathrm{vec}(B),$$

where A, B, and C are matrices conformable to the matrix product ABC and $\otimes$ denotes the Kronecker matrix product (Magnus and Neudecker, 1988, p.30).

B.2. Representations of matrix derivatives.

For k = 1, (B.1) and (B.2) become

$$\text{(B.4)} \qquad \partial_i A = \begin{bmatrix} \frac{\partial A_{11}}{\partial x_i} & \cdots & \frac{\partial A_{1q}}{\partial x_i} \\ \vdots & & \vdots \\ \frac{\partial A_{p1}}{\partial x_i} & \cdots & \frac{\partial A_{pq}}{\partial x_i} \end{bmatrix},$$

$$\text{(B.5)} \qquad dA = \sum_{i=1}^{n} \partial_i A \cdot dx_i .$$

Note that vectorization, summation, and differentiation operations are commutative, i.e., can be applied in any order. Therefore, we vectorize (B.5), to obtain

$$\text{(B.6)} \qquad \mathrm{vec}(dA) = [\partial_1 \mathrm{vec}(A), \ldots, \partial_n \mathrm{vec}(A)]dx,$$

where $dx = (dx_1, \ldots, dx_n)^T$, so that

$$\text{(B.7)} \qquad \mathrm{vec}(dA) = \nabla A \cdot dx,$$

$$\text{(B.8)} \qquad \nabla A = [\partial_1 \mathrm{vec}(A), \ldots, \partial_n \mathrm{vec}(A)].$$

Equations (B.7) and (B.8) relate the ∂, d, and ∇ forms of first-order derivatives of A to each other.

To obtain analogues of (B.7) and (B.8) for k = 2, we differentiate them to obtain

(B.9) $\quad \mathrm{vec}(d^2A) = d(\nabla A)dx,$

(B.10) $\quad d(\nabla A) = [d(\mathrm{vec}(\partial_1 A)), \ldots, d(\mathrm{vec}(\partial_n A))]$

$$= \sum_{j=1}^{n} [\partial_j(\mathrm{vec}(\partial_1 A)), \ldots, \partial_j(\mathrm{vec}(\partial_n A))]dx_j.$$

Then, we vectorize (B.10) to obtain

(B.11) $$\mathrm{vec}(d(\nabla A)) = \sum_{j=1}^{n} \begin{bmatrix} \partial_j(\mathrm{vec}(\partial_1 A)) \\ \vdots \\ \partial_j(\mathrm{vec}(\partial_n A)) \end{bmatrix} dx_j = \begin{bmatrix} \partial_1(\mathrm{vec}(\partial_1 A)) & \cdots & \partial_n(\mathrm{vec}(\partial_1 A)) \\ \vdots & & \vdots \\ \partial_1(\mathrm{vec}(\partial_n A)) & \cdots & \partial_n(\mathrm{vec}(\partial_n A)) \end{bmatrix} dx.$$

Then, because $\mathrm{vec}(\nabla A) = \begin{bmatrix} \partial_1(\mathrm{vec}(A)) \\ \vdots \\ \partial_n(\mathrm{vec}(A)) \end{bmatrix} = \begin{bmatrix} \mathrm{vec}(\partial_1 A) \\ \vdots \\ \mathrm{vec}(\partial_n A) \end{bmatrix}$, we obtain

(B.12) $\quad \mathrm{vec}(d(\nabla A)) = [\partial_1 \mathrm{vec}(\nabla A), \ldots, \partial_n \mathrm{vec}(\nabla A)]dx.$

Continuing in this manner for k = 2, ..., K, we obtain

(B.13) $\quad \mathrm{vec}(d(\nabla^{k-1}A)) = \nabla^k A \cdot dx,$

(B.14) $\quad \mathrm{vec}(d^k A) = [(\Pi_{k-1} \otimes dx^T) \otimes I_{pq}]\nabla^k A \cdot dx,$

where $\Pi_{k-1} \otimes dx^T$ denotes k-2 successive Kronecker products of dx^T ($\Pi_0 \otimes dx^T = dx^T$, $\Pi_1 \otimes dx^T = dx^T \otimes dx^T$, ...), and

(B.15) $\quad \nabla^k A = [\partial_1 \mathrm{vec}(\nabla^{k-1}A), \ldots, \partial_n \mathrm{vec}(\nabla^{k-1}A)].$

Applied for k = 1, ..., K, (B.15) recursively organizes a Jacobian form of derivatives of A up to order K as matrices. That is, $\nabla^k A$ is the

Jacobian matrix of the vectorization of $\nabla^{k-1}A$, is the Jacobian matrix of the vectorization of $\nabla^{k-2}A$, etc.

B.3. Rules of matrix differentiation.

Let $A(x) \in \mathbf{D}: \mathbf{R}^n \rightarrow \mathbf{R}^p$ and $B(y) \in \mathbf{D}: \mathbf{R}^p \rightarrow \mathbf{R}^q$ be differentiable vector functions (or vectorizations of matrix functions). Let $C(x) \in \mathbf{D}: \mathbf{R}^n \rightarrow \mathbf{R}^q$ be the differentiable composite vector function $C(x) = B(A(x))$. Then, the Jacobian form of derivatives of $C(x)$ is given by the chain rule of differentiation,

(B.16) $$\nabla C(x) = \nabla B(A) \cdot \nabla A(x).$$

Let $A(x) \in \mathbf{D}: \mathbf{R}^n \rightarrow \mathbf{R}^{p \times q}$ and $B(x) \in \mathbf{D}: \mathbf{R}^n \rightarrow \mathbf{R}^{p \times q}$ be differentiable matrix functions conformable to the ordinary matrix product $C(x) = A(x) \cdot B(x)$. Then, the differential form of derivatives of $C(x)$ is given by the product rule of differentiation,

(B.17) $$dC(x) = dA(x) \cdot B(x) + A(x) \cdot dB(x).$$

Rules (B.16) and (B.17) are quickly proved by elementwise application of the scalar chain rule of differentiation and the scalar product rule of differentiation. See Magnus & Neudecker (1988).

Appendix C: MSP computation of GCES/COLI.

Appendix B explains definitions and rules of matrix differentiation and should be skimmed before reading this appendix C. Equations in this appendix are derived in differential form and as few as possible are further converted to Jacobian form. Infinitesimal differentials are strictly uncomputable but finite-valued Jacobians are computable. However, differentials considered as finite approximations of infinitesimal changes are computable. We could

convert all expression in differentials to expression in Jacobians, but doing so would introduce much more complicated algebraic expressions to compute, with multiple Kronecker products that are more difficult to program. Chen and Zadrozny (2003) follow this approach. To avoid these complications, in this paper, we compute entirely in differential form, which corresponds to computing in forward mode of automatic or algorithmic differentiation, the preferred method for computing accurately and quickly (Griewank, 2000).

We express first-order conditions (FOC) of the representative consumer's primal optimization problem in differential form of equation (2.11) more compactly as

(C.1) $F(y(s),x(s))dy(s) = G(y(s),x(s))dx(s)$,

$y(s) = (\tilde{q}_1(s), \ldots, \tilde{q}_n(s), \lambda(s))^T = nx1$, $x(s) = (p_1(s), \ldots, p_n(s), a_1(s), \ldots, a_n(s), r_1(s), \ldots, r_n(s))^T = 3nx1$, and the details of $F(\bullet) = (n+1)x(n+1)$ and $G(\bullet) = (n+1)x3n$ can be seen in system (2.11). We say that $y(s)$ is "endogenous" because $dy(s)$ is determined by solving and integrating equation (C.1) using MSP, for given exogenous $x(s)$ and $dx(s)$. Because GCESUF is differentiable any desired number of times, so are $F(\bullet)$ and $G(\bullet)$.

Equation (C.1) has the unique solution

(C.2) $dy(s) = H(y(s),x(s))\,dx(s)$,

where $H(y(s),x(s)) = F(y(s),x(s))^{-1}G(y(s),x(s))$ or $H(s) = F(s)^{-1}G(s)$. Dividing $dy(s)$ and $dx(s)$ by ds converts them to continuous-time derivatives, but, even without explicit division by ds, $dy(s)$ and $dx(s)$ can be considered continuous-time derivatives.

For $s \in [t,t+1)$, we want to compute $\Delta y_t = y_{t+1} - y_t = \int_{s=t}^{t+1} dy(s)$. Let $y(s)$ and $\hat{y}(s)$ denote true and computed (estimated, approximate) solution paths of y, for a given path of $x(s)$. The implicit function

theorem (Apostol, 1974, pp. 373-375), upon which the MSP method is based, implies that because GCESUF is differentiable any number of times and satisfies the second-order conditions (SOC), the solution of the consumer's optimization problem has a unique interior solution. Then, the solution path is differentiable the k+1 desired number of times, integrates as $\int_{s=t}^{t+1} dy(s)$, and, for $s \in [t,t+1)$, has the kth-order polynomial (Taylor series) approximation

$$\text{(C.3)} \qquad \hat{y}(s) = y_t + \nabla y_t(s-t) + (1/2!)\nabla^2 y_t(s-t)^2 + \ldots + (1/k!)\nabla^k y_t(s-t)^k,$$

where y_t, ..., $\nabla^k y_t$ are (n+1)×1 coefficients to be determined in terms of exogenous observed prices and quantities and estimated preferences and qualities. We treat the true y and x processes and their polynomial interpolates as differentiable continuous-time processes in $s \in [t,t+1)$ and as discrete-time processes in t = 1, ..., T. Ford (1955) discusses connections between continuous and discrete processes and their polynomial interpolates. Following equation (C.4), we could approximate Δy_t as $\Delta\hat{y}_t = \int_{s=t}^{t+1} d\hat{y}(s) = \nabla y_t + (1/2!)\nabla^2 y_t + \ldots + (1/k!)\nabla^k y_t$, which has the accuracy $\varepsilon = |\Delta y_t - \Delta\hat{y}_t| = (1/(k+1)!)|\nabla^{k+1} y(\theta)|$, for $\theta \in (t,t+1)$ (Apostol, 1974, pp. 241-242).

For each period t and a chosen positive integer h, we partition the period into h subperiods of equal length 1/h as $[t,t+1) = \bigcup_{i=1}^{h} [t_i, t_i+(1/h))$, where $[t_i, t_i+(1/h)) = [t+(i-1)/h, t+(i/h))$, for i = 1,...,h. For each subperiod $t_i = t_1,\ldots,t_h$ in period t, we compute the y coefficients, $\nabla y_{t_i}, \ldots, \nabla^k y_{t_i}$, recursively and approximate Δy_t as

$$\text{(C.4)} \qquad \Delta\hat{y}_t = \sum_{i=1}^{h} \Delta\hat{y}_{t_i},$$

$$\Delta\hat{y}_{t_i} = \int_{s=t+(i-1)/h}^{t+(i/h)} d\hat{y}(s) = \sum_{j=1}^{k} (1/j!)\nabla^j y_{t_i} h^{-j}.$$

We now explain the MSP method for computing the y coefficients, $\nabla y_{t_i},\ldots,\nabla^k y_{t_i}$, when k = 4. We consider approximate y process (C.3) for k = 4, differentiate it four times, and obtain differentials (C.5). We differentiate differential (C.2) of the unknown true y process three more times and obtain differentials (C.6) of y in terms of the differentials of H(s) and dx. We derive differentials of H(s) in equations (C.6) in terms of differentials of F(s) and G(s). We consider 1st-order exogenous process (C.9) with two coefficients. Finally, we discuss combining these results and computing the coefficients of approximate y(s) process (C.3), hence, computing Δy_t according to equation (C.4).

For k = 4, $s \in [t_i, t_i+h)$, and $t_i = t_1,\ldots,t_h$, differentiating approximate y process (C.3) four times with respect to s implies

$$(C.5)\quad d\hat{y}(s) = \nabla y_{t_i} + \nabla^2 y_{t_i}(s-t_i) + (1/2)\nabla^3 y_{t_i}(s-t_i)^2 + (1/6)\nabla^4 y_{t_i}(s-t_i)^3,$$

$$d^2\hat{y}(s) = \nabla^2 y_{t_i} + \nabla^3 y_{t_i}(s-t_i) + (1/2)\nabla^4 y_{t_i}(s-t_i)^2,$$

$$d^3\hat{y}(s) = \nabla^3 y_{t_i} + \nabla^4 y_{t_i}(s-t_i),$$

$$d^4\hat{y}(s) = \nabla^4 y_{t_i}.$$

For each $s = t_i = t_1,\ldots,t_h$, we compute the y coefficients, $\nabla y_{t_i},\ldots,\nabla^4 y_{t_i}$, so that they are equal to the 1st to 4th differentials of true y process (C.2).

Using the product rule of differentiation (B.17) to differentiate equation (C.2) three times and assuming that the exogenous x(s) process is linear in s, so that its differentials beyond the first one are zero, implies that

$$(C.6)\quad d^2y(s) = dH(s)dx(s),$$

$$d^3y(s) = d^2H(s)dx(s),$$

$$d^4y(s) = d^3H(s)dx(s),$$

where $H(s) \equiv F(y(s),x(s))^{-1}G(y(s),x(s))$.

Repeatedly applying the product rule of differentiation to $F(s)H(s) = G(s)$, implies that $F(s)H(s) + F(s)dH(s) = dG(s)$, $d^2F(s)H(s) + 2dF(s)dH(s) + F(s)d^2H(s) = d^2G(s)$, and $d^3F(s)H(s) + 3d^2F(s)dH(s) + 3dF(s)d^2H(s) + F(s)d^3H(s) = d^3G(s)$. Then, solving for $dH(s)$, $d^2H(s)$, and $d^3H(s)$, implies that

(C.7)
$$dH(s) = F(s)^{-1}[dG(s) - dF(s)H(s)],$$

$$d^2H(s) = F(s)^{-1}[d^2G(s) - d^2F(s)H(s) - 2dF(s)dH(s)],$$

$$d^3H(s) = F(s)^{-1}[d^3G(s) - d^3F(s)H(s) - 3d^2F(s)dH(s) - 3dF(s)d^2H(s)].$$

Equations (C.7) are recursive. For given GCESUF, we compute $F(s)$, $G(s)$, and $H(s)$. Then, we compute $dF(s)$, $dG(s)$, and $dH(s)$. Then, we compute $d^2F(s)$, $d^2G(s)$, and $d^2H(s)$. Finally, we compute $d^3F(s)$, $d^3G(s)$, and $d^3H(s)$.

For $k = 1$, 2, and 3, $d^kF(s)$ and $d^kG(s)$ are

(C.8)
$$d^kF(s) = \begin{bmatrix} 0_{nxn} & 0_{nx1} \\ d^kS(s)^T & 0_{1x1} \end{bmatrix}, \quad d^kG(s) = \begin{bmatrix} 0_{nx2n} & 0_{nxn} \\ 0_{1x2n} & d^kS(s)^T \end{bmatrix}.$$

We assume $x(s)$ follows a 1st-order polynomial for $s \in [t,t+1)$ and $t = 1,\ldots,T$,

(C.9)
$$x(s) = x_t^! + \nabla x_t(s-t),$$

with n×1 coefficient $x_t^!$ and ∇x_t. Whereas exogenous coefficients remain at initial values, indexed at $t_1 = t$, throughout computations in period t, y coefficients $\nabla y_{t_i}, \ldots, \nabla^4 y_{t_i}$ are indexed by t_i and updated at each iteration $i = 1, \ldots, h$. From x(s) process (C.9), we require only that it passes through initial and ending x(s) in period t, i.e., $x(t) = x_t^!$ and $x(t+1) = x_{t+1}$.

For $k = 4$ and $s \in [t, t+1)$, differentiating x(s) process (C.9) with respect to s, implies that

(C.10) $\quad dx(s) = \nabla x_t dx.$

In seven numbered steps, we now discuss the details of computing Δy_t iteratively, for $t_i = t_1, \ldots, t_h$. The steps pertain to any period $t = 1, \ldots, T$ and are recursive, so that every step can be completed as long as the previous steps have been completed.

Step 1: Initialize y_t and x_t and their differentials.

For $s = t_1 = t$, compute $y(s) = y_t$, $x(s) = x_t$. By equation (C.10), set $dx(s) = \nabla x_t ds$.

Step 2: Compute 1st-order y coefficient.

Because true and approximate y processes are evaluated at the same times, $s = t_i = t_1, \ldots, t_h$, in all remaining steps y differentials and coefficients equate as $d^k y(t_i) = \nabla^k y_{t_i} ds$. Thus, using equation (C.2), the first equation of (C.5), equation (C.10), and $\nabla x_t = \Delta x_t = x_{t+1} - x_t^!$, compute

(C.11) $\quad H(y_t, x_t) = F(y_t, x_t)^{-1} G(y_t, x_t),$

$$\nabla y_t = H(y_t, x_t)\, \Delta x_t .$$

Step 3. Compute 2nd-order y coefficient.

Compute

$$(C.12) \qquad dF(s) = \begin{bmatrix} 0_{nxn} & 0_{nx1} \\ dS(s)^T & 0_{1x1} \end{bmatrix}, \quad dG(s) = \begin{bmatrix} 0_{nx2n} & 0_{nxn} \\ 0_{1x2n} & dS(s)^T \end{bmatrix}.$$

Following the first equation of (C.7), compute

$$(C.13) \qquad dH(y_t, x_t) = F(y_t, x_t)^{-1}[dG(y_t, x_t) - dF(y_t, x_t)H(y_t, x_t)].$$

Following the second equation of (C.5), the first equation of (C.6), equation (C.15), and, using $d^2y(t_1) = \nabla^2 y_t$, compute

$$(C.14) \qquad \nabla^2 y_t = dH(y_t, x_t)dx_t + H(y_t, x_t)d^2x_t.$$

Step 4: Compute 3rd-order y coefficient.

Following equation (C.8), compute

$$(C.15) \qquad d^2F(s) = \begin{bmatrix} 0_{nxn} & 0_{nx1} \\ d^2S(s)^T & 0_{1x1} \end{bmatrix}, \quad d^2G(s) = \begin{bmatrix} 0_{nx2n} & 0_{nxn} \\ 0_{1x2n} & d^2S(s)^T \end{bmatrix}.$$

Following the second equation of (C.7), compute

$$(C.16) \qquad d^2H(y_t, x_t) = F(y_t, x_t)^{-1}[d^2G(y_t, x_t) - d^2F(y_t, x_t)H(y_t, x_t) - 2dF(y_t, x_t)dH(y_t, x_t)].$$

Following the third equation of (C.5), the second equation of (C.6), equation (C.15), and, using $d^3y(t_1) = \nabla^3 y_t$, compute

$$(C.17) \qquad \nabla^3 y_t = d^2H(y_t, x_t)\,\Delta x_t$$

Step 5: Compute 4th-order y coefficient and update y.

Following equation (C.8), compute

$$(C.18) \qquad d^3F(s) = \begin{bmatrix} 0_{nxn} & 0_{nx1} \\ d^3S(s)^T & 0_{1x1} \end{bmatrix}, \quad d^3G(s) = \begin{bmatrix} 0_{nx2n} & 0_{nxn} \\ 0_{1x2n} & d^3S(s)^T \end{bmatrix}.$$

Following the third equation of (C.7), compute

$$(C.19) \qquad d^3H(y_t, x_t) = F(y_t, x_t)^{-1}[d^3G(y_t, x_t) - d^3F(y_t, x_t)H(y_t, x_t) - 3d^2F(y_t, x_t)dH(y_t, x_t) - 3dF(y_t, x_t)d^2H(y_t, x_t)].$$

Following the fourth equation of (C.5), the third equation of (C.6), equation (C.15), and, using $d^4y(t_1) = \nabla^4 y_t$, compute

$$(C.20) \qquad \nabla^4 y_t = d^3H(y_t, x_t)\,\Delta x_t \ .$$

Following equations (C.4), compute

$$(C.21) \qquad \Delta\hat{y}_t = \nabla y_t h^{-1} + (1/2)\nabla^2 y_t h^{-2} + (1/6)\nabla^3 y_t h^{-3} + (1/24)\nabla^4 y_t h^{-4}$$

and update y as $y_{t_2} = y_t + \Delta\hat{y}_t$.

Step 6: Update x_{t_2} and y.

For $s = t_2 = t + 1/h$, following equations (C.9) and (C.10), update exogenous variables and their differentials, as

(C.22) $x_{t_2} = x_t^{!} + \nabla x_t(1/h)$,

such that coefficient ∇x_t remains at initially computed values. For $s = t_2$, repeat steps 2 to 5 and update y coefficients to $\nabla y_{t_2}, \ldots, \nabla^4 y_{t_2}$. Following equation (C.21), compute $\Delta\hat{y}_{t_2} = \nabla y_{t_2} h^{-1} + (1/2)\nabla^2 y_{t_2} h^{-2} + (1/6)\nabla^3 y_{t_2} h^{-3} + (1/24)\nabla^4 y_{t_2} h^{-4}$ and update y as $y_{t_3} = y_{t_2} + \Delta\hat{y}_{t_2}$.

Step 7: Repeat steps 2 to 6.

For $s = t_3 = t + 2/h$, update exogenous variables and their differentials, as

(C.23) $x_{t_3} = x_t^{!} + \nabla x_t(2/h)$.

For $s = t_3$, repeat steps 2 to 5 and update y coefficients to $\nabla y_{t_3}, \ldots, \nabla^4 y_{t_3}$. Compute $\Delta\hat{y}_{t_3} = \nabla y_{t_3} h^{-1} + (1/2)\nabla^2 y_{t_3} h^{-2} + (1/6)\nabla^3 y_{t_3} h^{-3} + (1/24)\nabla^4 y_{t_3} h^{-4}$ and update y as $y_{t_4} = y_{t_3} + \Delta\hat{y}_{t_3}$. Repeat these steps for $s = t_4 = t + 3/h, \ldots, s = t_h = t + (h-1)/h$. At the last step, compute $\Delta\hat{y}_{t_h}$.

MSP computations are carried out with the two nested loops: outer loop indexed by sample periods t = 1,...,T and inner loop indexed by number of MSP steps per period i = 1,...,h.

Appendix D: Proof of MSP global accuracy of order h^{-k}.

Let f(x) denote a real-valued scalar function of real number x. The same proof holds with more complicated notation if x and f(x) are

finite-dimensional vectors. For simplicity, but without losing its essence, we give the proof in the purely scalar case.

Suppose that $f(x)$ is $k+1$ times differentiable for $x \in [x_0, x_h] = \{x | x_0 \leq x \leq x_h\}$, that $f(x)$ can be evaluated at some center point of approximation $x = x_0$, and that derivatives $f^{(j)}(x)$ of $f(x)$, for $j = 1,\ldots,k+1$, can be evaluated for $x \in [x_0, x_h]$. Then, $f(x_h)$ has the kth-order Taylor approximation centered at x_0,

$$\text{(D.1)} \qquad \hat{f}^{(SSP,k)}(x_h) = f(x_0) + \sum_{j=1}^{k} (1/j!) f^{(j)}(x_0)(x_h - x_0)^j,$$

with approximation error

$$\text{(D.2)} \qquad \hat{\varepsilon}^{(SSP,k)}(x_h) = (1/(k+1)!) f^{(k+1)}(\omega)(x_h - x_0)^{k+1},$$

for $\omega \in (x_0, x_h) = \{x | x_0 < x < x_h\}$ (Apostol, 1974, pp. 113-114). We denote (D.1) with superscript SSP because it's a "single-step perturbation" (Chen & Zadrozny, 2003).

Partition $[x_0, x_h]$ into h subintervals as $[x_0, x_h] = \bigcup_{i=1}^{h} [x_{i-1}, x_i]$ of equal length $|x_i - x_{i-1}| = |x_h - x_0|/h$, for $i = 1,\ldots,h$. Then, the kth-order MSP approximation of $f(x_h)$ centered at x_0 is

$$\text{(D.3)} \qquad \hat{f}^{(MSP,k,h)}(x_h) = f(x_0) + \sum_{i=1}^{h} \sum_{j=1}^{k} (1/j!) f^{(j)}(x_{i-1})(x_i - x_{i-1})^j,$$

with approximation error

$$\text{(D.4)} \qquad \hat{\varepsilon}^{(MSP,k,h)}(x_h) = \sum_{i=1}^{h} (1/(k+1)!) f^{(k+1)}(\omega_i)(x_h - x_0)^{k+1} h^{-(k+1)},$$

for $\omega_i \in (x_{i-1}, x_i)$. If $h = 1$, then, MSP (D.3)-(D.4) reduce to SSP (D.1)-(D.2). Suppose, for given k, h, and $\omega_i \in (x_{i-1}, x_i)$, for $i = 1,\ldots,k$, that there is an upper bound $K > 0$ such that

$$\left|\frac{f^{(k+1)}(\omega_i)(x_h - x_0)^{k+1}}{(k+1)!}\right| \leq K. \tag{D.5}$$

Then, the triangle inequality applied to equation (D.4) implies that

$$|\hat{\varepsilon}^{(MSP,k,h)}(x_h)| = |f(x_h) - \hat{f}^{(MSP,k,h)}(x_h)| \leq Kh^{-k}. \tag{D.6}$$

If $f(x)$ is analytic, then, for given $(x_h - x_0)$, $\hat{\varepsilon}^{(SSP,k)}(x_h)$ can be made arbitrarily small only by increasing k, which is expensive because it involves deriving, programming, and computing with higher-order derivatives of $f(x)$. Therefore, SSP is "expensively dependent" on $(x_h - x_0)$ and is a "local approximation".

Whether $f(x)$ is analytic or not, for given $(x_h - x_0)$, k, and K, $\hat{\varepsilon}^{(MSP,k,h)}(x_h)$ can be made arbitrarily small either by increasing k or by increasing h, or both. Increasing h involves only passing through an already programmed loop more times, which requires no extra derivations or programming. Therefore, MSP is "inexpensively dependent" on $(x_h - x_0)$ and is a "global approximation" accurate of order h^{-k}.

REFERENCES

Apostol, T. (1974), *Mathematical Analysis*, second edition, Reading, MA: Addison-Wesley.

Arrow, K.J., H.B. Chenery, B.S. Minhas, and R.M. Solow (1961), Capital-Labor Substitution and Economic Efficiency", *Review of Economics and Statistics* 43: 225-250.

Barnett, W.A. (2011), *Getting It Wrong*, Cambridge, MA: MIT Press.

Braithwait, S.D. (1980), "The Substitution Bias of the Laspeyres Price Index: An Analysis Using Estimated Cost-of-Living Indexes," *American Economic Review* 70: 64-77.

Breslaw, J.A. and J.B. Smith (1995), "A Simple and Efficient Method for Estimating the Magnitude and Precision of Welfare Changes," *Journal of Applied Econometrics* 10: 313-327.

Chen, B. and P.A. Zadrozny (2003), "Higher Moments in Perturbation Solution of the Linear-Quadratic Exponential Gaussian Optimal Control Problem", *Computational Economics* 21: 45-64.

Chen, B. and P.A. Zadrozny (2009), "Multi-Step Perturbation Solution of Nonlinear Differentiable Equations: Applied to an Econometric Analysis of Productivity," *Computational Statistics and Data Analysis* 53: 2061-2074.

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

Coats, A.T. (1939), "Hedonic Price Indexes with Automobile Examples," pp. 99-119 in *The Dynamics of Automobile Demand*, New York, NY: General Motors Corp.

Cobb, C.W. and P.H. Douglas (1928), "A Theory of Production", *American Economic Review* 18: 139-165.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers," *Journal of Econometrics* 4: 115-145.

Divisia, F. (1928), *Economique Rationelle*, Paris, France: Gaston Doin.

Dumagan, J.C. and T.D. Mount (1997), "Approximating Compensated Income from Ordinary Demand Functions," *Economics Letters* 55: 191-201.

Fisher, I (1922), *The Making of Index Numbers*, Boston, MA: Houghton-Mifflin.

Ford, L.R. (1955), *Differential Equations*, New York, NY: McGraw-Hill.

Gabor-Toth, E. and P. Vermeulen (2018), "The Relative Importance of Taste Shocks and Price Movements in the Variation of Costs of Living: Evidence from Scanner Data", working paper, European Central Bank, Frankfurt, Germany.

Golub, G.H. and C.F. Van Loan (1983), Matrix Computations, Baltimore, MD: Johns Hopkins.

Griewank, A. (2000), *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, Philadelphia, PA: Society for Industrial and Applied Mathematics.

Griliches, Z. (1961), "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," pp. 173-196 in *The Price Statistics of the Federal Government*, New York, NY: National Bureau of Economic Research.

He, J.H. (2006), "Some Asymptotic Methods for Strongly Nonlinear Equations", *International Journal of Modern Physics* 10: 1141-1199.

Hoerl, A.E. and R.W. Kennard (1970a), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics* 12: 55-67.

Hoerl, A.E. and R.W. Kennard (1970b), "Ridge Regression: Application to Nonorthogonal Problems," *Technometrics* 12: 69-82.

Klick, J. (2018), "Improving Initial Estimates of the Chained Consumer Price Index," *Monthly Labor Review* February 2018: 1-10, Washington, DC: Bureau of Labor Statistics.

Konus, A.A. (1939), "The Problem of the True Index of the Cost of Living," *Econometrica* 7: 10-29.

Laspeyres, E.L.E. (1871), "Die Berechnung einer Mittleren Warenpreissteigerung," *Jahrbueher fuer Nationaloekonomie und Statistik* 16: 296-315.

Magnus, J.R. and H. Neudecker (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York, NY: J. Wiley & Sons.

Mann, H.B. (1943), "Quadratic Forms with Linear Constraints," *American Mathematical Monthly* 7: 430-433.

Paasche, H. (1874), "Ueber die Preisentwicklung der letzten Jahre nach den Hamburger Boersennotirungen", *Jahrbueher fuer Nationaloekonomie und Statistik* 23: 168-178.

Redding, S. and D.E. Weinstein (2020), "Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences," *Quarterly Journal of Economics* 135: 503-560.

Samuelson, P.A. (1947), *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.

Sun, Z. and Y. Xie (2013), "Error Analysis and Comparison of Two Algorithms Measuring Compensated Income," *Computational Economics* 42: 433-452.

Theil, H. (1971), *Principles of Econometrics*, New York: Wiley.

Theil, H. and A.S. Goldberger (1961). "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review* 2: 65-78.

Törnqvist, L. (1936) "The Bank of Finland Consumption Price Index," *Bank of Finland Monthly Bulletin* 10: 1-8.

Vartia, Y.O. (1983), "Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions," *Econometrica* 51: 79-98.

Zadrozny, P.A. (1988), "Gaussian Likelihood of Continuous-Time ARMAX Models When Data Are Stocks and Flows at Different Frequencies", *Econometric Theory* 4: 108-124.