Scanner Data and the Construction of Inter-Regional Price Indexes.

July 14, 2025

NBER Summer Institute.
Boston, United States.

Erwin W. Diewert(The University of British Columbia & UNSW),
Naohito Abe (Hitotsubashi University),
Akiyuki Tonogi (Toyo University),
and
Chihiro Shimizu (Hitotsubashi University).

一橋大学 Hitotsubashi University

1. Motivations and Outlines.

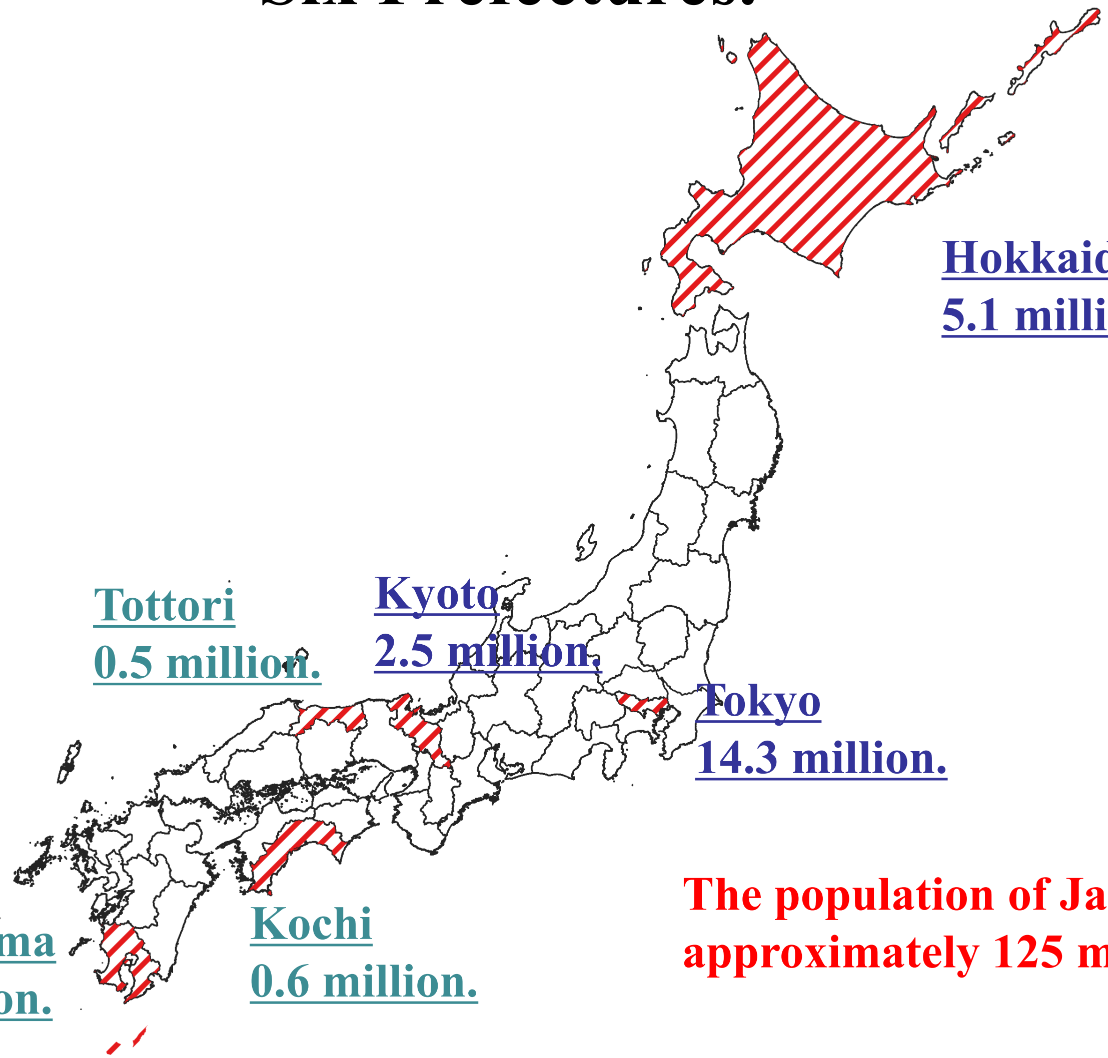
- **The paper addresses two main questions:**
  - How can we construct interregional price (and quantity) indexes for a country at the first stage of aggregation that are **"*transitive*" over "*time*" and "*space*"**?
  - How can we measure **the impact of differences in product availability on price levels and welfare between large cities and smaller towns**? How can we measure the effects on price levels and welfare of **smaller choice sets** for regions that **have a limited availability** of products?

- **Many alternative multilateral indexes:**
  - Bilateral ***Fisher Indexes*** and ***GEKS Multilateral Indexes***.
  - Weighted and Unweighted Time Product Dummy ***Hedonic Regressions***.
  - ***Geary Khamis Multilateral Indexes***.
  - The Estimation of Systems of ***Inverse Demand Functions***.
  - The Econometric Estimation of ***Linear Preferences***.
  - The Estimation of ***CES Preferences***.
  - The Econometric Estimation of ***KBF Preferences*** with a Rank One Substitution Matrix.

2. Data: Scanner Data-Rice.

- We use **household-level Rice price and quantity data for the 80 top-selling rice products over a 24-month period(2021 Jan – 2022 Dec) across six prefectures in Japan**.
  - Given that rice is the staple food in Japan and is consumed by virtually the entire population across all regions, we consider rice data to be particularly suitable for advancing regional price index research.

Six Prefectures.



The population of Japan is approximately 125 million.

The transitivity problem.

- A bilateral index number formula that provides an estimate of the price level in one region or period to the price level in another region or period is basically **a weighted average of ratios of product prices** where the price of each product n in one region-month is compared to the same product n price in another region-month.
  - Suppose we want to compare the prices in ***period 3*** with the same prices in period 1 for the ***same region***. Then we could construct a fixed base bilateral index number that directly compared **the period 3 prices to the period 1 prices**. Call this index **P(3/1).**
  - Alternatively, we could do a series of comparisons, **comparing the prices of period 2 to period 1**, obtaining the index **P(2/1),** and then comparing the prices of period 3 to the corresponding period 2 prices, obtaining the index **P(3/2).**
  - ***The chained index between periods 3 and 1*** is the product of **the two chain link indexes, P(2/1) times P(3/2).** We would like **P(3/1) to equal P(2/1)×P(3/2)** but this path independence **or *transitivity property frequently fails***.
  - When this property fails, we say that we have ***a chain drift problem***.

Chain drift problem.

- **The *chain drift problem* was not a problem before 2008.**
  - Before ***scanner data*** became available to National Statistical Offices, consumer (and producer) price indexes were produced in a very different way.
  - At the ***first stage of aggregation***, a sample of prices in a particular product category was collected in each month and these prices were compared to the ***same prices*** in the ***base month*** and either the arithmetic or geometric average of these product prices was taken as an estimate of the ***average price level of the current month to the price level of the product category in the base month***.
- **Solutions.**
  - The paper by de Haan (2008) led researchers to look for solutions to ***the chain drift problem***. Thus Ivancic, Diewert and Fox (2009) (2011) suggested using multilateral index number theory on ***a rolling window* of observations to mitigate *the chain drift problem*.** This strategy was eventually implemented by the Australian Bureau of Statistics (2016).

3. Comparison ***Multilateral methods***.

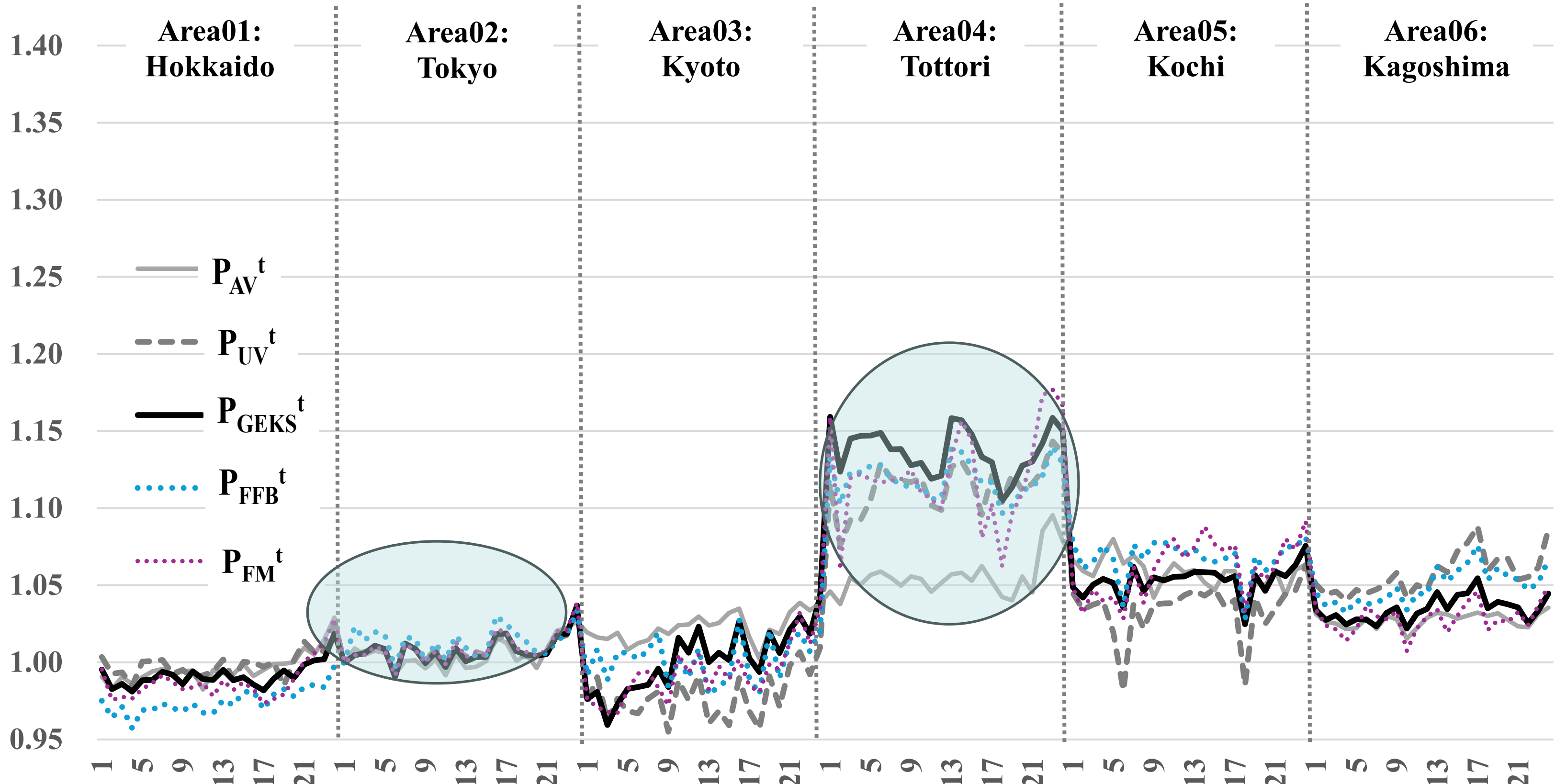
- **Multilateral indexes :**
  - Our goal is to calculate ***various multilateral indexes*** using our Japanese panel data on sales of **rice products for six Prefectures and the 24 months** in the years 2021-2022. We will use ***multilateral indexes*** which are ***transitive, invariant*** to changes in the units of measurement and satisfy a strong identity test for ***quantities***.
- **GEKS and Fixed Base Fisher indexes:**
  - The multilateral GEKS method is due to Gini (1924) (1931) and was further developed by Eltetö and Köves (1964) and Szulc (1964).
  - Our first multilateral method, the **GEKS** indexes(**P_{GEKS}**). We also calculate **fixed base Fisher indexes (P_{FFB} & P_{FM})** that take the first month in 2021 for Tokyo (the biggest Prefecture) as the base period.
  - For comparison purposes, we also **compute a simple index that uses the arithmetic average of prices in each region-month (P_{AV})** as an estimate of the price level for given month in the given region.
  - We also compute **unit value price indexes(P_{UV})** for comparison purposes.

Alternative multilateral indexes.

Index	Transitive	Substitution Flexibility	Handles Missing Prices	Other Comments
Bilateral Fisher (FB)	**× No**	Fully flexible (if no missing prices)	× No – relies on matched prices	**Not invariant to base period**
GEKS	**○Yes**	Fully flexible (if no missing prices)	× No – relies on matched prices	**Same flexibility issues as bilateral Fisher**
Unweighted TPD	○Yes	Linear preferences	○ Yes	Ignores expenditure shares and economic background
Weighted TPD	○Yes	Cobb–Douglas & linear preferences	○ Yes	Cobb–Douglas assumption is unrealistic
Geary-Khamis (GK)	○Yes	Leontief & linear preferences	○ Yes	Leontief assumption is unrealistic
Linear Utility Index	○Yes	Linear preferences	○ Yes	Simple but restrictive utility structure
CES Preferences	○Yes	CES preferences	○ Yes	Requires single substitution parameter
KBF Rank 1	○Yes	Flexible (Rank 1 substitution matrix)	○ Yes	Hard to estimate; high data requirements

***Transitivity implies resistance to chain drift.**

Chart 1: Average Price, Unit Value, GEKS, Fixed Base Fisher and Mizobuchi Fisher Price Indexes for Six Japanese Prefectures.



- It can be seen the 5 Tokyo indexes (observations t = 25-48) are ***all fairly close.***
- The ***Average Price index levels***, P_{AV}^t, tend to be ***smoother than the other indexes***.
- The Unit Value indexes, P_{UV}^t, are ***very volatile*** and for the most part.
- The ***bilateral Fisher indexes*** are somewhat close to each other but there is a great deal of volatility in these indexes, particularly for the low population Prefectures.
- The ***GEKS price levels*** are transitive and are invariant to changes in the units of measurement. However, they do not satisfy a strong version of Walsh’s (1901) (1921) Multiperiod Identity Test.

Summary.

- Price levels were somewhat stable on average for each Prefecture until the last 5 months of 2023 when prices rose quite rapidly.
- **Price levels differed substantially across Prefectures; the higher population Prefectures 1-3 had similar fairly stable price levels** for the first 20 months in our sample and then experienced rapid inflation.
- **The smaller population Prefectures 4-6 had substantially higher price levels compared to the Tokyo levels throughout the sample period**.
  - We will exclude ***the Average Price and Unit Value Price indexes, P_{AV}^t and P_{UV}^t***, from further consideration as ***"best" indexes*** due to the unrepresentative nature of P_{AV}^t and the volatility of P_{UV}^t.

4. Unweighted and weighted Time Product Dummy (TPD) *Hedonic price* indexes.

- We compute ***unweighted and weighted Time Product Dummy (TPD) Hedonic price indexes:*** $\mathbf{P_{TPD}}$, $\mathbf{P_{WTPD}}$
- The regional **Implicit Weighted Time Product Dummy** price levels $\mathbf{P_{IWTPD}}$.

- We have converted the 144 region(6)-month(24) price, quantity and value vectors into 144 p^t, q^t and v^t vectors of dimension 80 that are indexed by an artificial time index t for t = 1,…,144.
- These models are based on the price data satisfying (to some degree of approximation) the following equations:

  (10) $p_{tn} \approx \pi_t\alpha_n$; **$t = 1,\ldots,144$; $n \in S(t)$ (24 month & 6 regions)**

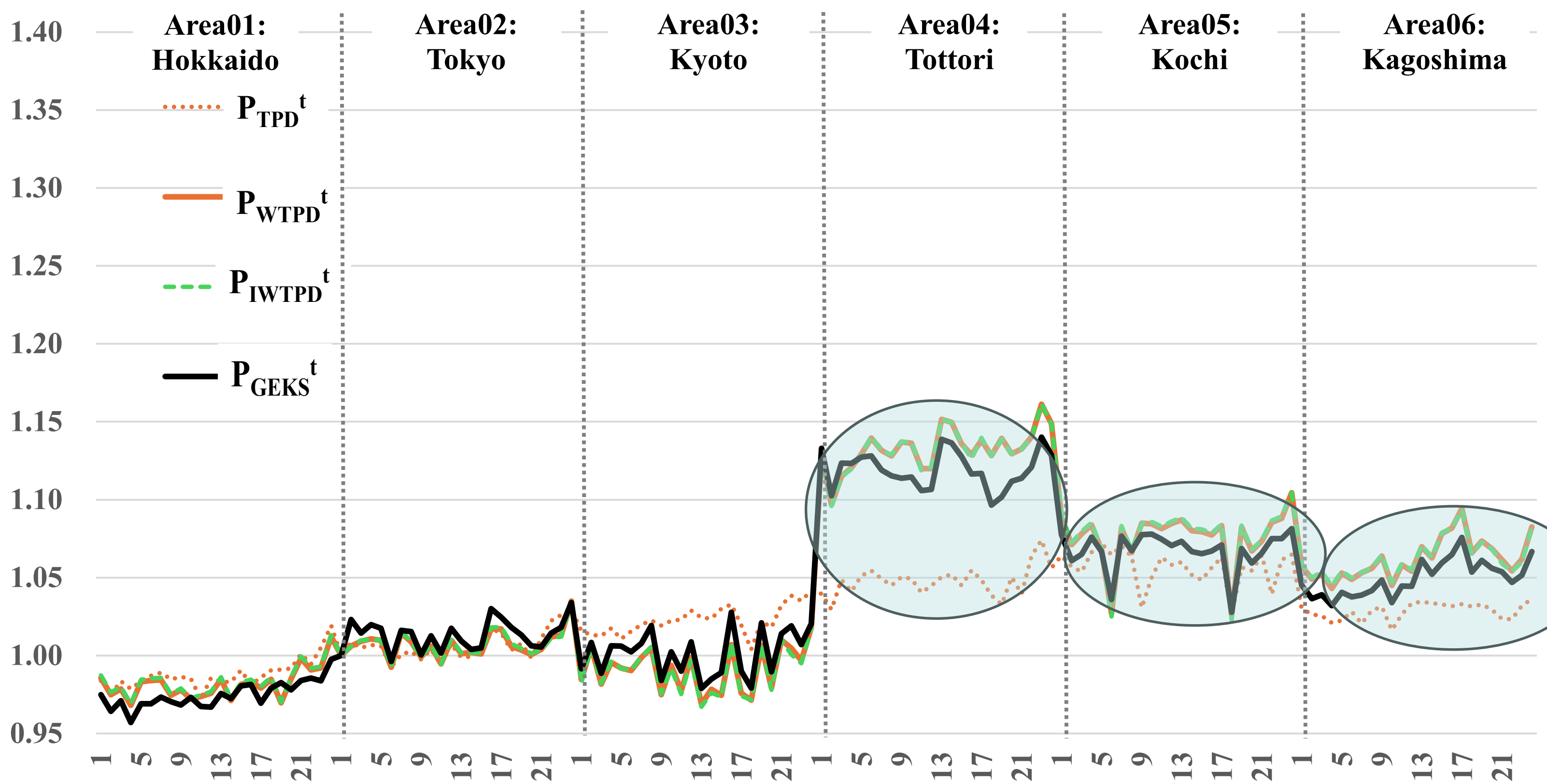
  – where π_t is interpreted as the period t price level and α_n is a parameter which reflects the quality (or marginal utility) of product n. Thus π_t is a summary measure for the level of prices in the region-month that corresponds to period t.

Alternative multilateral indexes.

Index	Transitive	Substitution Flexibility	Handles Missing Prices	Other Comments
Bilateral Fisher (FB)	× No	Fully flexible (if no missing prices)	× No – relies on matched prices	Not invariant to base period
GEKS	○Yes	Fully flexible (if no missing prices)	× No – relies on matched prices	Same flexibility issues as bilateral Fisher
Unweighted TPD	○Yes	**Linear preferences**	○ Yes	**Ignores expenditure shares and economic background**
Weighted TPD	○Yes	**Cobb–Douglas & linear preferences**	○ Yes	**Cobb–Douglas assumption is unrealistic**
Geary-Khamis (GK)	○Yes	Leontief & linear preferences	○ Yes	Leontief assumption is unrealistic
Linear Utility Index	○Yes	Linear preferences	○ Yes	Simple but restrictive utility structure
CES Preferences	○Yes	CES preferences	○ Yes	Requires single substitution parameter
KBF Rank 1	○Yes	Flexible (Rank 1 substitution matrix)	○ Yes	Hard to estimate; high data requirements

***Transitivity implies resistance to chain drift.**

Chart 2: Unweighted, Weighted and Implicit Weighted Time Product Dummy Price Indexes



- It can be seen that the **(unweighted) Time Product Dummy indexes P_{TPD}^t are not close to the more appropriate weighted indexes**, P_{WTPD}^t, P_{IWTPD}^t and P_{GEKS}^t, particularly for Prefectures 3-6 (observations 49-144).
- **The GEKS indexes, P_{GEKS}^t, are more volatile than the two Weighted TPD indexes**, P_{WTPD}^t and P_{IWTPD}^t, which are very close and cannot be distinguished from each other on the Chart.
- In particular, **the GEKS indexes are below the two Weighted Time Product Dummy indexes** for the smaller Prefectures (observations 73-144).

5. *Geary Khamis* Multilateral Indexes.

- The ***GK multilateral*** method was introduced by Geary (1958) in the context of **making international comparisons of prices**. Khamis (1970) showed that the equations that define **the method have a positive solution under certain conditions**.

- A modification of this method has been adapted to the time series context and is being used to construct some components of the Dutch CPI; see Chessa (2016).
  - The GK index was the multilateral index chosen by the Dutch to ***avoid the chain drift problem*** for the segments of their CPI that use scanner data.

Alternative multilateral indexes.

Index	Transitive	Substitution Flexibility	Handles Missing Prices	Other Comments
Bilateral Fisher (FB)	× No	Fully flexible (if no missing prices)	× No – relies on matched prices	Not invariant to base period
GEKS	○Yes	Fully flexible (if no missing prices)	× No – relies on matched prices	Same flexibility issues as bilateral Fisher
Unweighted TPD	○Yes	Linear preferences	○ Yes	Ignores expenditure shares and economic background
Weighted TPD	○Yes	Cobb–Douglas & linear preferences	○ Yes	Cobb–Douglas assumption is unrealistic
Geary-Khamis (GK)	○Yes	**Leontief & linear preferences**	○ Yes	**Leontief assumption is unrealistic**
Linear Utility Index	○Yes	Linear preferences	○ Yes	Simple but restrictive utility structure
CES Preferences	○Yes	CES preferences	○ Yes	Requires single substitution parameter
KBF Rank 1	○Yes	Flexible (Rank 1 substitution matrix)	○ Yes	Hard to estimate; high data requirements

***Transitivity implies resistance to chain drift.**

- Recall that S(t) was the set of products n that were purchased in region month t. Define **$S^*(n)$ as the set of periods t where product n was sold.**
- As was the case for **the Time Product Dummy multilateral system of price and quantity levels**, the equations which define **the GK price and quantity levels involve 144 price levels π_t and 80 quality adjustment parameters α_n** (recall equations (10) above). Define the vector q as the sum of the 144 observed quantity vectors q^t for each region -month t:
- (22) $q \equiv \Sigma_{t=1}^{144} q^t$.
-
- The equations which determine **the *GK price levels* $\pi_1,\ldots, \pi_{144}$ and *quality adjustment factors* $\alpha_1,\ldots,\alpha_{80}$** (up to a scalar multiple) are the following ones:
- (23) $\alpha_n = \Sigma_{t\in S^*(n)} [q_{tn}/q_n][p_{tn}/\pi_t] = \Sigma_{n=1}^{80} [1/q_n][p_{tn}q_{tn}][1/\pi_t]$; $n = 1,\ldots,80$
- (24) $\pi_t = p^t\cdot q^t/\alpha\cdot q^t = e^t/\alpha\cdot q^t$; $t = 1,\ldots,144$
- where **$\alpha \equiv [\alpha_1,\ldots,\alpha_N]$ is the vector of GK quality adjustment factors** and **$e^t \equiv p^t\cdot q^t$ is region-period t expenditure on the 80 rice products**. Once a solution α and $\pi_{1,\ldots,}$ π_{144} to equations (23) and (24) has been found, the period t price levels P^t can be set equal to the corresponding π_t and the period t quantity levels are defined as follows:
- (25) $Q^t \equiv \alpha\cdot q^t$; $t = 1,\ldots,144.$

- The GK indexes defined by (23)-(25) are exact for ***linear preferences (products are perfect substitutes)*** and ***for Leontief preferences (products are not substitutable at all)***.
- The first result is obvious from definition (25), i.e., utility in period t, u^t, is defined to be equal to the aggregate quantity $Q^t \equiv \alpha \cdot q^t$ for $t = 1,\ldots,T$ where T is equal to 144 in our empirical work.
- The second result was established by Diewert (1999; 58-60) but his proof is quite complicated. It is possible to establish ***that the GK indexes are exactly consistent with all purchasers having Leontief preferences*** by using the simple proof below.
  - Consider the case where there are N products and T observations. Assume that the period t price and quantity vectors, p^t and q^t for $t = 1,\ldots,T$ are consistent with purchasers of the N products all having Leontief preferences. If we have Leontief preferences, then every product that is purchased in one period must be purchased in all periods. This means that there exists an N dimensional vector of positive constants, β, which has components $\beta_1,\ldots,\beta_N$, utility levels $Q^1,\ldots,Q^T$ and unit cost price levels π^t such that the following equations are satisfied:

- (30) $q^t = \beta Q^t$; $\quad t = 1,\ldots,T$;
- (31) $\pi^t = \beta \cdot p^t$; $\quad t = 1,\ldots,T$.

- **Denote $D(q^t)$ as the diagonal matrix with the elements of q^t on the main diagonal.** With all products being positive in this case, equations (23)-(25) become the following equations:
- (32) $\alpha = [\Sigma_{t=1}^{T} D(q^t)]^{-1}[\Sigma_{t=1}^{T} D(q^t)p^t/\pi^t]$;
- (33) $\pi^t = p^t \cdot q^t/\alpha \cdot q^t$; $t = 1,\ldots,T$;
- (34) $Q^t \equiv \alpha \cdot q^t$; $t = 1,\ldots,T$.
-
- Substitute equations (30) and (31) into equations (32) and we obtain the following vector equation:
-
- (35) $\alpha = [\Sigma_{t=1}^{T} D(q^t)]^{-1}[\Sigma_{t=1}^{T} D(q^t)p^t/\pi^t]$
- $= [\Sigma_{t=1}^{T} D(\beta)Q^t]^{-1}[\Sigma_{t=1}^{T} D(\beta)Q^t p^t/\beta \cdot p^t]$
- $= [\Sigma_{t=1}^{T} Q^t]^{-1}[D(\beta)]^{-1}[D(\beta)][\Sigma_{t=1}^{T} Q^t p^t/\beta \cdot p^t]$
- $= [\Sigma_{t=1}^{T} Q^t]^{-1}[\Sigma_{t=1}^{T} Q^t p^t/\beta \cdot p^t]$.

- Thus the vector α is well defined by (35), given that we know the variables that appear in (30) and (31). Take the inner product of both sides of equations (35) with q^r for $r = 1,\ldots,T$. Using equations (30), we obtain the following T equations:
-
- (36) $\alpha\cdot\beta Q^r = [\Sigma_{t=1}^{T} Q^t]^{-1}[\Sigma_{t=1}^{T} Q^t p^t/\beta\cdot p^t]\cdot\beta Q^r$; $\quad r = 1,\ldots,T;$
- $= [\Sigma_{t=1}^{T} Q^t]^{-1}[\Sigma_{t=1}^{T} Q^t p^t\cdot\beta/\beta\cdot p^t]Q^r$
- $= [\Sigma_{t=1}^{T} Q^t]^{-1}[\Sigma_{t=1}^{T} Q^t]Q^r$
- $= Q^r$.

- Now normalize the α_n defined by (35) so that they satisfy the following constraint:
-
- (37) $\alpha\cdot\beta = 1$.

- The resulting GK indexes defined by (32)-(34) are exact for ***Leontief preferences***.
-

6. The Estimation of Systems of ***Inverse Demand Functions***.

- Traditional consumer demand theory in the case of ***homothetic*** **or linearly homogeneous preferences** works as follows:
  - Assume a once ***differentiable functional form for the household unit cost function c(p)*** (which is dual to the household linearly homogeneous utility function f(q)).
  - Assume that in period t, ***all households have the same preferences*** and face the vector of period t prices p^t.
  - Suppose each ***household maximizes utility subject to a budget constraint***. Let q^t be the observed vector of total purchases of the N products in scope and further assume that q^t is strictly positive. Let $e^t > 0$ be observed period t total expenditure on the products in scope. Then it can be shown that q^t, p^t and e^t satisfy the following system of consumer demand functions:
- (38) $q^t = e^t\nabla c(p^t)/c(p^t)$; $\quad t = 1,\ldots,T$
  - where $\nabla c(p^t)$ is the vector of first order partial derivatives of the unit cost function evaluated at p^t.

- We assume that f(q) is a ***linearly homogeneous function*** so that the resulting price index is independent of the scale of the quantity vectors q^t.
- We think that this is a reasonable assumption at the first stage of aggregation. It would be difficult for statistical agencies to produce price indexes that were conditional on the scale of purchaser demands.

- ***However, if there are missing products in one or more periods in the sample period then there are problems with the above traditional consumer demand methodology***. Suppose product n is not purchased in period t so that $q_{tn} = 0$. Then the nth component in equations (38) for period t becomes:
- (39) $q_{tn} = 0 = e^t[\partial c(p_{t1},\ldots,p_{tn},\ldots,p_{tN})/\partial p_n]/c(p_{t1},\ldots,p_{tn},\ldots,p_{tN})$.
-
- ***The problem is that we cannot observe the price of product n in period t, p_{tn}. Conceptually, it is the Hicksian reservation price*** which is just high enough to deter households from purchasing the product. Thus for every missing product in the sample of periods, we need to estimate an unknown reservation price in order to apply traditional consumer demand theory. This is not workable in practice. Hausman (1996) (1999) used variants of this cost function methodology to estimate reservation prices but it is not known how he solved this estimation problem.
- ***We turn to the estimation of the utility function, f(q), instead of estimating the dual unit cost function. When we make this switch, it turns out that we get a "practical" system of estimating equations***.

- ***The inverse demand function estimation methodology*** starts with the assumption that the observed period t quantity vector q^t is a solution to the following period t ***utility maximization problem***:
-
- (40) $\max_q \{f(q) : p^t \cdot q = e^t \, ; \, q \geq 0_N\}$; $t = 1,\ldots,T.$
- **It is workable if the functional form for the unit cost function is a CES (Constant Elasticity of Substitution) function because the reservation prices are known** (and equal plus infinity); see Feenstra (1994).
- Assuming that the linearly homogeneous function f is differentiable, the first order conditions for the observed q^t to solve the period t purchaser utility maximization problem are the following conditions:
-
- (41) $\nabla f(q^t) = \lambda_t p^t$; $t = 1,\ldots,T;$
- (42) $p^t \cdot q^t = e^t$; $t = 1,\ldots,T.$

- Take the inner product of both sides of (41) with q^t and solve the resulting equation for the Lagrange multiplier λ_t. We find that
- (43) $\lambda_t = q^t \cdot \nabla f(q^t)/e^t \qquad t = 1,\ldots,T$
- $\qquad = f(q^t)/e^t$
  - where the second line in (43) follows from Euler's Theorem on homogeneous functions which (using our assumption that f(q) is linearly homogeneous in q) implies that $f(q^t) = q^t \cdot \nabla f(q^t) = \Sigma_{n=1}^N q_{tn} \partial f(q^t)/\partial q_n$ for $t = 1,\ldots,T$. Substitute λ_t defined by (43) into equations (41) and after a bit of rearrangement, we obtain the following system of estimating equations:
- (44) $p^t = e^t \nabla f(q^t)/f(q^t)\ ; \qquad t = 1,\ldots,T.$
- ***The above equations assume that all products were purchased in each period t. However, equations (44) can be generalized to deal with the case of missing products***. When product n is missing in period t, we simply set q_{tn} equal to 0 and drop product n from the utility maximization problem defined by (40). This leads to the smaller system of estimating equations defined by (45):
- (45) $p_{tn} = e^t[\partial f(q^t)/\partial q_n]/f(q^t)\ ;\ \ t = 1,\ldots,T;\ n \in S(t).$
- **Equations (45) define a *system of inverse demand functions*.**

- We could assume a suitable functional form for the utility function f(q), add error terms of the right hand sides of these equations and use the resulting system of equations as estimating equations to determine the unknown parameters that characterize the function f(q). We also require a normalization on the parameters that define f(q) in order to obtain a unique function.
- ***It is usual in estimating systems of consumer demand equations to assume no missing prices and also to assume that the errors in the N equations pertaining to a single period are correlated so that a variance covariance matrix with N(N+1)/2 unknown parameters is also estimated.***
- ***In our present context where we have 80 products, this strategy becomes unworkable***. One strategy to solve this problem is to stack the estimating equations into a single estimating equation with only one variance parameter to deal with.
- However, this problem runs into a difficulty for National Statistical Offices: in general, the resulting parameter estimates are not invariant to the units in which we measure the products. Thus the ***resulting price and quantity indexes will also not be invariant to changes in the units of measurement***.

- A solution to these problems is to switch from prices as the dependent variables to expenditure shares.
- Thus multiply both sides of equation tn in equations (45) by q_{tn} and divide by period t expenditure e^t. This leads to the ***nth expenditure share in period t, s_{tn}***, as the dependent variable. These operations lead to the following system ***of inverse demand share estimating equations*** where e_{tn} is an error term:
-
- (46) $s_{tn} = q_{tn}[\partial f(q^t)/\partial q_n]/f(q^t) + e_{tn}$; $t = 1,\ldots,T;\ n \in S(t)$.
  - When product n in period t is not available, $s_{tn} = q_{tn} = 0$ so equations (46) are valid for $t = 1,\ldots,T$ and $n = 1,\ldots,N$. However, note that the error term e_{tn} is equal to 0 when $q_{tn} = 0$.
- We stacked the resulting augmented equations (46) into a single estimating equation. In particular, rather than specifying an explicit error structure for equations (46), we assumed that the unknown parameters which characterize the chosen utility function f(q) are estimated by solving the nonlinear least squares minimization problem (47) below with respect to the choice of these parameters:
- (47) $\min_{\text{parameters of } f(q)} \Sigma_{t=1}^{T} \Sigma_{n=1}^{N} \{s_{tn} - [q_{tn} f_n(q^t))/f(q^t)]\}^2$
  - where $f_n(q^t) \equiv \partial f(q^t)/\partial q_n$. A normalization on the parameters which characterize f(q) is also required in order to obtain unique parameter estimates.

- Once the unknown parameters characterizing f(q) have been estimated, we can calculate period t aggregate quantities Q^t and the corresponding price levels P^t using the following definitions:
-
- (48) $Q^t \equiv f(q^t)$; $P^t \equiv e^t/f(q^t)$; $t = 1,\ldots,T$.
- Note that the resulting quantity levels Q^t will satisfy the strong identity test for quantities: if $q^r = q^t$, then $f(q^r) = f(q^t)$, and hence $Q^r = Q^t$.
-
- The bottom line is this: ***it is virtually impossible to estimate systems of direct consumer demand functions when there are missing prices but it is reasonably straightforward to estimate systems of inverse demand functions. It is possible to estimate the utility function directly when there are missing prices but very difficult to estimate the corresponding dual unit cost function.***
-
- In the following three sections, we will work through the algebra presented in this section for three specific functional forms for f(q).

7. The Econometric Estimation of *Linear Preferences*.

- The case ***where the utility function is a homogeneous linear function of the quantities consumed***.
- Thus we assume that $f(q,\alpha)$ has the following functional form:
- (49) $f(q,\alpha) \equiv \Sigma_{n=1}^{N} \alpha_n q_n = \alpha \cdot q.$
-
- ***The least squares minimization problem (49)*** becomes the following problem:
-
- (50) $\min_{\alpha\text{'s}} \Sigma_{t=1}^{T} \Sigma_{n=1}^{N} \{s_{tn} - [q_{tn}\alpha_n/\alpha \cdot q^t]\}^2.$
  - If $\alpha^* \equiv [\alpha_1^*,\ldots,\alpha_N^*]$ is a solution to (50), then it can be seen that $\lambda\alpha^*$ is also a solution to (50) **where λ is any positive number.** This non-uniqueness always occur when we attempt to estimate utility functions.
  - The scale of utility is arbitrary so we need to impose at least one normalization on the estimated parameters in order to obtain a cardinal measure of utility.
  - There is another possible problem with **the minimization problem defined by (50)**: it can be the case that there is no solution to (50). For example, suppose that there are only 2 periods and 2 products in scope. Suppose further that product 1 is only available in period 1 and product 2 is only available in period 2. In this case, there are only 2 independent estimating equations for the nonlinear minimization problem defined by (50):

- (51) $1 = \alpha_1 q_{11}/(\alpha_1 q_{11} + \alpha_2 0) = 1$;
- (52) $1 = \alpha_2 q_{22}/(\alpha_1 0 + \alpha_2 q_{22}) = 1$.
- It can be seen that ***it is not possible to obtain estimates for the quality adjustment parameters α_1 and α_2*** in this situation. We need some product overlap between the periods in order to obtain solutions to (50).
  - In order to solve the problems of non-uniqueness and non-existence in general, we assume that there is a product that is present in **all 144 periods** and we assume that each product in scope is purchased in at least one period.
  - In our rice products data set, there were **11 products that were present in all 144 region-months**. Product 4 was the lowest number product that was present in all periods so we set $\alpha_4 = 1$.

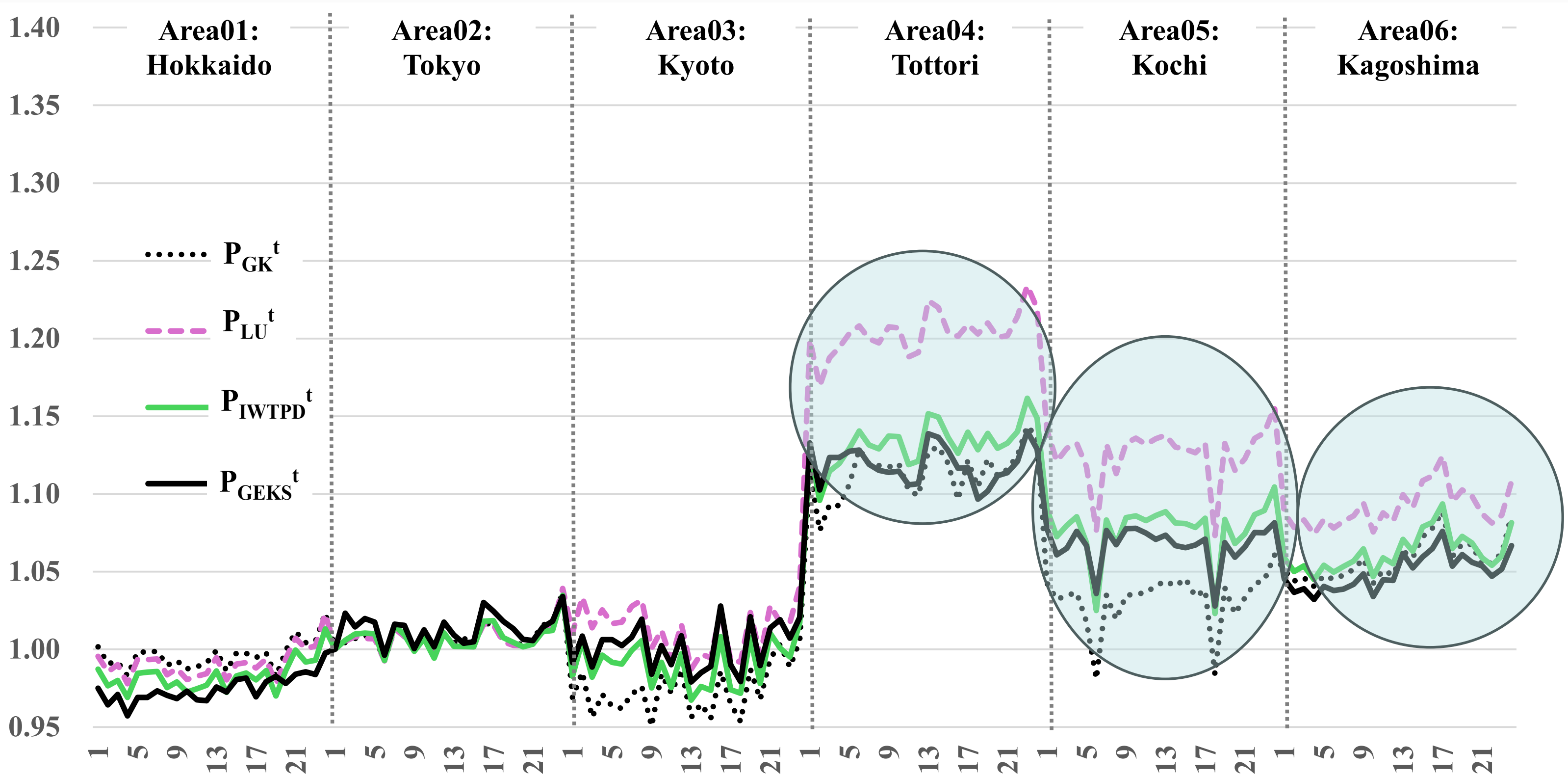
- Denote the estimated α_n by α_n^* except define $\alpha_4^* \equiv 1$. Define the vector $\alpha^* \equiv [\alpha_1^*,\ldots,\alpha_{80}^*]$ and define preliminary quantity and price levels, Q^{t*} and P^{t*} for period t (a region-month), as follows:
-
- (53) $Q^{t*} \equiv \alpha^* \cdot q^t$; $P^{t*} \equiv e^t/Q^{t*}$; $t = 1,\ldots,144$.
-
- Normalize the sequence of price levels P^{t*} into the series P^{t**} which is such that the normalized sequence of price levels equals 1 for t = 25 (month 1 for Tokyo):
-
- (54) $P^{t**} \equiv P^{t*}/P^{25*}$; $t = 1,\ldots,144$.
-
- Finally define the ***econometric linear utility price levels* for regions 1-6** for m = 1,…,24 as follows:
-
- **(55) $P_{LU}^{1,m} \equiv P^{m**}$; $P_{LU}^{2,m} \equiv P^{(24+m)**}$; $P_{LU}^{3,m} \equiv P^{(48+m)**}$; $P_{LU}^{4,m} \equiv P^{(72+m)**}$; $P_{LU}^{5,m} \equiv P^{(96+m)**}$;**
- **$P_{LU}^{6,m} \equiv P^{(120+m)**}$.**

Alternative multilateral indexes.

Index	Transitive	Substitution Flexibility	Handles Missing Prices	Other Comments
Bilateral Fisher (FB)	× No	Fully flexible (if no missing prices)	× No – relies on matched prices	Not invariant to base period
GEKS	○Yes	**Fully flexible (if no missing prices)**	**× No – relies on matched prices**	Same flexibility issues as bilateral Fisher
Unweighted TPD	○Yes	Linear preferences	○ Yes	Ignores expenditure shares and economic background
Weighted TPD	○Yes	**Cobb–Douglas & linear preferences**	**○ Yes**	Cobb–Douglas assumption is unrealistic
Geary-Khamis (GK)	○Yes	**Leontief & linear preferences**	**○ Yes**	Leontief assumption is unrealistic
Linear Utility Index	○Yes	**Linear preferences**	**○ Yes**	Simple but restrictive utility structure
CES Preferences	○Yes	CES preferences	○ Yes	Requires single substitution parameter
KBF Rank 1	○Yes	Flexible (Rank 1 substitution matrix)	○ Yes	Hard to estimate; high data requirements

***Transitivity implies resistance to chain drift.**

Chart 3: Geary Khamis, Linear Preferences, Implicit Weighted TPD and GEKS Price Indexes.



- The Linear Preferences price indexes, P_{LU}^t, are ***generally higher than the other indexes*** and very much higher for the **smaller population Prefectures**.
- The **Geary Khamis indexes**, P_{GK}^t , tended to ***be lower than the other indexes***.
- All four indexes were ***very close to each other for the highest population Prefecture***.
- The Implicit Weighted Time Product Dummy indexes, P_{IWTPD}^t, and the GEKS indexes, P_{GEKS}^t, ***were generally in the middle and fairly close to each other***.

Summary.

- It is interesting that the ***first 3 indexes*** **are all consistent with linear preferences** but they turned out to be quite different for our particular data set.

- What is striking is the fact that ***price levels in the 3 lowest population Prefectures*** **were generally much higher than price levels in the first 3 higher population Prefectures**.

- ***These differences indicate that there may be a problem in using national price indexes in order to deflate consumer expenditures into real consumption aggregates since national Consumer Price Indexes do not take differing interregional price levels into account in their construction.***

- Thus poverty measures and measures of national real consumption may be inaccurate to a significant degree.

8. The Estimation of ***CES Preferences***.

- ① P_{CES}: ***Direct*** from CES ***utility*** function, ② P_{ACES}: ***Indirect*** from CES ***Cost*** function (Feenstra type), ③ P_{CCES}: ***Direct*** estimation of CES ***Cost*** Function.

  – Our second example of the methodology explained in section 6 is the case where **the *utility* function is a CES (Constant Elasticity of Substitution) function**, f(q) defined as follows in the case of N products:

- **(56)** $f(q) \equiv [\Sigma_{n=1}^{N} \alpha_n (q_n)^{k}]^{1/k}$

  – where the α_n are positive parameters and the parameter k satisfies the following inequalities:

- (57) $0 < k \le 1$.

- **Note that if the parameter *k equals 1*, then the CES utility function defined by (56) becomes the linear utility function that was discussed in the previous section.**

  – Recall from section 6 that the observed period t vector q^t solves **the period t utility maximization problem** if $p_{tn} = e^t[\partial f(q^t)/\partial q_n]/f(q^t)$ for all $n \in S(t)$. If we multiply both sides of equation n by q_{tn}, then these first order necessary conditions become the following estimating equations:

  – (58) $s_{tn} = p_{tn}q_{tn}/e^t = q_{tn}[\partial f(q^t)/\partial q_n]/f(q^t) = \alpha_n (q_{tn})^k/\Sigma_{i\in S(t)} \alpha_i (q_{ti})^k$; $t = 1,\dots,T$; $n \in S(t)$.

- If $q_{tn} = 0$, then $s_{tn} = 0$. Thus the equations (58) can be replaced with the following equations:
- (59) $s_{tn} = \alpha_n(q_{tn})^k/\Sigma_{i=1}^N \alpha_i(q_{ti})^k$; $t = 1,\ldots,T$; $n = 1,\ldots,N$.

- We require that $k \le 1$ to ensure that the utility function is concave in the components of q and we require that $k > 0$ in order to ensure that the utility function is well defined if any component of the q^t vector happens to be equal to 0.
- The restrictions $0 < k < 1$ are also required in order to apply Feenstra's (1994) methodology for measuring the welfare effects of increased (or decreased) product choice.
- In our particular case, $N = 80$ and $T = 144$. We obtained estimates for the CES utility function by solving the following nonlinear least squares minimization problem:
- (60) $\min_{\alpha\text{'s}} \Sigma_{t=1}^{144} \Sigma_{n=1}^{80} \{s_{tn} - [\alpha_n(q_{tn})^k/\Sigma_{i=1}^N \alpha_i(q_{ti})^k]\}^2$.
  - Note that if $q_{tn} = 0$, then both s_{tn} and $\alpha_n(q_{tn})^k$ equal zero. If $\alpha^* \equiv [\alpha_1^*,\ldots, \alpha_N^*]$ and k is a solution to (60), then it can be seen that $\lambda\alpha^*$ and k is also a solution to (60) where λ is any positive number. Thus we imposed the normalization $\alpha_4 = 1$ because product 4 was the first product on our list of products that was present in all 144 region-periods.

- The estimated k was **$k^* = 0.95668$** with an estimated standard error equal to 0.00125. The corresponding elasticity of substitution σ^* was equal to:
- (61) **$\sigma^* \equiv 1/(1 - k^*) = 23.086.$**
- Denote the estimated α_n by $\alpha_n{}^*$ and define $\alpha_4{}^* \equiv 1$. Define the vector $\alpha^* \equiv [\alpha_1{}^*,\ldots,\alpha_{80}{}^*]$ and define preliminary CES quantity and price levels, Q^{t*} and P^{t*} for period t (a region-month), as follows:
- (62) $Q^{t*} \equiv [\Sigma_{n=1}{}^N \alpha_n{}^*(q_{tn})^{k*}]^{1/k*}$; $P^{t*} \equiv e^t/Q^{t*}$; $\quad t = 1,\ldots,144.$
  - Note that the P^{t*} are defined *indirectly* using the product test, $P^{t*}Q^{t*} = e^t$. Normalize the sequence of price levels P^{t*} into the series $P_{CES}{}^t$ which is such that the normalized sequence of price levels equals 1 for t = 25 (month 1 for Tokyo):
-
- (63) **$P_{CES}{}^t \equiv P^{t*}/P^{25*}$** ; $\quad t = 1,\ldots,144.$
-
- Finally define the econometric ***CES utility function price levels*** *for regions 1-6* as follows:
- (64) $P_{CES}{}^{1,m} \equiv P_{CES}{}^m$; $\quad P_{CES}{}^{2,m} \equiv P_{CES}{}^{(24+m)}$; $P_{CES}{}^{3,m} \equiv P_{CES}{}^{(48+m)}$;
- $P_{CES}{}^{4,m} \equiv P_{CES}{}^{(72+m)}$; $P_{CES}{}^{5,m} \equiv P_{CES}{}^{(96+m)}$; $P_{CES}{}^{6,m} \equiv P_{CES}{}^{(120+m)}$; $\quad m = 1,\ldots,24.$
- **Our CES price indexes $P_{CES}{}^t$ were defined *indirectly* using the estimated utility levels to deflate *actual expenditure levels* into *aggregate price levels*.**

- **CES cost function:**
  - There is another indirect method that could be used to define CES price levels given that we have estimated the CES utility function: we could use the estimated utility function to solve the following ***period t unit cost minimization problem*** for each period t:
- (65) $\min_q \{\Sigma_{n\in S(t)} p_{tn}q_n : f(q_1,q_2,\ldots,q_{80}) \geq 1;\ q_n = 0 \text{ if } n \notin S(t)\} = c(p^t)\ ;\ \ t = 1,\ldots,144.$
  - Suppose for the moment that there are no missing products for the period t cost minimization problem defined by (65) and our f(q) is defined by (62); i.e., $f(q) \equiv [\Sigma_{n=1}^N \alpha_n^*(q_n)^{k^*}]^{1/k^*}$. Then it can be shown that the CES unit cost function has the following functional form:
- (66) $c(p^t) = [\Sigma_{n=1}^N \beta_n^* (p_{tn})^{\kappa^*}]^{1/\kappa^*}$
- where the parameters β_n^* and κ^* are defined as follows:
- (67) $\beta_n^* \equiv (\alpha_n^*)^{1/(1-k^*)}$ for n = 1,...,80 and $\kappa^* \equiv -k^*/(1-k^*) = -22.0856$.
- In order to deal with the case where some products are not available in period t, **Feenstra (1994) assumed that the parameter κ^* which appears in definition (66) satisfies $\kappa^* < 0$**.

- This allowed Feenstra to set the reservation prices for the **missing products equal to $+\infty$ and thus when $\kappa^* < 0$, an infinite price p_{tn} raised to a negative power generates a zero**; i.e., if product n is unavailable in period t, then $(p_{tn})^{\kappa^*} = 1/(+\infty)^{|\kappa^*|} = 0$.
- Thus with **infinite *reservation prices* for *missing products***, period t unit cost is equal to:
- (68) $c(p^t) = [\Sigma_{n=1}^{N} \beta_n^* (p_{tn})^{\kappa^*}]^{1/\kappa^*} = [\Sigma_{n\in S(t)} \beta_n^* (p_{tn})^{\kappa^*}]^{1/\kappa^*} \equiv P^{t*}$; $t = 1,\ldots,144$
  - where the β_n^* and κ^* are defined by (67). Normalize the resulting period t unit costs P^{t*} into the following Alternative CES price levels, $P_{ACES}^{t} = P^{t*}/P^{25*}$ for $t = 1,\ldots,144$.
- Counterparts to definitions (64) are used to decompose P_{ACES}^{t} into the 6 regional indexes, $P_{ACES}^{1,m} - P_{ACES}^{6,m}$, which are listed in Table 5 in the Appendix. It should be noted that the use of the definitions in (67) led to β_n^* that were tiny if $\alpha_n^* > 1$ or β_n^* that were huge if $\alpha_n^* < 1$. This in turn led to estimated unit costs which exhibited excessive fluctuations 4.
- This method for ***forming CES price indexes is not recommended if the parameter κ^* is large in magnitude or if the elasticity of substitution $\sigma \equiv 1 - \kappa^*$ is large***.

- ***Our final CES set of regional price indexes is obtained by directly estimating the unit cost function defined by (68).*** Shephard's Lemma can be used to obtain cost minimizing quantities as functions of prices when preferences are represented by a differentiable unit cost function. Thus if preferences are represented by the CES utility function that is dual to a CES unit cost function that is defined by (66) in the case of no missing products, then $q_{tn} = Q^t\partial c(p^t)/\partial p_n$ for $n = 1,\ldots,N$. This approach can be generalized to the case of missing products and it leads to the following system of estimating equations:
- (69) $s_{tn} = \beta_n(p_{tn})^\kappa/\Sigma_{j\in S(t)}\, \beta_j(p_{tj})^\kappa + e_{tn}$; $t = 1,\ldots,144;\ n\in S(t)$.
-
- ***It proved technically difficult to set up the nonlinear least squares minimization problem that is associated with equations (69)*** so we used the following approach that is often used in the literature: take logarithms of both sides of equations (69) for $n\in S(t)$, subtract the resulting logarithmic equation for product 4 in period t from the corresponding log s_{tn} equation and set $\beta_4 = 1$ (so that the logarithm of β_4 equals 0). We obtain the following system of estimating equations where $\alpha_n \equiv \ln(\beta_n)$ for $n = 1,\ldots,80$ and e_{tn} is an error term:
- (70) $y_{tn} = \alpha_n + \kappa x_{tn} + e_{tn}$; $t = 1,\ldots,144;\ n\in S(t)$
  - where the e_{tn} are error terms, $y_{tn} \equiv \ln s_{tn} - \ln s_{t4}$ and $x_{tn} \equiv \ln p_{tn} - \ln p_{t4}$ for $t = 1,\ldots,144;\ n\in S(t)$.
  - See Samuelson and Swamy (1974) or Diewert (1974) (1976) for the details on how this dual approach works.

- The sum of absolute errors for the present regression was 5285.3 whereas the sum of absolute errors for the direct estimation of the CES utility function was **7.2587**.
- This is a very large difference in fit. Our new estimate for the parameter κ was $\kappa^* = -5.4994$ with a standard error equal to 0.1036. Thus our new estimate for the elasticity of substitution is:
-
- (71) $\sigma^* = 1 - \kappa^* =$ **6.4994**.
  - **Recall that our earlier estimate for the elasticity of substitution was equal to 23.086. This is a huge difference. Our preferred estimate for the elasticity of substitution is the estimate that results from the direct estimation of the CES utility function since this the utility function regression *fits the data much better than the cost function regression.***
  - This better fit for the utility function model is likely to carry over to other product classes where substitution between products is large.
  - The model that estimates the CES utility function reduces to a linear utility function if the parameter k in definition (56) is equal to 1 (and the resulting σ equals plus infinity) and the CES unit cost function reduces to a linear cost function if the parameter κ in (66) is equal to 1 (and the resulting σ equals 0).
  - Thus the CES regression that estimates the unit function will tend to have a hard time fitting the data if the products are highly substitutable because the starting point for the nonlinear regression defined by (69) is the case where $\kappa = 1$ and this is the unit cost function that corresponds to Leontief (no substitution) preferences.

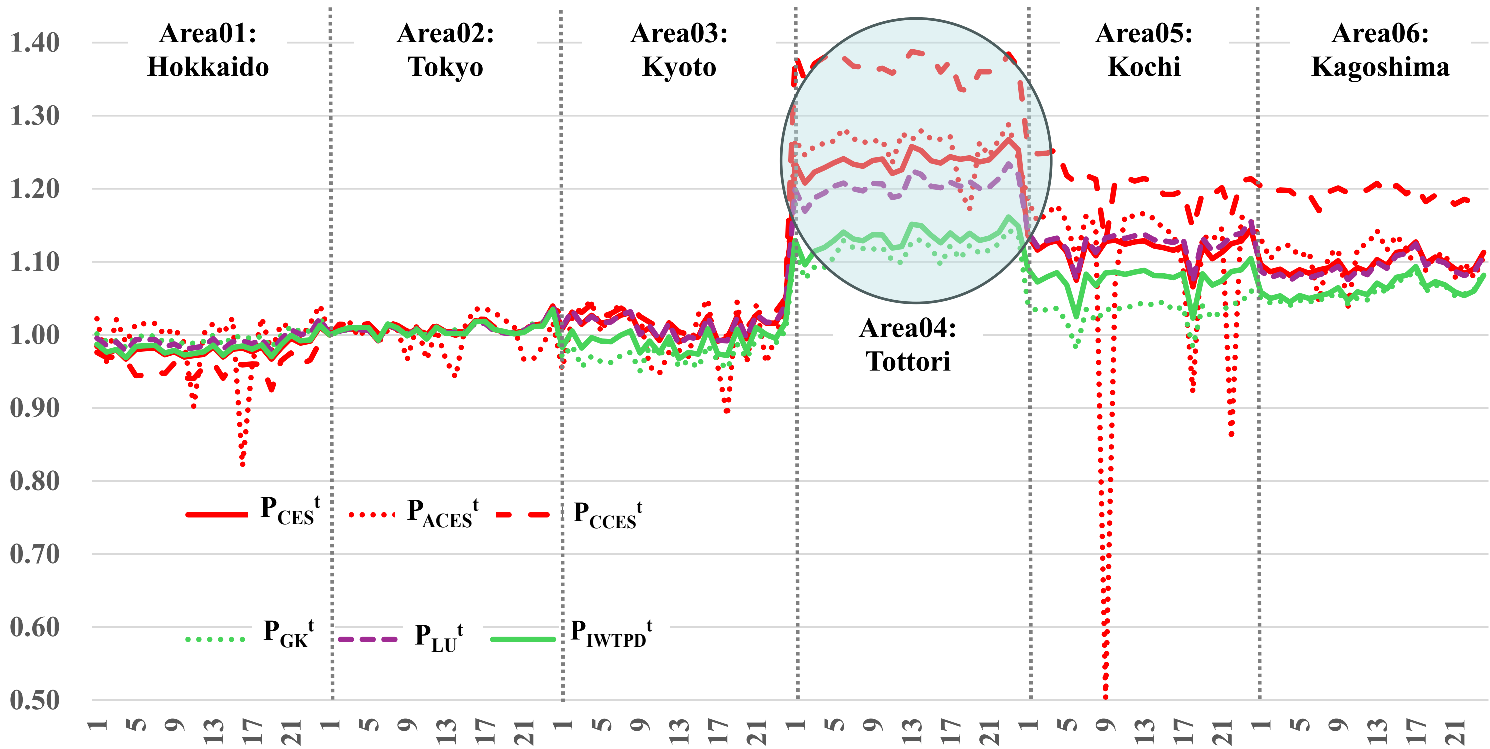
- Denote the estimated α_n by α_n^* and define $\alpha_4^* \equiv 1$. Define the vector $\alpha^* \equiv [\alpha_1^*,\ldots,\alpha_{80}^*]$ and define preliminary ***cost function based CES quantity and price levels*, Q^{t*} and P^{t*} for period t** (a region-month), as follows:
-
- (72) $P^{t*} \equiv [\Sigma_{n\in S(t)}\, \alpha_n^*(p_{tn})^{\kappa*}]^{1/\kappa*}$; $Q^{t*} \equiv e^t/P^{t*}$; t = 1,…,144.
  - Note that the Q^{t*} are defined *indirectly* using the product test, $P^{t*}Q^{t*} = e^t$. Normalize the sequence of cost function based price levels P^{t*} into the series P_{CCES}^{t} which is such that the normalized sequence of price levels equals 1 for t = 25 (month 1 for Tokyo):
-
- (73) $P_{CCES}^{t} \equiv P^{t*}/P^{25*}$; t = 1,…,144.
-
- Finally define the econometric ***Cost Function Based CES utility function price levels for regions 1-6*** as $P_{CCES}^{r,m}$ using P_{CCES}^{t} defined by (73) and an appropriate modification of definitions (64). The Cost Based CES indexes, $\mathbf{P_{CCES}^{t}}$.

Alternative multilateral indexes.

Index	Transitive	Substitution Flexibility	Handles Missing Prices	Other Comments
Bilateral Fisher (FB)	× No	Fully flexible (if no missing prices)	× No – relies on matched prices	Not invariant to base period
GEKS	○Yes	Fully flexible (if no missing prices)	× No – relies on matched prices	Same flexibility issues as bilateral Fisher
Unweighted TPD	○Yes	Linear preferences	○ Yes	Ignores expenditure shares and economic background
Weighted TPD	○Yes	**Cobb–Douglas & linear preferences**	○ Yes	**Cobb–Douglas assumption is unrealistic**
Geary-Khamis (GK)	○Yes	**Leontief & linear preferences**	○ Yes	**Leontief assumption is unrealistic**
Linear Utility Index	○Yes	**Linear preferences**	○ Yes	**Simple but restrictive utility structure**
CES Preferences	○Yes	**CES preferences**	○ Yes	**Requires single substitution parameter**
KBF Rank 1	○Yes	Flexible (Rank 1 substitution matrix)	○ Yes	Hard to estimate; high data requirements

一橋大学 HITOTSUBASHI UNIVERSITY

Chart 4: Alternative CES, GK, Linear Utility and Implicit Weighted TPD Price Indexes.



- The six alternative price indexes for ***Tottori*** are **widely separated with almost 30 percentage points** difference between **the highest and lowest index**.
- The cost function based CES price index $P_{CCES}t$ is highest, followed by the unit cost function price levels .
- The two lowest series were the Geary Khamis price indexes P_{GK}^t and the Weighted Implicit Time Product Dummy indexes , P_{IWTPD}^t. For Prefectures 4-6, the GK Price indexes tended to be lowest.

Summary.

- It is clear that **P_{ACES}^{t}** is not a suitable index due to its extreme volatility.

- A linear utility function is likely to overstate substitution possibilities so it is not surprising that ***the more flexible CES based price indexes***, **P_{CES}^{t}** and **P_{CCES}^{t}**, are generally higher than the linear utility function price indexes **P_{LU}^{t}**.

- What is surprising is that the ***cost function based CES indexes*** **P_{CCES}^{t}** ***are so much higher than the utility function based CES price indexes*** **P_{CES}^{t}**.

- Since the ***utility function based indexes*** **P_{CES}^{t}** fit the data so **much better than the *cost function based indexes*** **P_{CCES}^{t}**, the former indexes are preferred.

- With estimates for the elasticity of substitution in hand, we can use **Feenstra's 1994 methodology** to estimate the effects on welfare of different degrees of product availability across the 6 Prefectures.
  - Suppose that the CES unit cost function for month m in region r is defined as $c(p^{r,m}) \equiv [\sum_{n\in S(r,m)} \alpha_n (p_{rmn})^{\kappa}]^{1/\kappa}$ where S(r,m) is the set of rice products n that are purchased in month m in region r and the parameter κ is less than 0.
  - The unit cost $c(p^{r,m})$ represents the rice price level for region r in month m. Thus the rice consumer price index for month m in region r relative to the price level in Tokyo for the same month m is the ratio of unit costs, $c(p^{r,m})/c(p^{2,m})$. Feenstra obtained the following decomposition of $c(p^{r,m})/c(p^{2,m})$:

- (74) $P_{CES}{}^{r,m}/P_{CES}{}^{2,m} \equiv c(p^{r,m})/c(p^{2,m})$; r = 1,…,6; m = 1,…,24
- $\equiv [\sum_{n\in S(r,m)} \alpha_n (p_{rmn})^{\kappa}]^{1/\kappa} / [\sum_{n\in S(2,m)} \alpha_n (p_{2mn})^{\kappa}]^{1/\kappa}$
- $= [A^{r,m}]\times[B^{r,m}]\times[C^{r,m}]$
- where the three indexes in the last line of equations (74) are defined as follows:
- (75) $A^{r,m} \equiv [\sum_{n\in S(r,m)\cap S(2,m)} \alpha_n(p_{rmn})^{\kappa}]^{1/\kappa} / [\sum_{n\in S(r,m)\cap S(2,m)} \alpha_n(p_{2mn})^{\kappa}]^{1/\kappa}$; r = 1,…,6 ; m = 1,…,24
- (76) $B^{r,m} \equiv [\sum_{n\in S(r,m)} \alpha_n(p_{rmn})^{\kappa}]^{1/\kappa} / [\sum_{n\in S(r,m)\cap S(2,m)} \alpha_n (p_{rmn})^{\kappa}]^{1/\kappa}$; r = 1,…,6 ; m = 1,…,24
- (77) $C^{r,m} \equiv [\sum_{n\in S(r,m)\cap S(2,m)} \alpha_n(p_{2mn})^{\kappa}]^{1/\kappa} / [\sum_{n\in S(2,m)} \alpha_n (p_{2mn})^{\kappa}]^{1/\kappa}$; r = 1,…,6 ; m = 1,…,24.

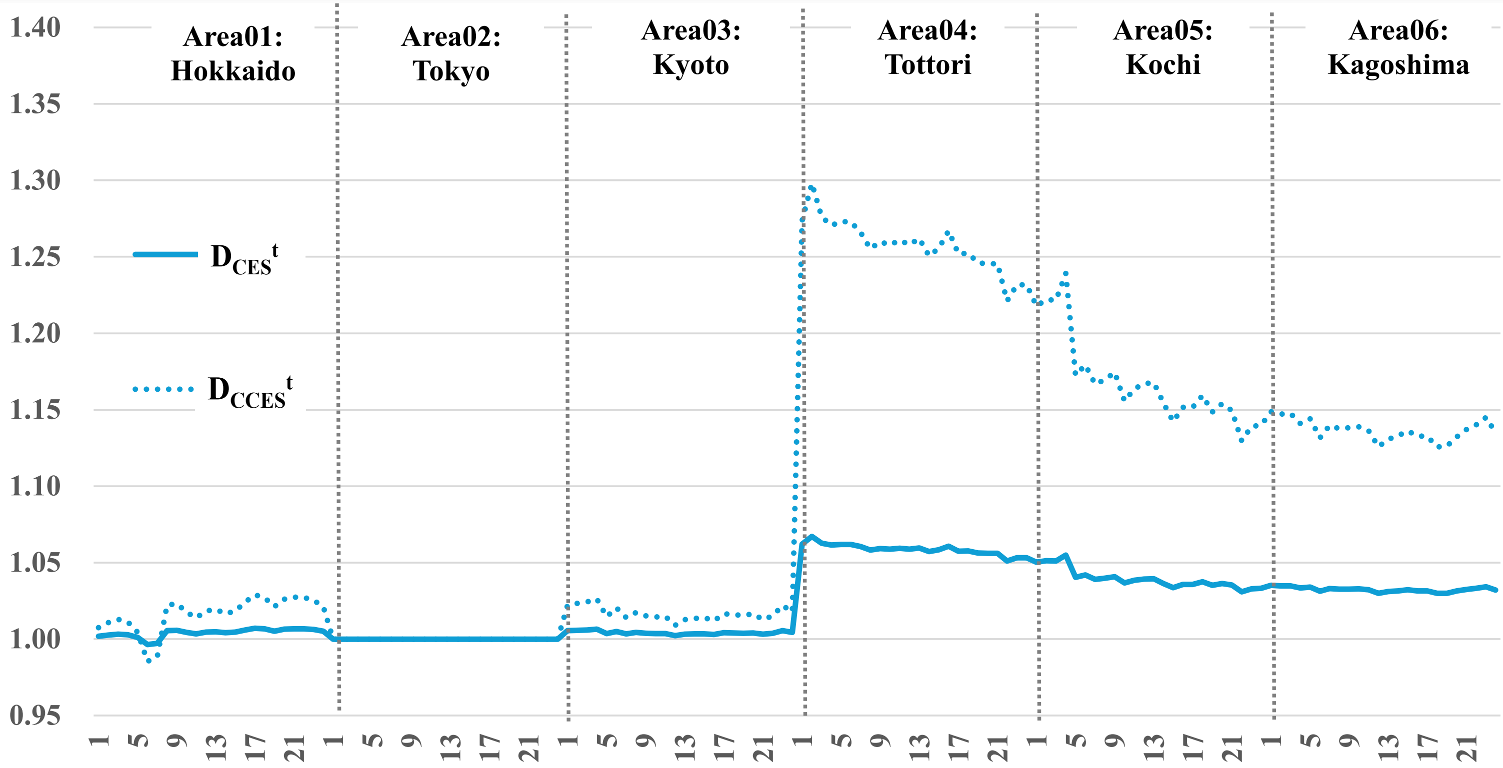
- The left hand side of (74) is $P_{CES}{}^{r,m}/P_{CES}{}^{2,m} \equiv c(p^{r,m})/c(p^{2,m})$ which is the overall CES rice ***Cost of Living index*** for Region r relative to Toyko in month m.
- The index $A^{r,m}$ is the relative cost of achieving ***the same utility level*** if purchasers faced the prices of rice products that are common to both Regions r and Tokyo (region 2) in month m with the Region r price level in the numerator and the Region 2 prices in the denominator.
- The index $B^{r,m}$ has the Region r cost of attaining one unit of utility if purchasers faced the actual prices of month m in Region r in the numerator and the denominator is the hypothetical Region r cost of attaining one unit of utility if only products found in Regions r *and* 2 were available.
- Thus $B^{r,m} \leq 1$. The lower $B^{r,m}$ is, the bigger is the benefit to purchasers in Region r of having extra products that were not available in Tokyo in month m. The Region r prices in month m are used in both numerator and denominator.
- The ***reciprocal*** of the index $C^{r,m}$ is again equal to the ratio of two unit costs: ***the cost of achieving one unit of utility in Region 2 in month m if region 2 products were available and the cost of achieving one unit of utility in Region 2*** if purchasers faced Region 2 prices in month m but were restricted to purchasing products that were available in both Regions 2 and r in month m. Region 2 prices in month m are used in the numerator and denominator.

- Thus $1/C^{r,m} \leq 1$ or $C^{r,m} \geq 1$ and the bigger $C^{r,m}$ is, the bigger is the benefit to Region 2 to having its choice set S(2,m) relative to the more restricted choice set $S(2,m) \cap S(r,m)$. We define the month m net cost to Region r of having its choice set relative to the corresponding month m, Region 2 choice set to be the product of $B^{r,m}$ and $C^{r,m}$:
-
- (78) $D^{r,m} \equiv [B^{r,m}] \times [C^{r,m}]$; $r = 1,\ldots,6$; $m = 1,\ldots,24$.
-
  - If $D^{r,m} > 1$, then the difference in choice sets between Region r and Region 2 adds to the Region r cost of living.
  - Of course, the product of $D^{r,m}$ and the matched product CES index $A^{r,m}$ is equal to the actual Cost of Living index between Regions r and 2 for month m, $P_{CES}{}^{r,m}/P_{CES}{}^{2,m}$. Thus $D^{r,m}$ can be interpreted as an adjustment to the matched product index that takes into account differences in product availability. For more details on the Feenstra methodology, see Feenstra (1994), Balk (1999), Melser (2006), Diewert and Feenstra (2017) (2022) and Diewert (2020b, 41-44).

- **Feenstra (1994)** showed that if Regions r and Tokyo (region 2) have a product in common for month m, then it is possible to estimate the indexes $B^{r,m}$ and $C^{r,m}$ without estimating the **CES cost function**, provided that we have an estimate for the parameter κ or equivalently for the elasticity of substitution, $\sigma \equiv 1 - \kappa$. His method starts by defining the following ***observable expenditure or sales ratios***:
-
- (79) $\lambda^{r,m} \equiv \sum_{n\in S(r,m)} p_{rmn}q_{rmn}/\sum_{n\in S(r,m)\cap S(2,m)} p_{rmn}q_{rmn}$; r = 1,...,6 ; m = 1,...,24
- (80) $\mu^{r,m} \equiv \sum_{n\in S(r,m)\cap S(2,m)} p_{2mn}q_{2mn}/\sum_{n\in S(2,m)} p_{2mn}\, q_{2mn}$; r = 1,...,6 ; m = 1,...,24.
  - $\lambda^{r,m}$ is the ratio of rice expenditures in Prefecture r in month m relative to rice expenditures in the same month restricted to the set of products that are available in *both* Prefecture r and Prefecture 2 (Tokyo).
  - Thus this ratio must satisfy the inequality $\lambda^{r,m} \geq 1$. $\mu^{r,m}$ is the reciprocal of the ratio of rice expenditures in Prefecture 2 in month m relative to rice expenditures in the same month restricted to the set of products that are available in *both* Prefecture 2 and Prefecture r.
  - Thus $\mu^{r,m}$ must satisfy the inequality $\mu^{r,m} \leq 1$. Of course, when r = 2, it is the case that $\lambda^{r,m} = \mu^{r,m} = 1$ for m = 1,…,24.

- The expenditure ratios defined by $\lambda^{r,m}$ and $\mu^{r,m}$ are listed in Table 6 in the Appendix. Finally, Feenstra (1994) showed that:
-
- (81) $B^{r,m} = [\lambda^{r,m}]^{1/\kappa}$ and $C^{r,m} = [\mu^{r,m}]^{1/\kappa}$; $r = 1,\ldots,6$; $m = 1,\ldots,24$.
  - Thus if κ (or the elasticity of substitution $\sigma = 1 - \kappa$) is known or has been estimated, then $B^{r,m}$ and $C^{r,m}$ can readily be calculated as simple ratios of sums of observable expenditures raised to the power $1/\kappa$.
  - Thus the measures of price level change due to changes in product availability in Prefecture r relative to Prefecture 2 for month m, $D^{r,m}$ defined as $B^{r,m}$ times $C^{r,m}$, can be calculated.
- **We have two estimates for the elasticity of substitution, $\sigma = 23.0856$ from our first CES model that estimated the CES utility function and $\sigma = 6.4994$ from our CES model that estimated the CES unit cost function.**
- **These alternative estimates for the elasticity of substitution lead to the two alternative estimates for κ equal to -22.0856 and -5.4994.**
  - Thus we can generate two sets of indexes $D^{r,m}$ using these two estimates for κ and the above definitions. These indexes are stacked and plotted as $D_{CES}{}^{t}$ and $D_{CCES}{}^{t}$ on Chart 5.

Chart 5: Two Measures of the Increase in the Price Level of Six Prefectures Relative to Tokyo Prefecture due to Differences in Product Availability.



- $D_{CES}{}^{t}$ and $D_{CCES}{}^{t}$ are equal to 1 for t = 25,…,48 since these indexes compare availability of products in Prefecture r = 2 with the same Prefecture (Tokyo).
- The average increase in rice prices in Hokkaido and Kyoto due to differences in the availability of products in these two Prefectures relative to availability in Tokyo was only **0.41** percentage points on average using the estimates for σ from our estimation of the CES utility function but this average estimate increased to **1.66** or **1.67** percentage points using the lower estimate of σ that came out of our estimation of the CES unit cost function.

Summary.

- **The increase in the cost of living index due to limited availability of products was much greater for the smaller population Prefectures**. Here are the average increases in cost due to limited product availability generated by the two CES models for the smaller Prefectures in percentage points: Tottori: 5.87 and 25.76; Kochi: 3.94 and 16.82; Kagoshima: 3.24 and 13.68.
- Thus for the smaller Prefectures, the cost function based indexes $\mathbf{D_{CCES}}^{t}$ lie far above the utility function based indexes $\mathbf{D_{CES}}^{t}$.
- It can be seen that it is very important to obtain **accurate and realistic estimates for the elasticity of substitution** when applying the Feenstra methodology.
- Since our **utility function based method for estimating the elasticity of substitution fit the data far better than the cost function based method**, the smaller estimates for the increase in the cost of living due to smaller choice sets are our preferred estimates.

9. The Econometric Estimation of *KBF Preferences* with a Rank One Substitution Matrix.

- Konüs and Byushgens (1926) introduced the **functional form for a *linearly homogeneous utility function***:
- (82) $f(q) \equiv (q^T \cdot \mathbf{A} q)^{1/2} = (\Sigma_{i=1}^N \Sigma_{j=1}^N a_{ij} q_i q_j)^{1/2}$; $a_{ij} = a_{ji}$; $1 \le i \le j \le N$.
  - Thus **A** is an **N by N symmetric matrix** that contains $(N+1)N/2$ **unknown a_{ij} parameters**.
- The matrix **A** satisfies certain restrictions which are spelled out in Diewert (1976). Konüs and Byushgens and Diewert showed that ***this utility function is exact for the Fisher (1922) Ideal quantity and price indexes*** so we call preferences defined by (82) **KBF preferences**.

- Using the utility maximization framework which was described in section 6, the possible estimating equations (45) become the following ***system of inverse demand functions***:
- (83) $p_{tn} = e^t d_{tn} (\Sigma_{j=1}^N a_{nj} q_{tj}) / (\Sigma_{i=1}^N \Sigma_{j=1}^N a_{ij} q_{ti} q_{tj})$; $t = 1,\ldots,144$; $n = 1,\ldots,80$.
  - where the dummy variable is defined as before; i.e., $d_{tn} \equiv 1$ if $n \in S(t)$ and define $d_{tn} \equiv 0$ if product n is not available in period t for $t = 1,\ldots,144$ and $n = 1,\ldots,80$.

- We will not attempt to estimate all (N+1)N/2 unknown parameters a_{ij} in the **KBF utility function** defined by (82). In order to reduce the number of parameters in the A matrix, we define A as the following matrix which has rank 2:
- (84) $A \equiv \alpha\alpha^T - \beta\beta^T$
  - where the transposes of the column vectors α and β are defined as $\alpha^T \equiv [\alpha_1,\ldots,\alpha_{80}]$ and $\beta^T \equiv [\beta_1,\ldots,\beta_{80}]$. Thus we have reduced the number of unknown parameters in A from (80+1)×80/2 to 2×80.
-
- With A defined by (84), the system of inverse demand ***share*** equations (46) becomes the following system of estimating equations:
- (85) $s_{tn} = q_{tn}[\alpha_n\alpha\cdot q^t - \beta_n\beta\cdot q^t]/[(\alpha\cdot q^t)^2 - (\beta\cdot q^t)^2] + e_{tn}$; $t = 1,\ldots,144$; $n = 1,\ldots 80$.
-
- Equations (85) are valid even when there are **missing products** because when product n is **missing in period** t, $s_{tn} = q_{tn} = 0$.
- The utility function, $f(q,\alpha,\beta)$ is defined as follows:
- (86) $f(q,\alpha,\beta) \equiv [q^T(\alpha\alpha^T - \beta\beta^T)q]^{1/2} = [(\alpha\cdot q^t)^2 - (\beta\cdot q^t)^2]^{1/2}$.
  - Note that if $\beta = 0_N$, then $f(q,\alpha,0_N) = [(\alpha\cdot q^t)^2]^{1/2} = \alpha\cdot q^t = \Sigma_{n=1}^N \alpha_n q_n$; i.e., the utility function collapses down to the linear utility function that was studied in section 7.
  - This is an important point because it implies that starting **coefficients α_n and β_n for the nonlinear least squares minimization problem** that is defined below can be set equal to the estimates of the α_n^* that result in the estimation of linear preferences with the starting coefficients for the β_n^* set equal to 0.

- There are some tricky aspects to the new utility function as compared to the case of a linear utility function.
- We need to ensure that $(\alpha\cdot q^t)^2 - (\beta\cdot q^t)^2 > 0$ so that we can calculate the positive square root, $[(\alpha\cdot q^t)^2 - (\beta\cdot q^t)^2]^{1/2}$. We also need to set $\beta_n = 0$ if product n is available in only one period.
  - However, in our data set, **<u>all products are available for at least 14 periods</u>**. In order to identify all of the parameters, we impose our usual normalization so that our present model contains our linear utility function model as a special case:
- **(87) $\alpha_4 = 1$**
  - Define the total sample consumption vector q^* as $\Sigma_{t=1}^{144} q^t$. In order to prevent multicollinearity between the α_n and β_n parameters, we imposed the following normalization on the β_n parameters:
- **(88) $\beta\cdot q^* = 0$.**
  - Estimates for the α_n and β_n parameters are obtained by solving the following nonlinear least squares minimization problem subject to the normalizations (87) and (88):
- (89) $\min_{\alpha,\beta} \Sigma_{t=1}^{T} \Sigma_{n=1}^{N} \{s_{tn} - q_{tn}[\alpha_n\alpha\cdot q^t - \beta_n\beta\cdot q^t]/[(\alpha\cdot q^t)^2 - (\beta\cdot q^t)^2]\}^2$.
  - If a product n appears in only one period in the sample of observations, then **our KBF model will be able to estimate the parameter α_n but it will not be able to estimate the parameter β_n**; see the Appendix in Diewert (2024).
  - This normalization also helps to ensure that $(\alpha\cdot q^t)^2 - (\beta\cdot q^t)^2 > 0$ so that we can define $f(q^t)$ as the positive square root of $(\alpha\cdot q^t)^2 - (\beta\cdot q^t)^2$. This normalization ensures that our estimated KBF Prefecture price and quantity indexes are invariant to changes in the units of measurement. In our regression, used the constraint $\Sigma_{n=1}^{80} q_n^*\beta_n = 0$ to solve for $\beta_4 = -\Sigma_{n=1}^{3} [q_n^*/q_4^*]\beta_n - \Sigma_{n=5}^{80} [q_n^*/q_4^*]\beta_n$.

- **Taking into account the normalizations (87) and (88), there are 158 free parameters to estimate.**
- The starting coefficient values for the α_n were the final coefficient estimates for the linear utility function model discussed in section 7 and the starting coefficient values for the β_n were set equal to 0.00001 or -0.00001.
- As a check on our code, the starting log likelihood for the model defined by (89) was equal to the final log likelihood for the linear model defined by (50) in section 7. Shazam took 393 iterations and 26 minutes to converge to a solution.
- The gain in log likelihood was 4289.18 points for adding 79 new β_n parameters. The R^2 for the new model was 0.9983, an increase over the R^2 for the linear model in section 7 ($R^2 = 0.9967$) and for the CES utility function model in the previous section ($R^2 = 0.9870$).
- The sum of absolute errors for the present model was 5.630 and for the linear model and the CES model, the sums were 7.9178 and 7.2587 respectively.
- <u>**Thus the present KBF model fits the data significantly better than previous models.**</u>

- Once the solution (α^*,β^*) to the nonlinear least squares minimization problem (89) has been obtained, the preliminary period t aggregate quantity and price levels, Q^{t*} and P^{t*}, are defined as follows:
-
- (90) $Q^{t*} \equiv f(q^t,\alpha^*,\beta^*) = [(\alpha^*\cdot q^t)^2 - (\beta^*\cdot q^t)^2]^{1/2}$; $P^{t*} \equiv e^t/Q^{t*}$; $t = 1,\ldots,144.$
-
- Normalize the sequence price levels P^{t*} into the series $P_{KBF}{}^t$ which is such that the normalized sequence of price levels equals 1 for t = 25 (month 1 for Tokyo):
-
- (91) $P_{KBF}{}^t \equiv P^{t*}/P^{25*}$; $t = 1,\ldots,144.$
-
- Finally define **the *KBF utility function price levels for regions 1-6* as $P_{KBF}{}^{r,m}$ using $P_{KBF}{}^t$ defined by (91) and an appropriate modification of definitions (64).** The KBF price indexes, $P_{KBF}{}^t$, are plotted on Chart 6 below and the regional KBF indexes $P_{KBF}{}^{1,m}$ – $P_{KBF}{}^{6,m}$ are listed in Table 8 in the Appendix.

- We note that it is of interest to calculate the reservation prices that the estimated KBF utility function generates.
  - With the solution (α^*,β^*) to (89) in hand, we can calculate Hicksian reservation prices p_{tn}^* for the products n that were *not* present in period t using equations (83) for our BF functional form for products n that are not available in region-period t:
-
- (92) $p_{tn}^* \equiv e^t f_n(q^t,\alpha^*,\beta^*)/f(q^t,\alpha^*,\beta^*)$; $\quad t = 1,\ldots,T;\ n \notin S(t).$
-
- The average reservation price for our estimated KBF utility function turned out to equal 0.43135 while the average predicted price for products that were present in each period was equal to 0.32779.
- **Thus on average, reservation prices were 0.43135/0.32779 = 1.316 or 31.6 percent higher than predicted prices.**
- Note that **the CES model generates infinite reservation prices which is a problem with the CES model**.

- The N by N matrix of second order partial derivatives of $f(q,\alpha^*,\beta^*)$ evaluated at $q = q^t$ is denoted by $\nabla^2 f(q^t,\alpha^*,\beta^*)$ and it is called the *period t inverse substitution matrix*. For a general linearly homogeneous and concave utility f(q), it must be a negative semidefinite matrix that satisfies the restrictions $\nabla^2 f(q^t)q^t = 0_N$.
- Thus the rank of $\nabla^2 f(q^t)$ is at most N−1. For our particular functional form for $f(q,\alpha^*,\beta^*)$ defined by (90), the period t inverse substitution matrix is defined as follows:
-
- (93) $\nabla^2 f(q^t,\alpha^*,\beta^*) \equiv -[f(q^t,\alpha^*,\beta^*)]^{-3}[\alpha^*(\beta^*\cdot q^t) - \beta^*(\alpha^*\cdot q^t)][\alpha^*(\beta^*\cdot q^t) - \beta^*(\alpha^*\cdot q^t)]^T$.
-
- If $\beta^* = 0_N$, then $\nabla^2 f(q^t,\alpha^*,\beta^*) = 0_N 0_N{}^T$ which is an N by N matrix of zeros. If α^* and β^* are both nonzero vectors and $\alpha^* \neq \beta^*$, then the period t substitution matrix defined by (93) will have rank equal to one.

- Diewert and Wales (1988) called a functional form for a cost function defined over N products a ***semiflexible functional form of rank k*** if its matrix of second order partial derivatives had rank k. Using this terminology, our $f(q,\alpha,\beta)$ defined by (90) is a ***semiflexible functional form of rank 1***.

Key Characteristics of the KBF Index (with Rank 1 Utility).

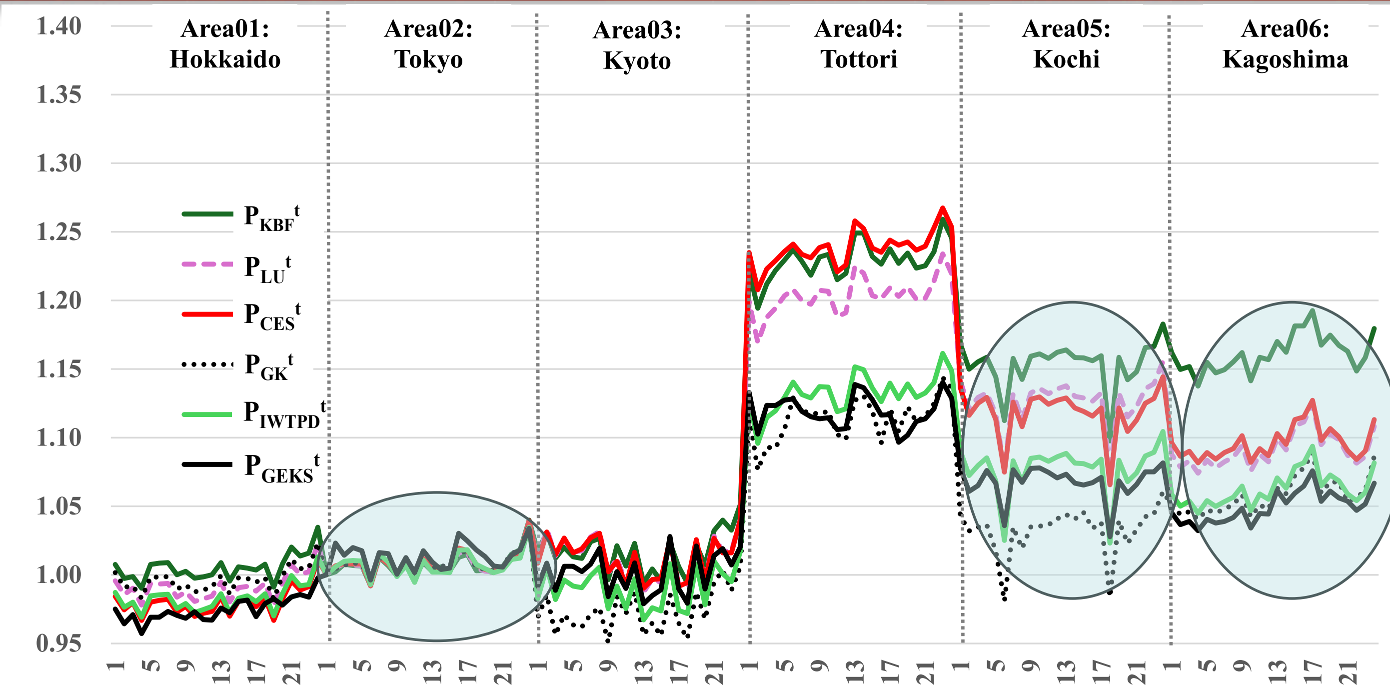
Item	Description
Theoretical Foundation	Based on Konüs-type **expenditure functions**, the KBF Index derives a theoretically consistent price index from **consumer utility maximization behavior**. The name combines elements from **Konüs, Byushgens, and Fisher**.
Utility Function Features	- **homothetic** - **Linearly homogeneous** - **Substitution structure is constrained to Rank 1**, allowing for flexibility while simplifying estimation.
Index Definition	Defined as the ratio of **expenditure functions**, representing **the cost of achieving the same utility level under different price vectors** (i.e., a Konüs price index).
Substitution Elasticity	Enables structural modeling of demand response to price changes. The **substitution matrix is of rank one**, simplifying estimation while capturing key substitution patterns.
Treatment of Missing Prices	It is possible to construct a cost-of-living index that reflects variety effects without relying on **matched samples**. While the KBF model is capable of generating ***reservation prices*** for **unobserved products**, these do not need to be explicitly estimated in order to construct the index.
Flexibility	Compared to Linear Utility or CES indices, the KBF Index allows **for greater flexibility in consumer behavior.**
Estimation Complexity	Requires estimation of **inverse demand systems** and typically involves **nonlinear least squares methods; computationally more demanding**.
Practical Applications	Suitable for constructing regionally or temporally consistent price indices, especially in settings with **heterogeneous consumption patterns**.

Alternative multilateral indexes.

Index	Transitive	Substitution Flexibility	Handles Missing Prices	Other Comments
Bilateral Fisher (FB)	× No	Fully flexible (if no missing prices)	× No – relies on matched prices	Not invariant to base period
GEKS	○Yes	**Fully flexible (if no missing prices)**	× No – relies on matched prices	**Same flexibility issues as bilateral Fisher**
Unweighted TPD	○Yes	Linear preferences	○ Yes	Ignores expenditure shares and economic background
Weighted TPD	○Yes	**Cobb–Douglas & linear preferences**	○ Yes	**Cobb–Douglas assumption is unrealistic**
Geary-Khamis (GK)	○Yes	**Leontief & linear preferences**	○ Yes	**Leontief assumption is unrealistic**
Linear Utility Index	○Yes	**Linear preferences**	○ Yes	**Simple but restrictive utility structure**
CES Preferences	○Yes	**CES preferences**	○ Yes	**Requires single substitution parameter**
KBF Rank 1	○Yes	**Flexible (Rank 1 substitution matrix)**	○ Yes	**Hard to estimate; high data requirements**

***Transitivity implies resistance to chain drift.**

Chart 6: KBF Price Indexes and Other Indexes for Six Japanese Prefectures.



- The six series of price indexes are quite close to each other for the **Tokyo Prefecture** months and somewhat close for Hokkaido and Kyoto but very different for the three smaller population Prefectures, Tottori, Kochi and Kagoshima.
- It can be seen that the KBF indexes, P_{KBF}^t, are clearly higher for **the Kochi and Kagoshima** time periods and in general, are the highest price indexes.
- The CES indexes P_{CES}^t and the Linear Utility function indexes P_{LU}^t are in general quite close and are the second highest indexes.
- The Implicit Weighted Time Product Dummy indexes P_{IWTPD}^t are well below P_{KBF}^t, P_{CES}^t and P_{LU}^t PLUt but above P_{GEKS}^t and P_{GK}^t PGKt for the smaller Prefectures.

10. Conclusions:

- **We considered 7 indexes**: *WTPD (Weighted Time Product Dummy), GK (Geary Khamis) and GEKS indexes (these are the main multilateral indexes used by statistical agencies at the first stage of aggregation), LU (linear utility function estimation), CES(f) (direct utility function estimation), CES(c) (estimation of unit cost function) and the estimation of a KBF utility function with a rank 1 substitution matrix*.
- **The last 4 indexes are academic indexes** that make use of econometric methods. ***Which index is "best"***?
  - **GEKS can work well if there is not too much product turnover in the time series context or if there is not too much variation in product availability in the interregional context.** But GEKS is not reliable if there is a lack of product matching across time or space since GEKS depends on a high degree of product matching. In our context, **GEKS was not reliable**.
  - WTPD, GK and LU are all based on an underlying assumption that purchasers have linear preferences to some degree of approximation. We expected a priori that these indexes would give similar results but this was not the case. P_{LU}^t was in general the highest of these 3 indexes and P_{GK}^t was in general the lowest.
  - **The differences in these 3 indexes in the smaller Prefectures was very large. We have no clear theoretical reason to prefer any one of these indexes over the other two indexes.** Obviously, the choice of the method which is used to estimate linear preferences matters. A major problem with linear preferences (from the viewpoint of the economic approach to index number theory) is that they are not very flexible and hence will have a tendency to have a downward bias.

- Turning to the two CES indexes, ***the CES index based on the estimation of a unit cost function was clearly not preferred to the CES index*** that was based on the direct estimation of a CES utility function. We noted that since most applications of the Feenstra methodology rely on estimates for the elasticity of substitution that are based on unit cost function estimation, **the Feenstra estimates for the benefits of new products are likely to have a substantial upward bias**.
- **The two indexes that do not suffer from a lack of matching bias and are more flexible than linear preferences are the CES index (based on the estimation of a CES utility function) and the estimation of KBF preferences with a rank 1 substitution matrix.** It turned out that P_{KBF}^t was in general higher than P_{CES}^t which in turn was higher than P_{LU}^t which is consistent with our a priori expectations.
- Since **KBF preferences are more flexible than CES preferences** which in turn are more flexible than linear preferences, we preferred P_{KBF}^t over P_{CES}^t and we preferred P_{CES}^t over the 3 indexes that were consistent with linear preferences.

• However, price indexes that ***rely on econometric estimation may be subject to a lack of reproducibility across econometricians***: different assumptions about the error structure in an econometric model or differences in the choice of exogenous and endogenous variables can lead to very different indexes.
  - Thus National Statistical Offices have in general been reluctant to embrace econometric modeling in order to construct consumer price indexes. Moreover, it can turn out that the estimated elasticity of substitution is negative which means the CES modeling equations are not valid. Multicollinearity problems can arise when estimating KBF preferences and the choice of normalizations can affect the results.

- Here are some additional tentative conclusions that we draw on from our analysis:
- Indexes which do not weight prices by their economic importance can be unreliable. Their use should be avoided if possible.
  - In the context of forming ***inter-regional price indexes where choice sets are very different, unit value and average price indexes can be very unreliable.***
  - A robust method for dealing with rapid product turnover, quality adjustment and the chain drift problem is still to be found.

W. Erwin Diewert,
University of British Columbia and University of New South Wales;
Email: erwin.diewert@ubc.ca

Naohito Abe,
Hitotsubashi University;
Email: n.abe@r.hit-u.ac.jp

Akiyuki Tonogi,
Toyo University;
Email: tonogi@toyo.jp

Chihiro Shimizu,
Hitotsubashi University;
Email: c.shimizu@r.hit-u.ac.jp