# Scanner Data and Economic Statistics: A Unified Approach

Stephen J. Redding
*Princeton*

David E. Weinstein
*Columbia*

October, 2016

# Big Data Can Transform Economic Measurement

1. Volume: Scanner data contains information on every purchase in every store, Census collects information on every transaction of every importer, Expedia and other companies collect data on services transactions
2. Variety: We *simultaneously* observe quantities and sales
   – Enables us to produce *new* statistics?
3. Velocity: Data collected nearly immediately at high-frequency
4. Veracity: Scanner Data is accurate but coverage may be an issue; trade data contains errors, e.g., unit values instead of prices

# Bar-Code Data: Challenges

- Product turnover is phenomenal
  - In a typical year, 40% of household's expenditures are on goods that were created in the last 4 years; 20% are on goods that will not survive 4 years.
  - How do we measure prices when the set of goods is changing? Cannot use *ad hoc* approaches
- Different price indexes can yield very different cost-of-living (COL) measures
  - Need to be precise about assumptions and links between indexes
- No single product firms or industries
  - Need to be precise about what industry output or price means

# How Did We Get Here?

- Inflation indexes were initially designed for statistical properties
  - The first modern price index was created by Nicolas Dutot in 1738 to understand the impact of money creation on a simple average of prices
  - His "common-sense" or "axiomatic" approach laid the foundation for all price indexes that lie at the heart of all modern official measures of inflation and real income.
  - Carli (1764), Jevons (1865), Laspeyres (1871), Paasche (1875), Fisher (1922), Törnqvist (1936) indexes were all developed to have certain properties but were not based on economic theory

# The Economic Approach

- Konüs (1924) realized that there were deep problems with this approach:
- Theories of consumer behavior indicated that price indexes should not be based on the mathematical properties of particular index numbers, but should reflect a change in a consumer's cost of living.
  - The downside of the "economic approach" to measuring variables is that the measurement is not independent of a particular utility function or class of utility functions
  - Nevertheless, most economists believe price indexes should be based on unit expenditure functions

# Facts vs. Constructs

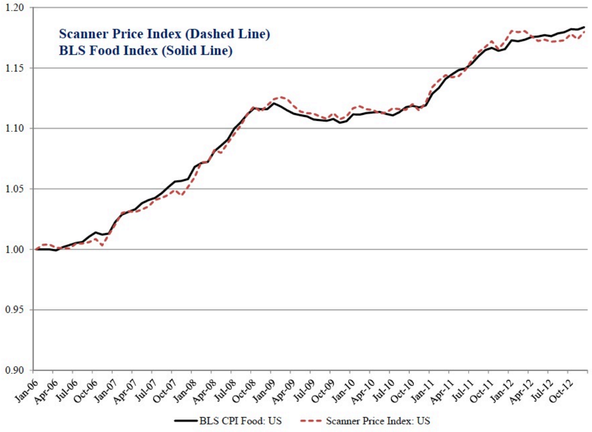
- "Facts" are nominal variables that can be measured without theory, e.g. *nominal* sales, expenditures, imports, exports, etc.
- "Constructs" are variables like *aggregate* price changes or real output that are based on theory
  - Bar-coded goods are tangible facts: (591mL bottle of Diet Coke")
    - Aggregates (Coke, industries, real output) are constructs
  - We agree about facts; we disagree about constructs
- The task of economic measurement is to deliver theory-consistent measures of constructs that are consistent with the facts

# Issues and Opportunities for Census

- Census gathers information on retail sales and imports
- Issues
  - How representative is scanner data?
  - Can these data be used to generate better facts (e.g., retail sales)
  - How accurate are trade-transactions data?
- Opportunities
  - Simultaneous collection of sales, output, and price information
  - Almost instantaneous collection of information
  - Capacity to exactly decompose output growth into quality, costs, markups, firm entry (entrepreneurship), innovation, etecetera

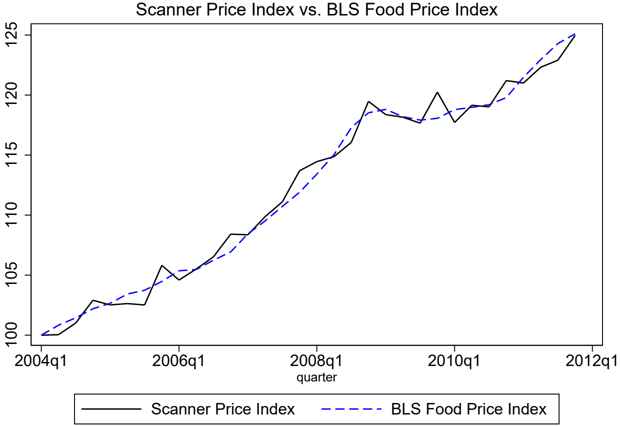
# Can Scanner Data Replicate Existing Indexes?

Figure 1: Nielsen Retail Price Index vs. CPI Food Price Index




Source: Beraja, Hurst, and Ospina (2016)

# Can HomeScan Data Replicate Existing Indexes?




Source: Kim (2016)

## But Can We Do More?

- "Quantifying the Sources of Firm Heterogeneity," with Colin Hottman and Stephen Redding, *QJE*, 2016.
  - Framework for output accounting with bar-code data
- "A Unified Approach to Estimating Demand and Welfare Changes," with Stephen Redding.
  - Unifies all major approaches to price measurement
  - Existing approaches are all special cases of the "Unified Price Index"
- "Trade and Welfare: A Big Data Approach," with Stephen Redding
  - Using Census trade transactions to decompose US import growth and welfare gains

# The Theory of “Measurement” Problems

- The measurement of price indexes, welfare, and demand parameters is based on three disjoint approaches
  1. Macroeconomic price indexes are based on *time-invariant* utility functions (e.g. Törnqvist index)
  2. Microeconomic demand estimation based on *time-varying* demand systems (stochastic error term)
  3. Actual price and real-output data often constructed using formulas that differ from either approach (e.g. Laspeyres)
- Problem
  - Assumptions underlying demand-system estimation imply that standard price indexes are incorrect
  - Assumptions underlying standard price indexes invalidate demand-system estimation
- *Micro and macro welfare estimates mutually inconsistent, and neither is consistent with statistical agencies’ data*

# Unified Price Index (UPI)

- We develop a "unified approach" that consistently estimates welfare and demand when demand for each good is time varying
  - Rationalizes observed data on prices and expenditure shares as an equilibrium of the model
  - Allows for entry and exit of goods over time
  - Identifies a unique elasticity of substitution ($\sigma$)
  - Satisfies money-metric utility (demand shocks cancel)
  - Yields consistent aggregation from micro to macro
  - Nests all major micro, macro, and statistical approaches to price measurement
  - Generalizes to heterogeneous groups of consumers
  - Shows that existing exact price indexes are biased in the presence of mean zero demand shocks

# Outline

- Unified Price Index
- Relation to Existing Price Indexes
- Estimating the Elasticity of Substitution
- Data and Results
- Conclusions

## CES Demand

- CES preferences over goods $k \in \Omega_t$ at time $t$:

$$\mathbb{U}_t = \left[ \sum_{k \in \Omega_t} \left( \varphi_{kt} C_{kt} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \qquad \sigma > 1, \qquad \varphi_{kt} > 0$$

- CES Unit Expenditure Function:

$$\mathbb{P}_t = \left[ \sum_{k \in \Omega_t} \left( \frac{P_{kt}}{\varphi_{kt}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}},$$

- Demand System Equation for Expenditure Shares:

$$S_{\ell t} = \frac{\left( P_{\ell t} / \varphi_{\ell t} \right)^{1-\sigma}}{\sum_{k \in \Omega_t} \left( P_{kt} / \varphi_{kt} \right)^{1-\sigma}} = \frac{\left( P_{\ell t} / \varphi_{\ell t} \right)^{1-\sigma}}{\mathbb{P}_t^{1-\sigma}}, \qquad \ell \in \Omega_t$$

# Product Turnover

- Share of each individual common good in total common good expenditure ($\Omega_{t,t-1} \subseteq \Omega_t$ and $\Omega_{t,t-1} \subseteq \Omega_{t-1}$):

$$S^*_{\ell t} \equiv \frac{P_{\ell t} C_{\ell t}}{\sum_{k \in \Omega_{t,t-1}} P_{kt} C_{kt}} = \frac{\left(P_{\ell t} / \varphi_{\ell t}\right)^{1-\sigma}}{\sum_{k \in \Omega_{t,t-1}} \left(P_{kt} / \varphi_{kt}\right)^{1-\sigma}}, \qquad \ell \in \Omega_{t,t-1}$$

- Share of common goods in total expenditure:

$$\lambda_t \equiv \frac{\sum_{k \in \Omega_{t,t-1}} P_{kt} C_{kt}}{\sum_{k \in \Omega_t} P_{kt} C_{kt}} = \frac{\sum_{k \in \Omega_{t,t-1}} \left(P_{kt} / \varphi_{kt}\right)^{1-\sigma}}{\sum_{k \in \Omega_t} \left(P_{kt} / \varphi_{kt}\right)^{1-\sigma}}$$

- CES unit expenditure function for common goods:

$$\mathbb{P}^*_t = \left[ \sum_{k \in \Omega_{t,t-1}} \left( \frac{P_{kt}}{\varphi_{kt}} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}, \quad \ln \tilde{\varphi}^*_t = \frac{1}{N_{t,t-1}} \sum_{k \in \Omega_{t,t-1}} \ln \varphi_{kt} = 0$$

# Change in Cost of Living

- CES Price Index:

$$\Phi_{t-1,t} \equiv \frac{\mathbb{P}_t}{\mathbb{P}_{t-1}} = \left(\frac{\lambda_t}{\lambda_{t-1}}\right)^{\frac{1}{\sigma-1}} \left[\frac{\sum_{k\in\Omega_{t,t-1}} \left(P_{kt}/\varphi_{kt}\right)^{1-\sigma}}{\sum_{k\in\Omega_{t,t-1}} \left(P_{kt-1}/\varphi_{kt-1}\right)^{1-\sigma}}\right]^{\frac{1}{1-\sigma}}$$

- Using the common good expenditure share:

$$\left[\frac{\sum_{k\in\Omega_{t,t-1}} \left(P_{kt}/\varphi_{kt}\right)^{1-\sigma}}{\sum_{k\in\Omega_{t,t-1}} \left(P_{kt-1}/\varphi_{kt-1}\right)^{1-\sigma}}\right]^{\frac{1}{1-\sigma}} = \frac{P_{\ell t}/\varphi_{\ell t}}{P_{\ell t-1}/\varphi_{\ell t-1}} \left(\frac{S^*_{\ell t}}{S^*_{\ell t-1}}\right)^{\frac{1}{\sigma-1}}, \qquad \ell \in \Omega_{t,t-1}$$

- Therefore we also have:

$$\Phi_{t-1,t} = \left(\frac{\lambda_t}{\lambda_{t-1}}\right)^{\frac{1}{\sigma-1}} \frac{P_{\ell t}/\varphi_{\ell t}}{P_{\ell t-1}/\varphi_{\ell t-1}} \left(\frac{S^*_{\ell t}}{S^*_{\ell t-1}}\right)^{\frac{1}{\sigma-1}}, \qquad \ell \in \Omega_{t,t-1}$$

- Taking logs and taking means, we have:

$$\ln \Phi_{t-1,t} = \frac{1}{N_{t,t-1}} \sum_{\ell\in\Omega_{t,t-1}} \left[\frac{1}{\sigma-1} \ln\left(\frac{\lambda_t}{\lambda_{t-1}}\right) + \ln\left(\frac{P_{\ell t}}{P_{\ell t-1}}\right) + \frac{1}{\sigma-1} \ln\left(\frac{S^*_{\ell t}}{S^*_{\ell t-1}}\right)\right].$$

# The Unified Price Index

## Proposition

*The "unified price index" (UPI)—which is exact for the CES preference structure in the presence of changes in the set of goods, demand shocks that do not affect aggregate utility, and discrete changes in prices and expenditure shares—is given by*

$$\Phi^{U}_{t-1,t} = \underbrace{\left(\frac{\lambda_t}{\lambda_{t-1}}\right)^{\frac{1}{\sigma-1}}}_{\text{Variety Adjustment}} \underbrace{\left[\frac{\tilde{P}^*_t}{\tilde{P}^*_{t-1}} \left(\frac{\tilde{S}^*_t}{\tilde{S}^*_{t-1}}\right)^{\frac{1}{\sigma-1}}\right]}_{\text{Common-Goods UPI}}.$$

- Note: $\lim_{\sigma\to\infty} \Phi^{U}_{t-1,t} = \Phi^{Jevons}_{t-1,t}$
- For $\sigma < \infty$, variety and heterogeneity terms

# Outline

- Unified Price Index
- Relation to Existing Price Indexes
- Estimating the Elasticity of Substitution
- Data and Results
- Conclusions

# Equivalences

- Three expressions for the change in the cost of living:

$$\Phi^F_{t-1,t} = \frac{\mathbb{P}_t}{\mathbb{P}_{t-1}} = \left(\frac{\lambda_t}{\lambda_{t-1}}\right)^{\frac{1}{\sigma-1}} \left[\sum_{k\in\Omega_{t,t-1}} S^*_{kt-1} \left(\frac{P_{kt}/\varphi_{kt}}{P_{kt-1}/\varphi_{kt-1}}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$$

$$\Phi^B_{t,t-1} = \frac{\mathbb{P}_{t-1}}{\mathbb{P}_t} = \left(\frac{\lambda_{t-1}}{\lambda_t}\right)^{\frac{1}{\sigma-1}} \left[\sum_{k\in\Omega_{t,t-1}} S^*_{kt} \left(\frac{P_{kt-1}/\varphi_{kt-1}}{P_{kt}/\varphi_{kt}}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$$

$$\Phi^U_{t-1,t} = \frac{\mathbb{P}_t}{\mathbb{P}_{t-1}} = \left(\frac{\lambda_t}{\lambda_{t-1}}\right)^{\frac{1}{\sigma-1}} \left[\frac{\tilde{P}^*_t}{\tilde{P}^*_{t-1}} \left(\frac{\tilde{S}^*_t}{\tilde{S}^*_{t-1}}\right)^{\frac{1}{\sigma-1}}\right]$$

$$\ln \tilde{\varphi}^*_t = \frac{1}{N_{t,t-1}} \sum_{k\in\Omega_{t,t-1}} \ln \varphi_{kt} = 0$$

# Theorem

- It is possible to show that if taste shocks are small, there exists a unique elasticity such that all three expressions converge to the same value
  - Under these conditions, there is a unique "reverse-weighting" elasticity that renders the forward and backward differences identical to the UPI
  - Deviations in these expressions emerge when you use the wrong elasticity

## "Statistical" Indexes

- Laspeyres ($\sigma = 0$, $\lambda_t/\lambda_{t-1} = 1$ and $\varphi_{kt} = \bar{\varphi}_k$):

$$\Phi^L_{t-1,t} \equiv \frac{\sum_{k\in\Omega_{t,t-1}} C_{kt-1}P_{kt}}{\sum_{k\in\Omega_{t,t-1}} C_{kt-1}P_{kt-1}} = \sum_{k\in\Omega_{t,t-1}} S^*_{kt-1}\frac{P_{kt}}{P_{kt-1}}.$$

- Paasche ($\sigma = 0$, $\lambda_t/\lambda_{t-1} = 1$ and $\varphi_{kt} = \bar{\varphi}_k$):

$$\Phi^P_{t-1,t} = \frac{\sum_{k\in\Omega_{t,t-1}} C_{kt}P_{kt}}{\sum_{k\in\Omega_{t,t-1}} C_{kt}P_{kt-1}} = \left[\sum_{k\in\Omega_{t,t-1}} S^*_{kt}\left(\frac{P_{kt}}{P_{kt-1}}\right)^{-1}\right]^{-1}.$$

- Cobb-Douglas ($\sigma = 1$, $\lambda_t/\lambda_{t-1} = 1$, $\varphi_{kt} = \bar{\varphi}_k = S^*_k$):

$$\Phi^{CD}_{t-1,t} = \prod_{k\in\Omega_{t,t-1}}\left(\frac{P_{kt}}{P_{kt-1}}\right)^{S^*_k}$$

- Jevons ($\sigma \to \infty$)

$$\Phi^J_{t-1,t} = \prod_{k\in\Omega_{t,t-1}}\left(\frac{P_{kt}}{P_{kt-1}}\right)^{\frac{1}{N_{t,t-1}}}$$

# Economic Indexes

- Quadratic mean of order $2(1-\sigma)$

$$\Phi_{t,t-1} = \left(\frac{\lambda_t}{\lambda_{t-1}}\right)^{\frac{1}{\sigma-1}} \left[\frac{\sum_{k\in\Omega_{t,t-1}} S^*_{kt-1}\left(\frac{P_{kt}/\varphi_{kt}}{P_{kt-1}/\varphi_{kt-1}}\right)^{1-\sigma}}{\sum_{k\in\Omega_{t,t-1}} S^*_{kt}\left(\frac{P_{kt}/\varphi_{kt}}{P_{kt-1}/\varphi_{kt-1}}\right)^{-(1-\sigma)}}\right]^{\frac{1}{2(1-\sigma)}}$$

- Fisher ($\sigma = 0$, $\lambda_t/\lambda_{t-1} = 1$ and $\varphi_{kt} = \bar{\varphi}_k$):

$$\Phi^F_{t-1,t} = \left(\Phi^L_{t-1,t}\Phi^P_{t-1,t}\right)^{1/2},$$

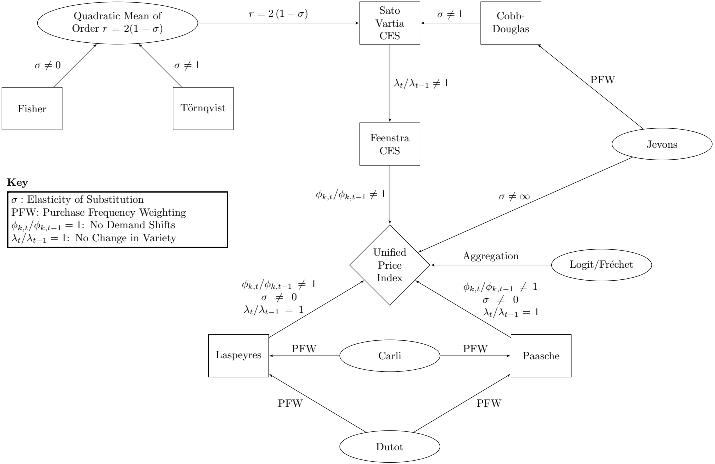
- Törnqvist ($\sigma \to 1$, $\lambda_t/\lambda_{t-1} = 1$ and $\varphi_{kt} = \bar{\varphi}_k$)

$$\Phi^T_{t-1,t} = \prod_{k\in\Omega_{t,t-1}} \left(\frac{P_{kt}}{P_{kt-1}}\right)^{\frac{1}{2}\left(S^*_{kt-1}+S^*_{kt}\right)}.$$

- Sato-Vartia CES ($\lambda_t/\lambda_{t-1} = 1$ and $\varphi_{kt} = \bar{\varphi}_k$):

$$\Phi^{SV}_{t-1,t} = \prod_{k\in\Omega_{t,t-1}} \left(\frac{P_{kt}}{P_{kt-1}}\right)^{\omega^*_{kt}}.$$

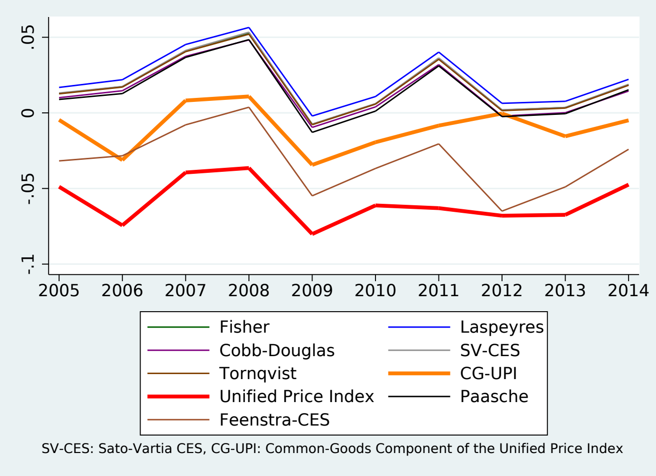
# The Big Picture




**Key**

$\sigma$ : Elasticity of Substitution
PFW: Purchase Frequency Weighting
$\phi_{k,t}/\phi_{k,t-1} = 1$: No Demand Shifts
$\lambda_t/\lambda_{t-1} = 1$: No Change in Variety

# Aggregate Price Indexes




Between 2004-14, $\Phi^{SV} - \Phi^{CG} = 2.8$ and $\Phi^{SV} - \Phi^{U} = 6.9$ averaged percentage points per year

Allowing for demand shifts matters enormously

# Can Decompose Imports and Firm Sales

$$\ln S_{ft} = \underbrace{\ln E_{gt}}_{\text{Expenditure}} + \underbrace{(\sigma_F - 1)\ln P_{gt}}_{\text{Competitive Pressure}} + \underbrace{(\sigma_F - 1)\ln \varphi_{ft}}_{\text{Firm Quality}} + \underbrace{\frac{1-\sigma_F}{1-\sigma_U}\ln N_{ft}}_{\text{Product Scope}}$$

$$\underbrace{\left[-(\sigma_F - 1)\,\widetilde{\gamma}_{ft}\right]}_{\text{Ave. Marginal Cost}} + \underbrace{\frac{1-\sigma_F}{1-\sigma_U}\ln\left(\frac{1}{N_{ft}}\sum_{u\epsilon U_{ft}}\left(\frac{\gamma_{ut}/\widetilde{\gamma}_{ft}}{\varphi_{ut}}\right)^{1-\sigma_U}\right)}_{\text{Cost Dispersion}} + \underbrace{(1-\sigma_F)\ln \mu_{ft}}_{\text{Markup}}$$

- Results akin to Solow Residuals, but much finer grained decompositions are possible
- Currently using Census trade transactions data to decompose trade flows into cost, quality-upgrading, firm entry and exit, etcetera.
  - Preliminary results using Chilean data suggest that the rise of Chinese exports can almost entirely be explained by quality upgrading, not cheaper products

# Conclusions

- We develop a "unified approach" that consistently estimates welfare and demand when demand for each good is time varying
  - Rationalizes observed data on prices and expenditure shares as an equilibrium of the model
  - Allows for entry and exit of goods over time
  - Identifies a unique elasticity of substitution ($\sigma$)
  - Satisfies money-metric utility (demand shocks cancel)
  - Yields consistent aggregation from micro to macro
  - Nests all major micro, macro, and statistical approaches to price measurement
  - Generalizes to heterogeneous groups of consumers
- Consumer valuation bias around as large as variety bias

# Thank You