Reassessing the Welfare Gains from Changes in Product Variety

Daniel Melser

Monash University

UNSW-ESCoE Conference: 9 Dec 2025

Email: daniel.melser@monash.edu

Overview

1. Introduction

1. Motivation and Outline

Motivations:

- A large literature that has looked at the consumer gains from product variety.
- The gains estimated are quite large.
- The approaches used have been quite restrictive.

What I do:

- Implement a range of new and existing approaches to estimating the gains from new and disappearing goods.
- This involves both using analytic and parametric approaches.
- Illustrate the results using a large US scanner dataset.

1. Data

Methods are illustrated using a large US scanner dataset:

- Supermarkets in New York.
- Data from 2006 to 2020.
- Stores must be in the sample for the whole period.
- 10 product categories: beer, canned soup, cereal, frozen pizza, ice cream, liquid soap, milk, peanut butter, tea bags, and toilet tissue.
- Assortment sizes vary from peanut butter (~ 600 varieties) to cereal and ice cream ($\sim 3,000$ varieties).

1. Background

Constant Elasticity of Substitution (CES):

- Robert Feenstra (1994). "New Product Varieties and the Measurement of International Prices". In: *American Economic Review* 84.1, pp. 157–77
- Analytic cost-of-living index can be derived
- $N + 1$ parameters, infinite reservation prices

Symmetric Translog (ST):

- Robert C. Feenstra and David E. Weinstein (2017). "Globalization, Markups, and US Welfare". In: *Journal of Political Economy* 125.4, pp. 1040–1074
- Analytic cost-of-living index can be derived
- $N + 1$ parameters, finite reservation prices

1. Background

Generalized Symmetric Translog (GST):

- W. Erwin Diewert (2024). "A Generalization of the Symmetric Translog Functional Form". In: *Journal of International Economics* 151, p. 103821
- No analytic cost-of-living index can be derived
- $2N$ parameters, finite reservation prices

Flexible Functional Forms:

- W. Erwin Diewert and Robert C Feenstra (2019). *Estimating the Benefits of New Products*. Working Paper 25991. National Bureau of Economic Research
- No analytic cost-of-living index can be derived
- Up to $(N-1)N/2$ parameters, finite reservation prices

1. Background

Key dimensions along which the methods differ:

- Flexibility and regularity of the functional form.
- Whether 'analytic' formulas are available (e.g., Feenstra 1994).
- Number of parameters to estimate.
- Complexity of estimation (linear/non-linear).
- Reservation prices (finite/infinite).

2. CES Analytic and Parametric Approaches

2. Analytic vs Parametric CES Approaches

Analytic CES Cost Function

$$c(\mathbf{p}_t | I_t) = \left(\sum_{i \in I_t} b_i p_{it}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$

- Estimate the parameter: σ.
- Derive an analytic expression for the COLI following Feenstra 1994:

$$P_{vt}^{CES} = \left(\frac{1 - \sum_{i \notin I_{vt}} s_{it}}{1 - \sum_{i \notin I_{vt}} s_{iv}} \right)^{-\frac{1}{1-\hat{\sigma}}} P_{vt}^{SV}$$

Parametric CES Utility Function

$$U(\mathbf{x}_t) = \left(\sum_{i \in I_t} a_i x_{it}^r \right)^{\frac{1}{r}}$$

- Estimate parameters: a_i, $r = 1 - \frac{1}{\sigma}$
- Derive the COLI by deflating relative expenditure by relative utility:

$$P_{vt}^{CES} = \left(\frac{\sum_{i \in I_t} p_{it} x_{it}}{\sum_{i \in I_v} p_{iv} x_{iv}} \right) \Bigg/ \left(\frac{\sum_{i \in I_t} \hat{a}_i x_{it}^{\hat{r}}}{\sum_{i \in I_v} \hat{a}_i x_{iv}^{\hat{r}}} \right)^{\frac{1}{\hat{r}}}$$

2. Estimation of CES Parameters

Estimation of σ:

$$\sigma = -\frac{\log\left(\frac{x_{it}}{x_{it-1}} \Big/ Q^{T}_{t-1t}\right)}{\log\left(\frac{p_{it}}{p_{it-1}} \Big/ P^{T}_{t-1t}\right)} = \hat{\sigma}_{itt-1}, \quad \forall t, \quad i \in I_t$$

- A slight extension of the double-differencing approach is proposed.
- Key difference is that only observations where the RHS > 1 are used.

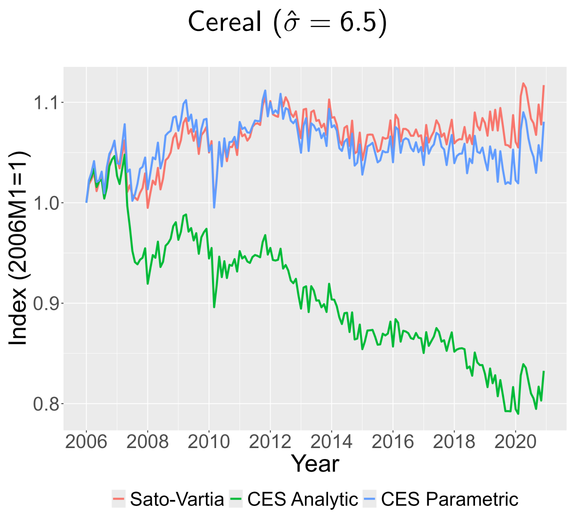
Estimation of a:

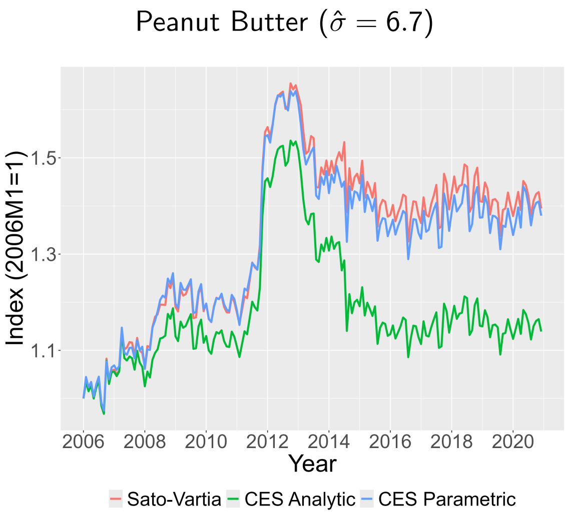
$$s_{it} = \frac{a_i x^r_{it}}{\sum_{j \in I_t} a_j x^r_{jt}} \equiv \hat{s}_{it}, \quad \forall t, \quad i \in I_t$$

$$error_{it} = s_{it} - \hat{s}_{it}$$

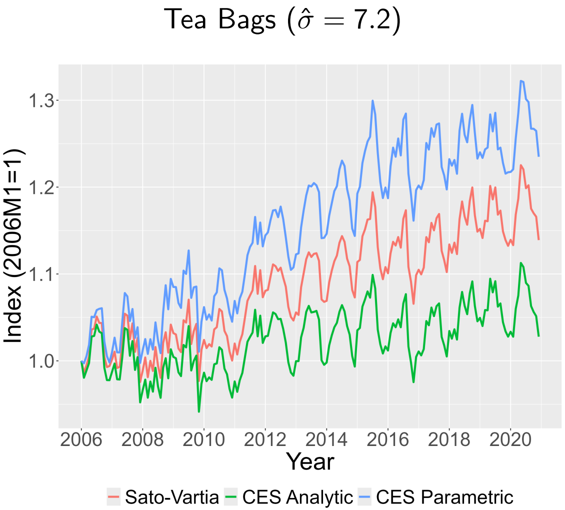
- r is set equal to $1 - \frac{1}{\sigma}$.
- Estimation of **a** is based on minimizing $error_{it}$ with the equaiton being derived from Wold's Identity / FOCs.

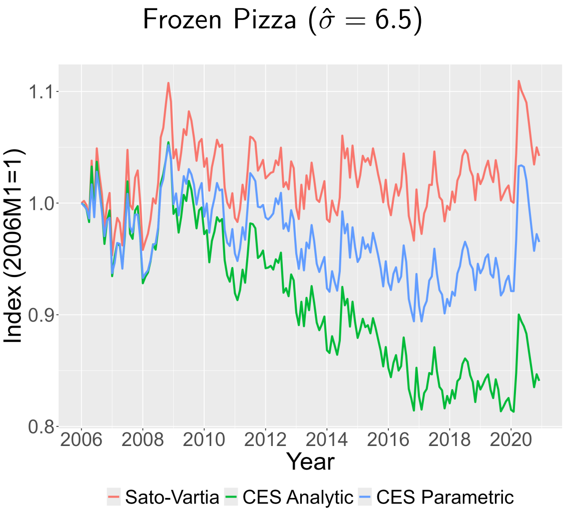
2. Results of Analytic vs Parametric CES Approaches





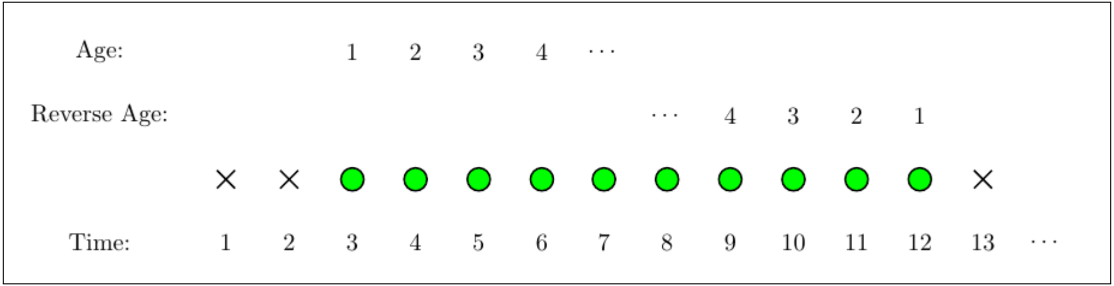
2. Results of Analytic vs Parametric CES Approaches





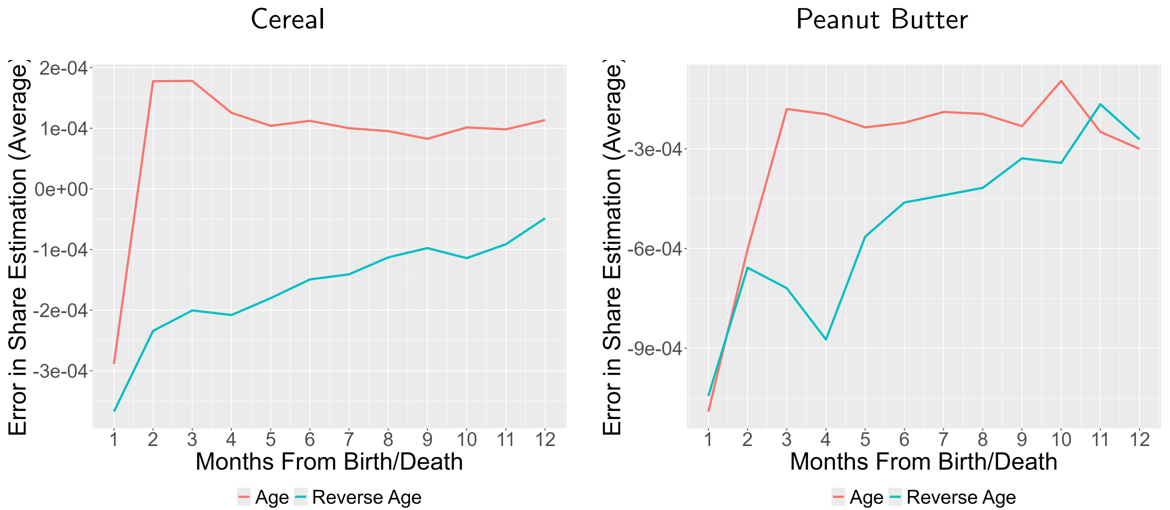
2. Results of Analytic vs Parametric CES Approaches

Age and Reverse Age

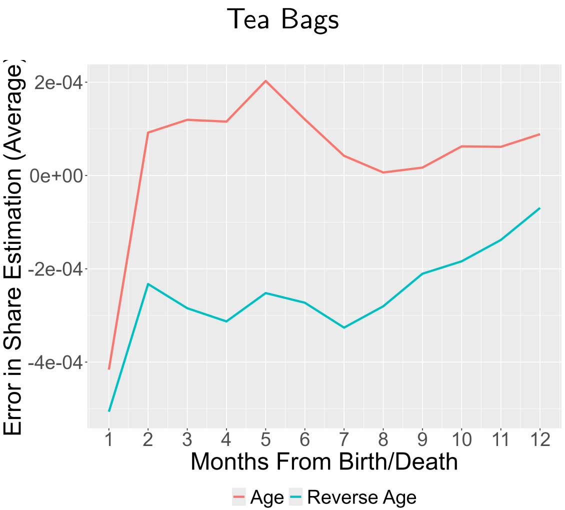


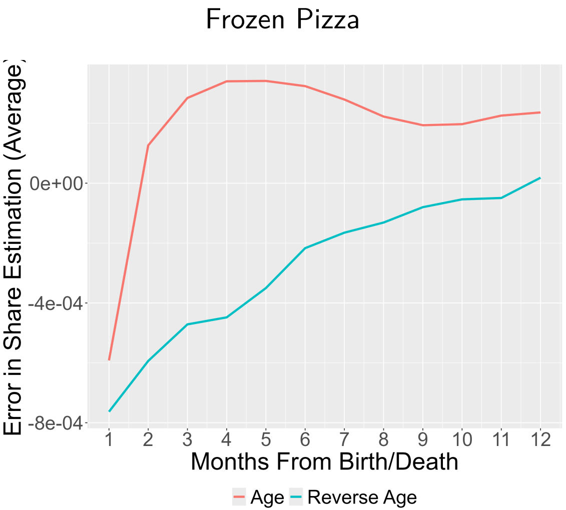
- Green circles denote when the product was sold
- **Age**: number of months the product has been on the market
- **Reverse Age**: number of months until the product disappears from the market

2. Errors in Estimated Shares: $\text{error}_{it} = s_{it} - \hat{s}_{it}$



2. Errors in Estimated Shares: $\text{error}_{it} = s_{it} - \hat{s}_{it}$





2. A Generalization of the CES Analytic Approach

'Unusual' first/last shares are problematic for the CES Analytic approach:

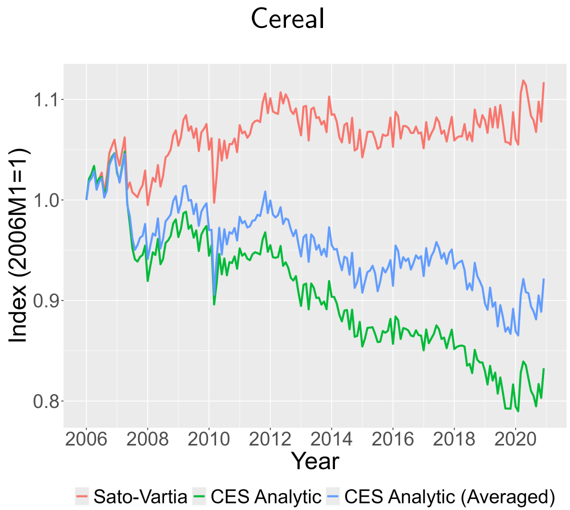
- Using Shepard's Lemma, we develop an extension that 'averages' shares,

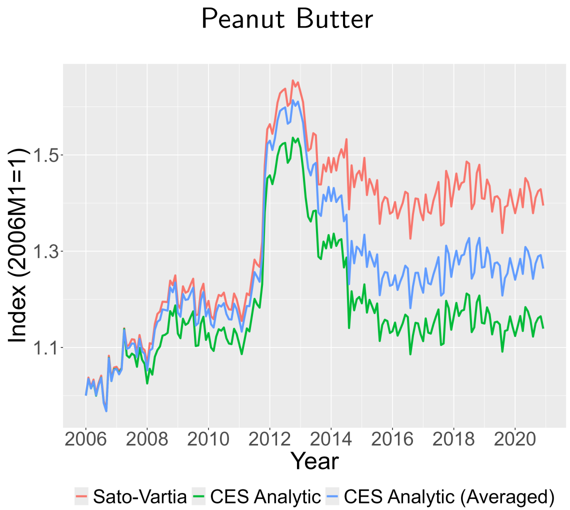
$$s_{it} = s_{iv} \left(\frac{s_{jt}}{s_{jv}} \right) \left(\frac{p_{it}}{p_{jt}} \frac{p_{jv}}{p_{iv}} \right)^{1-\sigma}$$

- We take the geometric mean across all possible combinations,

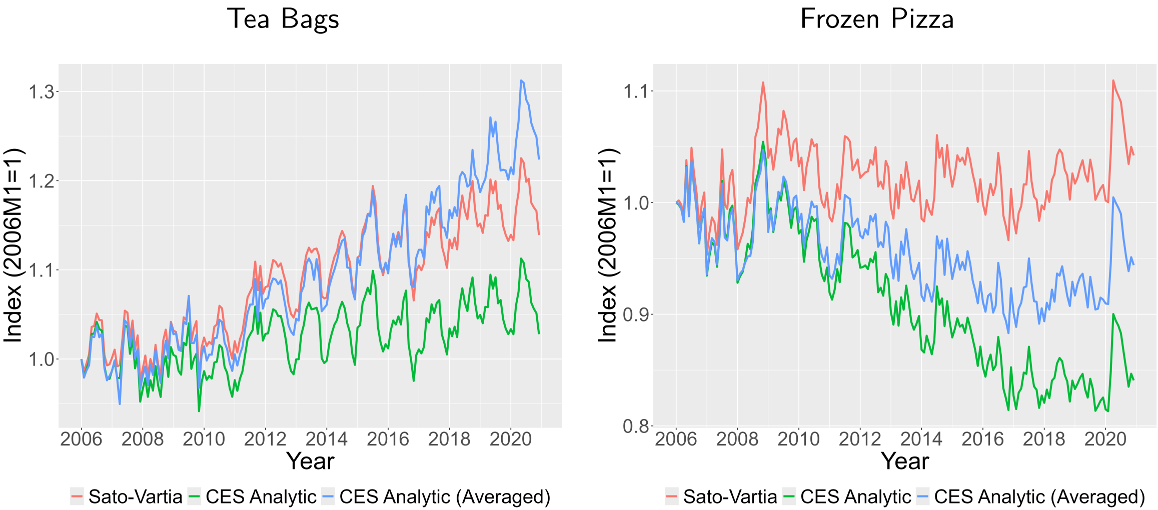
$$s_{it} = \prod_{j \in I_{vt}, v \in T_t^i} \left[s_{iv} \left(\frac{s_{jt}}{s_{jv}} \right) \left(\frac{p_{it}}{p_{jt}} \frac{p_{jv}}{p_{iv}} \right)^{1-\sigma} \right]^{\frac{1}{N_t^i}} = \bar{s}_{it}$$

2. Results of the CES Averaged Analytic Approach





2. Results of the CES Averaged Analytic Approach



2. The Bilateral CES Parametric Approach

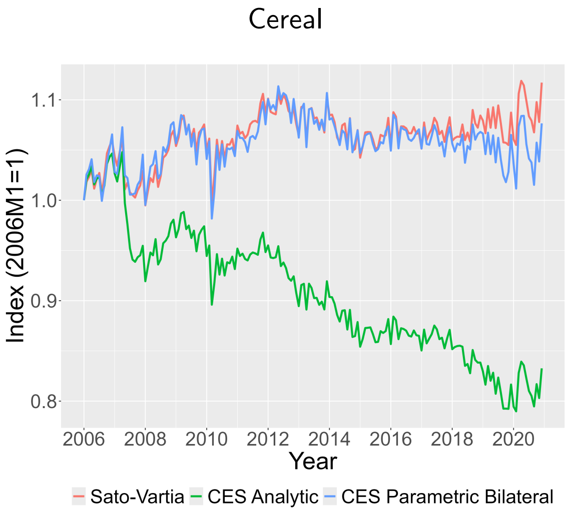
To take account of changes in preferences we can estimate a CES Parametric index across just two periods:

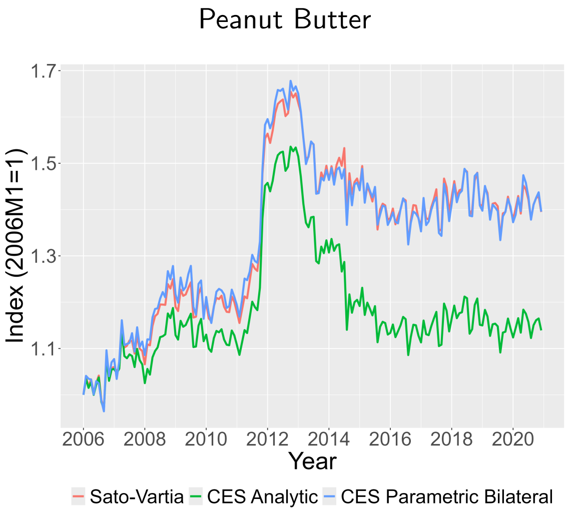
- Estimate σ_{vt} (r_{vt}) and $\boldsymbol{a}_{vt}$ that are consistent with the data in periods v and t.
- A bilateral comparison using common preferences can be constructed as,

$$P_{vt}^{CES}(\sigma_{vt}, \boldsymbol{a}_{vt}) = \left(\frac{\sum_{i \in I_t} p_{it} x_{it}}{\sum_{i \in I_v} p_{iv} x_{iv}} \right) \Big/ \left(\frac{\sum_{i \in I_t} \hat{a}_{ivt} x_{it}^{\hat{r}_{vt}}}{\sum_{i \in I_v} \hat{a}_{ivt} x_{iv}^{\hat{r}_{vt}}} \right)^{\frac{1}{\hat{r}_{vt}}}$$

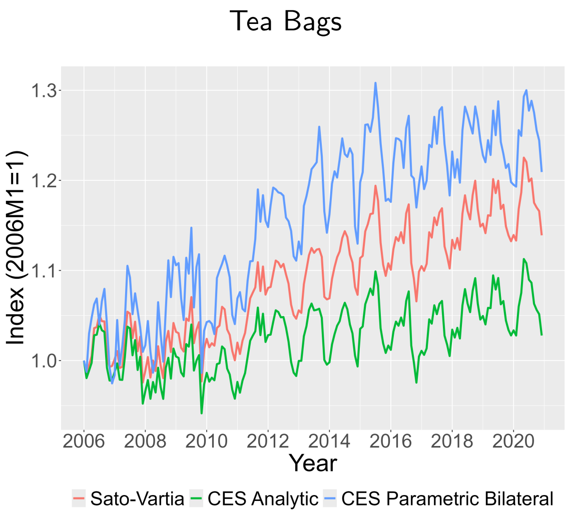
- Multilateral methods can be used to aggregate each of these bilateral indexes.
- This gives a different σ for each comparison and allows preferences to evolve

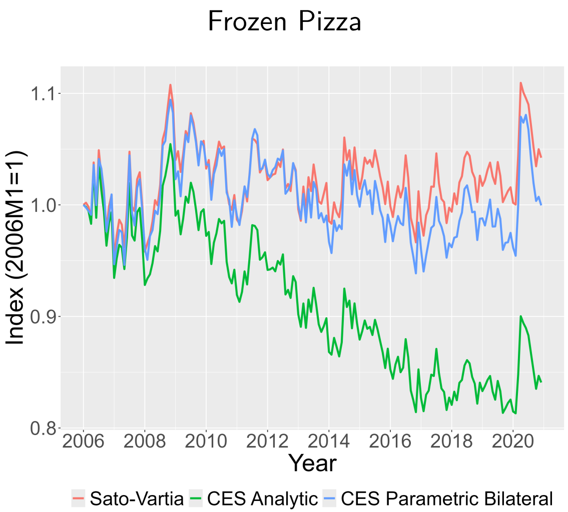
2. Analytic vs Parametric Bilateral CES: Indexes





2. Analytic vs Parametric Bilateral CES: Indexes





3. Approximate Flexible Functional Forms

3. The QMOR Utility Function

The Quadratic-Mean-of-Order-r utility function:

$$U(\mathbf{x}_t) = \left(\sum_{i \in I_t} \sum_{j \in I_t} a_{ij} x_{it}^{\frac{r}{2}} x_{jt}^{\frac{r}{2}} \right)^{\frac{1}{r}}, \quad a_{ij} = a_{ji} \geq 0 \; \forall i, j$$

- The QMOR utility function is concave for $r = 1$, and quasiconcave for $0 < r < 1$
- With estimated parameters, the cost-of-living index can be implicitly derived:

$$P_{vt}^{QMOR} = \frac{\left(\frac{\sum_{i \in I_t} p_{it} x_{it}}{\sum_{i \in I_v} p_{iv} x_{iv}} \right)}{Q_{vt}}, \quad Q_{vt} = \frac{U(\mathbf{x}_t)}{U(\mathbf{x}_v)}$$

3. Estimation of the QMOR Utility Function

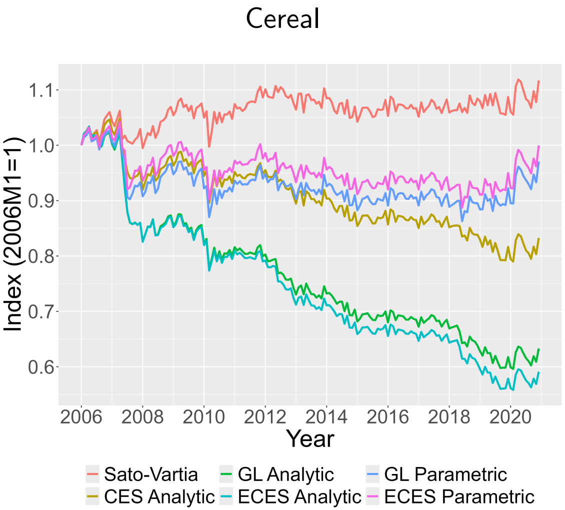
Regularization methods are used to approximate the QMOR utility function:

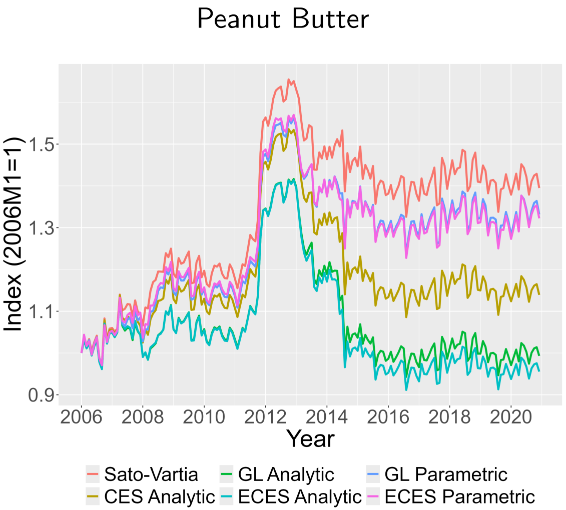
$$\min_{\mathbf{c}} \quad \frac{1}{2M}\sum_{t=1}^{T}\sum_{i\in I_t}\left[s_{it} - \hat{s}_{it}\right]^2 + \frac{\lambda}{2}\sum_{i=1}^{N}\sum_{j=i+1}^{N} c_{ij}^2$$

$$\text{s.t.} \quad a_{ij} = a_{ji} \geq 0 \;\; \forall\, i,j, \quad a_{ij} = b_i I(i=j) + c_{ij}, \quad \sum_{i=1}^{N} b_i = N, \quad \lambda \geq 0$$

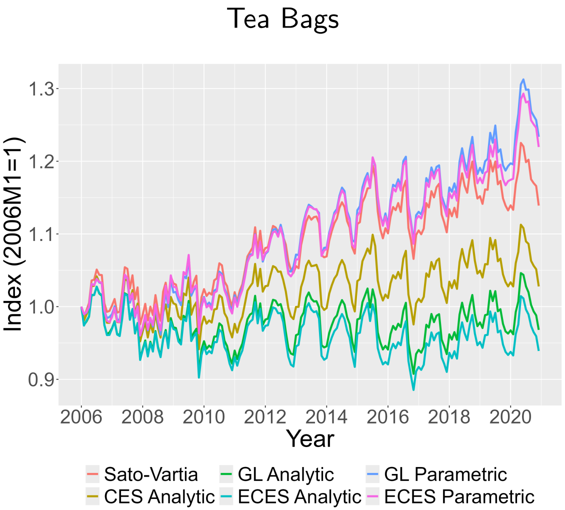
- 10 decreasing values of λ are used.
- In each case I iteratively solve a nonlinear optimization problem.
- The ‘best’ model has the λ with the lowest out-of-sample prediction error.

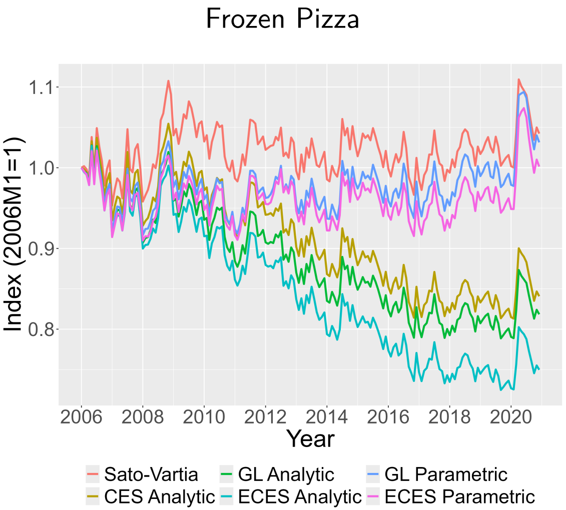
3. Results of the QMOR Utility Function Approach





3. Results of the QMOR Utility Function Approach





4. The Generalized Symmetric Translog Cost Function Approach

4. The Generalized Symmetric Translog Cost Function

The Translog cost function has the form,

$$\log c(\mathbf{p}_t) = \sum_{i \in I} \alpha_i \log p_{it} + \frac{1}{2} \sum_{i \in I} \sum_{j \in I} \gamma_{ij} \log p_{it} \log p_{jt}, \quad \alpha_i > 0, \quad \sum_{i \in I} \alpha_i = 1, \quad \gamma_{ij} = \gamma_{ji}$$

Feenstra and Weinstein 2017 proposed the Symmetric Translog functional form where,

$$\gamma_{ii} = -\gamma \left(\frac{N-1}{N} \right), \qquad \gamma_{ij} = \frac{\gamma}{N}, \qquad \gamma > 0$$

Diewert 2024 proposed a generalization where,

$$\gamma_{ii} = \beta_i^2 - \beta_i \left(\sum_{j \in I} \beta_j \right), \qquad \gamma_{ij} = \beta_i \beta_j, \qquad \beta > 0$$

4. Estimating the GST Cost Function

I start from the Symmetric Translog functional form in Diewert 2024 and use nonlinear optimization to minimize the error between the actual and fitted shares:

$$s_{it} = \alpha_i + \sum_{j \in I} \gamma_{ij} \log p_{jt}$$

- The unknown prices of products that are not sold can be eliminated, which results in a slightly more complex expression.
- With the model parameters we can solve for the reservation prices of products that were not sold.
- These can be used in a Törnqvist index across all varieties sold in either period.
- Key addition to the approach was stronger bounds on β

4. Estimating the GST Cost Function

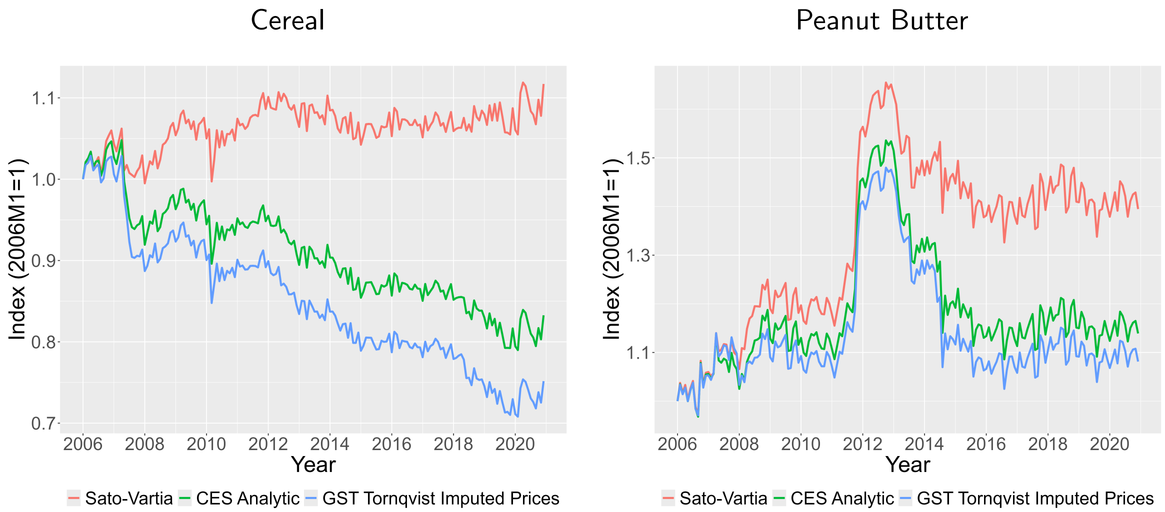
The β need to be positive and suitably far from zero otherwise reservation prices can get very large

- Note the own elasticity of substitution for this functional form is,

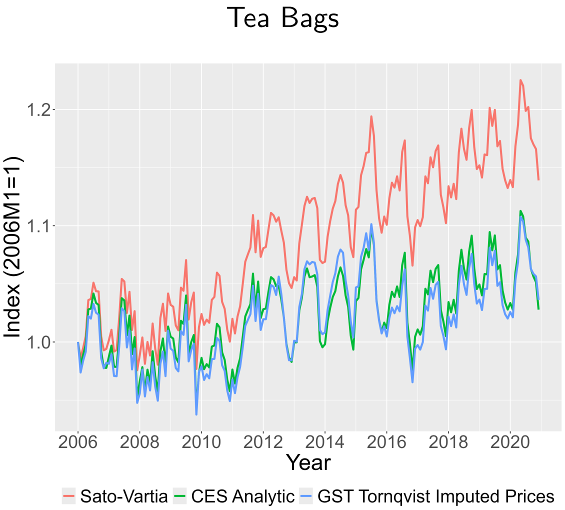
$$\sigma_{ii} = 1 - \frac{1}{s_i} + \frac{\beta_i^2 - \beta_i \left(\sum_{j \in I} \beta_j \right)}{s_i^2}$$

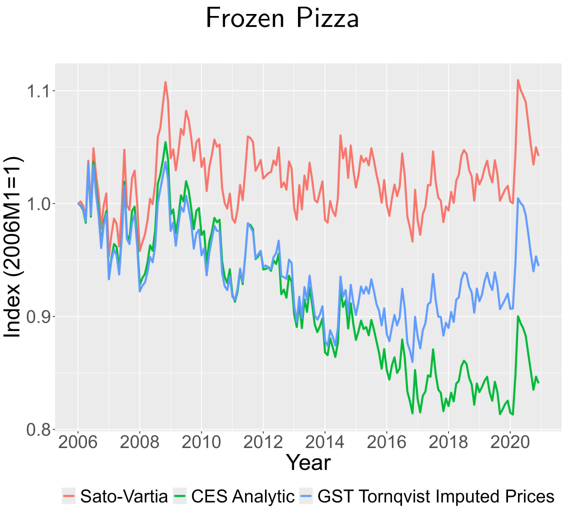
- I require β such that $|\sigma_{ii}| > 1$
- Lower bounds are set on β by solving a system of nonlinear equations where s_i is set equal to the product's maximum observed expenditure share.

4. Generalized Symmetric Translog Results



4. Generalized Symmetric Translog Results





5. Conclusion

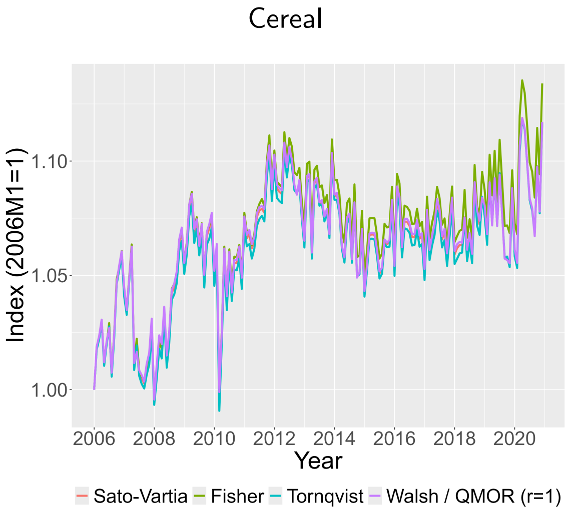
5. Conclusion

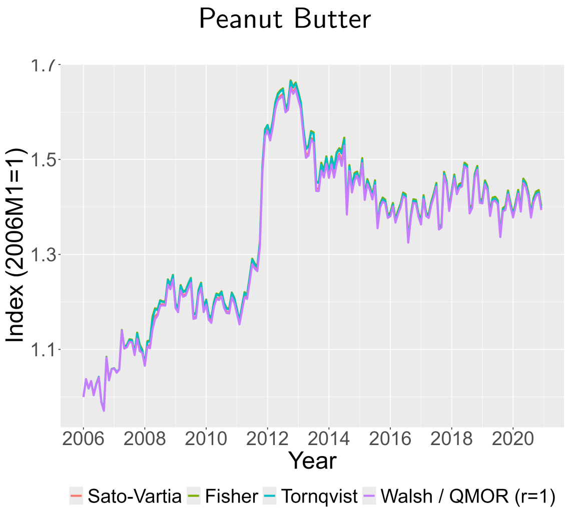
Key findings:

- The gains from variety are generally smaller using the parametric approach compared with the analytic approach
- The 'unusual' nature of first/last expenditure shares poses challenges for the analytic approaches
- New methods such as the Symmetric Translog and Generalized Symmetric Translog give broadly similar results to CES
- It is possible to estimate more flexible functional forms and this should increase our confidence in the results

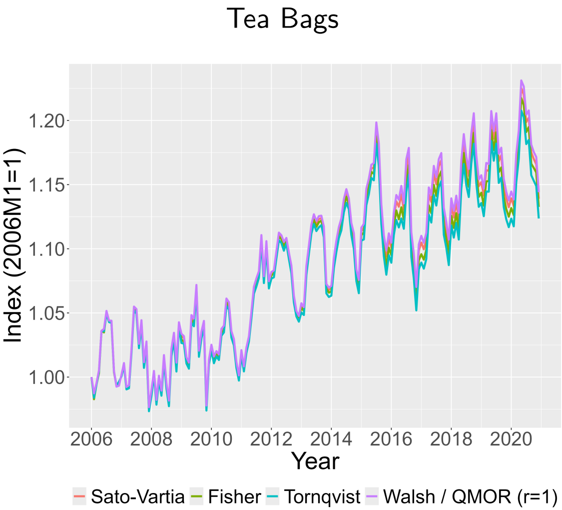
Appendix

Matched Price Indexes





Matched Price Indexes



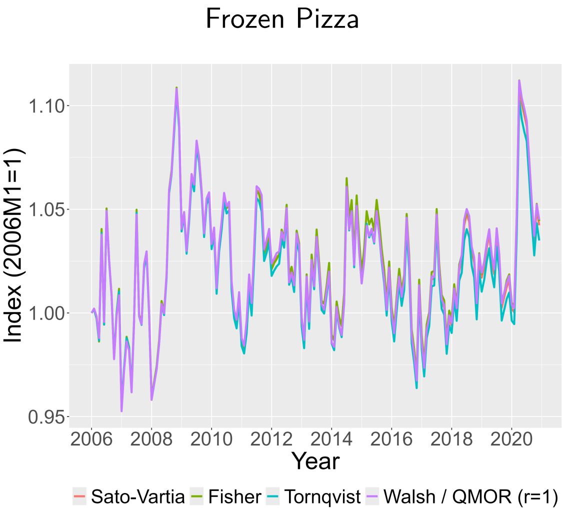
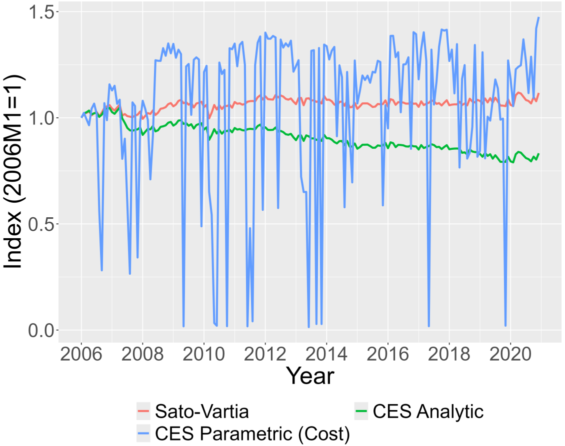


Table of σ Estimates

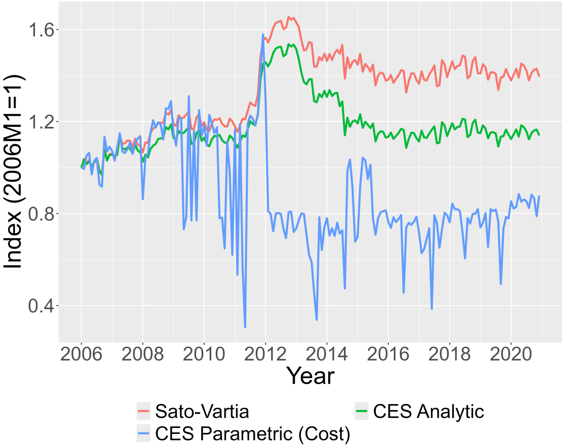
Product	% of $\hat{\sigma}_{itt-1}$ > 0	% of $\hat{\sigma}_{itt-1}$ > 1	Mean Unweighted	Mean Weighted	Geometric Mean Unweighted	Geometric Mean Weighted
Peanut Butter	68.0	61.8	34.0	19.1	6.7	5.0
Toilet Tissue	64.3	60.3	234.3	55.9	11.0	8.0
Teabags	67.5	61.8	58.3	21.3	7.2	4.8
Liquid Soap	62.1	56.2	136.0	111.4	9.1	5.7
Frozen Pizza	70.8	65.3	56.0	21.0	6.5	4.9
Milk	56.5	47.3	257.7	240.8	7.9	4.7
Canned Soup	71.1	63.4	37.9	12.9	5.2	3.8
Ice Cream	72.2	65.4	33.1	11.6	5.4	3.9
Beer	57.7	55.1	146.8	39.8	21.2	9.6
Cereal	68.2	62.0	59.0	25.3	6.5	4.5
Average	65.8	59.9	105.3	55.9	8.7	5.5

2. Analytic vs Parametric CES Approaches (Cost Fun.)

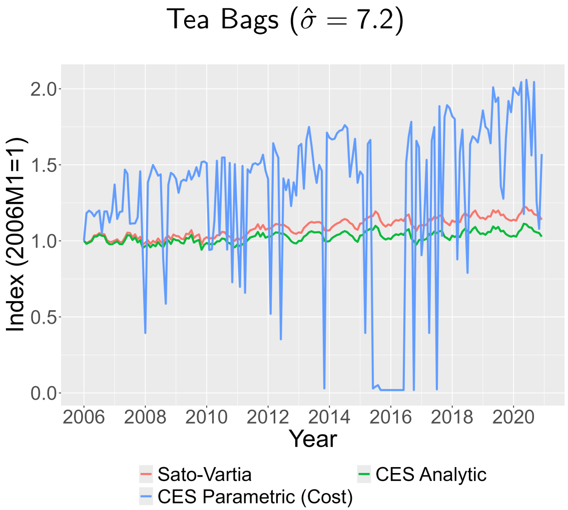
Cereal ($\hat{\sigma} = 6.5$)

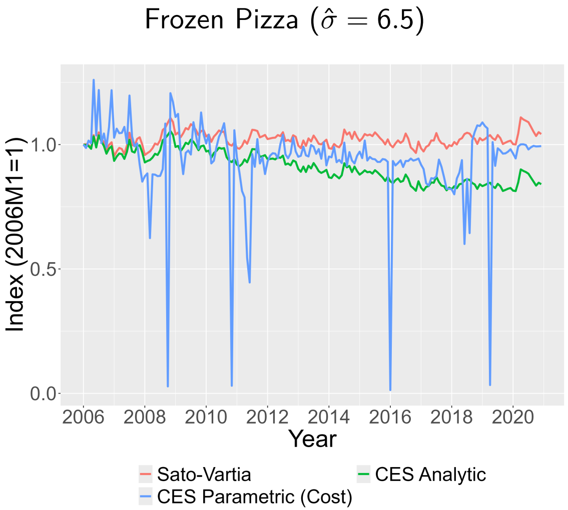


Peanut Butter ($\hat{\sigma} = 6.7$)



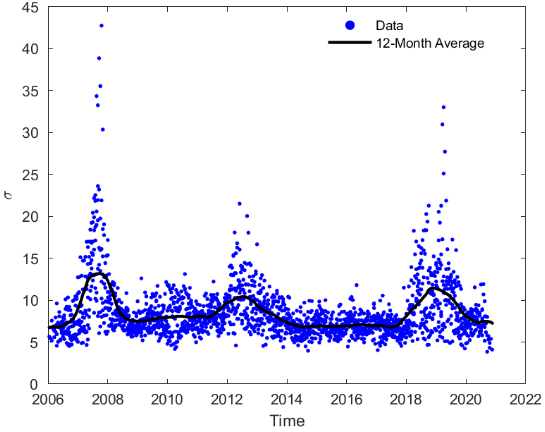
2. Analytic vs Parametric CES Approaches (Cost Fun.)



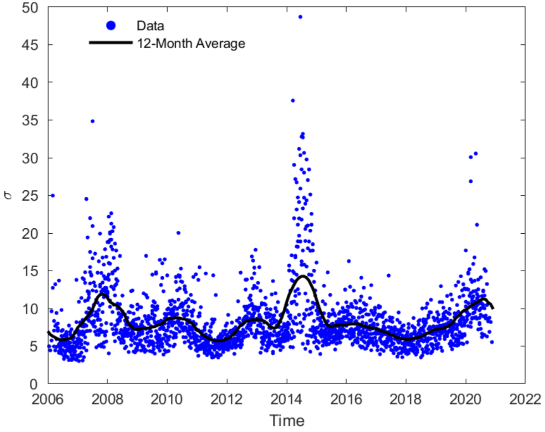


2. Analytic vs Parametric Bilateral CES: σ

Cereal

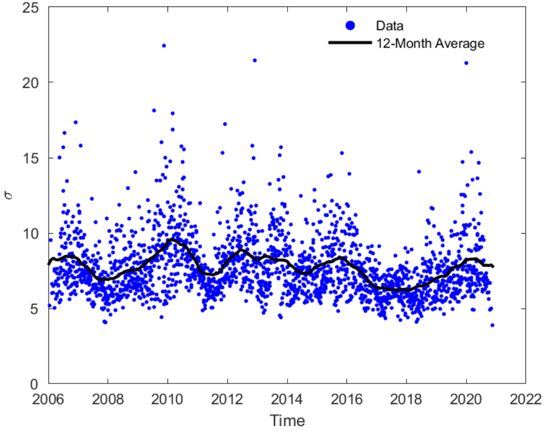


Peanut Butter



2. Analytic vs Parametric Bilateral CES: σ

Tea Bags



Frozen Pizza