UN ECE-ILO Group of Experts Meeting on Consumer Price Indices,
May 26, 2014, Geneva
Four problems:

- A temporary missing item problem
- A seasonal item problem
- A permanently missing item
  - Comparable replacement
  - Non-comparable replacement – imputation/linking or quality adjustment
  - Rapid turnover of models and characteristics, e.g. consumer electronics – new models are introduced and old ones disappear
- A new products problem
Price statisticians do take account of quality changes

- They use the matched models method:
  - sampled models (varieties) selected using detailed product descriptions on “initiation”
  - prices are recorded in initial month, and monitored in subsequent months
  - like is compared with like.
Importance of price collector

Price collectors need characteristic specifications, characteristics of replacements when permanently missing, and codes for:

- temporarily missing
- seasonal
- permanently missing: replacement variety’s price is:
  - comparable replacement
  - non-comparable replacement
    - previous period’s price (overlap)
    - explicit adjustment (specification)
Imputation techniques for temporary missing prices

- Calculate change in average price from *matched* observations of items from the same class using short-term (geometric mean) price changes.

- Estimate the price for the missing observations and mark them as “imputed.”
Induces undue stability into the index, especially for high inflation countries.

Not for use unless assured prices do not change.
Permanently missing varieties

- Comparable replacement
- Simply compare price of replacement with previous price
- Requires well trained price collectors and tightly-defined specifications
- There is an incentive to assume replacements are comparable.
Permanently missing: no comparable replacement

- Alternatives: Direct or Indirect Quality Adjustments
  - Indirect Quality Adjustments with Imputations
    - overlap price available – store manager;
    - if not need to impute overlap price (not carry forward)
  - Direct Quality Adjustment
    - data collector or analyst knowledge,
    - information from producers,
    - differences in production/option costs
    - hedonic regression models
Differentiated items with high rates of model turnover:

<table>
<thead>
<tr>
<th></th>
<th>jan</th>
<th>feb</th>
<th>mar</th>
<th>apr</th>
<th>may</th>
<th>jun</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>B</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>X</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
Say a 300 tablet box of a specified store brand of aspirins, sold in January at $12.00 is replaced by a 365 tablet box, sold for $15.00 in February – a non-comparable replacement. Assume linearity in the price-size relationship.

The rescaling of the January price to the February size is $365/300 \times 12 = 14.6$. The constant-quantity size price change is $15/14.6 = 1.0274$, i.e. 2.74 percent. The assumption here is that every extra pill cost $12/300 = $0.04.

Phrased another way: $\hat{p}_{365}^{Jan} = p_{300}^{Jan} + \beta \Delta size$

$12 + 0.04 (365-300) = 14.6$
Hedonic regressions

A set of \((z_k = 1, \ldots, K)\) characteristics of the models are identified and data over \(i=1, \ldots, N\) models are collected. A hedonic regression of the (log) price of model \(i\), \(p_i\), on its set of quality characteristics \(z_{ki}\) is given by:

\[
\ln p_i = \beta_0 + \sum_{k=1}^{K} \beta_k z_{ki} + \varepsilon_i
\]

The \(\beta_k\) are estimates of the marginal valuations the data ascribes to each characteristic.
A simple case of one explanatory variable: price of washing machines and their load capacity in kg.
The slope coefficient

Used to estimate the equation of the line that best fits the data.

\[ Y = a + b \, X \]

- \( a \) is the intercept – the value of \( Y \) when \( X \) is zero
- \( b \) is the slope – the change in \( Y \) arising from a unit change in \( X \)
A line of best fit

- Least squares criterion: minimises the sum of squared vertical differences between the points and the line.

- A one kg. increase in capacity increases price by 5.88 percent.

\[ \ln \text{ price} = 5.3961 + 0.058808 \text{capacity, kg} \]

\[ R^2 = 0.139 \]

- R-bar squared is the proportion of variation in (the log of) P that is explained by the right-hand-side variables.

- A t-statistic to test the null hypothesis of the coefficient being zero was 2.543: the null hypothesis of capacity having a zero effect on log price was rejected at a 5 percent level.
Hedonic regressions: an illustrative example of washing machines, continued

- Data from the UK Which website on advertised characteristics and prices

- All models of 29 brands compared in source data: 3 brands taken for illustration April 2014: LG, Hotpoint and AEG comprising 42 models in all.

- Log of price (pounds sterling) on:
Explanatory characteristics

- Capacity k.
- Spin speed rpm
- Stainless steel
- Non white
- Energy label (5 categories)
- Energy costs
- Brand LG
- Brand AEG
- Warranty years
- Time for standard load, mins

- Easy iron
- Hand wash
- Extra rinse
- Variable spin
- Delayed start
- Time remaining display
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPACITYKG</td>
<td>0.046096</td>
<td>0.035383</td>
<td>1.302791</td>
<td>0.2041</td>
</tr>
<tr>
<td>ENERGYCOSTS</td>
<td>0.02356</td>
<td>0.007973</td>
<td>2.955101</td>
<td>0.0066</td>
</tr>
<tr>
<td>ENERGYLABE</td>
<td>0.022671</td>
<td>0.047903</td>
<td>0.473263</td>
<td>0.64</td>
</tr>
<tr>
<td>SPINSPEEDRPM</td>
<td>-0.000154</td>
<td>0.000396</td>
<td>-0.390167</td>
<td>0.6996</td>
</tr>
<tr>
<td>TIMELOADMINS</td>
<td>-0.003533</td>
<td>0.001622</td>
<td>-2.178362</td>
<td>0.0386</td>
</tr>
<tr>
<td>WARRANTYYEARS</td>
<td>2.500903</td>
<td>0.299634</td>
<td>8.346537</td>
<td>0</td>
</tr>
<tr>
<td>DUMMYAEG</td>
<td>0.323514</td>
<td>0.134174</td>
<td>2.411148</td>
<td>0.0233</td>
</tr>
<tr>
<td>DELAYEDSTART</td>
<td>0.049834</td>
<td>0.204742</td>
<td>0.243397</td>
<td>0.8096</td>
</tr>
<tr>
<td>EASYIRON</td>
<td>0.14711</td>
<td>0.097878</td>
<td>1.502997</td>
<td>0.1449</td>
</tr>
<tr>
<td>EXTRARINSE</td>
<td>0.045771</td>
<td>0.096656</td>
<td>0.473544</td>
<td>0.6398</td>
</tr>
<tr>
<td>TIMEREMAININGDISPLAY</td>
<td>0.164096</td>
<td>0.21402</td>
<td>0.76673</td>
<td>0.4501</td>
</tr>
<tr>
<td>HANDWASHPROGRAM</td>
<td>2.623436</td>
<td>0.322079</td>
<td>8.145317</td>
<td>0</td>
</tr>
<tr>
<td>VARIABLESPIN</td>
<td>0.125619</td>
<td>0.233856</td>
<td>0.537163</td>
<td>0.5957</td>
</tr>
<tr>
<td>NONWHITE</td>
<td>0.006836</td>
<td>0.089611</td>
<td>0.076284</td>
<td>0.9398</td>
</tr>
<tr>
<td>STAINLESSSTEEL</td>
<td>0.157542</td>
<td>0.155097</td>
<td>1.015763</td>
<td>0.3191</td>
</tr>
</tbody>
</table>

R-squared 0.722656  Mean dependent var 5.876148
Adjusted R-squared 0.573316  S.D. dependent var 0.265973
S.E. of regression 0.173736  Akaike info criterion -0.38632
Sum squared resid 0.784792  Schwarz criterion 0.240593
Log likelihood 22.91964  Hannan-Quinn criter. -0.15804
### Dependent Variable: LOGPRICE

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPACITYKG</td>
<td>0.052047</td>
<td>0.018858</td>
<td>2.760026</td>
<td>0.0094</td>
</tr>
<tr>
<td>ENERGYCOSTS</td>
<td>0.018567</td>
<td>0.006003</td>
<td>3.092679</td>
<td>0.004</td>
</tr>
<tr>
<td>TIMEOLOADMINS</td>
<td>-0.002898</td>
<td>0.001151</td>
<td>-2.517834</td>
<td>0.0168</td>
</tr>
<tr>
<td>WARRANTTYEARS</td>
<td>2.552257</td>
<td>0.147547</td>
<td>17.29798</td>
<td>0</td>
</tr>
<tr>
<td>DUMMYAEG</td>
<td>0.356404</td>
<td>0.101557</td>
<td>3.509394</td>
<td>0.0013</td>
</tr>
<tr>
<td>EASYIRON</td>
<td>0.15951</td>
<td>0.078837</td>
<td>2.023292</td>
<td>0.0512</td>
</tr>
<tr>
<td>HANDWASHPROGRAM</td>
<td>2.592142</td>
<td>0.104254</td>
<td>24.86368</td>
<td>0</td>
</tr>
<tr>
<td>TIMEREMAININGDISPLAY</td>
<td>0.192852</td>
<td>0.069758</td>
<td>2.764597</td>
<td>0.0093</td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared</td>
<td>0.699144</td>
<td>Mean dependent var</td>
<td>5.876148</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.635326</td>
<td>S.D. dependent var</td>
<td>0.265973</td>
<td></td>
</tr>
<tr>
<td>S.E. of regression</td>
<td>0.160616</td>
<td>Akaike info criterion</td>
<td>-0.64642</td>
<td></td>
</tr>
<tr>
<td>Sum squared resid</td>
<td>0.851322</td>
<td>Schwarz criterion</td>
<td>-0.31206</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>21.25153</td>
<td>Hannan-Quinn criterion</td>
<td>-0.52466</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson stat</td>
<td>1.312802</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

- Selection of brands in appropriate market segment: e.g. up-market brands.
- Three outliers (AEG).
- Bundling of features: a few can represent many; multicollinearity.
- Additional kg. of capacity increases price by 5.2 percent.
- Scatter diagram shows a lot of price variation for higher-capacity models.
- AEG valued at $100 \times (\exp(0.356404 - 1)) = 42.82$ percent more than Hotpoint. Limited sample of high-priced AEG.
- More observations needed.
How to use hedonics

- Use individual coefficients as estimates of the value of changes in the quality of individual features – not advised

- Include a time dummy:

\[
\ln p_i = \beta_0 + \sum_{k=1}^{K} \beta_k z_{ki} + \gamma \text{TimeDum}_i^t + \varepsilon_i
\]

- Use predicted prices: single vs double imputation:

\[
\frac{\sum_i \hat{p}_{hed}^t q_{char}^t}{\sum_i \hat{p}_{hed}^0 q_{char}^t} = \frac{\sum_i \hat{p}_{hed}^t q_{char}^0}{\sum_i \hat{p}_{hed}^0 q_{char}^0}
\]

- Use the web to see difference in pricing of a similar model for different specifications.

- Use some else's results.
Statistical issues in hedonic regressions

- Data often readily available on websites; just use mark-ups for characteristics?
- Software makes estimation easy – even spreadsheets have regression.
- Need for diagnostic tests and expertise to validate model: heteroskedasticity, normality of residuals, multicollinearity, non-linear estimators – available in statistical/econometric software.
- Use of predicted values rather than coefficients for quality adjustment. Use of dummy time variable
- Functional form
- Inclusion of which variables