Use of novel data sources for consumer prices

Catherine McGirr
Office for National Statistics, UK
Quality dimensions

Coherence and comparability
Comparability over e.g. time or region, and with other statistics and analysis.

Timeliness and punctuality
Time gap between publication and the reference period

Accuracy and reliability
Closeness between estimated result and true value
Closeness of early estimates to subsequent estimates

Accessibility and clarity
Ensuring the data and analysis is easy to access, and accompanied by insightful advice and metadata

Relevance
Meeting user needs in coverage and content
What are alternative data sources?

- Scanner data
- Web scraped data
- Used in conjunction with data as currently collected
Why use alternative data sources?

• Improved coverage – for example regional
• High frequency of collection
• Reduced burden
• More granular information (eg expenditure)
Comparison of different data sources

<table>
<thead>
<tr>
<th>Data dimension</th>
<th>Traditional</th>
<th>Scanner data</th>
<th>Web scraping</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data acquisition</strong></td>
<td>Manual</td>
<td>Automated</td>
<td>Automated</td>
</tr>
<tr>
<td><strong>Completeness/scope</strong></td>
<td>Sample items from sample of retailers</td>
<td>All transactions (census) from medium to large retailers</td>
<td>Bulk or sample from online retailers</td>
</tr>
<tr>
<td><strong>Metadata</strong></td>
<td>Item description + some attributes</td>
<td>Item description + (some) attributes</td>
<td>Item description + attributes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Backseries may be available</td>
<td></td>
</tr>
<tr>
<td><strong>Quantity data</strong></td>
<td>None</td>
<td>Quantities sold</td>
<td>None</td>
</tr>
<tr>
<td><strong>Timing</strong></td>
<td>Single collection day</td>
<td>Daily</td>
<td>Daily/weekly</td>
</tr>
</tbody>
</table>
Ensuring quality of the data

- Initial data checks
- Standardising the data
- Monitoring the regular feeds
- Data sharing agreements
- Quality Assurance of Admin Data (QAAD)
Ensuring quality of the outputs

- Robust statistical systems
- Research into new methods
- Impact analysis
- Incorporate QA into regular monthly production
- Advisory panels – stakeholder and technical
- International best practice