Optimisation Applications at the Australian Bureau of Statistics

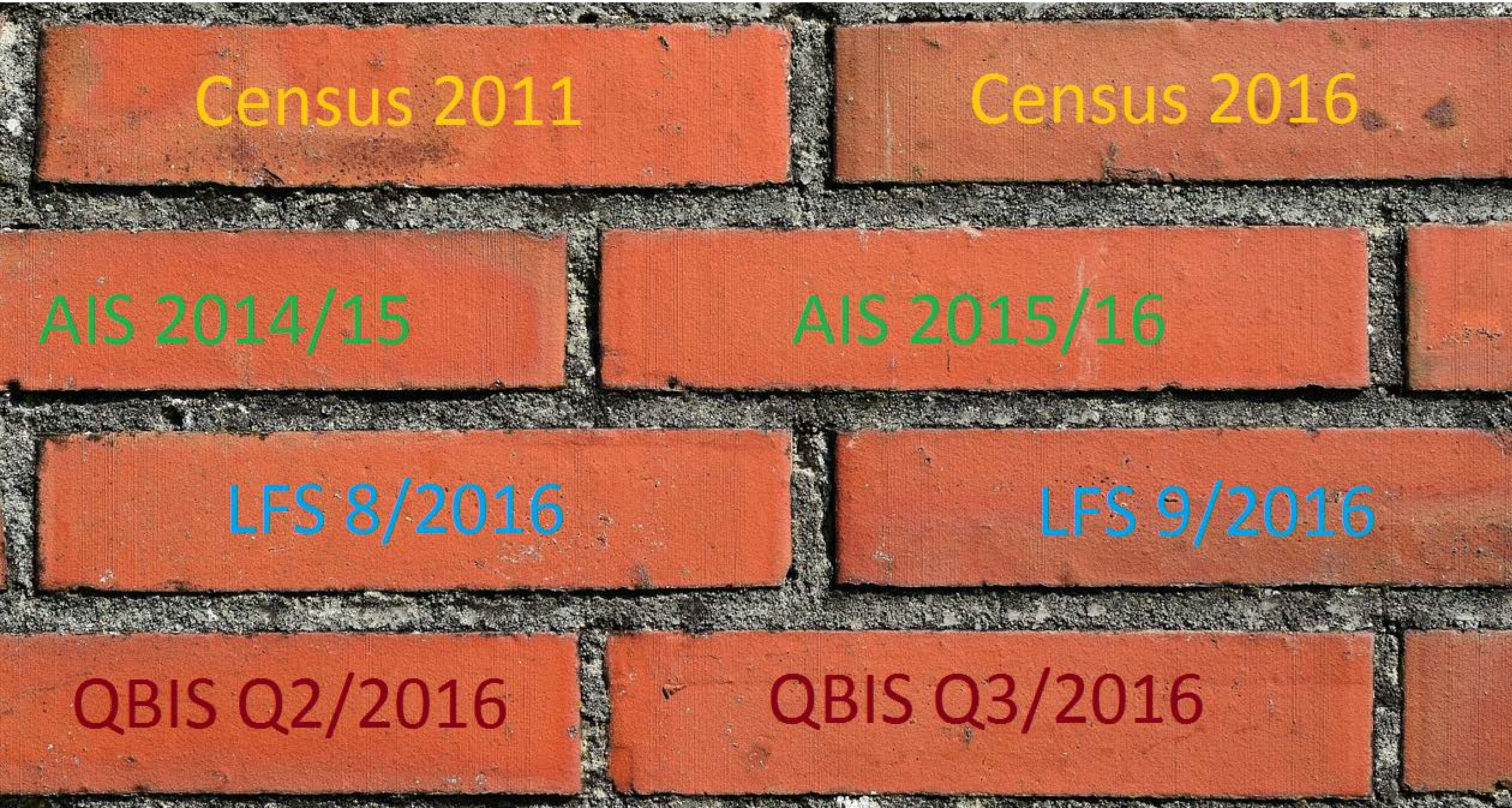
Geoffrey Brent, ABS National Accounts Branch
geoffrey.brent@abs.gov.au

About the ABS

We produce a range of social and economic statistics e.g.:

- Economic accounts
- Environmental accounts
- Employment figures
- Population estimates (used to determine electoral representation)
    - etc. etc. etc.

The past (idealised)



The future (exaggerated)

Job website A
Job website B
ATO
Centrelink
Scanner data

Work allocation problem

- ABS has field staff for household surveys and Census.
  - Shifted to “web first, phone second” approach but still have field work.
- Need to form workloads & allocate to staff, subject to various considerations:
  - Minimise costs e.g. travel.
  - Minimum/maximum workloads etc.
- Textbook OR problem.

Data editing

Respondents may give inconsistent, implausible, or missing data:

- Person born in 2016 is listed as parent of person born in 1965.
- Company reports turnover of $30,000 on tax return but $30,000,000 to ABS for same reference period.
- Data items left blank.

Data editing (2)

- Not always possible to query responses with data provider.
- Need to “edit” data: attempt to correct it.
- Can frame this as an MIP-type problem: what is the “cheapest” edit that satisfies consistency rules?
  - Can apply at group level: e.g. 51% of people are female but may not want to treat all blank responses as female.

Confidentiality

- ABS has legal and ethical responsibility to protect confidentiality of our data providers (individual and business).
- Sometimes need to withhold data to preserve confidentiality.
- Fictionalised example based on real issues…

Confidentiality (2)

We want to publish total sales of widgets, sprockets, and doohickeys by region:

Total sales ($M)	Product			
State	Widgets	Sprockets	Doohickeys	**Total**
NSW	45	20	5	**70**
Vic	15	20	30	**65**
Qld	5	80	25	**110**
Others	25	35	15	**75**
Total	**90**	**155**	**75**	**320**

Confidentiality (3)

- Queensland only has one sprocket manufacturer.
- To preserve their confidentiality, we cannot publish the value for Qld sprocket sales.
- We still need to publish the table. So…

Confidentiality (4)

Total for Qld sprocket sales is “suppressed”:

Total sales ($M)	Product			
State	Widgets	Sprockets	Doohickeys	**Total**
NSW	45	20	5	**70**
Vic	15	20	30	**65**
Qld	5	*	25	**110**
Others	25	35	15	**75**
Total	**90**	**155**	**75**	**320**

But suppressed value can be recovered…

Confidentiality (5)

So we need to apply "secondary suppression", e.g.:

Total sales ($M)	Product			
State	Widgets	Sprockets	Doohickeys	**Total**
NSW	45	*	*	**70**
Vic	15	20	30	**65**
Qld	5	*	*	**110**
Others	25	35	15	**75**
Total	**90**	**155**	**75**	**320**

Confidentiality (6)

- Secondary suppression is undesirable – reduces value of the information.
- Want to find the “cheapest” suppression solution.
- Need to ensure that readers can’t use rules of the table to recover confidential info.
- This becomes a tough LIP/MIP.

Table balancing

ABS compiles large demographic and economic tables e.g.:

- Estimated Resident Population: approx. 2000 regions x 180 age/sex classes.
- Supply-Use: supply and use of 301 products by 67 industries + household, government sectors.

Supply-Use

- S-U measures flows of products (goods/services) between sectors (industry, government, household etc.).
- Flows are measured from more than one perspective.
  - When I buy a pizza, somebody else sells a pizza.
  - We aim to record both the “sale” and the “purchase” sides of that activity.

Supply-Use (2)

The industry *supplies* $20 of food:

	2013-14	**Australian production**			
Supply-Use Product Code	Product name	… …	450 Food and beverage services	… …	Total Supply
…	…	…	…	…	…
45010	Takeaway food	…	+$20	…	**+$20**
…	…	…	…	…	…
Total		**…**	**+$20**	**…**	**+$20**

Supply-Use (3)

The household sector *uses* $20 of food:

	2013-14	**Final demand**			
SUPC		…	Household final consumption expenditure	…	Total Use
…	…	…	…	…	…
45010	Takeaway food	…	+$20	…	**+$20**
…	…	…	…	…	…
Total		**…**	**+$20**	**…**	**+$20**

Supply-Use (2)

- One transaction shows up in ~ 8 cells in the table – implies internal rules.
- Total value supplied for each product should match total value used.
- Total value supplied by each industry should match total cost of inputs plus value-add.
- Various other expectations, e.g. most items non-negative.

Table balancing (2)

SU tables are compiled from many sources including:

- Surveys of businesses
- Surveys of households
- Tax and excise records

Sampling error and other source issues create inconsistencies.

Table balancing (3)

- Tables need to be consistent.
- Large discrepancies are investigated and addressed by subject-matter experts.
- Many small discrepancies remain.
- Need an automated method for balancing them.
- Want to avoid distorting the economic picture while balancing.

Table balancing (4)

- Richard Stone *et al.* identified weighted least squares balancing as an option for accounts balancing in 1942.
- Computing limitations made this infeasible for large tables.
- Iterative methods (RAS) were used as a substitute.
- Computationally cheaper but have weaknesses.

Table balancing (5)

- Advances mean that WLS-type balancing is now achievable even for large National Accounts data sets.
- Several agencies have already adopted modern optimisation tools for balancing work.
- ABS is currently developing optimisation methods.

Table balancing (6)

- Other agencies have generally adopted a commercial optimisation solver (CPLEX or Xpress) and programmed directly for that solver.
- Encountering MiniZinc on Coursera suggested benefits of working via solver-independent platform.
- ABS currently using AMPL x Gurobi.

Table balancing (7)

A bit about the problem…

- Most constraints are straightforward linear constraints: $x + y = z$.
- A few nonlinear constraints: $price * volume = value$
  - – Need to use some tricks here since Gurobi doesn't support this kind of rule.
- Challenge here is number of constraints…

Table balancing (8)

- Balancing a single year of Supply-Use data involves ~ 300,000 individual constraints.
- These can be specified in ~ 50 AMPL statements.
  - Set expressions are very useful here!
  - e.g. define "set of government industries"
- About 90% can be eliminated in presolve.

Table balancing (9)

Big challenge: what should our objective function be?

Most approaches focus on preserving attributes of the unbalanced data:

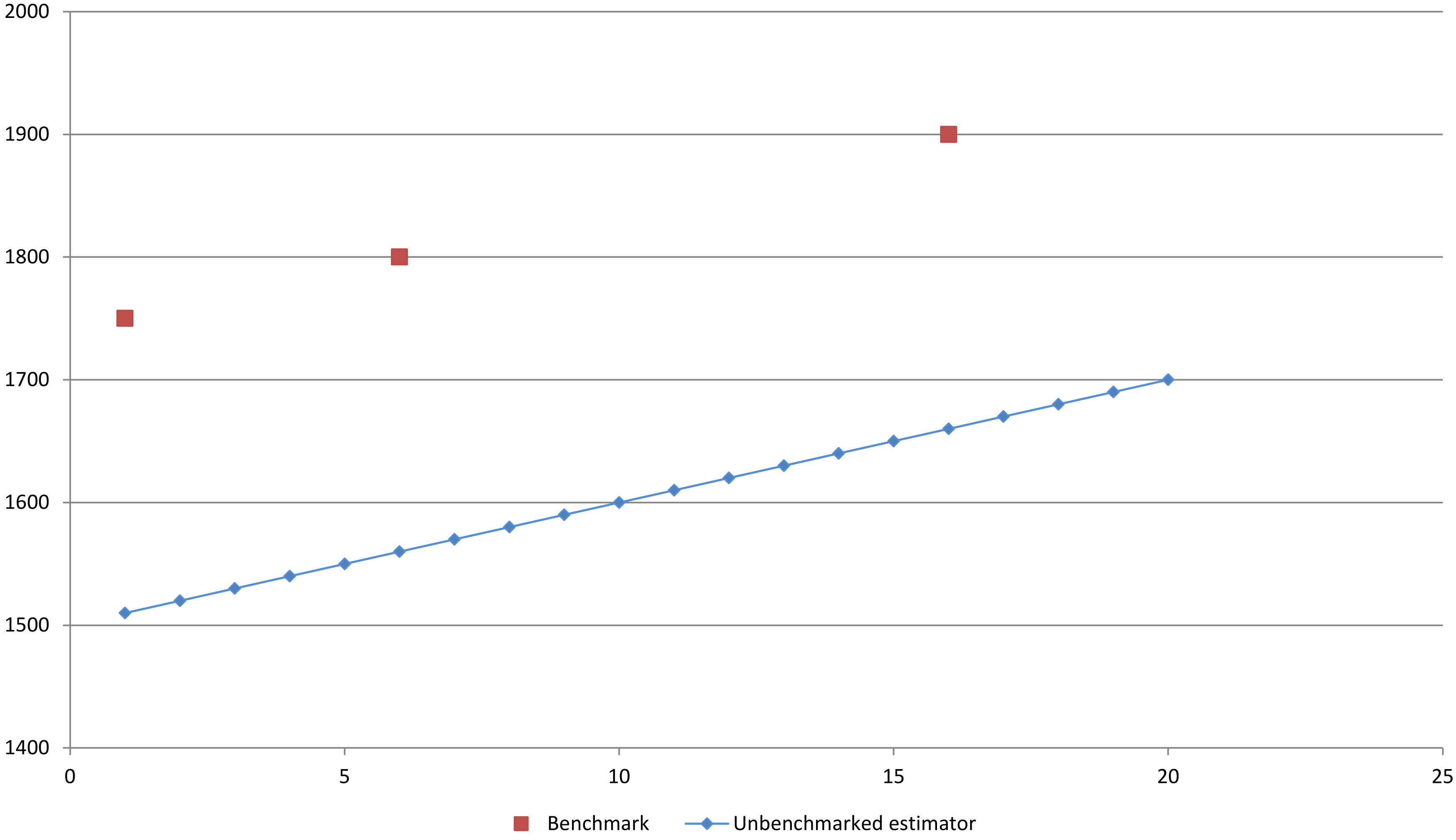
- "Levels": e.g. total household consumption of takeaway food for each year.
- "Movements": e.g. growth/decline in takeaway consumption.
- Some weighted combination of the two.

Table balancing (10)

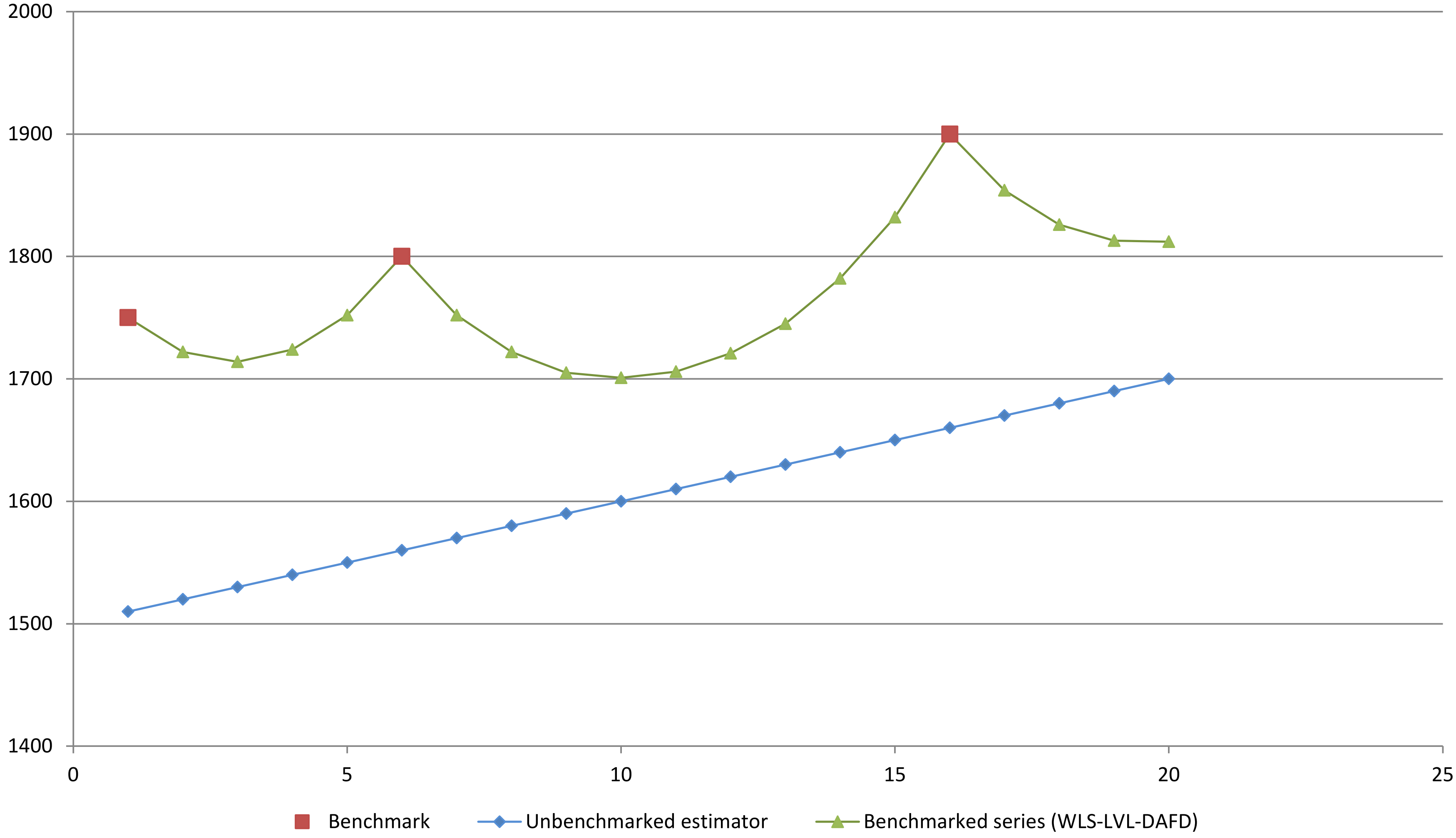
The “preservation” approach seems intuitive but has drawbacks:

- Discrepancies imply errors in the unbalanced data.
- Ideally we would *remove* all errors, not preserve them.
- Combined movement/level preservation can cause weird results…

Benchmarking artefacts



Benchmark artefacts (2)



Alternate approach

- Observed = true + error
- Specify model for the form of errors: e.g. Gaussian white noise, random walk, …
- Find the most plausible errors (under this model) that are consistent with observations.
  - Maximum likelihood estimate (MLE)
- Then subtract these errors to get balanced estimates.

MLE approach (2)

- MLE approach can be transformed into a quadratic objective function.
- For simple cases, the MLE method gives the same solutions as the “preservation” approach.
  - Different approach for justifying same methods.
  - Helps understand limitations of these methods.

MLE approach (3)

- For complex cases, MLE gives different results.
- Under MLE approach, adding level- and movement-preservation objective functions is not justifiable.
  - Implies some impossible assumptions.
- Instead, we estimate two components of error and apply different OF to each component.

Closing notes

- ABS has been using optimisation ad-hoc for a long time but is now coordinating optimisation work.
- Not many staff have optimisation backgrounds.
- We are working to build our optimisation knowledge:
  - Theory
  - Practical

Questions?

ABS 2018 graduate program is opening for applications shortly. See ABS website for details.