Australian Bureau of Statistics

1352.0.55.157

Research Paper

Quadratic Optimisation for Table Balancing in Official Statistics

New Issue

Research Paper

Quadratic Optimisation for Table Balancing in Official Statistics

Geoffrey Brent

Methodology Transformation Branch

Methodology Advisory Committee

30 June 2016, Canberra

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) WED 14 SEP 2016

ABS Catalogue no. 1352.0.55.157

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper. For further information, please contact Mr Ruel Abello, Director, Methodology Transformation Branch, on Canberra (02) 6252 6307 or email <analytical.services@abs.gov.au>.

QUADRATIC OPTIMISATION FOR TABLE BALANCING IN OFFICIAL STATISTICS

Geoffrey Brent
Methodology Transformation Branch

QUESTIONS FOR THE COMMITTEE

1. In cases where estimates of variance are not available, does the Committee have advice on whether variance should be approximated proportional to 1/cell magnitude, or as 1/magnitude^2? (Section 4.1)
2. Is it appropriate to use maximum likelihood estimation (MLE) methods to determine weighting parameters, as a substitute for subjective ratings of data accuracy? (Section 4)
3. Is it appropriate to use a MLE-oriented approach for balancing tables with a time series component, as an alternative to previously-published methods based on adding movement- and level-preservation objective functions? (Section 5)
4. Is the method described in Section 5.2 appropriate for reducing bias in forward-series estimates e.g. for quarterly National Accounts benchmarking?
5. Are the diagnostics identified for balancing appropriate, and are there others that should be considered? Is it possible to adapt leverage-type diagnostics for this application? (Section 6)
6. Are approximation/iteration approaches preferable to penalty function methods for dealing with nonlinear constraints/non-quadratic objective functions? (Section 7)
7. Are the proposed strategies for handling large problems appropriate, including rolling estimates subject to revision? (Section 8)
8. Does the Committee have any other advice on the methods discussed here?

CONTENTS

QUADRATIC OPTIMISATION FOR TABLE BALANCING IN OFFICIAL STATISTICS

Geoffrey Brent
Methodology Transformation Branch

ABSTRACT

National Accounts and other statistical outputs frequently require "balancing" or "benchmarking": adjusting estimates to satisfy internal consistency constraints and/or to reconcile information from multiple sources. Mathematical optimisation techniques such as weighted least squares balancing (WLS) are used for this purpose by other agencies and offer advantages over current ABS methods. The additional flexibility of WLS methods will improve the production and quality of National Accounts estimates, by making better use of data quality information and saving time currently spent in manual processes.

Ideally, weighting of adjustments would be based on the variance of the unbalanced estimate. When variance data is unavailable (e.g. non-survey sources), a fall-back method is required. This paper offers arguments for fall-back weighting according to the magnitude of unadjusted value (as opposed to squared magnitude).

The literature on WLS balancing has often approached these problems with the goal of minimising changes to levels and/ or changes to time series movements. This paper considers the problem of producing a maximum likelihood estimator (MLE) for the true data values, and shows that for certain important cases, WLS movement- or level-preservation is in fact equivalent to a MLE-based solution. It also offers a method, usable in both MLE- and movement-preservation frameworks, for reducing bias in forward-series estimates for balanced time series.

WLS balancing requires setting weighting parameters that indicate expected error behaviour. This has generally been done through numerical models which require parameters to be set subjectively, e.g. through estimation by subject matter experts. This paper suggests an alternative way for setting parameters: by framing balancing as a MLE problem, a MLE approach can be applied to estimating these parameters, reducing subjectivity and need for manual involvement in the process.

This paper then extends the MLE approach to give a WLS method covering cases where both movements and levels are of interest, and shows that a previously-published WLS method for handling such cases is *not* equivalent to the MLE solution.

This paper also discusses options for balancing diagnostics, for handling nonlinear constraints with software that only supports linear constraints, and for dealing with very large problems whose size may present computational difficulties.

1. INTRODUCTION

The Australian Bureau of Statistics has recently started a program of infrastructure renewal. This program includes updates to a range of ABS systems and processes, offering an opportunity for re-evaluation and improvement of these processes.

Several of these processes can be framed as numerical optimisation problems, e.g. making the most cost-effective use of ABS field interviewers, creating an efficient sample design that meets multiple objectives, or reconciling inconsistent data sources to create high quality estimates that combine the strengths of those sources.

The availability of high quality commercial and open source optimisation software on an off-the-shelf basis presents an opportunity for ABS to improve these processes while reducing maintenance and support burdens, and where possible bringing different applications into the same conceptual framework.

This paper concentrates on National Accounts applications, in particular balancing economic activity tables, because improving efficiency and output quality for economic statistics is a priority in ABS' current transformation plan. However, many of the issues discussed here will also be relevant to other ABS applications.

Major objectives of our work are:

- Increase versatility of National Accounts balancing systems, reducing the need for manual balancing and time taken to balance tables.
- Improve quality of outputs by applying consistent framework for initial balancing of tables, and making better use of knowledge about data sources.
- Share knowledge with other statistical institutions using similar methods and improve understanding of methods used by these institutions.

The success of optimisation methods depends on being able to design an "objective function" that adequately quantifies one's goals. This paper offers a theoretical basis for doing so, by reviewing currently-used implementations of optimisation and reinterpreting them as maximum likelihood estimators (MLEs).

This MLE interpretation offers a theoretical justification for some currently-used methods, and offers a more objective way to resolve some of the decisions made in their implementation e.g. setting weights for objective functions. Section 2 gives the context for the National Accounts balancing problem, and presents a simple example.

Section 3 presents a theoretical framework for basic balancing problems, showing an equivalence between WLS level-preservation and maximum likelihood estimation, and identifies limitations and assumptions of that framework.

Section 4 discusses methods used to set weights for WLS level preservation, and offers a MLE-based approach to reduce the need for subjective choices in this process.

Section 5 extends the MLE approach to time series considerations, in particular movement-preservation adjustments. It shows that simple movement -preservation objective functions can also be interpreted as MLEs.

Despite the MLE basis for level- and movement-preservation objective functions when taken individually, adding these objectives together to form a combined objective function cannot be easily justified from a MLE perspective. However, this paper identifies another option: adding the *error models* implicit in those objectives to create a combined error model, then constructing a new MLE/WLS objective function appropriate to that model.

Sections 6, 7, and 8 discuss related issues: diagnostics for optimisation balancing, methods for dealing with nonlinear constraints, and options for scaling methods to very large problems.

Methods used for this work should be defensible, minimising the use of subjective decision making. The complexity of solutions should be appropriate for maintenance by staff without a specialist background in optimisation.

2. CONCEPTUAL BACKGROUND

This section discusses the background to National Accounts balancing work, with further theoretical detail to be provided in subsequent sections.

2.1 Motivation: need for balancing in National Accounts

An important part of National Accounts production is compiling tables of balanced estimates for various kinds of economic activity. The relationships between these items of interest imply certain rules about how their values should relate to one another.

For example, table 2.1 shows an estimated breakdown of the value of goods and services supplied in a year by sectors of a fictional economy.[1] Note that some industries may be associated with more than one product, and vice versa: e.g. a business whose primary activity is telecommunications may also provide financial services as a secondary activity.

2.1 Supply of goods and services

SUPPLY		*Industry*								
		Iron Ore Mining	*Steel Mfg*	*Car Mfg*	*Tele Svcs*	*Fin Svcs*	*Oth Svcs*	*Imports*	*Other Supply*	*Total Supply*
Product	Iron Ore	400	0	0	0	0	0	0	50	**450**
	Steel	50	500	0	0	0	0	160	100	**810**
	Cars	0	0	700	0	0	0	1,000	300	**2,000**
	Tele Svcs	0	0	0	400	50	20	150	80	**700**
	Fin Svcs	0	0	0	20	1,100	50	200	150	**1,520**
	Oth Svcs	10	0	100	0	80	120	50	40	**400**
	Total Output	**460**	**500**	**800**	**420**	**1,230**	**190**	1,560	720	**5,880**

Table 2.2 shows the value of goods and services *used* by these same sectors, along with gross value added by each industry (GVA) and changes to inventories.

1 In practice, accounts tables are combined not only in values but also in volumes, and values may be assessed under more than one set of prices (see e.g. Nicolardi, 2011). For simplicity this discussion assumes that tables are to be balanced in values; volumes and alternate prices can be incorporated through ratio constraints.

2.2 Use of goods and services

USE		Industry									
		Iron Ore Mining	*Steel Mfg*	*Car Mfg*	*Tele Svcs*	*Fin Svcs*	*Oth Svcs*	*Exports*	*Invent-ories*	*Other Use*	*Total Use*
Product	Iron Ore	20	250	0	0	0	0	190	40	0	**500**
	Steel	0	0	350	0	0	70	150	30	0	**600**
	Cars	0	0	0	0	0	0	300	–100	1,800	**2,000**
	Tele Svcs	100	120	150	200	80	50	50	0	250	**1,000**
	Fin Svcs	50	60	110	180	450	20	100	0	500	**1,470**
	Oth Svcs	30	40	30	40	150	10	20	0	180	**500**
	Total Use	200	470	640	420	680	150	810	–30	2,730	**6,070**
	GVA	245	33	154	10	540	35				
	Total Input	**445**	**503**	**794**	**430**	**1,220**	**185**				

Together, these estimates of supply and use constitute a Supply-Use ("SU") table.

Economic definitions impose several constraints on the values for these items, e.g.:

- For any product in the economy, total supply must match ("balance") total use.
- For any industry in the economy, total input must match total output.
- Most items must be non-negative. Exceptions exist e.g. GVA and changes to inventories.
- Some items are conceptually zero, e.g. inventories for services; these are called "structural zeroes".
- Many other items are assumed small enough that they can be approximated as zero, even if non-zero values are possible (e.g. supply of wool by mining businesses); these are also considered structural zeroes.

Note that in the example shown above, the requirements on total supply/total use and total input/total output are *not* satisfied: e.g. total supply of steel is 810 but total use is only 600.

This example has been simplified for convenience. A real SU table for Australia contains 301 products and 67 industries, along with some extra terms e.g. taxes, subsidies, private financial and non-financial, household and government sectors, but the same kinds of requirements apply. In some cases it may be desirable to balance several years of data simultaneously; this will be discussed in more detail in Section 6.

Data for such tables are compiled from a variety of sources including government taxation records, ABS industry surveys, and import/export declarations. Many of these sources are subject to some degree of error, e.g. due to misreporting or sample-based estimation, and these errors lead to inconsistency within tables.

While some error in these tables is inevitable, inconsistency can create serious problems for analysis (as per the example, starting from the premise that 600=810, literally anything can be proven!)

Inconsistency also implies that the tables could be made more accurate through better use of the available data. Referring to the example above, the discrepancy between supply and use of steel indicates that some of the use estimates may be too low, and hence overall accuracy could be improved by increasing them, or alternately that some supply estimates are too high and should be reduced.

Major discrepancies may be best resolved by subject matter experts who can investigate causes and estimate an appropriate correction, but the size of these tables makes it time consuming to do all balancing through manual processes. This is further complicated by the multi-dimensional nature of such tables: an adjustment that balances a row may unbalance a column, and the balancing must reconcile three different measures of gross domestic product.

These factors create a requirement for an automated balancing process ("auto-balancing") that adjusts table entries to satisfy applicable constraints.

Usually there will be many different ways to balance the table. In the example above we could increase estimates for steel use items, decrease estimates for steel supply, or some combination of the two, so that total supply aligns with total use. We could even satisfy the constraints listed above by adjusting all values to zero, but this would not be a realistic solution.

The object then becomes to find a solution that balances the table, with the least possible disruption to the data overall; later sections will discuss how exactly to quantify overall disruption.

These tables are typically compiled from multiple sources that may differ greatly in their accuracy. For example, data on tax receipts can be obtained directly from taxation records with very high accuracy; imports and exports are also available from government data. Other items may be somewhat less accurate, having been obtained from sample surveys or similar sources. Furthermore, since most industries use a wide range of products but supply only a few, data on the Supply side will often be more accurate than that on the Use side.

A good balancing process would factor in these differences in accuracy, using weighting or other methods to encourage smaller adjustments for items known to be reliable.

Business needs may impose other requirements on the balancing process. For example, there may be a need to maintain consistency with a previous publication; e.g. if we have already published an estimate that the total value of financial services supplied is $1.25 billion, then we may require that the balanced cells are consistent with this total, while allowing individual cells to vary within that.

2.2 Related problems

The focus of this paper is on what are considered “balancing” problems, such as the provided Supply-Use example.

However, many of the same concepts and tools can be applied to other problems within National Accounts. An important example is the benchmarking/temporal disaggregation problem, where we have two sources of data relating to (e.g.) a quarterly time series for an industry of interest. The “benchmark” data gives very accurate totals at an annual level; the “indicator” series has larger errors but is known to correlate with the quarterly movements. These two sources can be combined to give an accurate quarterly series, by using annual benchmarks to calibrate the quarterly indicators (cf. Dagum and Cholette, 2006; International Monetary Fund, 2014).

There are also combination benchmarking/balancing problems, e.g. multivariate benchmarking where we seek to estimate multiple time series for various industries, requiring consistency with a quarterly-level benchmark across multiple industries as well as the annual industry-level benchmarks already discussed. Section 6.3 presents an example of such a problem.

Another case (not further discussed in this paper) is the need to publish estimates for supply and use of commodities in both current prices and volumes, with the requirement that both versions balance individually while preserving a relationship between them.

These problems can be expressed in the same way as those set out in Section 3: continuous optimisation problems with linear constraints and a quadratic objective function.

Similar problems also arise in official statistics outside National Accounts work. For example, calculation of Estimated Resident Population also requires adjusting fine level population estimates for consistency with known aggregates, and could be approached through the same methods. Although this paper focuses on National Accounts and specifically on table balancing applications, the principles discussed here should also be understood as relevant to those other applications.

Other important ABS optimisation problems require choosing solutions from a discrete set of options and are better approached as linear integer problems (LIPs), e.g. field interviewer allocation where the variables represent yes/no decisions. LIP methods are outside the scope of this paper, but in many cases the software tools that would be used for quadratic optimisation are also suitable for LIPs, and the theoretical frameworks are similar enough that familiarity with one class of problem will be beneficial when approaching the other.

2.3 Approaches to solving balancing problems

Lenzen *et al.* (2009) identify two major classes of methods used for balancing problems:

2.3.1 RAS balancing and entropy theory approaches

RAS[2] (aka "iterative proportional fitting", "matrix raking", "iterative scaling", "cross-entropy") relies on iteratively balancing subsets of the problem, or numerically equivalent techniques.

For example, in the balancing problem shown in Section 2.1, unbalanced Total Use for iron ore is 500, but Total Supply is 450. Taking the Total Supply data as authoritative, we would scale all Use items for iron ore by 450/500 = 0.9, ensuring that the Use total matches Supply. We would then repeat the process for other rows, e.g. for steel (supply=810, use=600) we multiply all use items by 810/600.

Next we would use the same process to balance columns. This unbalances rows, but the discrepancies will be less than in the original inputs. We then iterate row and column scaling until the data converges to a balanced table.

Currently ABS uses manual adjustment for Supply-Use tables, with RAS then used to balance finer level Input-Output data to SU benchmarks. A separate implementation of RAS is used elsewhere in ABS for calculation of Estimated Resident Population.

Naïve RAS methods are relatively simple to implement and their computational requirements are modest, which can be important when balancing large data sets. However, they have some important limitations:

- Zero values will never be adjusted to non-zero, even when this might be appropriate, and sign changes are impossible even when this might be appropriate (Lenzen *et al.*, 2014).
- RAS does not converge when constraints based on primary data are conflicting.
- Missing values need to be replaced by non-missing before balancing.

---

2 Not an acronym. Named for Stone's notation $\hat{r}A\hat{s}$; see Bacharach (1965).

- When a row contains both positive and negative values, RAS may produce unnecessarily large adjustments. For example, suppose a given item has unbalanced use of $110m and change in inventories of –$100m (adding to $10m Total Use) but is required to match a figure of $30m Total Supply. In this case RAS would triple the Use and Inventories items, resulting in adjustments of 200% (before any changes caused by rebalancing columns). A more sensible solution might be e.g. to adjust use to $120m and inventories to –$90m.
- Naïve RAS does not consider information about data reliability and cannot handle conflicting external data.

Various refinements have been proposed that address these issues (see e.g. Lenzen *et al*., 2006, 2009, 2012a, 2012b, 2013, 2014). However, these methods increase the complexity of a RAS-based solution and are likely to require a custom-built solver.

Various other approaches to balancing are based on concepts of information-theoretical entropy, e.g. minimising information loss as measured by a Kullback-Leibler divergence function. McDougall (1999) notes that RAS is equivalent to minimisation of cross-entropy, although some authors have not been aware of this relationship.

Detailed discussion of entropy-based approaches is out of scope for this paper. Without attempting to compare their theoretical properties to the WLS/MLE approach outlined below and in Section 3, it seems likely that WLS/MLE would be significantly easier to implement at ABS for practical reasons. In particular, the concepts involved in WLS/MLE will be more familiar to ABS staff and more easily explained, and in some cases have strong parallels to existing ABS processes.

2.3.2 Optimisation balancing

Subject to these limitations, RAS generally achieves balanced outputs with reasonably small adjustments, but these may not be the smallest possible. An alternative to RAS-type balancing is a more exact optimisation approach: we explicitly define the quality of potential solutions as a function of the adjusted values ("objective function"), and then use numerical methods to optimise on that objective function[3].

For example, in the Supply-Use scenario presented in Section 2.1 above, we might define the objective function as a weighted sum of squares of the adjustments to individual cells. Linear algebra methods can then be used to find a solution that minimises this sum of squares, subject to the specified constraints.

3 RAS and entropy methods can also be framed as optimisation, but here I follow Lenzen *et al.* in making a distinction between such methods and those that support a more customisable objective function.

Optimisation methods have several strengths:

- Optimisation methods do not assume any particular structure to the data to be balanced; rather, structure is specified by individual constraints. This allows flexibility in what sort of constraints can be imposed; for example, tables can easily be balanced in more than two dimensions, a key requirement for National Accounts.
- The objective function can incorporate information about accuracy of contributing estimates: e.g. cells known to be accurate can be weighted heavily in the objective function, discouraging large adjustments to these cells. Where appropriate, values can be fixed to zero adjustment.
- Negative and zero values do not require special treatment.
- Missing values and their weights would ideally be modelled or otherwise imputed as per the literature on missing data methods, but in cases where it is not possible to do so, they can be accommodated by deleting the corresponding terms from the objective function. Under the MLE framework discussed in Section 3, this is equivalent to assuming an arbitrarily-chosen unbalanced value with an infinitely large variance (i.e. a flat prior).
- Quadratic optimisation problems are important in a wide range of commercial applications (e.g. costs in electricity transmission networks can be described by a weighted least squares objective function) so many commercial and open-source products have been developed for solving them, and a great deal of research has gone into developing efficient methods for solving large problems quickly.[4]

The history of optimisation balancing goes back to the description of a weighted least squares framework by Stone *et al*. (1942). At the time and for many years afterwards, lack of computing power made this approach infeasible; Lenzen *et al.* (2009) note that optimisation methods have high computing requirements and programming complexity when compared to RAS.

More recently, improvements in computing power and off-the-shelf availability of powerful general purpose constrained optimisation solvers have brought these methods within reach. For example, the U.S. Bureau of Economic Analysis now uses CPLEX software for reconciliation and balancing work (Rassier *et al.,* 2007a and 2007b) and Statistics Netherlands uses Xpress-MP for the same purpose. The Australian Bureau of Statistics and other national statistical organisations are currently evaluating and developing optimisation methods.

4 In the related field of mixed integer programming, Koch *et al.* (2011) report that between 1990 and 2010, algorithm advances gave a 55,000× improvement in solution speed.

3. THEORETICAL FRAMEWORK

This section sets out a framework for mathematical discussion of optimisation balancing.

3.1 Derivation of weighted least squares objective function

It is well known (see e.g. Bradley, 2009) that weighted least squares estimators for a regression model are equivalent to maximum likelihood estimators ("MLEs") when certain conditions are satisfied. A similar result applies in balancing-type problems:[5]

Let $\underline{x}$ be a vector $\left(x_1, x_2, \ldots, x_n\right)^t$ containing the true (possibly unobservable) values for quantities to be estimated in a table, with n being the length of $\underline{x}$.

Considerations discussed in Section 2 place some constraints on the possible values of $\underline{x}$. For the time being, we assume that these constraints are of two kinds:

1. Linear equality constraints e.g. Supply=Use, structural zeroes, or other fixed values/subtotals within the table.
2. Linear inequality constraints e.g. values that cannot be negative.

These constraints can be collectively expressed in matrix-vector form as:

$$A^{EQ}\underline{x} = \underline{c}^{EQ}$$

$$A^{INEQ}\underline{x} \geq \underline{c}^{INEQ}$$

where A^{EQ} and A^{INEQ} are known matrices of dimensions $k^{EQ} \times n$ and $k^{INEQ} \times n$ respectively, and $\underline{c}^{EQ}$ and $\underline{c}^{INEQ}$ are known constant vectors of length k^{EQ} and k^{INEQ} respectively, with k^{EQ} and k^{INEQ} being the number of equality and inequality constraints. The "feasible region" is the set of all possible solutions for $\underline{x}$ that would satisfy these constraints.

From some source we obtain "unbalanced" estimates $\hat{\underline{x}}$ that approximate the true series, but with measurement errors:

$$\hat{\underline{x}} = \underline{x} + \underline{\varepsilon}$$

$\underline{\varepsilon} = \left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\right)^t$ with estimation errors $\{\varepsilon_i\}$ being random variables. In general, these errors mean that $\hat{\underline{x}}$ will not satisfy the balancing constraints.

5 Stone *et al*. (1942) observed that a weighted average of GDP measures satisfies the principle of maximum likelihood, but there is little if any mention of this MLE/WLS connection in more recent balancing literature.

If we have information about the joint distribution of $\{\varepsilon_i\}$, we can use this to produce a balanced estimate $\underline{\tilde{x}}$ that is close to $\underline{\hat{x}}$ but satisfies the constraints $A^{EQ}\underline{\tilde{x}} = \underline{c}^{EQ}$ and $A^{INEQ}\underline{\tilde{x}} \geq \underline{c}^{INEQ}$. This can be described as adjusting the unbalanced values, with the vector of adjustments[6] being $\underline{r} = \underline{\tilde{x}} - \underline{\hat{x}}$, but it is also useful to think of it in terms of producing an estimate of the errors, $\underline{\tilde{\varepsilon}} = \underline{\hat{x}} - \underline{\tilde{x}}$. Within this section $\underline{r} = -\underline{\tilde{\varepsilon}}$, but in some cases (see e.g. Section 6) we may have additional error terms, in which case this identity may not hold.

If we assume that $\{\varepsilon_i\}$ are independent normally distributed variables having distributions $\varepsilon_i = N\left(0, \sigma_i^2\right)$ it follows that for any possible choice of $\underline{\tilde{\varepsilon}}$, the prior probability density for $\underline{\varepsilon} = \underline{\tilde{\varepsilon}}$ is equal to:

$$f\left(\underline{\tilde{\varepsilon}}\right) = \prod_i \left(\frac{\exp\left(-\tilde{\varepsilon}_i^2 / 2\sigma_i^2\right)}{\sigma_i \sqrt{2\pi}} \right).$$

We can then select $\underline{\tilde{x}}$ through a maximum likelihood estimation approach, choosing the value from the feasible region that maximises this probability.

Since ln is a monotonically increasing function, maximising *f* is equivalent to minimising a negative log-transform:

$$-\ln\left(f\left(\underline{\tilde{\varepsilon}}\right)\right) = \frac{1}{2}\sum_i \frac{\tilde{\varepsilon}_i^2}{\sigma_i^2} + \sum_i \ln\left(\sigma_i \sqrt{2\pi}\right).$$

The second term in this expression is constant, so this is the same as minimising the quadratic function $\sum_i \tilde{\varepsilon}_i^2 w_i$ i.e. a weighted sum of squared adjustments, where the weight w_i for a given cell is the reciprocal of its variance: $w_i = 1/\sigma_i^2$.

We will refer to functions of this type as *objective functions* or "OFs". Later we will examine other objective functions and their corresponding error models. For clarity this particular function will be referred to as OF_{Lvl} since its form emphasises preservation of levels:

$$OF_{Lvl} = \sum_i w_i \left(\tilde{x}_{i,t} - \hat{x}_{i,t}\right)^2 \quad \Leftrightarrow \quad \hat{x}_i = x_i + \varepsilon_i,\ \varepsilon_i = N\left(0, \sigma_i^2\right),\ w_i = 1\big/\hat{\sigma}_i^2.$$

OF_{Lvl} can also be expressed in matrix form: $OF_{Lvl} = \underline{\tilde{\varepsilon}}' W \underline{\tilde{\varepsilon}}$ where $W = \mathrm{diag}\left(\underline{w}\right)$.

6 Strictly speaking, it would be more accurate to consider these as *residuals*, to encompass e.g. cases where $\hat{x}$ includes two independent measurements of the same x_i. For simplicity the notation here assumes a one-to-one correspondence between elements of $\underline{\hat{x}}$ and of $\underline{x}$, but the method can easily be adapted for other possibilities.

This formulation has some desirable characteristics:

- It is mathematically convenient. In cases where the inequality constraints are ignorable, the problem can be reduced to a closed form matrix inversion. In cases where this solution would violate inequality constraints, a solution can be obtained by iterative application of closed form methods, usually requiring only a few iterations. Solving the problem can be computationally expensive, when a large number of variables are involved, but various commercial and free software products are designed for this purpose.
- It is easy to interpret and supports fine level examination of balancing results. Because the OF breaks down into a sum of terms that each relate to a single adjustment, it can be understood as assigning a cost to each individual adjustment. These costs can then be used for fine level diagnostics, to be discussed in Section 4.
- It is flexible. For example, if the unbalanced estimate for one of the x-variables is missing and we have no good basis for imputing it, this can be handled by setting the corresponding term in the objective function to zero (equivalent to having an infinitely large variance). If multiple independent sources are available for the same x-variable, the objective function can include a term for each of them.
- Quadratic optimisation problems with linear constraints are a very important class of problem in operations research, so powerful tools have been developed to solve them.

Some implications of the assumptions made in the above derivation:

- Assumption of non-biased errors: where possible, known sources of systematic error (bias) should be corrected before balancing, to satisfy the assumption that errors have expected value zero. Eurostat recommends that even though optimisation balancing is able to eliminate large discrepancies, these should instead be addressed and resolved by specialists. This becomes particularly important when using methods that assume small adjustments relative to initial values, as discussed in Sections 6 and 7.
- Assumption of independence: inputs should be examined for relationships likely to lead to correlation between error terms, and ramifications of such correlation should be considered (see Section 3.2 below).

- Assumption of normality: in cases where non-normally distributed errors are expected, the WLS estimate may not be equivalent to a MLE. This may arise e.g. where large outliers make Central Limit Theorem approximations unreliable. A least squares framework is mathematically and computationally convenient; simulation studies may be desirable in order to assess how sensitive WLS balancing methods are to non-normality.
- Flat prior assumption (implicit in the choice of a MLE): in cases where we have strong information about expected distributions of true values (other than that observed in $\hat{x}$), it may be possible to produce a better anterior estimate by modifying the objective function according to that prior. Such methods are outside the scope of this paper.

3.2 Covariance between cells

The discussion above assumes that $\{\varepsilon_i\}$ are independent. In some cases this assumption does not hold, in particular:

1. Cases where some dimensions of $\underline{\hat{x}}$ are derived from others without being measured separately: e.g. the table includes unbalanced estimates for "use of steel" by industry, and the "total use of steel" item has been calculated as the sum of these other estimates. In such cases, the derived $\hat{x}_i$ and corresponding ε_i will be completely dependent on the dimensions they are derived from. (In other words, the mechanism that produces these data automatically enforces an equality constraint.)
2. Cases where dimensions are measured separately, but those measurements are not completely independent, due to effects in the measurement process, leading to some correlation in the $\{\varepsilon_i\}$.

Note that a balancing constraint that applies to the true values does not automatically create dependence in the unbalanced estimates. For example, we might have some data source on total use of steel that is independent of the sources used to compile the use-by-industry estimates.

Here we assume that it is possible to divide an unbalanced table up into two groups of cells having the following properties:

- "basic cells": cells representing (approximately) independent measurements, whose ε_i are hence independent of one another.
- "derived cells": cells whose unbalanced values and corresponding ε_i are completely determined by the basic cells and their corresponding ε_i.

(For the time being, we ignore the more complex case of weaker but non-zero dependence.)

Since the values for basic cells completely determine the values of derived cells, we obtain the correct maximum likelihood objective function by weighting basic cells according to their variance, and weighting all derived cells at zero.

If we ignore these dependencies, e.g. by weighting all cells (both basic and derived) at 1/ variance, the objective function is effectively over-weighting the data that contributes to derived cells. This can lead to sub-optimal adjustments, and would mean that the results of the balancing can be changed by adding extra sub-total rows or columns, even though these contain no new information.

As a simple illustration, suppose that we expand the Use table presented in 2.1 to include a new "manufacturing industries subtotal" column (table 3.1) with unbalanced values derived as the sum of unbalanced steel and car manufacturing activity:

3.1 Supply of goods and services, with added "subtotal" column

SUPPLY		*Industry*									
		Iron Ore Mining	*Steel Mfg*	*Car Mfg*	***Mfg subtotal***	*Tele Svcs*	*Fin Svcs*	*Oth Svcs*	*Imports*	*Other Supply*	*Total Supply*
Product	Iron Ore	400	0	0	**0**	0	0	0	0	50	**450**
	Steel	50	500	0	**500**	0	0	0	160	100	**810**
	Cars	0	0	700	**700**	0	0	0	1,000	300	**2,000**
	Tele Svcs	0	0	0	**0**	400	50	20	150	80	**700**
	Fin Svcs	0	0	0	**0**	20	1,100	50	200	150	**1,520**
	Oth Svcs	10	0	100	**100**	0	80	120	50	40	**400**
	Total Output	**460**	**500**	**800**	**1,300**	**420**	**1,230**	**190**	1,560	720	**5,880**

Previously the objective function included a single term for each item created by manufacturing industries. Showing only the terms related to supply for manufacturing industries:

$$OF_{Supply|Mfg} = \frac{\left(\tilde{x}_{Steel|SteelMFG} - 500\right)^2}{\sigma^2_{Steel|SteelMFG}} + \frac{\left(\tilde{x}_{Cars|CarMFG} - 700\right)^2}{\sigma^2_{Cars|CarMFG}} + \frac{\left(\tilde{x}_{OthSvcs|CarMFG} - 100\right)^2}{\sigma^2_{OthSvcs|CarMFG}}.$$

But if we include the “manufacturing industries subtotal” entries in the objective function without regard for covariance, this becomes:

$$\begin{aligned} OF_{Supply|Mfg} &= \frac{\left(\tilde{x}_{Steel|SteelMFG} - 500\right)^2}{\sigma^2_{Steel|SteelMFG}} + \frac{\left(\tilde{x}_{Steel|MFGSubtotal} - 500\right)^2}{\sigma^2_{Steel|MFGSubtotal}} \\ &+ \frac{\left(\tilde{x}_{Cars|CarMFG} - 700\right)^2}{\sigma^2_{Cars|CarMFG}} + \frac{\left(\tilde{x}_{Cars|CarMFGSubtotal} - 700\right)^2}{\sigma^2_{Cars|MFGSubtotal}} \\ &+ \frac{\left(\tilde{x}_{OthSvcs|CarMFG} - 100\right)^2}{\sigma^2_{OthSvcs|CarMFG}} + \frac{\left(\tilde{x}_{OthSvcs|MFGSubtotal} - 100\right)^2}{\sigma^2_{OthSvcs|MFGSubtotal}}. \end{aligned}$$

However, the derivation of the subtotal means that $\tilde{x}_{Steel|SteelMFG} = \tilde{x}_{Steel|MFGSubtotal}$, $\sigma^2_{Steel|SteelMFG} = \sigma^2_{Steel|MFGSubtotal}$ and so forth. Therefore:

$$OF_{Supply|Mfg} = 2\frac{\left(\tilde{x}_{Steel|SteelMFG} - 500\right)^2}{\sigma^2_{Steel|SteelMFG}} + 2\frac{\left(\tilde{x}_{Cars|CarMFG} - 700\right)^2}{\sigma^2_{Cars|CarMFG}} + 2\frac{\left(\tilde{x}_{OthSvcs|CarMFG} - 100\right)^2}{\sigma^2_{OthSvcs|CarMFG}}.$$

Hence, the effect of including this subtotal in the objective function is equivalent to doubling the weights on each item supplied by manufacturing industries. The effect will be to reduce adjustments to Supply-Manufacturing items, while increasing adjustments elsewhere, resulting in output that is no longer a maximum likelihood estimate.

The effects will be more complex in cases where more than one cell contributes non-zero values to a subtotal (e.g. a “services industries” subtotal in the above example) but in general, weighting derived cells such as subtotals has the effect of reducing adjustments to the cells that contribute to those subtotals, causing errors in the balancing.

If all cells are affected equally, these errors may largely cancel out, but this will not always be the case. For example, in the table presented in 2.1, Use-Steel-Car Manufacturing contributes to four different totals (Total Use for product, Total Use for industry, Total Input for industry, Total Use for all products and industries) but Use-Iron Ore-Exports contributes to only three totals (no Total Input). This effectively over-weights the Car Manufacturing data at the expense of Exports data, leading to larger-than-optimal adjustments in Exports.

Hence, when setting weights we should distinguish between basic and derived cells, and set zero weights for derived cells. (The weights on their contributing cells will still act to limit adjustments to derived cells.)

The second type of dependence can arise due to mechanisms such as sample rotation and selection methods.

For example, the ABS' Economic Activity Survey (EAS) is an important contributor to industry supply and use data. EAS is an annual survey, and the sample selection is designed with the objective that where possible, businesses selected for the survey will remain in sample for three years before rotating out. In cases where we're balancing multiple years of data simultaneously, this could lead to correlation between sampling errors for the same industry and product.

However, this problem is partly mitigated by EAS' use of regression weighting that reduces the impact of sampling error, and previous ABS investigations have found that the sample rotation strategy causes only a small covariance in year-to-year estimates.

Sampling issues can also cause covariance within a single year's data. For example, a single chemical manufacturing classification may cover several different specialties, each with different inputs and outputs. This could lead to covariance between some supply and use items within this industry.

In theory, given full data on variance and covariance of the epsilon terms, the MLE approach discussed above might be generalised to produce estimates based on the exact likelihood for a given estimate. However, this would be a very complex approach requiring a large number of parameters to be estimated.[7] As discussed in Section 5 it is difficult to obtain good data even for the variance of individual items. Quantifying interactions would likely be considerably more difficult; without good quality inputs it is unlikely that the complexity of such an approach would be justified.

As a general rule it seems reasonable to assume that such covariance between basic cells will be small relative to the variances of those cells, and hence can be ignored. Section 6 addresses an important exception, the case of balancing a time series where movement preservation is required.

7 For a single-year Supply-Use table with ~10,000 non-zero terms, the variance/covariance matrix would contain approximately 50,000,000 unique parameters. Even this would not be enough to calculate the likelihood for a given estimate, because pairwise covariances don't contain enough information to fully define a multivariate distribution.

3.3 Modifications to weighting

The MLE framework described above is based solely on the (estimated) error function. In some cases it may be appropriate to modify weights based on other factors. In particular, if one particular group of cells within the table is unusually important (e.g. crucial to policy-making) it might be desirable to improve the accuracy for these cells, even at the cost of reduced accuracy elsewhere in the table[8].

This can be achieved by fixing their values, increasing their weights, and/or restricting adjustments to these cells. However, excessively strict weighting/restriction may actually worsen accuracy for these cells. Balancing combines direct information about a cell (its unbalanced estimate) with indirect information (the rest of the unbalanced table); putting too much priority on the direct information risks losing accuracy through ignoring indirect information. Simulation studies may be required to determine where over-weighting becomes counter-productive.

Even where weights are adjusted to prioritise specific parts of the table, it's desirable to be transparent about the distinction between weighting based on source accuracy and weighting based on output priorities. This can be achieved by defining weights relative to the "pure MLE" values. Using MLE balancing for a first cut solution also provides a starting point for discussion about output priorities.

8 In general, important data items will be prioritised in data collection, so we may expect that these will have high accuracy and hence large weights even under pure MLE weighting, but this depends on the nature of the inputs.

4. WEIGHTING STRATEGIES

Under the restrictions discussed in Section 3, we achieve a maximum likelihood estimate by setting cell weights equal to (or proportional to) 1/variance. When estimates of variance are available (e.g. data item comes from a sample survey amenable to standard methods of variance estimation) this provides a straightforward rule for setting weights.

However, many items in National Accounts tables come from other sources, e.g. administrative data, which do not provide error estimates. This requires a fall-back weighting strategy for use in such cases, i.e. a method for producing an approximate estimate of variance.

Information that could be used for this purpose includes:

- Magnitudes of unbalanced estimates;
- Expert judgement on accuracy;
- Cell type e.g. tax data vs. supply vs. intermediate use, how collected.

Since expert opinion is often based on cell type, these two may be considered together: e.g. subject matter experts may give an accuracy rating for each type of cell.

A simple approach to weighting can be expressed as:

$$\frac{1}{w_i} = k \cdot m_i \cdot \tau_i \,.$$

Here w is the weight assigned to a cell, m is a function of cell magnitude (estimated from unbalanced magnitude, or other sources – generally a power of the magnitude), τ is a function of cell type/expert judgement, and k is a constant, with all terms positive. The product $km_i\tau_i$ can be taken as an estimate of variance for cell *i.*

If all weights are set through this method, then the choice of k has no effect on the solution, and it can be set to 1. If some are set through other methods e.g. for data sources with known variance, then k can be used to calibrate the relative weights for these different methods. The methods used at Statistics Netherlands and the U.S. Bureau of Economic Analysis are along these lines, with some differences in the formulation (see Eurostat, 2013 and Chen, 2012).

4.1 Weighting by magnitude

We considered two candidates for this term:

Method 1: $m_i = \hat{x}_i^2$

Method 2: $m_i = |\hat{x}_i|$

Method 1 is equivalent to assuming that cells of the same type have similar coefficients of variance, with standard errors are in proportion to the magnitude of the cell. In the single-constraint balancing scenario discussed in 6.2, this results in adjustments proportional to the square of cell magnitude.

Method 2 is equivalent to assuming that standard errors vary in proportion to the square-root of magnitude (hence, disregarding differences in cell type, larger cells have larger SEs but smaller RSEs). In the single-constraint scenario, this results in adjustments proportional to the unbalanced cell magnitude. If all cells have the same sign, this is effectively proration.

Statistics Netherlands originally used Method 1 as discussed in Eurostat (2013) but now also uses Method 2 (Dr. Reinier Bikker, personal communication). The BEA has used both approaches in different parts of their balancing work (cf. Chen, 2012; Chen *et al.,* 2014). Di Fonzo and Marini (2009) discuss both options and favour Method 1. Fortier and Quenneville (2009) also discuss both, and state that Statistics Canada's TSRAKING reconciliation procedure uses Method 2.

Some arguments in favour of Method 2:

- In economic outputs, larger values tend to be of greater importance and so are collected with smaller RSEs. For example, ABS industrial collections prioritise sample allocation to give high accuracy for major industries such as mining.
- In non-economic contexts, while smaller values may often be of great interest, these tend to be associated with larger RSEs. For example, in estimating counts of various subpopulations from a simple random sample, the numbers counted will follow a Poisson distribution with variance for each subpopulation proportional to its total numbers.
- Proration is a familiar concept to many users of National Accounts data, and Method 2 can be understood as similar in effect to pro-ration.
- Under the assumption of independent errors, if we aggregate two or more cells to form a single total, the variance of their sum should match the sum of their variances. Method 2 is consistent with this (assuming all aggregated items have the same sign, which would usually be the case); Method 1 is not.

- This is particularly important in cases where we need to balance the same data under different aggregations: e.g. a change to industry classifications may require producing tables for the same reference period under both old and new classifications, or considerations of problem size may require balancing a table at a broad classification before rebalancing at finer levels. In these cases, Method 2 is likely to give better consistency between outputs; for the single-constraint scenario it gives perfect consistency.
- Even in the single-constraint scenario, Method 1 can produce paradoxical results that greatly distort relativities between values in the same constraint group.
  - Consider a table with 100 entries of unbalanced value $10m each, and one entry of unbalanced value $100m (i.e. unbalanced total $1100m), adjusted to match a total of $900m. Under Method 1, the small entries would each be adjusted by –10% to $9m, but the large entry would be adjusted by –100% to $0. Under Method 2, all entries are adjusted by –18%, preserving relative sizes. Fortier and Quenneville (2009) describe what appears to be a similar problem with simulated weekly data under Method 1.
  - As well as being implausible, adjusting values to zero may increase computational difficulty of the problem by activating inequality constraints; a method that approximates pro-ration is much less likely to do this.

However, given that Method 1 is commonly used, ABS should conduct further investigation to determine which method is preferable, e.g. by comparing behaviour of the two on test cases and by further research into the reasons behind other agencies' choices.

4.1.1 Graphical comparison of weighting methods

As a preliminary investigation, our team experimented with balancing actual Australian Supply-Use tables using the two methods discussed above.

This investigation used the 2010–2011 Supply-Use table, taking a version balanced by the current RAS-based system and also an unbalanced version from before the RAS process. Subject matter experts gave subjective reliability ratings for different groups of data within the unbalanced table: e.g. Intermediate Use items were typically rated as "40% reliable", most Supply items were rated as "80% reliable", and export/import and household and government final consumption expenditure items were rated as "90% reliable".[9]

9 These ratings were made on an ad-hoc basis, to allow us to experiment with different balancing methods. They are not intended as a precise assessment of data quality for these different items and should not be interpreted as such.

Cells were then weighted at:

Method 1: $$w_i = \frac{1}{\hat{x}_i^2 (1 - r_i)}$$

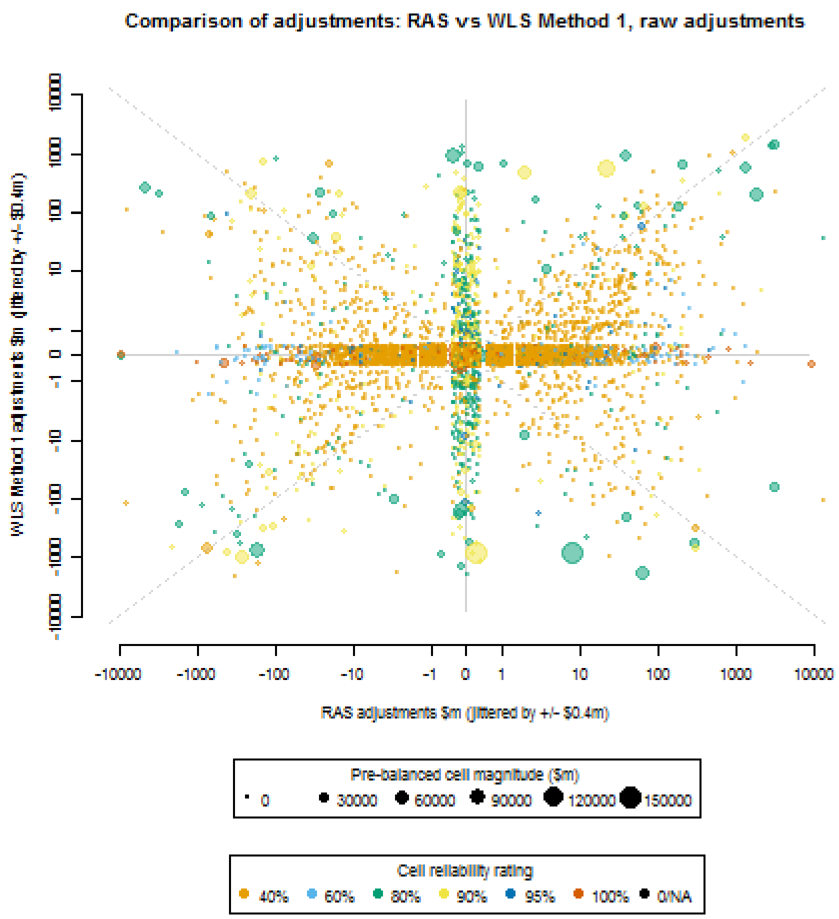
Method 2: $$w_i = \frac{1}{|\hat{x}_i| (1 - r_i)}$$

where r_i is the reliability rating discussed above: hence e.g. a cell with reliability rating of 0.9 would have twice the weight of another cell with reliability 0.8, if their unbalanced magnitudes are equal. (This is not the only way to translate such ratings into weights, and not necessarily the best; the object here was to explore the ability of WLS balancing to incorporate reliability measures, rather than to set those measures.) Reliability of 100% is interpreted as a non-adjustable cell.

We then balanced the 2010–11 data via WLS, using Method 1 and 2 weights. For these balanced tables and for the RAS output, we calculated adjustments for each cell.

Figure 4.1 shows a comparison of the adjustments for RAS and for WLS Method 1.

4.1 Comparison of raw adjustments, RAS vs WLS Method 1



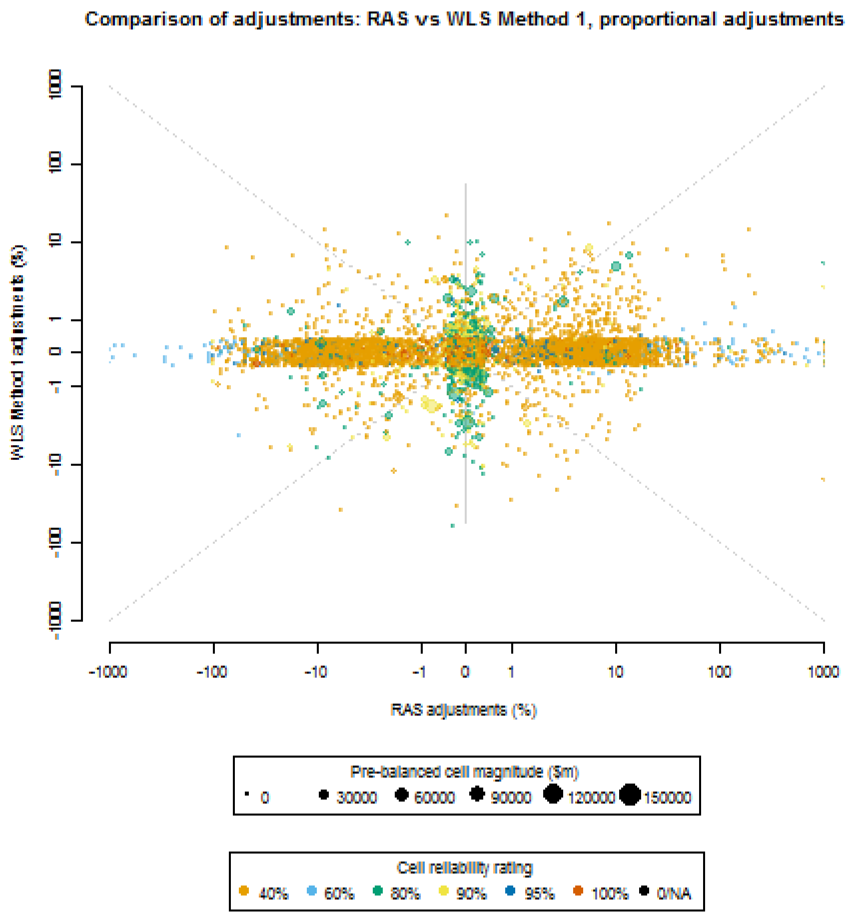
In this plot, each dot represents a single cell in the table. The size of the dot indicates its unadjusted magnitude, the colour indicates its reliability rating, and its location indicates the adjustments it received: the x-coordinate shows the adjustment under RAS, and the y-coordinate shows the adjustment under WLS Method 1. Cells receiving no adjustment, and total/subtotal items, are omitted.

Balancing outputs were rounded to multiples of $1m, with the result that many points would be plotted on top of one another; in these graphics those values have been randomly jittered by up to $0.4m to separate such points and help in visualisation. Both axes use a double-sided log scale.

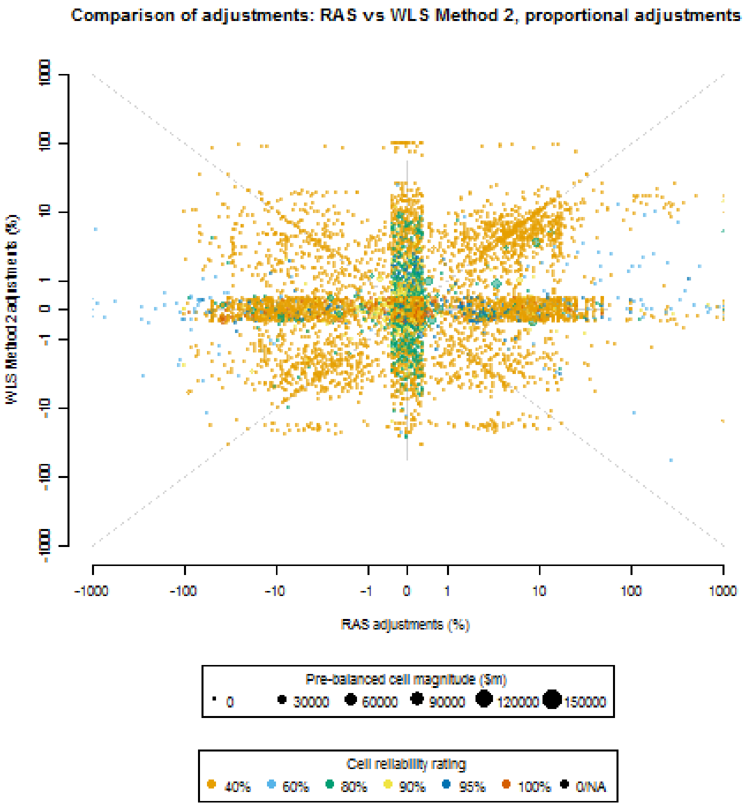
Points in the top and bottom quadrants of the plot are those where WLS Method 1 produced larger adjustments than RAS, and vice versa for those in the left and right quadrants.

Adjustments can also be represented as a proportion of unadjusted cell magnitudes:

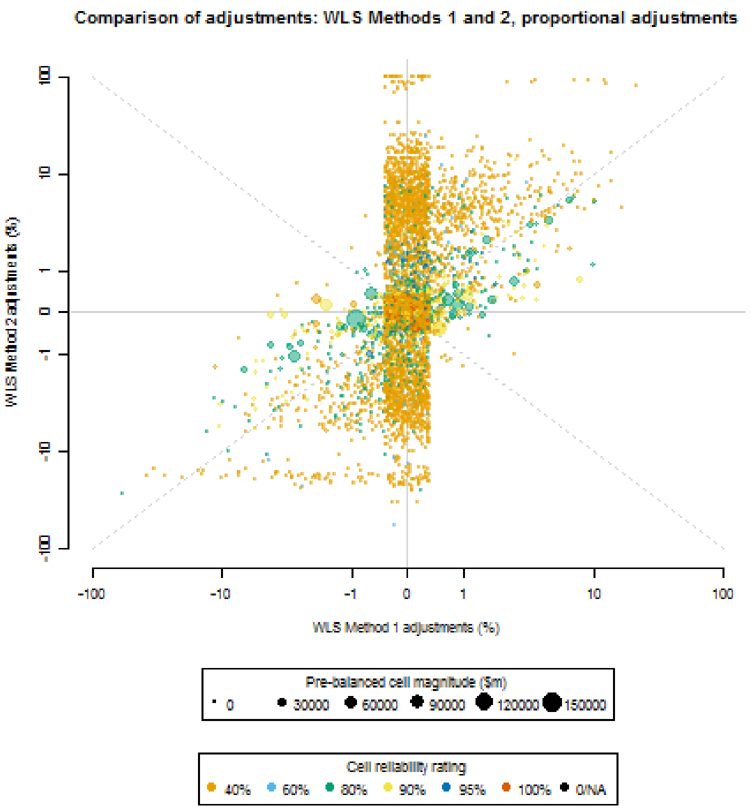
4.2 Comparison of proportional adjustments, RAS vs WLS Method 1



4.3 Comparison of proportional adjustments, RAS vs WLS Method 2



4.4 Comparison of proportional adjustments, WLS Method 1 vs Method 2



Some points of note:

- Large cells often get large adjustments in dollar terms, but in proportional terms their adjustments are usually small.
- Method 2 tends to make smaller adjustments to large cells, but larger adjustments to small cells, compared to Method 1 (as would be expected from theory)
- Individual cells may receive larger adjustments under WLS balancing, but overall WLS (both methods) tend to produce smaller adjustments than RAS, especially for those cells rated at 90–95% reliability.
- This is especially noticeable when looking at the largest adjustments: e.g. WLS rarely adjusts even a low reliability cell by more than ~20%, whereas such adjustments are much more common under RAS.
- WLS and RAS often adjust the same cell in different directions (i.e. points in top-left/bottom-right quadrants); the two WLS methods almost always adjust in the same direction, although the size of adjustment may be very different.

Appendix A shows RAS vs WLS adjustments, split by cell reliability. In general, even at low reliability, WLS adjustments tend to be smaller than RAS.

4.1.2 Special cases for magnitude

In some cases, the unbalanced cell magnitude may not be a good proxy for variance. This is particularly a concern with "net" items that have been estimated as the difference of two large positive quantities. In this case, the standard error may be very large compared to non-"net" items of similar magnitude.

Many of these "net" items will fall under the "dependent by definition"/"derived cells" case discussed in 3.2, in which case their weight can simply be set to zero. If this is not the case (e.g. the data used to derive the "net" item doesn't appear separately in the table) then it may be necessary to consider a substitute for magnitude, e.g. choose a cell expected to have comparable magnitude to those contributing to the net.

There may also be cases where a data item is believed to be small but non-zero, but has not been measured (e.g. for reasons of collection cost or respondent burden) and hence the unbalanced value is zero. Normally a zero input would be treated as non-adjustable (estimated zero standard error). In such cases, we might be willing to believe a small adjustment from zero—but in practice, it seems better to keep them fixed at zero, since we have no defensible basis for estimating the adjustment, and allowing a small adjustment will have little effect on the balancing.

4.2 Expert-guided weighting

Even where no rigorous estimates of standard errors are available, National Accounts subject matter experts have a good understanding of the data sources feeding into these tables, and can provide qualitative information about their expected reliability. Although this introduces a subjective element to the weighting, it seems far better than ignoring this knowledge and basing weighting only on magnitude.

Making use of this knowledge requires translating a qualitative assessment into a numerical estimate of relative accuracy. We consider three options, of the sort that have previously been used by Statistics Netherlands and the BEA:

- Subject matter experts estimate the standard error for each type of cell, relative to other cell types (e.g. a supply cell has $\sim 5\times$ the standard error of a tax cell of the same magnitude).
- Subject matter experts estimate the magnitude of reasonable adjustments for each type of cell, relative to other cell types (e.g. we expect supply cells to receive $\sim 10\times$ the adjustment of tax cells for the same magnitude).
- Subject matter experts rate cells on a simple scale (e.g. $1 =$ least reliable to $10 =$ most reliable).

Ratings from the first of these methods can be translated to weights by setting τ_i such that $1/\tau_i$ varies in proportion to the square of the estimated SEs, relative to other cell types at the same magnitude.

For the second method, the single-constraint scenario suggests that it would be reasonable to set τ_i such that $1/\tau_i$ varies in proportion to the relative expected adjustment size.

For the third method, Statistics Netherlands convert the integer rating to a weight by an exponential translation.

The first two methods have the advantage of making a quantitative prediction about the data. These predictions could be compared to outcomes as a quality check, and can also be used to choose an appropriate value for *k* so that these cells are not over- or under-weighted relative to those whose SE is derived from other methods.

However, the third method may be simpler to apply, and makes the subjective nature of the data more visible.

4.3 MLE weighting for unknown errors

As an alternative to the methods discussed in Section 4.2, it may be possible to increase automation and reduce subjectivity of the weighting process by using maximum likelihood estimation to determine appropriate values for τ_i.

For example, suppose that the cells in our table for which standard errors are not available can be divided into sets *A*, *B*, and *C*, with the expectation that cells within the same sets will have similar levels of error. We then restrict τ_i to the form:

$$\tau_i = \begin{cases} \kappa_A, \text{if } i \in A \\ \kappa_B, \text{if } i \in B \\ \kappa_C, \text{if } i \in C \end{cases}$$

for some parameters $\kappa_A, \kappa_B, \kappa_C$.

For any given values of κ_A, κ_B, and κ_C, we can find the WLS/MLE solution for $\underline{\tilde{x}}$, and calculate the likelihood associated with that particular estimate; denote this as $f\left(\kappa_A, \kappa_B, \kappa_C\right)$. We may then optimise over possible values of $\left\{\kappa_A, \kappa_B, \kappa_C\right\}$ to find the parameter choices that maximise that likelihood.

In theory this method could be applied to generate new weights for every balancing problem. In practice, calculating weights is likely to be computationally intensive, and it may be desirable to keep weights stable from one reference period to the next, so it may be preferable to use the same set of weights for several years and update occasionally.

If weights are instead calculated by the methods discussed in Section 4.2, the same MLE approach could be used to determine the constant *k* that calibrates these weights against those derived from known standard errors. This would likely be less computationally demanding, since it only requires estimating one parameter.

Since expert judgement would still be required to classify cells into similar groups prior to applying the MLE process, this approach should be seen as *reducing* subjectivity in weighting rather than completely eliminating it.

4.4 Example of MLE method

As an example of how this technique might be applied, we will return to the two weighting options discussed in Section 4.1. Recall that cells were weighted at:

$$w_i = \frac{1}{\hat{x}_i^2\left(1-r_i\right)} \text{ (Method 1)}$$

$$w_i = \frac{1}{\left|\hat{x}_i\right|\left(1-r_i\right)} \text{ (Method 2)}$$

This corresponds to a variance model of:

$$\hat{\sigma}_i^2 = k\left(1-r_i\right)\left|\hat{x}_i\right|^\theta = k / w_i$$

where θ equals 1 (for Method 2) or 2 (for Method 1) and k is some positive scalar.

We would like some objective criteria to determine which of these two models is more appropriate. In the absence of other information, it seems reasonable to choose the one that gives the highest likelihood solution.

Returning to the likelihood function derived in Section 3 and substituting in this estimate for σ_i^2, we get an expanded likelihood function:

$$\begin{aligned} f\left(\underline{\tilde{\varepsilon}}, k, \theta, \underline{r}\right) &= \prod_i \left(\frac{\exp\left(-\tilde{\varepsilon}_i^2 / \left(2k\left(1-r_i\right)\left|\hat{x}_i\right|^\theta\right)\right)}{\left(2\pi k\left(1-r_i\right)\left|\hat{x}_i\right|^\theta\right)^{1/2}} \right) \\ &= \prod_i \left(\left(\frac{w_i}{2\pi k}\right)^{1/2} \exp\left(-\tilde{\varepsilon}_i^2 w_i / \left(2k\right)\right) \right), \\ w_i &= \frac{1}{\left(1-r_i\right)\left|\hat{x}_i\right|^\theta}. \end{aligned}$$

Given sufficient data and computing power, we could perhaps optimise over all of $\left(\underline{\tilde{\varepsilon}}, k, \theta, \underline{r}\right)$ and choose the values that maximise the resulting likelihood. However, to simplify this example, we will treat $\underline{r}$ as a fixed constant vector, and restrict θ to either 1 or 2, with k and $\underline{\tilde{\varepsilon}}$ as continuous.

As in Section 3, maximising the likelihood is equivalent to minimising the negative log of that likelihood:

$$\begin{aligned} OF &= -\ln\left(f\left(\underline{\tilde{\varepsilon}}, k, \theta, \underline{r}\right)\right) \\ &= \sum_i \left(\tilde{\varepsilon}_i^2 w_i / \left(2k\right) - \frac{1}{2}\ln\left(\frac{w_i}{2\pi k}\right)\right) \\ &= \frac{1}{2}\sum_i \left(\frac{\tilde{\varepsilon}_i^2 w_i}{k} + \ln\left(k\right) + \ln\left(2\pi\right) - \ln\left(w_i\right)\right) \end{aligned}$$

with k and θ now treated as variables within the optimisation, this is no longer a pure quadratic optimisation problem. It could still be solved through generic optimisation methods e.g. gradient descent, but this might be computationally expensive given the large number of variables involved.

However, in this case we can simplify the optimisation by observing that for any given choice of $\theta, \underline{r}$ (which together define $\underline{w}$) the optimal values of $\underline{\tilde{\varepsilon}}$ are independent of k. This can be seen by rewriting the objective function as:

$$OF = \frac{1}{2k}\sum_i \tilde{\varepsilon}_i^2 w_i + \frac{1}{2}\sum_i \left(\ln(k) + \ln(2\pi) - \ln(w_i)\right).$$

For fixed values of k, r, and θ (hence also fixed w) this expression is minimised if and only if $\sum_i \tilde{\varepsilon}_i^2 w_i$ is minimised, and the solution to that minimisation is independent of *k*.

Hence, for any given $\theta, \underline{r}$ we can find the optimal values for $\underline{\tilde{\varepsilon}}$ as a quadratic optimisation, and only then do we need to solve for optimal *k*.

With all other parameters now fixed, discarding constant terms shows that the optimal *k* will be that which minimises:

$$\sum_{i=1}^{n}\left(\frac{\tilde{\varepsilon}_i^2 w_i}{k} + \ln(k)\right) = n\ln(k) + \frac{1}{k}\sum_i \tilde{\varepsilon}_i^2 w_i \ .$$

Solving for *d*/*dk*=0 then shows that the optimum lies at $k = \sum_i \tilde{\varepsilon}_i^2 w_i \,/\, n$.

In this case, with $\underline{r}$ fixed and θ restricted to only two possible values, we can reduce the problem to two quadratic optimisation runs, rather than a large non-quadratic optimisation.

Doing this for the 2010–2011 Supply-Use data discussed in Section 4.1.1 gives the following values:

Method 1: *k*=0.00047, OF = 5283.7

Method 2: *k*=1.03, OF = 28705.6

Recalling that the objective function is the negative log of likelihood, lower values for the OF indicate higher likelihood. In this case, if we were satisfied with the other assumptions used for the balancing (in particular, we assume that the parameters in $\underline{r}$ have been set to their best possible values) then this would imply that Method 1 is preferable.

In practice, the values for $\underline{r}$ have not been set rigorously for this example, and it may be that optimising on $\underline{r}$ as well as on *k* and θ would lead to different conclusions. It may also be that allowing different values of θ for different types of inputs, and/or non-integer θ, will give a better fit to the behaviour of the inputs. Hence, this should be taken only as an example of how such decisions *might* be made.

5. BALANCING WITH MOVEMENTS

Many economic tables are compiled annually or quarterly, and changes in estimates over time ("movements") are of interest, often as much so as the point-in-time estimates ("levels"). When adjusting such data, it may be desirable to control the adjustments to movements.

Benchmarking theory quantifies adjustment to movements via the Denton Additive First Differences (AFD) and Proportionate First Differences (PFD) measures:

$$AFD_{i,t} = \left(\tilde{x}_{i,t} - \hat{x}_{i,t}\right) - \left(\tilde{x}_{i,t-1} - \hat{x}_{i,t-1}\right)$$

$$PFD_{i,t} = \frac{\tilde{x}_{i,t}}{\hat{x}_{i,t}} - \frac{\tilde{x}_{i,t-1}}{\hat{x}_{i,t-1}}$$

where $\hat{x}_{i,t}$ is the original estimate or indicator for cell *i* at time period *t* and $\tilde{x}_{i,t}$ is the adjusted value.

Minimising squared AFD or PFD adjustments is a well-known method for benchmarking individual time series (cf. Dagum and Cholette, 2006). Eurostat (2013) describes a balancing approach based on a combination of the WLS level-preservation function given in Section 3 of this paper, and weighted squared AFD or PFD terms.

This section will examine AFD/PFD-minimisation approaches within the same sort of MLE framework applied in Section 3. It will show that AFD and PFD-based movement-preservation objective functions correspond to maximum likelihood estimators for random walk error models (in the case of PFD, approximately random walk), and will offer a modification of such methods to improve forward-series estimation.

It will also demonstrate that these correspondences are *not* additive: even though a MLE for white-noise additive errors corresponds to a level-preservation objective function, and a MLE for random-walk additive errors corresponds to an AFD objective function, the linear combination of AFD and level-preservation functions (as used e.g. in Eurostat) does *not* correspond to the MLE for a combination of white-noise and random walk additive errors.

Hence, although the two approaches are equivalent for important simple cases, they are not equivalent in more general cases. This section will discuss the pros and cons of a "strict MLE" approach as compared to the Eurostat method.

5.1 MLE justification for AFD and PFD balancing

Section 3 discussed a scenario where the unbalanced values are related to true values by independent, normally distributed errors: $\hat{x}_i = x_i + \varepsilon_i$, with $\varepsilon_i = N\left(0, \sigma_i^2\right)$. In that case we established that a balanced estimate $\underset{\sim}{\tilde{x}}$ can be calculated as a maximum likelihood estimator for $\underset{\sim}{x}$ by minimising the weighted sum of squared adjustment.

Now consider a scenario where instead of the error structure assumed in Section 3, the errors behave as an additive random walk (ARW):

$$\hat{x}_{i,t} = x_{i,t} + \sum_{k=1}^{t} \eta_{i,k} .$$

Here the "innovations" $\eta_{i,t}$ are independent and normally distributed, with mean zero:

$$\eta_{i,t} = N\left(0, \rho_{i,t}^2\right).$$

Note that $\eta_{i,1}$ represents total estimator error at the start of the series, while $\eta_{i,t>1}$ represents only the additional error accrued between one period and the next. Hence we will assume that $\rho_{i,1}^2 \gg \rho_{i,t>1}^2$ and approximate $\eta_{i,1}$ as having a flat prior.

By definition, for $t > 1$,

$$\eta_{i,t} = \left(\hat{x}_{i,t} - \hat{x}_{i,t-1}\right) - \left(x_{i,t} - x_{i,t-1}\right).$$

Hence, for any given set of balanced estimates $\underset{\sim}{\tilde{x}}$, the implied estimates for the innovations are:

$$\begin{aligned} \tilde{\eta}_{i,t} &= \left(\hat{x}_{i,t} - \hat{x}_{i,t-1}\right) - \left(\tilde{x}_{i,t} - \tilde{x}_{i,t-1}\right), t > 1 \\ \tilde{\eta}_{i,1} &= \hat{x}_{i,1} - \tilde{x}_{i,1} . \end{aligned}$$

Because the innovations are independent, we can apply the same arguments used in Section 3 to show that a maximum likelihood estimate is obtained by selecting the values for $\underset{\sim}{\tilde{x}}$ within the feasible region that minimise a modified objective function:

$$OF_{ARW} = \sum_{i,t} \left(\tilde{\eta}_{i,t} \big/ \rho_{i,t} \right)^2 .$$

The flat-prior approximation for $\eta_{i,1}$ allows us to drop out the $t=1$ terms from the objective: $\tilde{\eta}_{i,1} / \rho_{i,1} \cong 0$ so can be ignored. (This requires using the fact that $\tilde{\eta}_{i,1}$ cannot grow very large without some other $\tilde{\eta}_{i,t}$ also growing very large, and hence causing a non-optimal objective that need not be considered.)

This then gives:

$$OF_{ARW} = \sum_i \sum_{t>1} \left(\tilde{\eta}_{i,t} \big/ \rho_{i,t} \right)^2 = \sum_i \sum_{t>1} v_{i,t} \left(\left(\tilde{x}_{i,t} - \tilde{x}_{i,t-1} \right) - \left(\hat{x}_{i,t} - \hat{x}_{i,t-1} \right) \right)^2$$
$$= \sum_i \sum_{t>1} v_{i,t} AFD_{i,t}{}^2$$

where movement weights $\underline{v}$ are set as $v_{i,t} = 1/\hat{\rho}_{i,t}^2$.

Hence we have shown that an AFD-based least squares objective function is equivalent to the MLE estimator obtained from an additive random walk:

$$OF_{ARW} = \sum_i \sum_{t>1} v_{i,t} AFD_{i,t}{}^2 \Leftrightarrow$$
$$MLE : \hat{x}_{i,t} = x_{i,t} + \sum_{k=1}^{t} \eta_{i,t},\ \eta_{i,t} = N\left(0, \rho_{i,t}^2\right),\ \rho_{i,1}^2 \gg \rho_{i,t>1}^2,\ v_{i,t} = 1 \big/ \hat{\rho}_{i,t}^2$$

This approach corresponds to AFD-based benchmarking methods, and to the additive movement-preservation component of the objective function presented in Eurostat.

In some cases, it may be more appropriate to assume that errors scale in proportion to the underlying estimate. The most straightforward form for such a model would be a proportional random walk (PRW):

$$\hat{x}_{i,t} = x_{i,t} \sum_{k=1}^{t} \eta_{i,k}$$

with $\eta_{i,t} = N\left(0, \rho_{i,t}^2\right)$ and independent, as before. This leads to:

$$\tilde{\eta}_{i,t} = \frac{\hat{x}_{i,t}}{\tilde{x}_{i,t}} - \frac{\hat{x}_{i,t-1}}{\tilde{x}_{i,t-1}}.$$

However, this form is inconvenient for an optimisation approach because the resulting objective function would require nonlinear constraints and/or a non-quadratic objective function.

To avoid this difficulty, benchmarking methods often use the PFD movement-preservation measure instead, which is linear in the decision variables and hence gives a quadratic objective function when squared.

In fact, the PFD measure corresponds to a modified version of the proportional random walk model (M-PRW) presented above:

$$OF_{M-PRW} = \sum_{i}\sum_{t>1} v_{i,t} PFD_{i,t}^{\,2}$$

$$\Updownarrow$$

$$MLE : \hat{x}_{i,t} = x_{i,t}\left(\sum_{k=1}^{t}\eta_{i,k}\right)^{-1},\ \eta_{i,t} = N\left(0, \rho_{i,t}^{2}\right),\ \rho_{i,1}^{2} \gg \rho_{i,t>1}^{2},\ v_{i,t} = 1\big/\hat{\rho}_{i,t}^{2}$$

Note that under the assumption that $\rho_{i,1}^{2} \gg \sum_{t>1}\rho_{i,t}^{2}$ (i.e. the probability of the random walk crossing zero is negligible) the reciprocal random walk is itself approximately a random walk. This can be proved by using the approximation $\left(1+h\right)^{-1} \cong 1-h$ for small h.

Hence a PFD-type least squares objective function seems a reasonable approximation for a MLE estimator with a proportional random walk error model, in the interests of keeping to a tractable objective function.

5.2 Addition of trend term

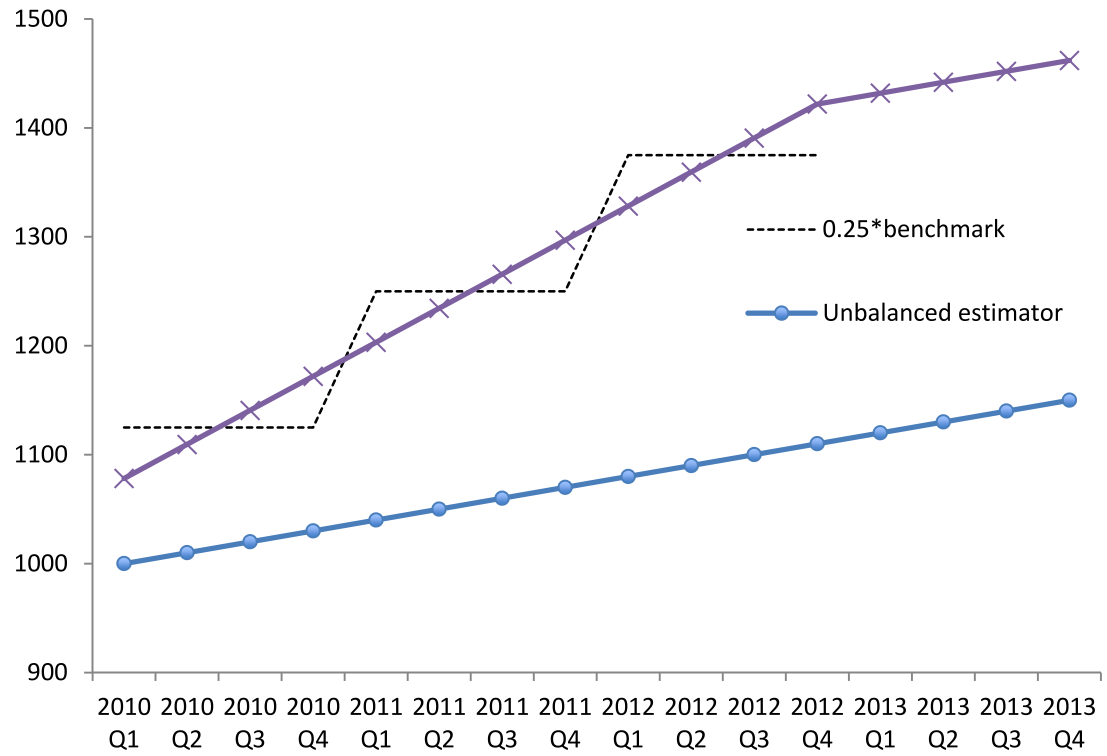
In some cases the estimation error may show a consistent trend over time. For example, our unbalanced estimator $\hat{x}_{i,t}$ may come from a source that only has partial coverage of the population of interest, with that coverage steadily increasing or decreasing from period to period.

In such cases, the measures discussed above may incur avoidable bias in estimation, because part of the estimator movement that they preserve is in fact due to systematic error.

In Brent, Stuckey, and Davidson (2015) we considered the problem in the context of adjusting a quarterly series to match annual benchmark totals (assumed to be exactly correct). Benchmarks help mitigate this error, since the net movement over long periods has to match the movement between benchmarks. However, it often happens that the most recent benchmark is not yet available and so the most recent estimates are not subject to this constraint.

Figure 5.1 below shows a fictional illustration of this situation. Both the annual benchmarks and the unbalanced series are increasing smoothly over time, but there's a steadily increasing discrepancy, suggesting a growing systematic error in the unbalanced estimator. In the benchmarked years, reconciling to benchmarks eliminates this error, so the balanced estimate shows higher growth than the indicator. But for the last year in the series, because no benchmark constraint yet applies, an AFD objective function requires it to match the unbalanced movement exactly, resulting in a decline in growth that will probably need to be corrected when the next benchmark does become available. As long as the relationship between benchmarks and unbalanced estimates continues to follow this pattern, the same error will recur every year.

5.1 Adjusting a quarterly series to an annual benchmark using AFD



In our previous benchmarking work we found that this kind of bias could be greatly reduced, without substantial increase in errors, by forecasting the growth of a benchmark indicator ratio from benchmarked to unbenchmarked periods – for instance, via a "random walk with drift" approach. This idea can be adapted to an optimisation framework by modifying the error models discussed previously.

We can amend the additive RW model used to justify the AFD-based objective function by adding a trend term:

$$\hat{x}_{i,t} = x_{i,t} + \sum_{k=1}^{t} \eta_{i,k} + s_i t$$

where s_i is the expected increase in error per period. In general this is not directly observable; instead we treat it as another unknown to be chosen/estimated as part of the optimisation.

With this modification, our estimate for the innovation terms becomes:

$$\tilde{\eta}_{i,t} = \left(\hat{x}_{i,t} - \tilde{x}_{i,t}\right) - \left(\hat{x}_{i,t-1} - \tilde{x}_{i,t-1}\right) - \tilde{s}_i .$$

If we assume flat priors for $\{s_i\}$, this then leads to a simple modification to the additive random walk/AFD-based objective function:

$$OF_{ARW+trend} = \sum_i \sum_{t>1} v_{i,t} \left(AFD_{i,t} + \tilde{s}_i \right)^2 .$$

Although this approach increases the size of the problem by adding new variables, the increase is not large – only as much as adding another period's worth of data. The fitted trend terms may also be useful as a quality-control measure for inputs; a large trend term may suggest e.g. that the coverage of the unbalanced estimator is changing rapidly.

The proportional versions of these functions may be modified in a similar way[10]. An error model of the form

$$\hat{x}_{i,t} = \left(-s_i t + \sum_{k=1}^{t} \eta_{i,k} \right)^{-1} x_{i,t}$$

leads to a modified PFD-type objective:

$$OF_{M-PRW+trend} = \sum_i \sum_{t>1} v_{i,t} \left(PFD_{i,t} + \tilde{s}_i \right)^2 .$$

When balancing over very short intervals, including a trend term may lead to overfitting-type problems; in these cases, it may be better to use the non-trend versions presented in Section 5.1 above. For very long intervals, it's possible that the trend behaviour will change; in such cases a "rolling window" approach may be helpful (e.g. fix older estimates five years after their benchmarks become available, and apply these objective functions only for recent years.)

10 The additive version represented trend as $s_i t$. The choice of $-s_i t$ for PFD is made to achieve consistency in resulting objective functions and in interpreting the sign of s_i; in both cases, positive values of s_i imply that the unbalanced estimate is increasing relative to true value.

5.3 Combining movement- and level-preservation approaches

In many cases the true error model may resemble some mix of the white-noise type errors considered in Section 3 (which led to an objective function based on minimising changes to levels) and the sort of auto-correlated errors considered in Sections 5.1 and 5.2 (which led to an objective function based on minimising changes to movements, modulo trend).

Those error models can be combined, e.g.:

$$\hat{x}_{i,t} = x_{i,t} + s_i t + \sum_{k=1}^{t} \eta_{i,k} + \varepsilon_{i,t}$$

(additive RW + trend + white-noise transients), or

$$\hat{x}_{i,t} = x_{i,t} \left(-s_i t + \sum_{k=1}^{t} \eta_{i,k} + \varepsilon_{i,t} \right)^{-1}$$

(approximately proportional RW + trend + transients)

with the same assumptions noted previously on distributions of $\underline{\varepsilon}$ and $\underline{\eta}$, and adding the assumption that $\underline{\varepsilon}$ and $\underline{\eta}$ are independent of one another.

The problem then becomes choosing values for $\underline{\tilde{s}}$, $\underline{\tilde{\varepsilon}}$ and $\underline{\tilde{\eta}}$ in a way that minimises

$$OF_{combined} = \sum_i \left(\sum_{t>1} v_{i,t} \tilde{\eta}_{i,t}^2 + \sum_t w_{i,t} \tilde{\varepsilon}_{i,t}^2 \right)$$

with weights $\underline{v}$, $\underline{w}$ defined as before and with the implied $\underline{\tilde{x}}$ satisfying all relevant constraints. The balanced table $\underline{\tilde{x}}$ is then defined as a linear function of these decision variables; constraints can be defined in terms of $\underline{\tilde{x}}$ or of $\underline{\tilde{s}}$, $\underline{\tilde{\varepsilon}}$ and $\underline{\tilde{\eta}}$ as preferred. Weighting parameters might be set by the same MLE approach discussed in Section 4.3.

Note that this approach increases the size of the optimisation problem, approximately doubling the number of free variables, since each $x_{i,t}$ has corresponding ε and η.

If appropriate, the error model could be extended further, e.g. to include weighted-average type terms, but this will further increase the size of the optimisation problem and the number of weighting parameters required, presenting a risk of overfitting.

An alternate approach, as presented in Eurostat, is to optimise on a linear combination of the objective functions used for these individual cases.[11] For example, one such form is:

$$OF_{AFD+lvl} = \underbrace{\sum_i \sum_{t>1} v_{i,t} AFD_{i,t}{}^2}_{movements} + \underbrace{\sum_i \sum_t w_{i,t} \left(\hat{x}_{i,t} - \tilde{x}_{i,t} \right)^2}_{levels} .$$

Although each of the two components corresponds individually to a MLE estimator based on models discussed previously, the sum of those components does *not* correspond to an MLE based on the sum of the two error models. For example, consider the following problem where a single series with one missing benchmark must be adjusted for consistency with benchmarks:

	Year		
	1	2	3
Unbalanced estimate	10	20	30
Benchmark value	110	–	130

Here the correct values for year 1 and 3 are known exactly, but we still wish to use the unbalanced estimates to help choose an appropriate value for year 2.

Under a strict MLE approach, summing white-noise and additive random walk error terms (without drift) gives the total error model:

$$\hat{x}_{i,t} = x_{i,t} + \sum_{k=1}^{t} \eta_{i,k} + \varepsilon_{i,t} .$$

Under the flat-prior assumption for $\eta_{1,1}$, the resulting objective function is:

$$OF_{combined} = \sum_{t>1} v_{i,t} \tilde{\eta}_{1,t}^2 + \sum_t w_{i,t} \tilde{\varepsilon}_{1,t}^2 .$$

Assuming weights are non-negative, the unique optimal solution for this problem is to set $\tilde{\eta}_{1,1} = 100$ and set all other $\hat{\eta}$ and $\tilde{\varepsilon}$ terms to zero, i.e. $\underline{\tilde{x}} = \left(110,120,130\right)^t$. This satisfies constraints and achieves an objective of zero, which clearly cannot be improved on, and cannot be matched by any other choice of variables. Note that this solution remains optimal regardless of choice of weights, as long as they remain non-negative.

11 The full objective function given in Eurostat contains options for both proportional and additive movement preservation, as well as level preservation and two additional terms not discussed here. It does not contain a trend term, but the movement-preservation components could easily be modified to include one.

However, if instead we generate an objective function by adding the WLS and AFD objectives, we have:

$$OF = \sum_{t>1} v_{1,t} AFD_{1,t}{}^2 + \sum_t w_{1,t} \left(\hat{x}_{1,t} - \tilde{x}_{1,t} \right)^2 .$$

Since constraints require $\tilde{x}_{1,1} = 110, \tilde{x}_{1,3} = 130$ this simplifies to:

$$\begin{aligned} OF_{AFD+lvl} &= v_{1,2} \left(\tilde{x}_{1,2} - 20 + 10 - 110 \right)^2 + v_{1,3} \left(130 - 30 + 20 - \tilde{x}_{1,2} \right)^2 \\ &\quad + w_{1,1} \left(110 - 10 \right)^2 + w_{1,2} \left(\tilde{x}_{1,2} - 20 \right)^2 + w_{1,3} \left(130 - 30 \right)^2 \\ &= \left(v_{1,2} + v_{1,3} \right) \left(\tilde{x}_{1,2} - 120 \right)^2 + w_{1,2} \left(\tilde{x}_{1,2} - 20 \right)^2 + 10000 w_{1,1} + 10000 w_{1,3} \end{aligned}$$

The optimal solution for this objective is:

$$\tilde{x}_{1,2} = \frac{120 \left(v_{1,2} + v_{1,3} \right) + 20 w_{1,2}}{v_{1,2} + v_{1,3} + w_{1,2}} .$$

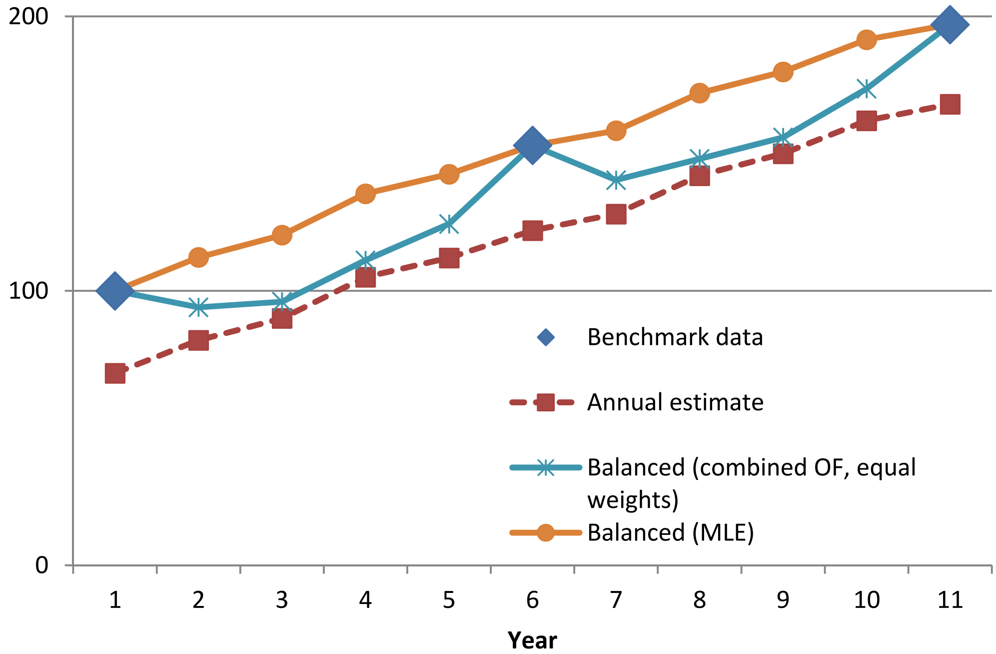
This solution varies with the weights and hence cannot be equivalent to the MLE solution, which does not. Thus it can be seen that in general, adding the objective functions derived from two error models does *not* yield an objective function that corresponds to the sum of those error models.

This difference arises because the two methods have somewhat different purposes. Objectives such as $OF_{AFD+lvl}$ are based in the philosophy that the aim of balancing is to satisfy constraints while *minimising change* from unbalanced movements and levels (see e.g. discussion in Eurostat) and the form of these methods makes them appropriate for that purpose, with weighting to manage necessary trade-offs between those two aims.

As presented here, the MLE method aims to give an estimate that is *most likely to be true* (subject to flat-prior approximations and simplifications as discussed previously), which is not quite the same thing as change minimisation. In the example shown above, a balanced table of (110,120,130) requires a large change to the middle value – but given the starting values of 110 and 130, and an indicator that shows smooth growth across the period, this seems like a very plausible choice.

Figure 5.2 shows another comparison between the $OF_{AFD+lvl}$ approach and a MLE based on the combined error model. In this scenario, an annual estimate series is supplemented by accurate benchmark data for years 1, 6, and 11; we wish to adjust the estimate to match these benchmarks. Two options are shown: $OF_{AFD+lvl}$ with both components weighted equally, and the MLE approach.

5.2 Comparing balancing methods, with five-year benchmark data



In years 1, 6, and 11 the solution is constrained to benchmark values, but in other years the level-preservation component of the additive objective function pulls estimates back towards the annual estimate. In between benchmarks, over years 3–4 and 8–9, this gives a solution that closely matches both movements and levels for the annual estimate, but at the cost of spurious peaks at the benchmark years.

A pure AFD objective function can be interpreted as accepting the possibility of systematic error and of random walk-type errors; under this objective a systematic correction is "free" and random walk corrections are cheap, making it easy to correct them. However, when a level-preservation function is added, all systematic and RW corrections are heavily penalised (since they require changes across much/all of the series, and each data point attracts a penalty) largely negating this feature of the AFD objective.

For this particular scenario, the problem could be resolved by reducing the level-preservation weighting to zero. But this results in a simple AFD function that may not be appropriate to other cases where independent errors are large.

By contrast, a MLE method based on combined (RW + independent) errors avoids these artefacts, without removing the capability to incorporate independent errors. It retains the ability to make "free" systematic corrections and cheap RW corrections, but doesn't assume that these are the only kind of errors. The MLE solution shown above is not sensitive to weighting choices; no matter how strongly or weakly the random walk component is weighted against the independent component, the solution to this particular problem is virtually unchanged because the systematic correction remains free.

Each method has some advantages and disadvantages:

- The Eurostat method may be easier to understand intuitively because it allows the balancing problem to be understood as a sum of "preserve movements" and "preserve levels". However, this separation requires treating movements and levels as independent, when in fact movements are completely dependent on levels.
- The MLE approach is likely to be more computationally expensive because it requires fitting more variables.
- The MLE approach allows use of MLE methods to automatically fit weighting parameters; it is less obvious whether this method would be appropriate with the Eurostat objective.
- The MLE approach may provide clearer insight into the nature of errors in the inputs, potentially guiding improvements to data collection.

Based on these considerations, it seems desirable to evaluate the MLE approach further, and compare its behaviour to that of the Eurostat-style approach.

5.4 Setting weights for movements

As for the level-preservation balancing discussed in previous sections, applying these methods require setting weights, or equivalently, estimating variances $\underline{\sigma}^2$ and $\underline{\rho}^2$.

The Statistics Netherlands weighting strategy as published in Eurostat varies the weights for AFD movements proportional to the inverse square of unbalanced magnitude : $v_{i,t} \propto 1/\left|\hat{x}_{i,t}\right|^2$. The exact value is also modified by categorical-type factors based on e.g. expected accuracy of unbalanced data, and by parameters that control the relative importance of movement preservation vs. level preservation vs. other considerations.

Weights for PFD movements have no dependence on unbalanced magnitude, and are determined solely by categorical and relative importance considerations.

Both of these approaches are consistent with an assumption that expected magnitude of errors will scale in proportion to the cell magnitude, which also matches the weighting for level adjustments.

Movements at time *t* are affected both by errors in levels at time *t* and at *t*–1 hence a more exact approach might base movement weights on both $\left|\hat{x}_{i,t}\right|$ and $\left|\hat{x}_{i,t-1}\right|$. In practice, these magnitudes will generally be similar, so ignoring $\left|\hat{x}_{i,t-1}\right|$ and defining weights relative to $\left|\hat{x}_{i,t}\right|$ seems a reasonable simplification under this weighting strategy.

Under the alternate assumption that magnitude of errors scales in proportion to the square root of magnitude, weights for AFD movements should be set proportional to $1/\left|\hat{x}_{i,t}\right|$ (modified by other factors, as above) and weights for PFD movements should be proportional to $\left|\hat{x}_{i,t}\right|$ (no reciprocal) because larger values are expected to have smaller proportionate errors, hence modifications to errors will be weighted more highly.

For the "strict MLE" approach discussed in 6.3, weights on the white-noise error terms may be set proportional to either $1/\left|\hat{x}_{i,t}\right|$ or $1/\left|\hat{x}_{i,t}\right|^2$:

$$w_{i,t} = \frac{1}{\tau_{i,t}\left|\hat{x}_{i,t}\right|} \text{ or } w_{i,t} = \frac{1}{\tau_{i,t}\left|\hat{x}_{i,t}\right|^2}$$

with proportionality constants $\tau_{i,t}$ set through MLE methods as discussed in Section 5 (i.e. classify cells into a small number of groups, assume $\tau_{i,t}$ equal for all cells within the same group, and then choose values that give the maximum likelihood solution).

Random walk terms are likely to correspond to systematic errors, or to other issues that can be treated in the same way (e.g. the use of an indicator that correlates to the variable of interest, but is measured in different units).

In general, it seems appropriate to parameterise these in a way that lets the magnitude of expected RW errors vary approximately in proportion to the magnitude of the estimates. (A lower order relationship implies that for series with very large magnitude, the proportion of systematic error approaches zero; a higher order relationship implies the same for series with very small magnitude. Neither of these seem desirable.)

For an additive RW error model, this implies $Var\left(\eta_{i,t>1}\right) = \rho_{i,t}^2 \cong \varsigma_{i,t}\left|x_{i,t}\right|^2$ i.e.

$$v_{i,t>1} = \frac{1}{\varsigma_{i,t}\left|\hat{x}_{i,t}\right|^2}$$

for some proportionality constants, $\varsigma_{i,t}$ to be chosen by the same approach as for $\tau_{i,t}$ above.

For the approximate proportional RW model, this implies that $Var\left(\eta_{i,t>1}\right)$ should be independent of $\left|\hat{x}_{i,t}\right|$, i.e. set

$$v_{i,t>1} = \frac{1}{\varsigma_{i,t}}.$$

5.5 Time series shocks

Time series may sometimes experience unusual shocks that require careful consideration in balancing and benchmarking.

Some statistical methods rely on assumptions about the time series nature of the true series, e.g. ARIMA-type models. Violations of those assumptions may cause large errors. The MLE approach in this paper does not make any assumptions about smoothness or continuity of the true series.[12] Hence, disturbances such as a level shift or trend break in the true series may not require any special treatment.

The MLE approach *does* rely on assumptions about the error structure. Changes to the error structure are more significant – e.g. a change to data sources that causes a sudden large change in systematic error, or an increase/reduction in the variance of independent errors. Options for addressing these would include allowing a break in the error model (e.g. fitting different parameters before and after the change) and/or limiting the time period used for fitting parameters, so that model expectations are not overly influenced by obsolete information.

12 Although it could be extended to do so, by using a model of the true series as a Bayesian prior.

6. DIAGNOSTICS FOR BALANCING

This section discusses diagnostics that might be used for evaluating the results of an auto-balancing process. This discussion focuses on level-preservation balancing as introduced in Sections 3–4, but approaches presented here could be extended to the time series methods discussed in Section 5 by considering individual error-model components rather than just net adjustments.

Diagnostics for the numerical optimisation itself are out of scope for this discussion; we expect that the optimisation will be performed by an off-the-shelf solver with its own diagnostics to identify problems within that process, so we will assume here that the balanced output $\underline{\tilde{x}}$ is indeed an optimal solution for the objective function and constraints as defined.

Our primary interest here is in identifying unusual results that might signal a problem in the way the problem has been posed (e.g. bad inputs), flagging these for expert scrutiny, and interpreting those results. A secondary objective is to compare performance of two different balancing methods.

A simple approach is to look for large adjustments (in dollar terms and/or relative to the cell's unbalanced magnitude) but on its own this is unlikely to be adequate since it does not incorporate knowledge about expected variances.

6.1 Important cases for diagnostics

Some important cases to consider in diagnostics:

- Cell weight is set too low (i.e. we have overestimated potential errors, which may result in larger adjustments than are appropriate for that cell).
- Cell weight is set too high (e.g. we have underestimated the variance of estimates for that cell, or failed to correct for known bias before auto-balancing).
- Cell weight is appropriate but some combination of inputs and constraints leads to poor behaviour (e.g. the matrix expression of the problem is ill-conditioned). Solver diagnostics may assist in identifying such cases.

These cases might apply to a single cell, or across a group of related cells (e.g. we underweight all supply items for a single industry).

In the first of the listed cases, the under-weighted cell/s will receive larger adjustments than they ought, while other cells closely related (via constraints) to the under-weighted cell/s will tend to receive smaller adjustments than they ought.

In the second case, the effect is reversed. An over-weighted cell will receive smaller adjustments than it ought, and its "constraint neighbours" will receive somewhat larger adjustments than they ought.

In the third, a cell may receive large adjustments in response to a discrepancy elsewhere in the table.

Without an independent source of information on input variances, detecting under-weighting for individual values is likely to be difficult. If we estimate that the standard error for an unbalanced cell is $100m and weight accordingly, but the final balancing outcome suggests a residual of only $10m for that cell, this *might* indicate that the true standard error was smaller than $100m (i.e. the cell was under-weighted). But it's also quite possible that the realised error just happened to be smaller than the standard error.

For larger groups of cells it may be possible to detect a pattern of errors: e.g. if we estimate that a specific measure consistently has standard error $100m, but over several years it never requires an adjustment of more than $10m, this may suggest that the standard error is less than estimated and the weights for the corresponding variables should be increased.

An important diagnostic in regression is *leverage*, measuring the influence of an individual data point on the fitted model. A similar approach might be applied here, e.g. by measuring how much a small change to the weight for a given cell changes estimates elsewhere in the table.

6.2 Use of objective function as a diagnostic

In assessing the results of balancing we may wish to examine the adjustments made, to gain insight into the errors that might be present in our inputs, along similar lines to the examination of residuals to evaluate a regression model. Unexpectedly large adjustments might indicate a problem with input data or problem specification, but defining "unexpectedly" requires some thought.

As noted previously, the objective function provides a good starting point for evaluating adjustments. In many operations research problems, individual terms in the objective function translate directly to costs or profits for a single component of the overall work (e.g. power loss on a particular line) and therefore they are likely to be of interest in their own right as part of a cost breakdown.

Here, the "cost" is the overall unlikelihood of the solution. Assuming weights w_i are set proportional to 1/variance (excluding dependent cells, which are weighted at zero), then the adjustment on the i-th cell can be considered to contribute a "cost" of $\left(\tilde{\varepsilon}_i / \sigma_i\right)^2 = \tilde{\varepsilon}_i^2 w_i$ to the objective function discussed in Section 3. (Recall that the standard deviation for ε_i is σ_i.)

Hence, the cells that make the largest contribution to the objective function represent the adjustments that should be considered least likely. These make good candidates for scrutiny: e.g. we might choose the ten or twenty cells that make the largest contributions to the score. This can also be aggregated to a group-level diagnostic, e.g. for comparing entire rows within a table, which corresponds to the likelihood of that group's adjustments.

A complication here is that even though adjustments attempt to estimate errors, the expected distribution of adjustments will not match the distribution of errors. Consider the simple case of balancing a group of cells to meet a single constraint S on their sum, with no other constraints active. As noted by Stone *et al.* (1942), a Lagrange multiplier calculation shows that the optimal solution has all cell adjustments proportional to the inverse of their weights – hence adjustments will be in proportion to variance (as implied in the weighting) and *not* to standard error.

Although this paper does not attempt a rigorous proof for more complex scenarios, this case suggests that cells with the highest standard error will tend to receive "less likely" adjustments (higher $\left(\tilde{\varepsilon}_i / \sigma_i\right)^2$). Using the contribution to objective function as a diagnostic may over-emphasise such cells while potentially overlooking problems with low error (high weight) cells. Hence we suggest the use of $\tilde{\varepsilon}_i w_i$ as a second metric for detecting unusual adjustment patterns that might otherwise be masked by large weights; in the simple case mentioned above, all cells would score equally on this metric. We are not aware of this metric being used in previous work.

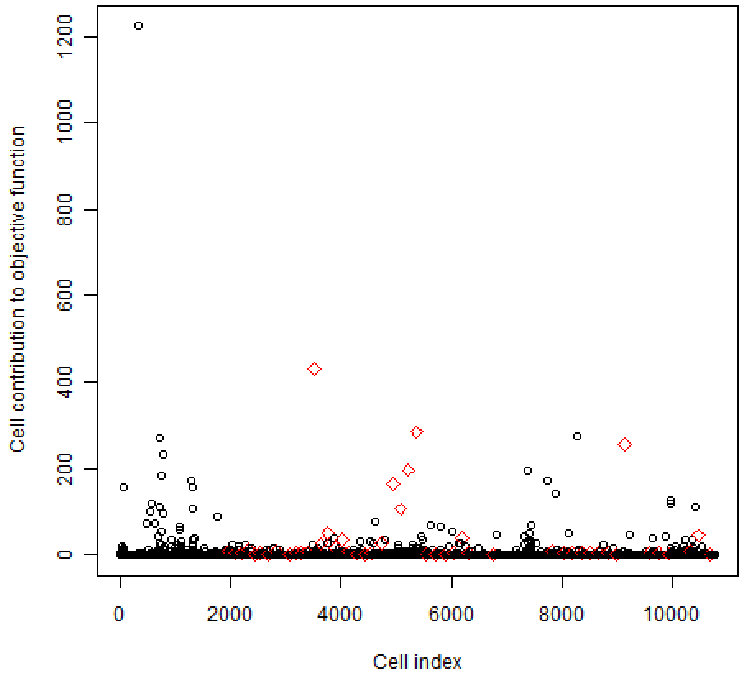
Under either metric, a high score does not necessarily signify a problem; it may simply be part of a trade-off that accepts large adjustment to one cell as a trade-off for smaller adjustments elsewhere. Nevertheless it may be helpful to flag such adjustments so that users are aware of the trade-off, and to inform decisions for future data collection.

Figures 6.1 and 6.2 show examples of these two metrics as applied to balancing on test data from a 2010–2011 Supply-Use table. In the unbalanced data one product had a large gap between supply and use, leading to large (proportional) adjustments for the relevant Use items (marked with red diamonds). However, because most of these had small initial values and hence large weights, these adjustments appear large but not exceptionally so when assessed by their contribution to the objective function $r_i^2 w_i$ (figure 6.1).

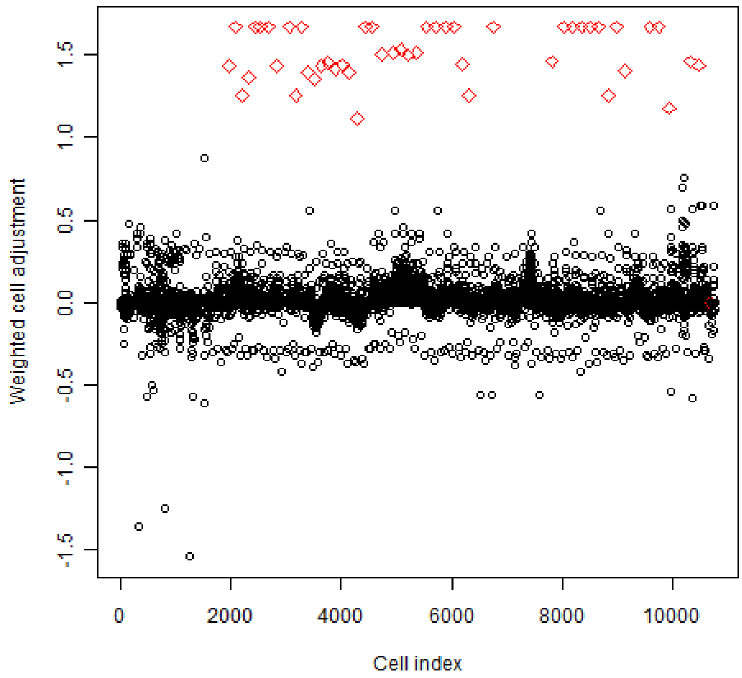
Plotting weighted adjustments $r_i w_i$ (figure 6.2) makes it much more obvious that these cells are receiving unusually large adjustments (in the context of their weights), suggesting that this discrepancy might benefit from further attention; for example, we might consider the possibility of systematic underestimation for these items, or that the accuracy on the Supply side has been overestimated.

In cases where a given cell or group of cells has a non-zero adjustment, increasing the weight for that cell will always increase the overall objective function, but can increase or decrease that cell's contribution to the objective function (i.e. $r_i^2 w_i$) and in the single-constraint case will decrease its *proportional* contribution to the objective function.

6.1 Adjustments scored by contribution to objective function $r_i^2 w_i$



6.2 Adjustments scored by weighted adjustment $r_i w_i$



6.3 Inequality-condition diagnostics

The main category of inequality conditions is that of sign constraints, e.g. items that cannot conceptually be negative.

Such a constraint will be considered "satisfied" if the x-variable in question is greater than or equal to zero. However, even when a zero or near-zero value is conceptually possible, it may be implausible. For example, if the unbalanced estimates show a non-zero value for production of some good, it's most likely because at least one business has reported producing that good, and so estimating a zero value for the entire economy would be hard to justify.

These expectations could be enforced by modifying constraints (e.g. replace "not negative" constraints with "balanced value cannot be less than 50% of unbalanced value"), noting that over-zealous constraints may increase the risk of creating an unsolvable problem, or of forcing unrealistically large adjustments elsewhere.

Whether or not this is done, a solution that lies near or on its inequality constraints may signal problems in some aspect of the inputs.

Hence it may be desirable to flag non-zero values that have been adjusted to zero (or exactly matching some non-zero constraint), and possibly those that have been adjusted to less than e.g. 50% of their unbalanced magnitude.

6.4 Summary of balancing diagnostics

In summary of the above, diagnostics for a balancing run might include:

- List of the top-N largest adjustments from unbalanced values (in dollar value) along with context on these adjustments: e.g. magnitude of adjustment as percentage of original value, magnitude of revisions to percentage growth, magnitude of discrepancies for relevant rows/columns, comparison to adjustments from recent years.
- List of the top-N largest adjustments in proportion to the unbalanced value, with context as above, excluding those where a large proportional change would not be of concern (e.g. the unbalanced value is very small).
- List of the top-N most "costly" adjustments as measured by the objective function $r_i^2 w_i$.
- List of the top-N most "costly" adjustments as measured by $r_i w_i$.
- List of rows, columns, and other constraint groups receiving unusually large adjustments (relative to other comparable entities) as measured by the objective function.

- List of inequality conditions that are exactly met (i.e. solution lies on boundary), excluding those that were already on the boundary before balancing—may be redundant with "large proportional adjustments" above.
- List of cells that have been adjusted to less than e.g. 50% of unbalanced magnitude—again, may be redundant with "large proportional adjustments".

These lists can then be examined to identify any results that seem questionable.

Section 4.1.1 shows examples of graphical diagnostics, as used to compare the behaviour of different balancing options.

6.5 Corrections following diagnostics

In some cases examination of diagnostics may identify problems, e.g. a table value that should have been corrected before auto-balancing, or has been given an incorrect weight.

In such cases it may be appropriate to correct the inputs and rebalance the table. However, in order to keep the balancing process as objective as possible, it's desirable that there should be clear guidelines in advance about when it's appropriate to do this.

In particular, cell weights should not be increased merely because the cells have received an unexpectedly large adjustment. Doing so risks biasing outputs to meet user expectations, and may also cause increased adjustments to other cells. In some cases, increasing weights will not substantially reduce the size of an adjustment (e.g. if constraints leave no other way to resolve a large discrepancy). In this case it may be appropriate to interpret the results as an indication that we have under-estimated variance for these cells, in which case weights should be *decreased* (which will not reduce the adjustments, but may prevent them from being flagged as anomalous).

Other areas of ABS work require similar considerations about whether to adjust an estimation process after viewing its outputs, e.g. decisions on whether to outlier an atypical survey respondent. These areas have developed processes to govern such decisions and reduce the risk of user expectations biasing results; it may be appropriate to adapt such processes here.

7. NONLINEAR CONSTRAINTS

Some balancing applications require the use of nonlinear constraints and/or non-quadratic objective functions. For example, balancing that incorporates past and present prices may require that the balanced data satisfy conditions such as $\tilde{x}_1\tilde{x}_4 = \tilde{x}_2\tilde{x}_3$.

This requirement cannot be expressed without either introducing a nonlinear constraint, or reformulating to eliminate the nonlinear constraints at the cost of introducing a non-quadratic objective function. Optimisation programs exist that are capable of handling nonlinear constraints and/or non-quadratic objectives. However, increasing generalisation often comes with costs, e.g.:

- Increased computing requirements;
- Complex and less transparent problem specification;
- Reduced choice in software options (ideally ABS balancing would use the same optimisation tools as other applications across the organisation).

For these reasons it may be desirable to solve such problems using tools designed only for linear constraints and quadratic objective functions. Here we discuss some options for doing so.

7.1 Approximating constraints/objectives

If we expect that our unbalanced inputs are reasonably close to the true values (and hence that the balanced outputs will also be close) then we can construct linear approximations to the constraints and/or quadratic approximations to the objective function.

For example, if adjustments are expected to be small relative to initial values, the requirement $\tilde{x}_1\tilde{x}_4 = \tilde{x}_2\tilde{x}_3$ can be approximated as:

$$\hat{x}_1\tilde{x}_4 + \hat{x}_4\tilde{x}_1 - \hat{x}_1\hat{x}_4 = \hat{x}_2\tilde{x}_3 + \hat{x}_2\tilde{x}_3 - \hat{x}_2\hat{x}_3 .$$

Where the adjustments are expected to be large, this method can be iterated. We can use the linear approximation above to calculate a first-run set of balanced estimates $\tilde{x}^1$ and then substitute in the estimates from $\tilde{x}^1$ to produce a more accurate set of constraints for a second balancing run:

$$\tilde{x}_1^1\tilde{x}_4^2 + \tilde{x}_4^1\tilde{x}_1^2 - \tilde{x}_1^1\tilde{x}_4^1 = \tilde{x}_2^1\tilde{x}_3^2 + \tilde{x}_3^1\tilde{x}_2^2 - \tilde{x}_2^1\tilde{x}_3^1 .$$

For this second run, $\tilde{x}^1$ is treated as constant and $\tilde{x}^2$ is the vector of variables to be optimised.

In some cases, iterative approaches of this sort can fail to converge and may even diverge: consider e.g. the "overshoot" behaviour of *Newton*'s method when attempting to find a zero for $y = x^{1/3}$). Use of iterative balancing will require further consideration of this issue.

In a case that requires a non-quadratic objective function, a similar approach can be used: we estimate some quadratic approximation to the OF, optimise for this approximation, then calculate derivatives of the exact OF at the approximate solution and use these to construct a new quadratic approximation, repeating as necessary.

7.2 Penalty function methods

As an alternative to "hard constraints" (enforced exactly), we can use "soft constraints" that allow some inexactness. This can be done by incorporating the constraint into the objective function by means of a penalty term.

For example, Statistics Netherlands (see Eurostat) includes linearised soft ratio constraints via an objective function term of the form $w\left(\tilde{x}_i - d\tilde{x}_j\right)^2$ where d is the required ratio. Note that as $\tilde{x}_i$ and $\tilde{x}_j$ become small this term approaches zero even when $\tilde{x}_i / \tilde{x}_j$ is not close to d, implying that this "constraint" becomes weaker for small values (unless weighted to counteract this).

One option here is to solve with weak or zero-weighted soft constraints, check for constraint violations, and iteratively increase the weight for constraints that show large violations until the solution is within acceptable tolerances.

8. SCALING CONSIDERATIONS

Balancing problems can be expanded almost indefinitely, to finer levels and additional dimensions of classification, more years of data, etc. etc. Sooner or later this will lead to problems that are too large to solve with available time and resources.

Several companies provide powerful cloud-based "optimisation as a service" solutions, where problems can be supplied to a large cluster of servers optimised for such work. However, cloud solutions may not be appropriate where confidential data is involved.

For such cases, a divide-and-conquer approach can be used: break the problem into smaller sub-problems that can be combined to give an approximate optimal solution. For example, Input-Output tables entail a similar balancing problem to Supply-Use, but at finer classifications resulting in more variables in the optimisation. If this makes the system too large to solve, one option might be to balance at Supply-Use level, then use the Supply-Use values as additional constraints on the Input-Output problem; this effectively reduces the number of independent variables in the optimisation, and depending on the solution algorithm may reduce the difficulty of the problem.

For multi-year balancing problems, rebalancing/rebenchmarking can potentially cause changes to data many years in the past. Through appropriate choice of constraints the back series can be left open to adjustment, or older years can be fixed, as preferred. However, a long-running benchmarking system with many variables can result in a very large optimisation problem.

One option for dealing with this is a "rolling window" approach: e.g. we balance the data for years 1–5, ignoring later years, then rebalance for years 1–6 but fixing year–1 data to the values previously calculated, then rebalance for years 1–7 with years 1–2 fixed, and so on.

Another option for reconciling large groups of time series that have both temporal and contemporaneous constraints is to use a two-step procedure that first benchmarks each annual series individually, optimising movement preservation within that series, and then adjusts to achieve contemporaneous balancing within each low-frequency time period (e.g. a year). See Di Fonzo and Marini (2009) and Fortier and Quenneville (2009) for details on such methods and discussion of other work in this area. Cycling methods may also be used (Lenzen *et al.,* 2012b).

9. CONCLUDING REMARKS

Optimisation-based adjustment has great potential for improving both the quality of ABS economic statistics and the efficiency of their production.

One of the major challenges in optimisation balancing for National Accounts and similar tables will be managing size and complexity. Some of this can be achieved through appropriate software tools (outside the scope of this paper) but this must also be considered on the theoretical side.

Over-engineering for theoretical accuracy increases the difficulty of maintenance and can in fact become a risk to real-world data quality, since complex systems are harder to interpret, debug, and maintain. Hence, optimisation balancing/benchmarking systems should favour simple approaches that produce reasonable outputs, rather than aiming for theoretical perfection.

A major objective of the ABS transformation program is to increase automation in the production of economic statistics. Other agencies have already used WLS adjustment methods for these purposes, but typically these methods still require subjective involvement in determining accuracy weights. This is likely to require considerable effort by subject matter experts and ongoing maintenance to ensure that the weighting models are still appropriate.

The MLE approach discussed in this paper needs further exploration to confirm its viability, but it has several attractive features:

- It offers an objective method for making weighting decisions, one which can largely be automated.
- It clearly identifies assumptions about the structure of measurement errors, and allows those assumptions to be changed.
- It is consistent with simple balancing/benchmarking methods already in use (WLS level-preservation *or* AFD/PFD movement-preservation).
- It suggests that simply adding movement- and level-preservation objective functions together may not be the most appropriate way to handle a combination of correlated and uncorrelated errors, and offers a solution that can still be applied within a WLS framework.

A possible next step would be to test this approach on realistic, synthesised data: for instance, start with a balanced SU table, perturb it under different error models, and then test how accurately the MLE method and other candidates recover the unperturbed table and estimate the parameters of the error model. This simulation approach can also be used for other investigations, e.g. testing sensitivity to normality assumptions.

For reasons related to the ABS transformation timetable, our work to date has focused on Supply-Use balancing, but we expect to extend these ideas to other economic outputs, and potentially to other ABS estimation processes with similar requirements.

ACKNOWLEDGEMENTS

I would like to thank team members past and present (Carl Mackin, Alex Stuckey, Tom Davidson, Bernadette Fox, Kin Chung, Linh Huynh) for their contribution to this work and their encouragement. In particular, Tom did the balancing for the 2010–2011 Supply-Use data shown in Section 4. Thanks also go to our managers (Philip Bell, Ruel Abello) for their encouragement.

Further thanks to our National Accounts colleagues including Simon Usback, Kieran Burgess, Katherine Keenan, and Olivier Brunet, for providing data and feedback on balancing requirements.

Kevin Fox and other members of the ABS Methodology Advisory Committee gave encouragement and made useful suggestions on further areas of investigation and assumptions to consider.

Our international colleagues provided a great deal of helpful information, including Baoline Chen at the U.S. Bureau of Economic Analysis; Jacco Daalmans, Nino Mushkudiani and Reinier Bikker at Statistics Netherlands (CBS); Jeroen Kole, Richard Penny, and John Créquer at Statistics New Zealand; and Craig McLaren at the U.K. Office for National Statistics.

Manfred Lenzen of the University of New South Wales provided information on RAS methods and corrected some errors in my discussion of these.

Peter Rossiter provided advice and assistance with editing and formatting this paper.

REFERENCES

Bacharach, M. (1965) "Estimating Nonnegative Matrices from Marginal Data", *International Economic Review*, 6(3), pp. 294–310. <http://www.jstor.org/stable/2525582>

Bradley, E.L. (2009) "The Equivalence of Maximum Likelihood and Weighted Least Squares Estimate in the Exponential Family", *Journal of the American Statistical Association,* 68(341), pp. 199–200.

Brent, G.; Stuckey, A. and Davidson, T. (2015) "Reducing Forward-Series Errors for Benchmarked Quarterly National Accounts", *EURONA – Eurostat Review on National Accounts and Macroeconomic Indicators*, 2015(2), pp. 49–70. <http://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-GP-15-002>

Chen, B. (2012) "A Balanced System of U.S. Industry Accounts and Distribution of the Aggregate Statistical Discrepancy by Industry", *Journal of Business and Economic Statistics*, 30(2), pp. 202–211.

Chen, B.; Di Fonzo, T.; Howells, T. and Marini, M. (2014) "The Statistical Reconciliation of Time Series of Accounts after a Benchmark Revision", *JSM 2014 – Business and Economic Statistics Section*, pp. 3076–3087.

Dagum, E.B. and Cholette, P.A. (2006) "Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series", *Lecture Notes in Statistics 186,* Springer-Verlag, New York.

Di Fonzo, T. and Marini, M. (2009) "Simultaneous and Two-step Reconciliation of Systems of Time Series", *Working Paper Series, N. 9,* Department of Statistical Sciences, University of Padua, Italy.

Eurostat (2013) "Annex C – Benchmarking and Balancing Methods in the Dutch National Accounts", *Handbook on Quarterly National Accounts*. <http://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-GQ-13-004>

Fortier, S. and Quenneville, B. (2009) "Reconciliation and Balancing of Accounts and Time Series: From Concepts to a SAS Procedure", *ASA Proceedings of the Business and Economic Statistics Section*, American Statistical Association.

International Monetary Fund (2014) "Chapter 6. Benchmarking and Reconciliation (Draft)", Update of the *"Quarterly National Accounts Manual: Concepts, Data Sources and Compilation"*. <https://www.imf.org/external/pubs/ft/qna/>

Koch, T. *et al.* (2011) "MIPLIB 2010: Mixed Integer Programming Library, Version 5." *Mathematical Programming Computation*, 3, pp. 103–163. <http://plato.asu.edu/ftp/miplib5.pdf>

Lenzen, M.; Gallego, B. and Wood, R. (2006) "A Flexible Approach to Matrix Balancing Under Partial Information", *Journal of Applied Input-Output Analysis*, 11/12, pp. 1–24.

Lenzen, M.; Gallego, B. and Wood, R. (2009) "Matrix Balancing Under Conflicting Information", *Economic Systems Research,* 21(1), pp. 23–44.

Lenzen, M.; Kanemoto, K.; Moran, D. and Geschke,A. (2012a) "Mapping the Structure of the World Economy", *Environmental Science & Technology*, 46(15), pp. 8374–8381.
<http://dx.doi.org/10.1021/es300171x>

Lenzen, M.; Moran, D.D.; Geschke, A. and Kanemoto, K. (2014) "A Non-Sign-Preserving RAS Variant", *Economic Systems Research*, 26(2), pp. 197–208.

Lenzen, M.; Moran, D.; Kanemoto, K. and Geschke, A. (2013) "Building EORA: A Global Multi-Region Input-Output Database at High Country and Sector Resolution", *Economic Systems Research*, 25(1), pp. 20–49.

Lenzen, M.; Pinto de Moura, M.C.; Geschke, A.; Kanemoto, K. and Moran, D.D. (2012b) "A Cycling Method for Constructing Input-Output Table Time Series from Incomplete Data", *Economic Systems Research*, 24(4), pp. 413–432.

McDougall, R.A. (1999) "Entropy Theory and RAS are Friends", *GTAP Working Papers, Paper 6*, Purdue E-Pubs, Purdue University.
<http://docs.lib.purdue.edu/gtapwp/6/>

Nicolardi, V. (2011) *Supply-Use Tables: Simultaneously Balancing at Current and Constant Prices. A New Procedure*, Paper presented to the 19th International Input-Output Conference, Alexandria, Virginia, 13–17 June 2011.
<https://www.iioa.org/conferences/19th/papers/files/296_20110429111_IIOAArticle.pdf>

Rassier, D.G.; Howells, T.F. III; Morgan, E.T.; Empey, N.R. and Roesch, C.E. (2007a) *Implementing a Reconciliation and Balancing Model in the U.S. Industry Accounts*, U.S. Department of Commerce, Bureau of Economic Analysis.
<http://www.bea.gov/papers/pdf/rassier_reconciliation_ind.pdf>

Rassier, D.G.; Howells, T.F. III; Morgan, E.T.; Empey, N.R. and Roesch, C.E. (2007b) *Integrating the 2002 Benchmark Input-Output Accounts and the 2002 Annual Industry Accounts*, U.S. Department of Commerce, Bureau of Economic Analysis.
<http://www.bea.gov/scb/pdf/2007/12%20December/1207_indyaccount.pdf>

Stone, R.; Meade, J.E. and Champernowne, D.G. (1942) "The Precision of National Income Estimates", *Review of Economic Studies,* 9(2), pp. 111–125.

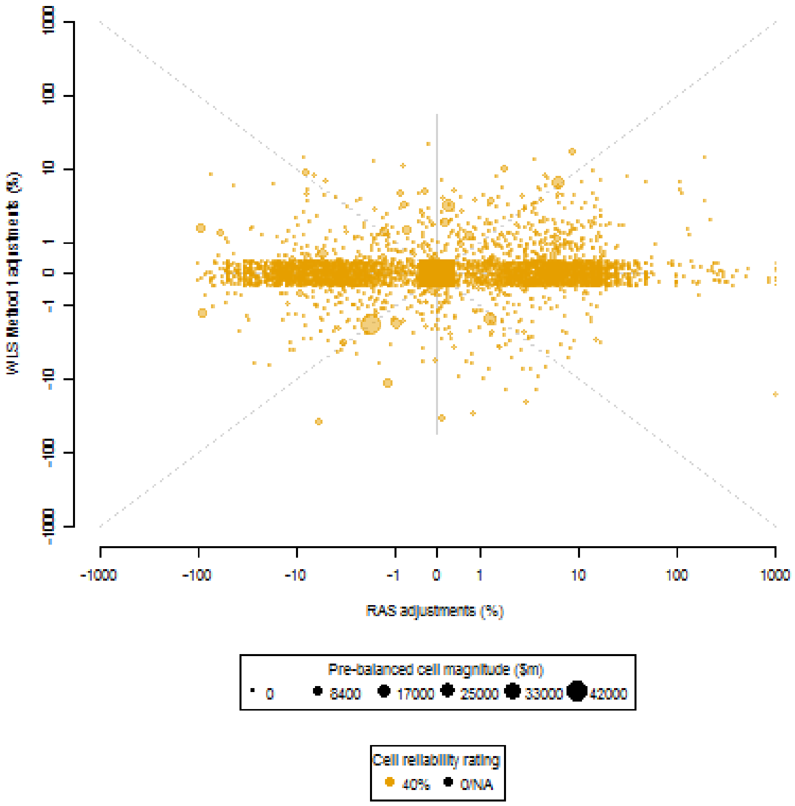
APPENDIX

A. DIAGNOSTIC PLOTS

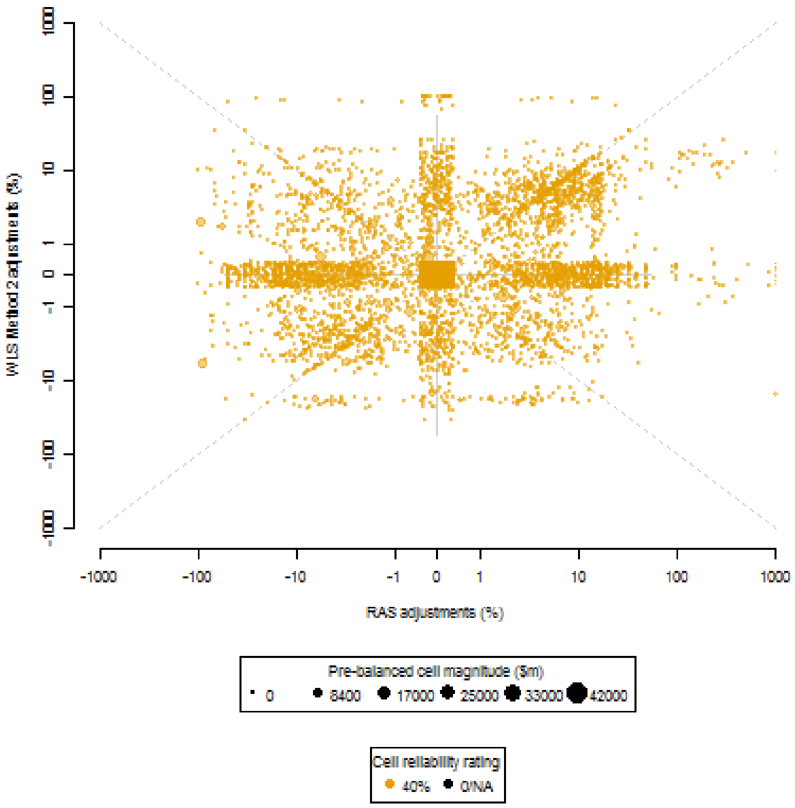
Plots in this section follow the form discussed in Section 4.1.1.

Comparisons of proportional adjustments under RAS and under WLS Methods 1 and 2, separated by cell reliability rating.

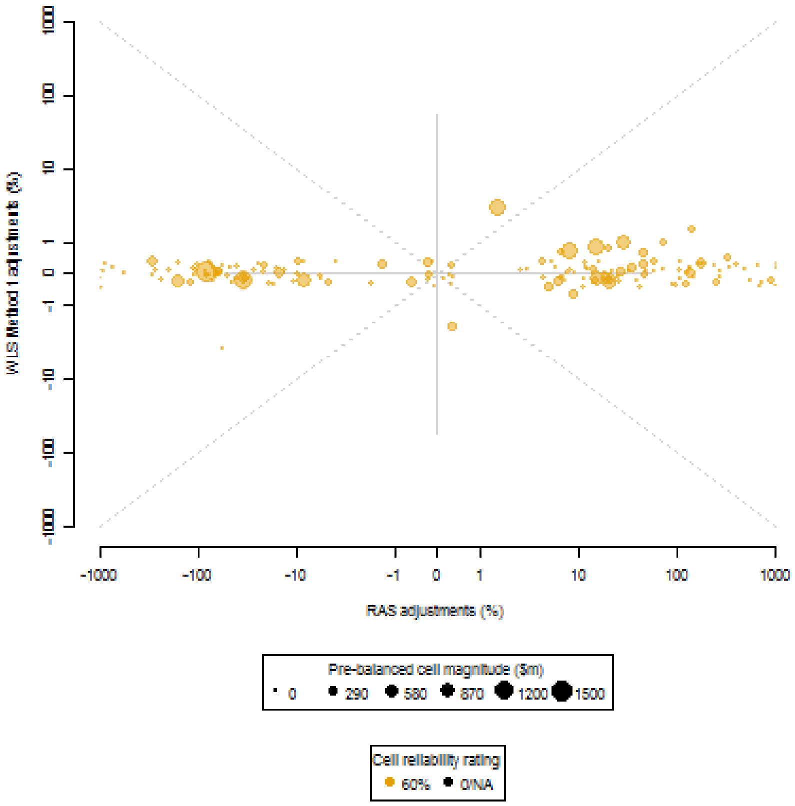
RAS vs WLS Method 1, proportional adjustments, 40% cells



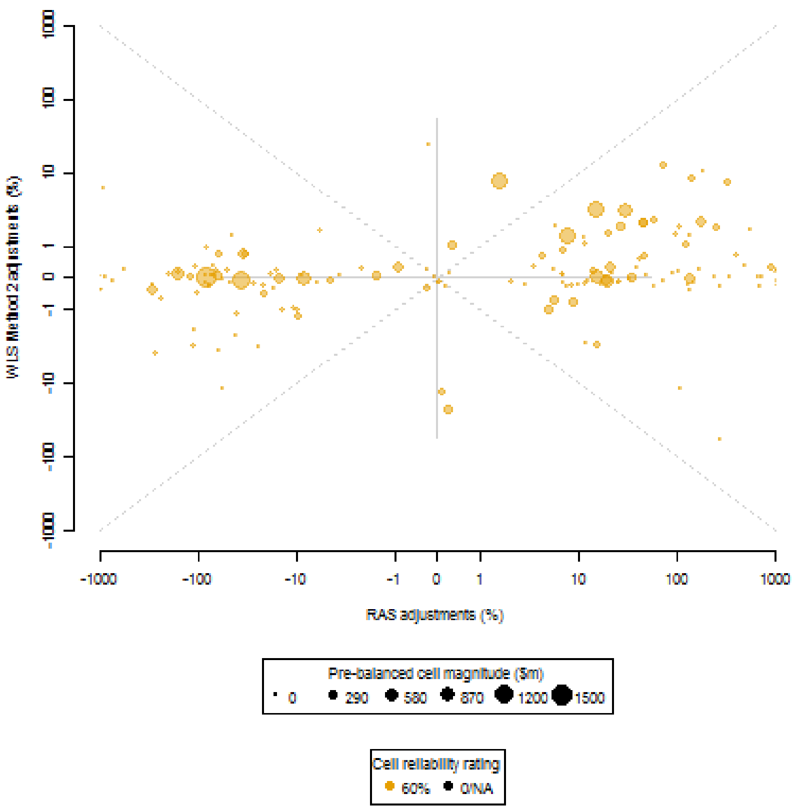
RAS vs WLS Method 2, proportional adjustments, 40% cells



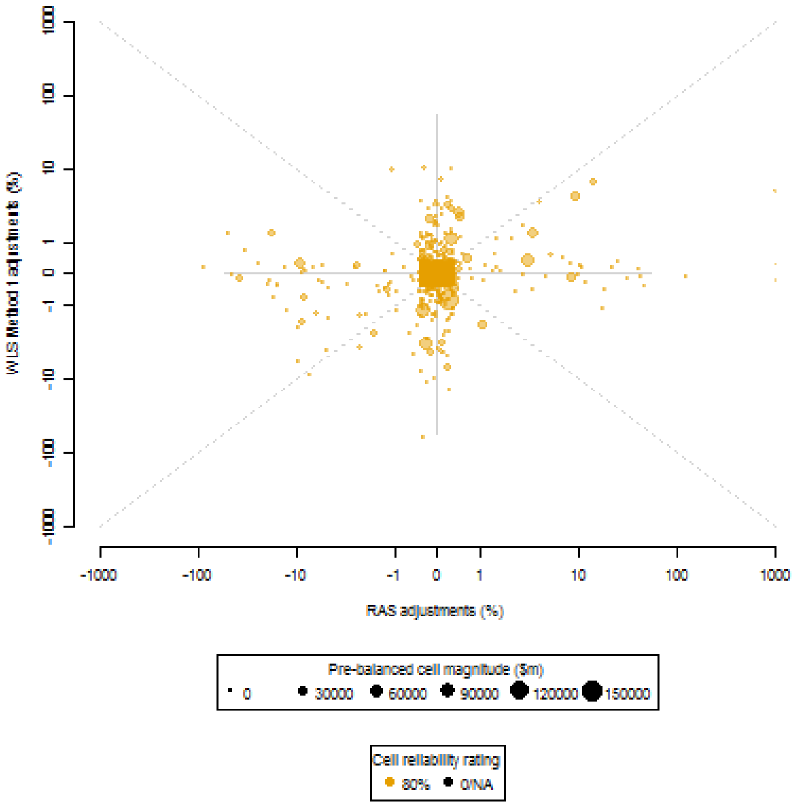
RAS vs WLS Method 1, proportional adjustments, 60% cells



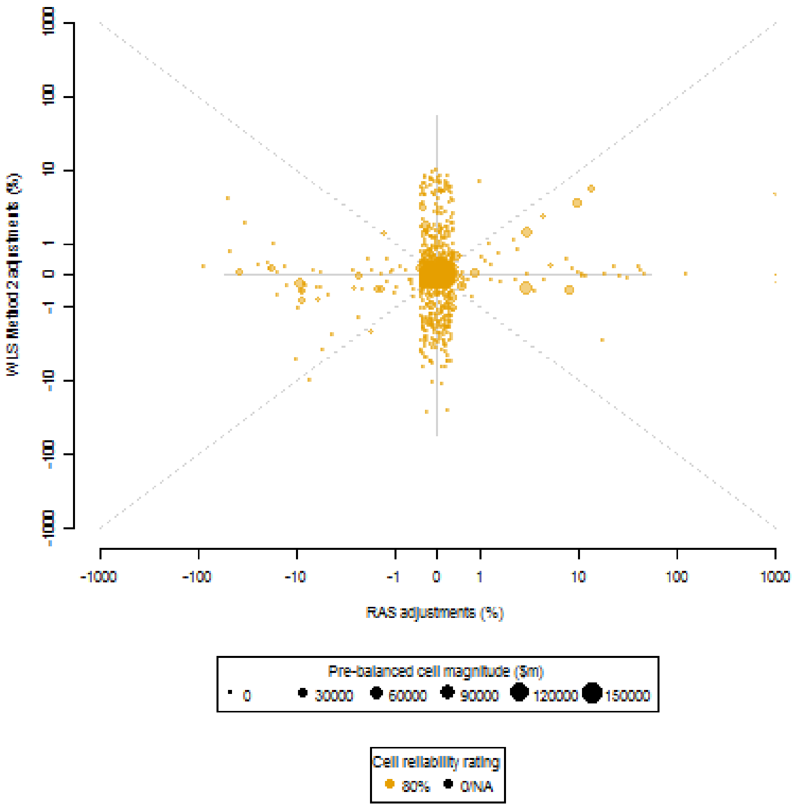
RAS vs WLS Method 2, proportional adjustments, 60% cells

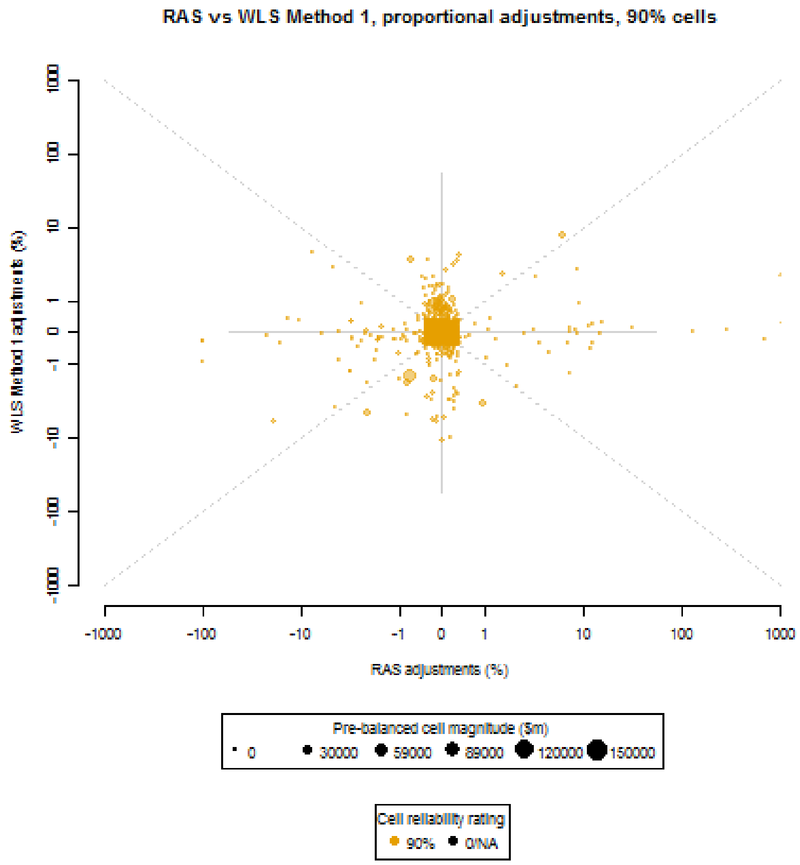


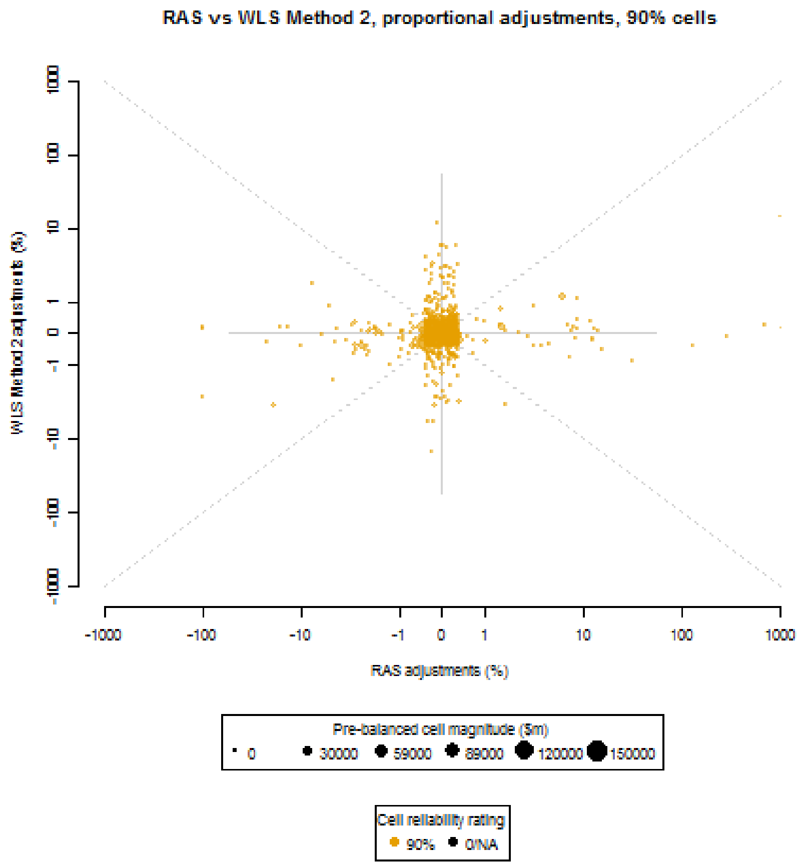
RAS vs WLS Method 1, proportional adjustments, 80% cells

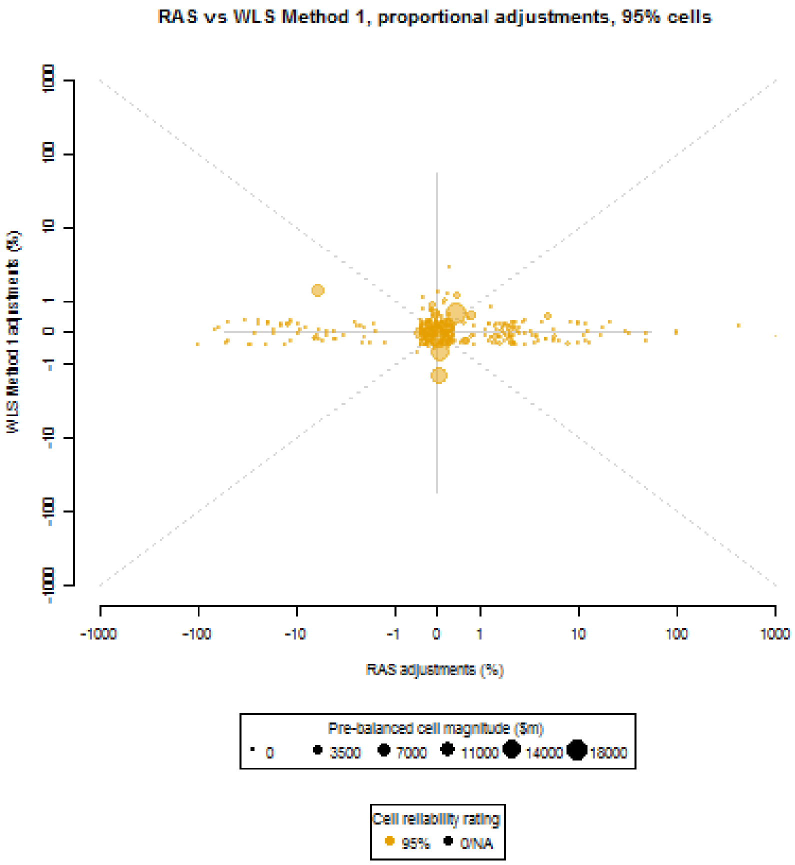


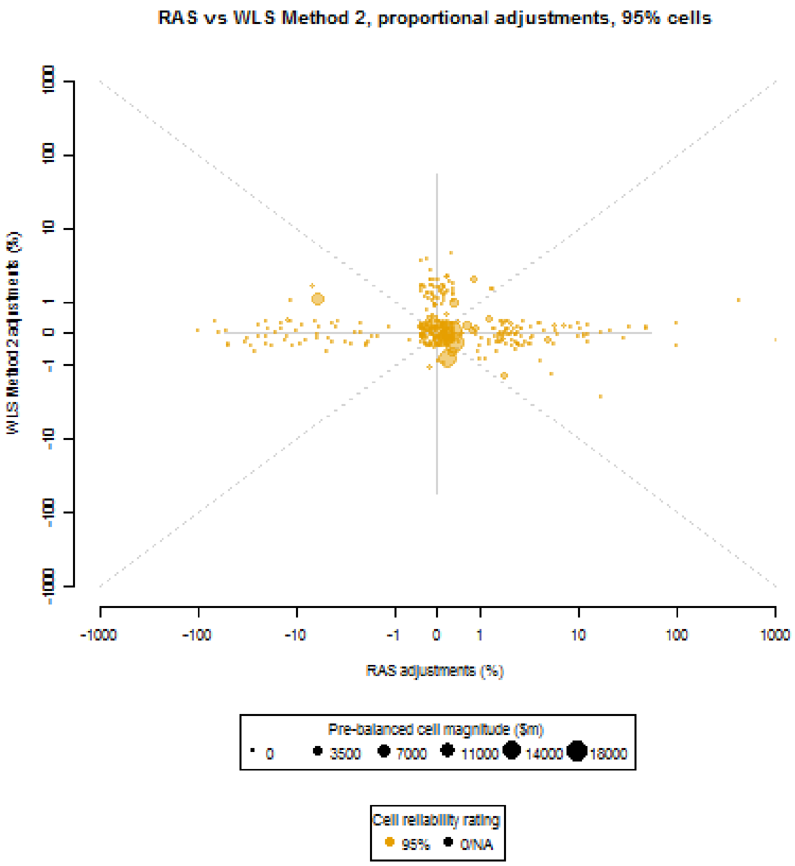
RAS vs WLS Method 2, proportional adjustments, 80% cells




RAS vs WLS Method 1, proportional adjustments, 90% cells
WLS Method 1 adjustments (%)
1000
100
10
1
0
-1
-10
-100
-1000
-1000
-100
-10
-1
0
1
10
100
1000
RAS adjustments (%)
Pre-balanced cell magnitude ($m)
0
30000
59000
89000
120000
150000
Cell reliability rating
90%
0/NA


RAS vs WLS Method 2, proportional adjustments, 90% cells
WLS Method 2 adjustments (%)
1000
100
10
1
0
-1
-10
-100
-1000
-1000
-100
-10
-1
0
1
10
100
1000
RAS adjustments (%)
Pre-balanced cell magnitude ($m)
0
30000
59000
89000
120000
150000
Cell reliability rating
90%
0/NA


RAS vs WLS Method 1, proportional adjustments, 95% cells
WLS Method 1 adjustments (%)
RAS adjustments (%)
-1000
-100
-10
-1
0
1
10
100
1000
Pre-balanced cell magnitude ($m)
0
3500
7000
11000
14000
18000
Cell reliability rating
95%
0/NA


RAS vs WLS Method 2, proportional adjustments, 95% cells
WLS Method 2 adjustments (%)
RAS adjustments (%)
-1000
-100
-10
-1
0
1
10
100
1000
Pre-balanced cell magnitude ($m)
0
3500
7000
11000
14000
18000
Cell reliability rating
95%
0/NA

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au** The ABS website is the best place for data from our publications and information about the ABS.

LIBRARY A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our website for a list of libraries.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS www.abs.gov.au