# 8 How should the Australian Energy Regulator use benchmarking?

|  |
| --- |
| Key points |
| * Recent rule changes are likely to increase and improve the AER’s use of benchmarking. The AER will be able to use benchmarking to consider whether network expenditure proposals are reasonable and, potentially, to estimate values for some cost categories. The AER will also begin to publish annual benchmarking reports. * At this stage, aggregate benchmarking models are ill-suited to setting regulatory revenue allowances (in place of building blocks). In the immediate future, benchmarking would be most useful: * as a diagnostic tool to help assess the reasonableness of bottom-up proposals * in providing information to consumers and others, thereby providing pressure for improved performance by network businesses. * Over time, benchmarking may take a larger role in determining revenue allowances. Where benchmarking is used to estimate substitute values for opex and capex allowances, this should involve: * demonstration that the results are robust through detailed publication and peer review * choosing a yardstick more akin to that applying in competitive markets — which would be a firm close to, but not at the efficiency frontier. * In the future, benchmarking may also facilitate negotiated arrangements that bypass the current costly and protracted regulatory processes. * The AER will need to adopt various processes to ensure the successful use and evolution of benchmarking, including: * the development of publicly available databases and full transparency in its processes and methods * the development of internal expertise, and strategies to maximise learning * international collaboration and peer review of its benchmarking practices (‘benchmarking of benchmarking’) * appropriate consultation with stakeholders about data and methods * effective communication of the results of benchmarking to its diverse audiences * regular checking to ensure that the benefits of its benchmarking practices exceed the compliance and resource burdens. |
|  |
|  |

A major international survey ranked Australia as a relatively unsophisticated user of benchmarking in electricity networks (Haney and Pollitt 2011). End users have also criticised the limited use of benchmarking in the current regulatory regime:

Our view is that benchmarking has generally had an insignificant role in the AER’s determination of expenditure allowances. We have observed that in most of its determinations there is no evidence that the AER has benchmarked capitalised expenditure allowances at all. The benchmarking that it has done of operating expenditures has not, in our opinion, been adequate. Even where there is some evidence of benchmarking by the AER, there is no evidence of how this information affected its view of the appropriate expenditure allowances. (EUAA, sub. 24, p. 4)

The AER has not used benchmarking effectively and yes it should adopt different practices. The reasons for this probably lie with the regulatory approach (propose/respond) the AER must implement. (MEU, sub. 11, p. 30)

This raises the question of the appropriate aspirations for benchmarking in an Australian context — recognising the limits raised in chapter 4. The appropriate type of benchmarking and the manner of its application will depend on the purposes of its use — the policy imperative is ‘do not use benchmarking on its own account’. There are many such purposes, with implications for the degree to which the tests set out in chapter 4 would need to be met.

Rule changes introduced in late 2012 require the AER to undertake routine benchmarking and give it the discretion, though not the obligation, to use benchmarking in making price and revenue determinations (AEMC 2012r; box 8.1; and box 4.2 in chapter 4). Given this greater discretion, it is particularly important to be clear about how the AER should use benchmarking, both now, and as the sophistication of the data and methods evolve — the subject of this chapter.

Some potential uses of benchmarking are alternatives, but some are complementary.

* Section 8.1 considers whether the AER should use benchmarking as the primary basis for its revenue determinations, and highlights some of the major drawbacks in that approach at this stage.
* Section 8.2 examines the degree to which higher-level benchmarking could indicate the overall effectiveness of the regulatory regime, which would have implications for policy settings, as well as provide a guide to whether the regime is achieving its intended purposes.
* The bottom-up approach in determining revenue allowances is often exhaustive in its detail — one of the reasons for the increasingly lengthy documents used in the propose-respond model. The AER could use benchmarking to determine the areas where detailed analysis is required, while avoiding excessive analysis in other areas — in effect a filter for more targeted analysis (section 8.3).

|  |
| --- |
| Box 8.1 Recent AEMC rule changes on benchmarking |
| *There will be a published benchmarking report each year*  The first most important change is that the new Rules stipulate that the AER will undertake and publish regular benchmarking reports, with the contents of the reports to be decided by the AER:  The AER must prepare and publish a network service provider performance report (an annual benchmarking report) the purpose of which is to describe, in reasonably plain language, the relative efficiency of each Distribution Network Service Provider in providing direct control services over a 12 month period. (s. 6.27 of the Rules, v. 54)  Such reports will require data of a sufficient detail, quality and comparability. The AEMC indicated that this will require an expansion of the AER’s data collection (2012r, p. 108) — an issue that this chapter covers.  The initial benchmarking report is due by September 2014, but the Rules are (understandably) quiet about the scope of the reports or any requirements for testing the validity and reliability of the results, or in meeting any of the other criteria set out in chapter 4. It is notable that in the Rule change report, the AEMC indicated that in undertaking the annual benchmarking analyses of network businesses, the AER should take ‘into account the exogenous factors that distinguish them’ (AEMC 2012r, p. 25). No such requirement exists in the Rules themselves, but such a principle would be important in any AER benchmarking exercise.  In preparation for the annual benchmarking report and for the regulatory use of benchmarking, the AER (2012y) released an issues paper in late December 2012 concerning guidelines for assessing efficient expenditure forecasts for network businesses.  *The AER may use benchmarking results to determine substitute estimates in revenue determinations*  The AER must accept reasonable proposals by network businesses. However, the processes by which it determines reasonableness — which may include benchmarking — can also be used to set alternative revenue allowances where a proposal is deemed unreasonable. As the AEMC put it:  While the AER must form a view as to whether a [network service] provider’s proposal is reasonable, this is not a separate exercise from determining an appropriate substitute in the event the AER decides the proposal is not reasonable. For example, benchmarking the [network service provider] against others will provide an indication of both whether the proposal is reasonable and what a substitute should be. Both the consideration of ‘reasonable’ and the determination of the substitute must be in respect of the total for capex and opex. (AEMC 2012r, p. 112)  Accordingly, the Rules would enable the AER to use benchmarking in a determinative sense. This clarification does not compel the AER to use the results of benchmarking models as substitute values — it may still rely on other methods (such as bottom-up analysis). |
| *Sources*: Version 54 of the Rules and AEMC (2012r). |
|  |
|  |

* Currently, end-users (whether households or commercial users) are disenfranchised from the regulatory process. While greater engagement should occur regardless of the form of the regulatory model (chapter 21), it may also be possible for end-users to play an active role in reaching negotiated settlements in regulatory determinations — avoiding the complex and protracted processes currently in place. Benchmarking would support such a framework (section 8.4).
* Benchmarking need not always directly inform regulatory decisions. In some cases, the publication of benchmarking results may itself create pressures for improved performance (section 8.5).
* While this chapter is sceptical of the degree to which benchmarking could play a major role in determining regulatory allowances in the near future, that may change over the longer-term with the increasing sophistication of the models and with better data. In that instance, benchmarking could assume a more prominent role (section 8.6).

Regardless of the particular purposes of benchmarking, the regulator has to develop competencies in benchmarking and follow processes that ensure that others can interpret and use the results, and that compliance and other costs associated with benchmarking are moderated. Section 8.7 spells out the processes that will achieve cost‑effective and useful benchmarking.

## 8.1 Should benchmarking be used in a mechanistic role to set revenue allowances?

The apparently greater simplicity and clarity of using aggregate benchmarking as the primary basis for price and revenue determinations lies behind its attraction for some parties. In principle, its aggregate nature would require less data than for bottom-up analysis, while using it mechanistically to set allowances might avoid the prolonged processes apparent in recent regulatory determinations. In some circumstances, it might create stronger incentives for cost minimisation than the current arrangements.

There are several options. The AER could use benchmarking to:

* set expenditure allowances within the existing building block framework
* set revenue growth, based on a total factor productivity (TFP) growth framework, but without building blocks
* set revenue, based purely on benchmarking results, without building blocks.

#### Setting aggregate capex and opex in building block models

A regulator (or indeed a business proponent) could use aggregate benchmarking models to determine forecasts of total efficient opex and capex, which would then be included as the key inputs of the standard building blocks model.[[1]](#footnote-1)

This use of benchmarking resembles that outlined by the AEMC (2012r), where it clarified the AER’s ability to determine estimates of total opex and capex using benchmarking models (box 8.1). It would still represent a significant departure from what was common practice in the first round of AER determinations, where deriving the total cost forecasts in the building blocks model involved aggregating many detailed sub-components of total spending (as described by Major Energy Users, sub. 11, p. 10), with benchmarking at best informing that process.

#### Setting revenue growth based on total factor productivity growth (CPI-x)

A second mechanistic approach would be to allow benchmarking results to determine the trajectory of a business’s annual price or revenue increases (under a CPI-x model).[[2]](#footnote-2) This would involve setting a starting price, P0, based on efficient or reasonable costs. The trajectory would be then determined by setting *x* to the *industry-wide* total factor productivity growth calculated using standard index methods.

This would not require a building blocks model at all, since it would focus at the highest possible level of aggregation (AEMC sub. 16, p. 1). The National Electricity Law already allows the AER to use the TFP approach as either a replacement for, or a complement to, the building block approach (clause 26J).

As the TFP approach uses industry-wide data to determine *x*, individual businesses would have strong incentives to cost minimise, and would have few opportunities for gaming, such as by exaggerating their efficient costs or by producing unrealistic demand forecasts (Kaufmann 2006, 2007; ESC 2006, 2009; DPI 2009, p. 4; Pacific Economics Group, sub. 35 and sub. DR48).

#### Setting revenue based on ‘supra- aggregate’ benchmarking model

A third approach would use ‘supra-aggregate’ benchmarking at the commencement of any regulatory period to set the required total revenue allowance. Similar to the CPI-x approach above, this would not require any reference to the role of the separate WACC, capex and opex allowances. However, in contrast to the CPI-x approach, there would not be any reference to the historical costs of the business (since this model would not set P0 and its trajectory).

Rather, the model could estimate revenues for a given regulatory period — similar to models used by Mountain (2011) and Mountain and Littlechild (2010).[[3]](#footnote-3) Each network business would then make all the relevant choices about how to provide its services, including choices between capex and opex. Furthermore, the regulator would not need to formally roll any capital into a regulated asset base (RAB) in the next period, but would need to ensure that the benchmark method related to the long-run marginal costs of supplying services. If the regulator took such a long-run perspective, the risk of asset stranding (devaluation of assets) would be low.[[4]](#footnote-4) (The National Electricity Objective’s long-term focus would oblige the AER to do this.)

This approach would still need to estimate future levels of demand at efficient prices (similar to the approach discussed in chapter 11), and take account of differences in network operating environments. Accordingly, it would not eliminate the need for some sophisticated analysis. Nevertheless, on the face of it, it would be simpler than the current building block approach.

### Problems with deterministic approaches

While apparently simpler than the approaches used by the AER thus far, all of these approaches have their own theoretical and practical difficulties:

In respect of the CPI-x method, the AEMC (2011b) did not believe that the available data were adequate yet, but that it could work as an alternative mechanism for setting allowances in the future. Even if the basis for x were productivity growth within the network industry, it may still be optimistic to expect CPI-x benchmarking to occur over the medium term, given the wave of prospective mergers of New South Wales distribution businesses (and uncertainties about future structural change in the industry), the potential privatisation of networks, and slowly developing data collection.[[5]](#footnote-5) Chapter 4 also noted that a ‘pure’ CPI-x approach could preserve inefficiency or perpetuate rents.

Neither is it clear how regulators would derive x in practice. Comprehensive reviews by Farrier Swier Consulting (2002), the AEMC (2011b) and London Economics (2008) found a wide range of methods used by regulators to actually set x. This includes setting the value of x:

* as an assumed value[[6]](#footnote-6)
* as a catch-up factor to improve business performance over several years
* as a means to freeze prices (x = CPI)
* based on business proposals, or
* as the result of technical analysis using index or econometric methods.

Many of these are pragmatic rather than scientific choices. However, so long as x is not too high,[[7]](#footnote-7) and adjustments to initial revenues are not too dramatic,[[8]](#footnote-8) CPI-x approaches provide strong incentives for productivity, while ensuring business viability and placing a (arbitrary) cap on network price changes for consumers. Accordingly, a reasonable operational rule would be to set an x sufficiently low that it gives the business some scope to keep some of the gains of higher productivity growth.

Given its positive features, ongoing development of TFP indexes would still be useful. If nothing else this would ensure that the AER collects the appropriate data and would allow the AER to assess the degree to which the indexes are reliable. Moreover, there may be greater benefits from using the CPI-x approaches if governments privatise networks across the NEM, as recommended by the Commission — an issue explored further in section 8.6. Nevertheless, as the AEMC notes, the practical realisation of the CPI-x approach is some way off.

Similar — if not greater problems — would beset any deterministic use of benchmarking models reliant on cross-sectional data.[[9]](#footnote-9) Their use would require strong confidence in the results, which would be misplaced given the findings in chapter 4. The deterministic approach has five other significant risks (to which the ‘supra’ aggregate approach would be particularly susceptible).

#### (a) Regulatory opportunism

Were electricity prices to be rising rapidly, there could be pressure on the regulator to choose tougher benchmarks (recognising that even were sophisticated benchmarking models to develop over time, it is likely that they could be tweaked to give higher or lower benchmarks).

As explained by Yarrow (2012), the assumption of regulatory impartiality may not hold given opportunism:

By way of further example of the difficulties with the arguments as they presently stand, consider the argument that the current ‘propose-respond’ process precludes the AER from substituting ‘impartial’ forecasts of costs for what are claimed to be the biased forecasts of costs that are submitted by the companies. This argument begs a fundamental question. As discussed above, the working presumption in the relevant economics is that a regulator with unconstrained discretion to set price controls will be tempted to opportunism, and that the temptation will be particularly great in circumstances of rate-shock. That is, at bottom, there is an underinvestment problem associated with the regulation of private monopoly. (Yarrow 2012, p. 9)

The extent of regulatory risk perceived by the industry would depend partly on the specific characteristics of the AER itself, including its objectivity; its independence; its abilities and funding; and its general reputation in the industry (a matter discussed further in chapter 21).

#### (b) Future network developments present a challenge for benchmarking models

Unlike some investments where, once made, the additional investment requirements are relatively modest (say a toll road), electricity networks require significant ongoing maintenance, replacement, new connections and other augmentation. Many network costs reflect expectations about the future characteristics of demand and supply, and these can change over time given the effect of climatic variations and peak loads. As such, many costs would be hard to incorporate into ex ante aggregate benchmark models (at least without significant further development). There would need to be a process for approving contingent projects.

#### (c) Benchmarking, information and efficient contracting

The AER’s revenue allowance determinations are effectively contracts over the regulatory period between network businesses and a regulator acting on behalf of end users. Setting efficient contracts requires shared and sufficiently rich information about the nature of costs and contingencies.

In the current regulatory regime, network businesses have large information advantages over the regulator, which they may be able to exploit to increase regulatory revenue allowances. For example, ex ante, they may claim the need for allowances to meet additional demand or to replace specific ageing assets, and then, ex post, invest at lower levels, taking the residual as a surplus. Ex ante, the regulator may not know enough information to challenge highly detailed ‘bottom-up’ costs and demand forecasts. However, the scope for gaming the regulator can be reduced (and the scope for efficient contracts strengthened) if a benchmarking model can be built that:

* uses a small set of verifiable data
* takes account of the key operating differences of the businesses
* adequately predicts efficient costs.

Few participants in this inquiry suggested that such benchmarking models were currently available.

Rather than forcing both regulators and businesses to act as blind players in their dealings with each other, a preferred approach would be to:

* use the recent Rule changes (chapter 5) giving the AER more regulatory discretion to avoid ‘line by line’ assessments of businesses’ regulatory proposals, thus eliminating the principal avenue for the businesses to exploit their informational advantages
* ensure adequate provision of relevant detailed information (and not rely on the exclusive use of aggregate benchmarking).

Reliance on too lean an information set would increase regulatory risk and require the regulator to pay a premium to cover that risk. There would then be a tradeoff between simplicity and cost.

#### (d) Detailed analysis would creep in via the back door

The checks and balances required to ensure aggregate benchmarking was reliable as the sole basis for setting revenue allowances would re-create the need for scrutiny of bottom-up data to test the results (thus losing simplicity after all). At the very least, these checks and balances would need to address capex, opex and the WACC separately (thus reinstating the RAB as an important facet of the regulatory arrangements).

#### (e) Uncertainties about the merit review process

Moreover, it is not clear how merit review would proceed if regulators used aggregate benchmarking to determine revenue allowances. The outcomes would depend on how merit review arrangements were structured (a matter still under consideration following the review by the Limited Merits Review Panel, 2012). But regardless, it is likely that merit reviews would involve:

* battles between econometricians.[[10]](#footnote-10) For example, significant tensions between alternative ways of estimating TFP emerged during the AEMC’s assessment of TFP growth as a viable benchmark (AEMC 2011b and Pacific Economics Group sub. DR48). Such battles would have some advantages to the extent that it encouraged good statistical processes, clear statements of how to interpret the results, better data collection, and the development of expertise (as set out in figure 4.7 in chapter 4). However, given the data and modelling problems afflicting benchmarking, it is doubtful that a merits review body could reach a well-based judgment on whether it is meaningfully possible to, say, disentangle ‘the heterogeneity from inefficiency in one step, using a latent class model for stochastic frontiers’ (Cullmann 2009). A tribunal might then seek its own expert advice, but this may only serve to broaden the contest rather than resolve it
* the re-admission of detailed data as the corroborating evidence on which to base the determination. The bottom-up approach might then simply be deferred to the ‘courtroom’. It is notable that in an appeal brought by EnergyAustralia, the Tribunal emphasised the relevance of the business’s detailed information:

EnergyAustralia is correct to submit that it is not the AER’s role to simply make a decision it considers best. It is also correct for it to say that the AER should be very slow to reject a DNSP’s proposal backed by detailed, relevant independent expert advice because the AER, on an uninformed basis, takes a different view. Nor, as EA submits, may the AER reject such a proposal merely because it has an expert opinion. The AER, based upon any expert advice, needs to make its own evaluation, an evaluation that is reviewable by this Tribunal. (Australian Competition Tribunal 2009, p. 56)

Currently, the arguments for the AER to use benchmarking to set allowances mechanistically are not compelling. The AER, the network businesses and some other stakeholders share this view (box 8.2). Responses on the Commission’s draft report did not repudiate this perspective.

## 8.2 Benchmarking the effectiveness of the regulatory regime

The AER and various stakeholders have claimed that regulatory arrangements have reduced efficiency in electricity networks. They cited several issues that are discussed in various areas of this report, including deficiencies in incentive regulations (chapters 1 and 5), prescriptive reliability settings (chapters 14 to 16), policy obstacles to demand management (chapters 9 to 12) and problems in the efficient utilisation of interconnectors (chapters 18 and 19). As shown in each of those chapters, the Commission agrees that there is a lot of room for improvement in the overall regulatory environment. The case for policy reform is much increased if analysis can reasonably demonstrate that a deficiency in a regulation has material consequences and that a specific change is likely to improve outcomes.

Benchmarking can play a useful role in these areas.

First, as emphasised in chapters 1 and 4, rigorous benchmarking analysis intended to measure business performance can incidentally highlight the efficiency impacts of the regulatory and policy environments facing network businesses. While most of the regulator’s interest in benchmarking is on the *residuals* from a model (the proxies for the inefficiency levels of businesses), policymakers should concentrate on the estimated *parameters* regarding policy variables in any model. In the current imperfect policy environment, those parameters may reveal the most lucrative direction for reform. These policy opportunities are the principal subjects of chapters 9 to 16.

|  |
| --- |
| Box 8.2 Participants said benchmarking should not be used deterministically |
| The AER highlighted concerns about using benchmarking in a determinative fashion:  Benchmarking is not a substitute for rigorous analysis and the exercise of judgement to determine expenditure allowances for a network business and cannot be used in a mechanistic fashion to directly determine expenditure allowances. However, when benchmarking is used prudently and carefully, and based on a robust specification that incorporates good quality data, it can be a very useful tool in the overall assessment of an expenditure proposal. (AER, sub. 13, p. 13)  The AER considers that at the current time it cannot establish revenue allowances based primarily on the outcome of comparative benchmarking against other firms. When more standardised and appropriate data becomes available as a result of the application of the AER's new framework, noted above, and benchmarking models give more consistent results, the weighting given to top down benchmarking as a part of the AER’s comparative analysis will likely increase. (AER 2010b, Appendices, p. 99)  A review of the international use of benchmarking in regulatory agencies worldwide indicated the dangers of using benchmarking to punish/reward businesses:  The significant uncertainties in efficiency estimates could have important undesired consequences especially because in many cases the estimated efficiency scores are directly used to reward/punish individual companies through regulation schemes such as price-cap formulas. (Farsi et al. 2007, p. 13)  Network businesses were (unsurprisingly) hostile to deterministic benchmarking:  The ENA considers that it is not yet possible to set efficient costs using pure statistical benchmarking at high levels of aggregation without regard to expert interpretation. Doing so would impose material regulatory risk on the businesses and deter much needed investment in the regulated network sector (ENA, sub. 17, p. 5)  The Businesses are not aware of any country that uses benchmarking exclusively to regulate DNSPs’ revenues and prices. (ETSA Utilities et al., sub. 6, p. 30)  Perhaps more importantly, high level aggregate benchmarking with the use of a few high level causal factors makes it impossible to ‘sanity check’ the results that come out of a statistical model with the real world engineering constraints facing a business. If the regulator determines that expenditure on zone substation is not efficient because less costly alternatives exist, then this is a finding that can be contested on the available facts. By contrast, consider an example where a regulator decides that five per cent of total expenditure is not efficient purely on the basis of a high level statistical comparison to other businesses. This reasoning provides no indication of what aspects of the expenditure proposal are imprudent. Consequently, the business has no recourse to defend its proposed asset investment program on the engineering needs of the business because this was not the basis of the regulator’s finding. (ENA, sub. 17, p. 27)  Others have been similarly cautious:  The CEC agrees with the Commission that benchmarking will be unable to supersede the current ‘building block’ approach taken by the AER. Rather, benchmarking methods should be able to be applied by the AER in order to provide supporting information to its determinations. Indeed, benchmarking should be another tool from the AER’s toolbox of revenue assessment tools at its disposal. Given the broad use of benchmarking in regulated markets globally the CEC sees no reason why this could not be the case. (Clean Energy Council, sub. 38, p. 3) |
|  |
|  |

Second, benchmarking can be aimed directly at measuring regulatory efficiency and effectiveness. Some examples could include:

* whether changes in the Rules or guidelines by the AER lead to cost-reflective network and retail prices for each distribution business. As chapter 11 notes, if they do not, then the efficiency benefits from demand management would be reduced, and it would be important to find out why the predicted outcomes were not eventuating. It would be relatively straightforward to collect data on the actual pricing behaviour of the various distribution businesses and assess whether it accorded with the objectives of the regulations
* the degree to which retail price regulation frustrated cost-reflective pricing across the NEM
* examining any gap between the value of customer reliability and the costs of investment intended to achieve given reliability improvements
* the links between various safety regulations and safety outcomes (an issue discussed in relation to fire risk in chapter 7)
* the costs borne by the AER, the businesses, the merit review body and other stakeholders associated with regulatory determinations, and the regulatory processes that most increase those costs. While any regulatory process must involve transactions costs for the contesting parties, benchmarking might be able to assess whether process reform could economise on these.

It appears likely that many of the greatest inefficiencies in networks lie outside the control of the businesses, but reflect the unintended consequences of a parochial and flawed set of regulations across the NEM. Regulatory benchmarking may be able to identify and quantify these inefficiencies, prompting reform. The AER, the AEMC and AEMO can all perform useful roles in these areas.

In the past, some claims about deficiencies in the Rules — such as those that led to the major changes in the Rules in late 2012 (AEMC 2012r) — were not strongly empirically based at the time. A proactive approach to regulatory benchmarking might help bolster cases for Rule changes (or cut them off early).

### An overall test of the effectiveness of the regulatory regime

While much regulatory benchmarking would target specific areas, an advantage of upper-level benchmarking of business’s performance would be to provide a bird’s eye view of the overall effectiveness of the regulatory regime. As well as cross-sectional benchmarking (described in chapter 4), the AER should also model growth rates of total factor productivity (and their constituent partial productivity rates) — for each business and for the industry as a whole. This would highlight how any business is performing compared with the entire electricity network industry, and with other industries in the Australian economy. This would also test the degree to which revenue determinations reflected ongoing negative (or low) TFP growth (and would assist in any progression to CPI-x benchmarking later).[[11]](#footnote-11)

These forms of benchmarking should be included in the AER’s regular benchmarking publications. And as discussed in chapter 4, any such indicative aggregated benchmarking analysis should control for the most important differences in the operating environments of businesses — such as customer density, line type and length, reliability requirements, and the capital vintage of relevant assets.

Recommendation 8.1

The Australian Energy Regulator’s regular aggregate benchmarking of the performance of network businesses should include comparisons of:

* multifactor productivity — the output of services for given inputs
* separate productivity of capital, labour and intermediate inputs.

The results should control, to the best extent available, for any significant variations in the operating environments of the businesses, including customer density, line type and length, reliability requirements, and the age of relevant capital assets.

## 8.3 Could more targeted analysis act as a filter?

Benchmarking could be used as a diagnostic tool to identify areas of a proposal that may require greater scrutiny, but without it assuming a determinative role. Many participants supported the use of benchmarking as an input into the regulatory process, rather than as a replacement for the building blocks framework (box 8.3).

#### From aggregate benchmarking to detailed analysis

The use of aggregate benchmarking as a filter would have advantages over detailed bottom-up approaches in that it could first identify those businesses that were more likely to be inefficient in their expenditure. The analysis could then bore down into the more disaggregated detail, but only down the branch where benchmarking revealed that costs were sufficiently distant from the efficiency frontier (figure 8.1).

|  |
| --- |
| Box 8.3 Benchmarking is a useful adjunct to other modelling |
| Many participants accepted the value of benchmarking as one input into the regulator’s determination of the appropriate aggregate or disaggregated costs of the regulated business:  The Businesses support the Rules’ requirement for the AER to use benchmarking as part of the building block approach to test the efficiency of DNSPs’ expenditure — and to choose which types of benchmarking techniques it will use — although this should recognise the inherent limitations of different benchmarking techniques and of the comparability of data. (ETSA Utilities et al., sub. 6, p. 15)  Overall, benchmarking of certain performance outcomes is useful only as an adjunct to the establishment of revenues using the existing cost build-up approach. The fundamental objective enshrined in the National Electricity Law will not be achieved with regulatory uncertainty, which will ultimately deter investment. (APA Group, sub. 2, p. 2)  Ergon Energy believes that benchmarking techniques are not robust enough to replace a detailed investigation of costs and should not be relied on entirely to set revenue allowances. Instead, benchmarking should be one of many assessment techniques adopted by the AER to determine efficient and prudent expenditure. (Ergon Energy, sub. 8, p. 9)  In the ENA’s view, pure statistical analysis is most likely to be useful: as a means of identifying anomalies in an expenditure proposal that require closer more detailed examination; or when applied at low levels of expenditure aggregation. (ENA, sub. 17, p. 5)  … benchmarking results can also be used to allow the regulator to request further information. Consumer Action believes an appropriate first step to sophisticated benchmarking approaches should be to ensure that the regulator has more information combined with the ability to request further information/evidence from network businesses that under-perform. The onus should then be on the network businesses to justify and prove their case (in relation to revenue proposals). (Consumer Action Law Centre, sub. 5, p. 2)  Some leading researchers are pessimistic about the usefulness of benchmarking in economic regulation. For instance, in looking at arrangements for Swiss distribution businesses Farsi and Filippini (2005, p. 1) concluded that benchmarking analysis should be used to support rather than to determine regulatory decisions. Similarly, Shuttleworth concluded that:  In practice, benchmarking has proven either troublesome or irrelevant to the regulatory process, but proponents continue to search for ‘better’ models that will be more useful. … I conclude that, at best, benchmarking can help to focus regulatory enquiries, but that it shows no prospect of becoming a substitute for detailed evaluation of each regulated utility’s own costs. (Shuttleworth 2005) |
|  |
|  |

Figure 8.1 Boring down through successive layers of network costsa

|  |
| --- |
| Figure 8.1 Boring down through successive layers of network costs. This figure shows an example of how benchmarking progresses through different layers of costs associated with capital expenditure and operating expenditure. |

a The chart is illustrative of the process, but not intended to depict the cost structures of distribution network businesses.

Where an expenditure class appeared close to best practice, further investigation of its constituent elements — of the forensic kind currently undertaken — would be avoided (box 8.4). The AER’s repex model for modelling replacement expenditure by distribution businesses exemplifies this approach:

The calibrated repex model, among other things (e.g. significance of expenditure), was used as a guide to whether or not we considered a detailed review of a specific asset category should be undertaken. (Nuttall Consulting 2010a, p. 31)

Any such analysis would have to consider substitution possibilities between certain classes of spending, particularly opex and capex.[[12]](#footnote-12)

|  |
| --- |
| Box 8.4 The bottom-up approach — getting into the detail |
| **Locks and keys**  In its regulatory proposal to the AER, Ergon Energy proposed an allowance for expenditure on 300 000 locks and keys. The AER asked its consultants, PB Associates to assess this claim. As a result of this, Ergon revised its budget for locks and keys and provided ‘a business case’ for this expenditure including an ‘options analysis’. PB Associates then assessed this claim by examining the number of locks per kilometre of track, and the number of keys to be provided. At the end of its review, PB Associates concluded that the scope of works was transparent and the cost estimate was well supported and so it decided that the revised budget for locks and keys was prudent and efficient. The AER then concluded that Ergon Energy provided a ‘well substantiated’ forecast for its revised keys and locks program in its revised regulatory proposal and so it accepted Ergon’s revised proposal. The lock and key budget was less than 0.2 per cent of the total allowed expenditure by Ergon during its regulatory period. (Mountain 2011, p. 55)  **Detailed information on the average span lengths of insulated conductors**  The [blacked out writing] also assumes an average span length of 50m for an insulated conductor. This assumed span length is actually 25 per cent greater than the actual average span length. The [blacked out] identifies that CitiPower has 191km of ABC [aerial bundled cable] and 4,703 spans. This works out as an average of 40.6m per span. The reason why the longer span assumption was used when actual span length information was available is not clear. (Nuttall Consulting 2010a, p. 296)  **Special trees**  As a result of discussions with the ESV, United Energy has revised down its resource requirement to one full time equivalent (FTE) to establish the ‘habitat’ tree register in the first year (2011), followed by 0.4 FTE in subsequent years to monitor and update the register, process questions and information requests, and provide on-going training to employees and vegetation contractors. (Nuttall Consulting 2010a, p. 357)  **Pole treatment processes**  Aurora submitted that the AER erred in comparing Aurora's historical pole lives to pole lives achieved by mainland distribution network service providers (DNSPs) because Aurora uses a different type of timber pole to mainland DNSPs. In its draft determination the AER considered that the treatment of Aurora's timber poles should result in similar lives to the untreated mainland timber poles despite Aurora using a different timber class to mainland DNSPs. Aurora submitted that the treatment process typically only impregnates the sapwood (outer layers) whilst the heartwood (inner core) remains untreated. Aurora submitted that although treatment may extend pole life, there is no engineering reason to expect that it would result in a pole with the same life as the poles used on the mainland. (AER 2012f, p. 60) |
|  |
|  |

Regardless of whether the regulator used a formal hierarchical modelling approach, aggregate opex/capex benchmarking could test the overall reasonableness of any bottom-up approach. If the two are sufficiently discordant, the regulator could ask the business to investigate the likely source of the difference, such as important omitted control variables that affect the benchmarking model. This would again focus attention on major rather than minor cost drivers, and improve the quality of the benchmarking models.

Moreover, if two similar network businesses have very different unit costs, then they should be able to explain the probable reasons and to quantify them. Some networks have already used this approach in the determination process:

Country Energy provided us with a comparison that had been undertaken with Ergon Energy’s vegetation management expenditure. The comparison showed that Ergon Energy had a similar profile of vegetation density and that after allowing for differences in cycles and size, Country Energy’s proposed expenditure was comparable to that incurred by Ergon Energy. (Wilson Cook and Co. 2008, p. 41)

Requesting feedback from businesses about why results might diverge would not require the business to *demonstrate* how the benchmark result was wrong (notwithstanding the views of one key participant — box 8.2). Until the methods and the data underpinning benchmarking are significantly improved, reversing the existing onus of proof in the Rules would increase business risk significantly (Lowry and Getachew 2009, p. 1325).

#### Benchmarking at a disaggregated level

In general, disaggregated benchmarking could be used to judge detailed bottom-up aspects of a business’s proposal (an approach supported by the ENA, sub. 17, p. 5). This would be possible under the current regulatory framework, as well as under the framework described in figure 8.1. Disaggregated benchmarking might relate to:

* vegetation management
* the linkages between asset vintages and replacement rates by major class of assets (such as poles and distribution substations)
* maintenance efficiency, such as time and resources to correct certain faults
* the efficient monitoring of assets.

Such specific benchmarking may be reasonably reliable because there are fewer confounding variables.

Management performance measures should not be overlooked. For example, benchmarking of distribution businesses in Europe suggested that higher performing businesses were more likely to outsource their network functions (figure 8.2). Among other factors, management performance could be measured as the adoption rates of best-practice commercial processes and equipment, including:

* the use of customer panels and surveys (as these can be important elements of customer engagement)
* employment and procurement practices
* outsourcing (a way in which the business itself can exploit competitive processes)
* work processes and occupational safety
* demand management (as a network alternative)
* information technologies (increasingly important as networks become ‘smarter’, and given the need to serve millions of customers and control complex networks efficiently)
* innovation
* financial controls
* project management.

Figure 8.2 Do not overlook management processes

Outsourcing and high performing firms in European distribution network businessesa

|  |
| --- |
|  |

a Outsourced network activities related mainly to preventative maintenance, and assembly and construction.

*Data source*: Chanel (2008).

Following Turvey’s advice (discussed in chapter 4 and 11), regulators need an ‘understanding of what an enterprise does and how it does it’ before collecting information. Consequently, in determining relevant benchmarking performance and control variables, the AER should consult directly with network businesses, as well as with others in the supply chain (including generators, retailers and network equipment suppliers).[[13]](#footnote-13)

It is equally critical to test whether the performance measures are relevant to customers (by the type of customer). For example, customers may give different weights to connection costs, fault rectification times, reliability, and call centre performance, and may sometimes directly observe practices they regard as inefficient. Finally, and as emphasised later, electrical engineers and other experts may provide strong guidance on the most appropriate measures and controls, including the pitfalls in their measurement and interpretation.

As much as possible, the AER’s annual benchmarking reports (box 8.1) and their regulatory determinations should include such disaggregated measures. Moreover, while there is a requirement under the Rules for the reports to use ‘reasonably plain language’, this should not preclude the use of, and full documentation of elaborate benchmarking analysis (which is reflected in recommendation 8.11).

However, it is unlikely, given the limited time available, that the first report due in September 2014 could cover all of the above matters.

Recommendation 8.2

Subject to compliance and other costs (recommendation 8.12), the Australian Energy Regulator should accompany aggregate analysis with detailed benchmarking of particular aspects of the performance of the businesses, including:

* the rate of investment relative to the age-weighted capital stock by asset class
* the efficiency of major maintenance activities
* the adoption rate of best-practice commercial processes and equipment, including the use of customer panels and surveys, outsourcing, demand management, information technologies, financial controls, procurement practices, occupational safety, and project management.

In determining relevant benchmarking performance and control variables, the Australian Energy Regulator should consult with:

* network businesses, generators, retailers and network equipment suppliers
* customer representatives
* relevant experts within Australia and internationally.

As discussed in chapter 4, the small number of network businesses in Australia poses an obstacle to authoritative benchmarking. This is because it limits the possibilities of making ‘like with like’ comparisons between networks, especially when there are significant differences in operating environments. There are several strategies for addressing this:

* use combined time-series, cross-sectional (or panel) data
* collect data about different businesses’ performances by feeder type, so that comparisons are more robust. For example, the AER (2008a, pp. 160ff) examined reliability by four feeder types — central business district, urban, short rural and long rural feeder
* collect data about the performance of different business units within businesses (for example, comparing in-house with contracted services, or monitoring activities in different areas).

Given that most people served by distribution networks are in Australia’s major cities, comparisons between business performances within metropolitan areas are likely to be the most useful.

Recommendation 8.3

The Australian Energy Regulator should periodically assess the comparative performance of network business units within particular sub-regions of the National Electricity Market, where:

* those sub-regions share similar physical operating environments
* the costs and informational requirements of doing this are not too great (recommendation 8.12).

The comparisons should relate to units within a particular business, as well as comparable units in different businesses.

The Australian Energy Regulator should place most emphasis on comparisons of the efficiency of distribution networks in metropolitan areas.

### Potential challenges of the targeted approach

This targeted approach may result in excessive rents for some businesses. Any benchmarking model will involve:

(i) false positives (firms that appear inefficient, but are not)

(ii) false negatives (firms that appear efficient, but are not).

Under the targeted approach, firms in (i) would be subject to more detailed scrutiny, which should correct the false positive error. However, businesses in (ii) would not be subject to such scrutiny, and so the false negative error would not be corrected. Overall, this would create an upward bias in revenue allowances.

This issue may not matter greatly over successive determinations as the regulator continues to learn. Moreover, if incentive regulation is functioning properly, businesses in (ii) have strong incentives to improve their performance. In that respect, the asymmetric treatment of false positives and negatives may be appropriate.

Alternatively, were the negotiated approach outlined in section 8.4 to be used, it would also help to reduce the incidence (or severity) of false positives, as customers would have some bargaining power to reduce rents.

A further challenge is that, notwithstanding its greater likely reliability and accuracy, the passage of disaggregated benchmarking through the limited merit review process still suggests that it can be problematic to control for the relevant differences in the operating environments of the businesses (box 8.5).

## 8.4 Benchmarking could be a trigger for negotiated settlements

As the analytical rigour of aggregate benchmarking develops, it could also encourage early settlement in determinations, short-circuiting the protracted, uncertain and costly processes currently applying under the Rules (figure 8.3). The current costs of participating in the determination process are up to $15 million for each network firm and up to $6 million for the AER (AEMC 2009a, p. 9). In considering the advantages of CPI‑x as a simpler approach to determinations, the AEMC (2011b, p. iii) estimated that the approximate cost of one complete cycle of revenue determinations using the current building blocks method was $330 million (of which the component paid by the businesses would mainly be passed onto consumers).

|  |
| --- |
| Box 8.5 The fiery hoop of merits review |
| CitiPower and Powercor each challenged the AER’s expenditure allowances for vegetation clearance (with the Tribunal considering their concerns in a single review). The AER’s ruling relied heavily on benchmarking the costs of Powercor and CitiPower against those of other distribution networks. The challenge provides a useful case study of how the limited merits review has applied to benchmarking.  The network’s proposal lacked explanatory evidence. The Tribunal agreed that it should not have been accepted. However, the AER also had to prove that its judgment was reasonable under the Rules. Similar boundaries are likely to apply to benchmarking in the future.  The Tribunal had agreed in principle with the AER’s treatment of the proposal, given the lack of explanatory information.  The AER was entitled to be suspicious of the quantum of the step change amounts claimed by each of CitiPower and Powercor given the shortcomings in the information provided and the significant increase over the 2009 base year. Furthermore, it was entitled to benchmark those rates against information provided by the other DNSPs.  In our view, CitiPower and Powercor had ample opportunity to provide greater assurance to the AER concerning the step change amounts which they had claimed. They must be taken to have understood that the AER would wish to look at the rates which underpinned those amounts carefully, would wish to benchmark them against the other DNSPs’ rates and would wish to cross-check them as against expenditure in prior periods. (Application by United Energy Distribution Pty Limited [2012] ACompT 1).  On the other hand, the Tribunal also considered the benchmarking undertaken by the AER, and found that it had not considered factors that were likely to be significant.  … the assessment made by Nuttall Consulting failed to pay proper regard to the differences between Powercor’s network and those of the other DNSPs and failed to take proper account of the differences between the work programs which had been put in place by Powercor, in particular, and those which the other DNSPs proposed to undertake. After all, the work programs which Powercor had put in place had been assessed as reasonable by ESV, at the behest of the AER. ESV had concluded that the Powercor work programs constituted a reasonable response to the new regulatory environment created by the Victorian Government as a result of the Black Saturday bushfires.  The AER was justified in not being satisfied with the VEMCO costings [Vemco Pty Ltd is an independent, third party vegetation management contractor]. However, its assessment of the costs of Powercor’s work programs was unreasonable. (ibid)  The Tribunal’s decision was to remit the matter back to the AER. |
|  |
|  |

Figure 8.3 Benchmarking as a short circuit

|  |
| --- |
| Figure 8.3 Benchmarking as a short circuit. This figure shows how benchmarking can be used to assess the merit of business proposals. Consumers, businesses and the regulator may sometimes be able to short-circuit the current laborious process if they could negotiate early in the regulatory process. |

Under a ‘short circuit’ approach, depending on the divergence between benchmarking and the business proposal, the AER could immediately accept a proposal as reasonable, or if the proposal was in the ‘ballpark’, commence negotiations with the network business, with the involvement of customers. These would be better informed because of published benchmarking analyses, and could act better on this information given the reforms to their resourcing set out in chapter 21.

The AER could also request further information (a ‘please explain’ notice) to assist the early resolution of an agreement. Failing a quick resolution, the AER would adopt the current forensic and protracted processes, with the risks and costs that this would involve for all parties.

#### There are precedents for benchmarking to assist consumer advocacy

In California, benchmarking analysis is used in the settlement process between the business and a consumer advocate (the Division of Ratepayer Advocates). For example, the DRA has assessed total factor productivity measures to examine the performance of utilities (DRA 2010) and to assess the reasonableness of a given utility’s calculations.

As noted by the ACCC/AER, the role of the DRA is

… to advocate on behalf of the customers of regulated public utilities. It represents consumers in the CPUC [California Public Utilities Commission] proceedings, including rate settings, investigations, and rule makings. The DRA also participates in CPUC-sponsored working groups, advisory boards, workshops, and other forums. The DRA also evaluates utility proposals, investigates issues, presents findings and formal testimony, litigates complaints, and makes recommendations to the CPUC and to other forums. The DRA must ‘represent and advocate on behalf of the interests of public utility customers and subscribers…to obtain the lowest possible rate for service consistent with reliable and safe service levels’. The DRA also has statutory rights to obtain information from utilities through discovery and other means. The CPUC is required to provide sufficient legal support for the DRA, and provide the DRA with its own lead counsel. (ACCC/AER 2012b, p. 176)

Ofgem has proposed the implementation of a similar approach for the regulation of electricity networks in the United Kingdom. In the Ofgem Information Quality Incentive framework, firms are initially judged on the quality of their proposals; their past performance; and benchmarking exercises. Firms are then categorised accordingly as:

* requiring a low level of scrutiny — business plans would typically be given a shorter assessment, and final determinations would be relatively early
* requiring a moderate level of scrutiny — assessments would focus on the deficiencies of the proposal, as well as past performance. Capital projects may be subject to random inspections. Determinations would likely run their normal duration
* requiring a high level of scrutiny — business plans would be subject to a full engineering and economic analysis, as well as follow up analysis. Firms may be required to supplement their business plans with further explanatory data. Determinations would likely run their normal duration (Ofgem 2010, p. 58).

Regulation of water in Italy has used a similar approach. The business proposes a tariff in the first step. The regulator estimates a price cap using benchmarking analysis, and approves the business’s proposal if it lies within an acceptable range around the estimated price cap. If it does not, then the tariff is renegotiated with the business required to justify its excessive tariff before any revision (Farsi et al. 2005, p. 25).

In Florida, the Florida Public Services Commission has encouraged negotiated settlements in telecommunications and energy utilities. A statutory consumer advocate, the Office of Public Counsel, has been at the forefront of negotiating settlements, replacing formal regulatory processes (Cunningham sub. DR84, p. 9; and the Consumer Action Law Centre, sub. DR79, p. 38). Cunningham noted:

Over the seven years to 2002, approximately 30 % of telecommunications and energy utility price reviews were settled, growing further since then. The available evidence suggests that the elements of settlements have been innovative and customers have tended to obtain better outcomes when compared with formal rate cases.

More broadly, Cunningham provides a comprehensive examination of settlement arrangements in a range of other utility and non-utility contexts, commenting on their (significant) advantages, but observing that all such models have limitations.

#### Further work on negotiated settlements

The short-circuit process outlined by the Commission in its draft report had in-principle support from many stakeholders.[[14]](#footnote-14) However, participants (and indeed the Commission itself in its draft report) recognised that some pre-conditions would need to be met.

* A consumer body would need to have a statutory right to participate in such negotiations.
* Benchmarking would have to be sufficiently reliable and informative to underpin such settlements.
* A customer group (or groups) would need to be formed and have sufficient expertise before it could occur. For example, the Public Interest Advocacy Centre (sub. DR65, p. 18) and the Total Environment Centre (sub. DR50) considered that consumer stakeholders did not yet have the experience and resources to effectively represent the interest of all consumers in a three‑way negotiation process. While these stakeholders recognised the future value of incorporating negotiated settlement processes, the Major Energy Users questioned whether any single consumer body could act as a representative negotiator (sub. DR66, p. 29). However, the experiences in other jurisdictions described above appear to belie that contention. The formation of effective advocacy is discussed further in chapter 21.
* Some of the potential deficiencies of the approach would need to be resolved, or at least examined more closely. For example, as observed by Cunningham (sub. DR84, p. 16), once a consensus is achieved in a negotiated settlement between the nominated parties, the process stops (that being one of its intentions), which narrows consultation about the outcome, and potentially procedural fairness. This raises issues about the transparency of the settlement process — a point also made by the AER (sub. DR92, p. 3). If nothing else, this reinforces the need for a consumer negotiating body to have credibility with its constituencies.

Among participants to this inquiry, the AER expressed the greatest misgivings about the general viability of negotiated settlements. Not only was it concerned about the transparency of the arrangements, but it suggested that:

The PC has also overestimated the potential role of expenditure benchmarking in terms of being able to ‘fast track’ an entire regulatory proposal, which includes a considerable amount of material non-expenditure items. (AER sub. DR92, p. 3)

It supported a more selective use of a negotiated settlement approach.

Overall, the Commission still considers that a negotiated settlement arrangement should be introduced — subject to the formation of a credible, well-resourced consumer body and sufficiently reliable benchmarking information. The AER itself would have to develop a capability to assist in such a process. As discussed above, years of experience of similar arrangements in the United States and other jurisdictions shows that they are workable and produce good outcomes.

Recommendation 8.4

When benchmarking is sufficiently reliable, the National Electricity Rules should be changed to allow the Australian Energy Regulator (AER) to have the discretion to initiate a three-way negotiation of a mutually acceptable settlement. This should involve itself, the network business and the representative and qualified customer body identified in recommendation 21.5:

* Negotiation would only be triggered if the AER judged that the divergence between aggregate benchmarking estimates of forecast spending and the business’s proposal were sufficiently narrow.
* Where an agreement was successfully negotiated using this process, the AER should not be obliged to go through the current formal draft/final determination processes.

## 8.5 Information and ‘moral suasion’

Benchmarking at any level of aggregation could be used to inform customers (and the media) about the relative performance of businesses, which provides indirect pressures on inefficient businesses and their shareholders (including state governments). In this respect, the Consumer Action Law Centre pointed out:

Enhanced information and transparency about regulated network businesses can benefit the regulatory process as well as improving the behaviour of network businesses. A better-informed regulator will produce more efficient price setting, while comparative analysis and reporting on the network businesses’ performance by the regulator can create an incentive for the network businesses to ‘self-discipline’ as a result of competition-by-comparison and brand protection. (sub. 5, p. 1)

Benchmarking would also facilitate the participation of consumer groups in the determination process (beyond the negotiated settlement arrangement discussed above). For instance, the EUAA cited an example where the availability of benchmarking might have assisted in the merit review process:

The [Australian Competition Tribunal] refused the EUAA leave to appeal on the basis that the EUAA could not demonstrate that the AER’s failure to have regard to benchmarks satisfied the financial threshold for appeals under the National Electricity Law. The Tribunal required the EUAA to have benchmarked the distributors’ expenditure and to show that, had the AER also done this, the expenditure allowance would have been significantly lower than the allowance determined by the AER. Obviously, the EUAA was not in a position to undertake extensive regulatory benchmarking itself and so could not satisfy the Tribunal’s criterion for leave to appeal the AER’s determination. (EUAA, sub. 24, p. 4).

Aside from contributing to determinations, the advantage of systematic benchmarking comparisons between network businesses at the micro and macro level is that it could identify groups of firms that are consistently more efficient, and highlight some of the potential causes of the efficiency gaps (an issue raised in chapter 4).

The requirement that the AER produce annual benchmarking reports in ‘reasonably plain language’ will achieve many of the above objectives, as will the formation of an effective consumer advocacy body (chapter 21).

## 8.6 The long-run application of benchmarking

As discussed in section 8.1, even though benchmarking may contribute to regulatory determinations, there is little immediate scope for benchmarking to play a decisive role. Nevertheless, as data and modelling improve, and with better-designed incentives arrangements, there may be greater scope to give more weight to aggregated benchmarking. As Jemena observed:

The AER now has extensive information gathering powers under the NEL and it is exercising those powers. Over time, that should produce a data-set that could support more extensive use of benchmarking and the use of more sophisticated benchmarking techniques; however that is some way off. (sub. 21, p. 10)

While the risk of regulatory error would still persist, the protection against regulatory error would be to set the benchmark at a level close to the competitive market standard — such as the 75th percentile — rather than at the frontier (Lowry and Getachew 2009, p. 1329). The Major Energy Users also recognised the need for the long-run viability of the network businesses when setting the benchmark:

If it is awarded too little revenue which is based on the efficient frontier, then the firm could be in financial trouble which would be a worse outcome for consumers. (MEU, sub. 11, p. 26)

As Jemena (sub. DR77) point out, any new benchmarking-based framework should be considered in light of both the Revenue and Pricing Principles in the NEL, and the National Electricity Objective.

Participants also noted that moving to a regime where the AER uses benchmarking as the primary basis for setting revenue determinations would be a significant change, and that the processes to reach that goal would require consultation and stakeholder confidence (SP AusNet, sub. DR69). The Commission recommends a consultative and rigorous process to achieve this goal, particularly in regard to data, model outcomes and methodologies (section 8.7).

Recommendation 8.5

In any of the next rounds of regulatory determinations, the Australian Energy Regulator should not use aggregate benchmarking as the exclusive basis for making a determination. Instead, it should use aggregate benchmarking as a diagnostic tool in responding to business cost forecasts.

### CPI-x may have some advantages when linked with privatisation

The evidence suggests that state-owned networks are less efficient than their private sector counterparts (chapters 6 and 7) and, without reform, would continue to face muted incentives to reach the efficiency frontier over time. Privatisation can be expected to significantly increase efficiency and strengthen the responsiveness of the businesses to incentive regulation. This could open the door to the use of the TFP methodology described in section 8.1 and chapter 4 (and strongly advocated by the Victorian Government in the Rule change it proposed in 2008).

This could best be achieved by:

* privatising the state-owned enterprises following the orderly approach described in chapter 7
* refining the methodology for deriving x from TFP, and the commencement of data collection. Much of the work for specifying the appropriate methodologies for TFP indexes has already been completed (AEMC 2011b and Pacific Economics Group, subs. 35 and DR48) and some data are already available
* examining the revealed costs of the privatised businesses after completion of the forthcoming round of building-block determinations. The AER would use these costs as the *partial* basis for setting the base year revenue amount, when the TFP approach commenced. (It would be important to set a base at a higher level than the actual revealed efficient costs because otherwise the private business would have an incentive to cost pad during the forthcoming determination period)
* ensuring that the TFP measure was reasonably reflective of likely future productivity trends. DPI (2009, p. 9) recommended that x should be calculated some years after privatisation, reflecting the short-term, positive impacts of privatisation on productivity. However, it may be possible to use the TFP estimates of the long-privatised Victorian and South Australian distribution businesses as a proxy for x across the NEM, noting that TFP *growth* rates are less likely to be affected by the environmental factors affecting efficiency *levels*.

The businesses might still charge higher than desirable prices under such a regulatory regime, but would have incentives to set these in a way that reduced their allocative inefficiencies. Moreover, subject to an effective privatisation process (chapter 7), the sale price would capitalise at least a part of the businesses’ future stream of rents.

Given the sequenced nature of the reforms required, the realisation of this option would be some way away. The Commission has not recommended the long-run adoption of a CPI-x approach based on TFP, since that decision does not need to be made now, and its desirability would depend on what actually transpires in the market and in the Rules. However, this option should be regarded seriously, and the preliminary analysis and data collection to realise it should be undertaken, a view echoed by the Victorian DPI (sub. DR94, p. 5).

## 8.7 The regulator’s benchmarking practices

The path to more sophisticated benchmarking requires supportive actions by the regulator.

### Acquiring, sharing and using data

Due to the rulings in AEMC (2012a), network operators are already collecting greater amounts of data specifically for benchmarking purposes, while the AER is standardising its data collection and benchmarking processes (box 8.6).

There remains a further question over the extent to which data should be made publicly available. Several commentators have emphasised the importance of putting data in the public domain in order to allow greater transparency, and to allow stakeholders to undertake their own analysis (Lawrence 2009). Information and data on network characteristics are often only publicly available in high-level aggregates, whereas detailed data is often subject to confidentiality concerns.[[15]](#footnote-15) However, network businesses are natural monopolies, where confidentiality is not as justified as it is for businesses operating in the competitive market. Accordingly:

* most data required to assess business performance should be made publicly available. This would allow rigorous analysis of network performance by customer groups and researchers. This would increase the capacity for customers to act in the negotiating role set out in section 8.4 and in creating public pressures for improved efficiency (section 8.5)
* even where data are genuinely commercial-in-confidence, consumer groups involved in negotiations under recommendation 8.4 and independent researchers should be able to access the data. This access should be subject to requirements that they do not divulge publicly the information either directly, or in a way that identifies specific businesses. Such arrangements are routine for other sensitive information (for example, survey data sets collected by the ABS and administrative records from major government agencies).

To give effect to the better dissemination and use of these data, the AER needs to develop systematic and easily used databases, and publish more information on how businesses are performing. The annual ‘State of Energy’ reports are useful, but the information provided has reduced over time. The 2008 report provided more than 320 pages of information, while the 2011 report was 120 pages in length. (This is in contrast to the exponential growth in the length of regulatory proposals and determinations.)

|  |
| --- |
| Box 8.6 The AER’s information strategy |
| The AER’s information strategy includes reviewing the data definitions of key information required to undertake economic regulation, and developing:   * benchmarking measures for electricity network capex, operating expenditure and, if appropriate, total expenditure * benchmarking measures to compare the relative efficiency of regulated energy businesses, with an initial focus on electricity distribution   Having expanded the electricity distribution performance report to include network businesses from the Australian Capital Territory, New South Wales, Queensland and South Australia for the 2010‑11 report, and Victorian businesses for the 2011‑12 report, it will look to include Tasmanian businesses in the 2012‑13 report.  In 2013, while the AER will continue the development of analytical tools and data requirements, it will also consult with stakeholders regarding a stable set of reporting requirements for the following regulatory period. |
| *Source*: AER (sub. 13, p. 18; 2012, AER Information Framework, Information Paper, June). |
|  |
|  |

### In-house expertise

While raw data is valuable, a major role for benchmarking is to transform complex data into meaningful performance measures. As discussed in chapter 4, this often entails high-level technical expertise. While the AER should continue to engage external consultants,[[16]](#footnote-16) an in-house capacity to undertake sophisticated analysis would have significant benefits. It would:

* better inform the collection of data relevant to best practice analysis
* improve the AER’s capacity for intelligent and demanding outsourcing. Outsourcing requires sufficient in-house capability to discriminate among the various consultants. Moreover, there are likely to be greater benefits from outsourcing if the AER can absorb the research results it receives[[17]](#footnote-17)
* allow the AER to interpret and communicate any technical results in a way that is accessible to non-experts.

### Maximising the AER’s capacity to learn

One particularly important aspect of acquiring and retaining internal expertise is a greater capacity for learning within the AER, and a more general capacity to interpret the revenue proposals of businesses.

A major problem besetting any method for determining the benchmark efficient costs of a business is that there is no recognised standard against which to test the accuracy and reliability of the estimates of efficient costs.[[18]](#footnote-18) In effect, there is no agreed benchmark for verifying *either* bottom-up or benchmarking models against the ‘true’ cost.[[19]](#footnote-19)

Given that businesses have strategic interests, it seems likely that, after considering the businesses’ processes and the detailed consultations that follow, the AER is in the best position to determine the best, unbiased estimate of their true costs. For the reasons outlined in chapter 5, that may be different from the AER’s actual determination reached under the Rules. However, recommendation 5.2 proposes that if there is any divergence between the two estimates, the AER would also publish its preferred estimate. Over time, the AER can compare the estimate derived from a more bottom-up approach with the simpler estimates derived from aggregate benchmarking analysis. If benchmarking models improve, the model results should converge on the true estimate (in the fashion depicted in figure 8.4), improving their wider use and displacing the need for as much bottom-up testing.

One immediately useful exercise in this vein would be for the AER to reveal its preferred estimate of capex and opex arising from *past* determinations and examine how these estimates compare with the predictions of simple benchmarking models. That would provide an early test of the value of ‘primitive’ benchmarking.

Figure 8.4 Convergence between benchmarking models and bottom-up cost estimates

A speculative illustration for a single business

|  |  |
| --- | --- |
|  |  |

a The graph on the left hand side shows the relationship between the bottom-up and benchmark estimate of costs from T1 to T8. Full convergence occurs if the two estimates are the same, which means they would lie on the 45 degree line. At T9, the two estimates have converged. The right hand chart shows the *ratio* of bottom-up cost to the benchmarking costs over time, showing how they move closer together over time (shown as reduced variance).

Similarly, given the importance of demand forecast errors for both transmission planning and for the outcomes of regulatory determinations, there would be benefits in examining the reasons for variations in demand forecasts produced by AEMO and distribution businesses. The two most important aspects of demand are total energy and maximum energy because these drive the requirements for network capacity.[[20]](#footnote-20) Currently, AEMO (2012a) produces top-down forecasts of demand at the regional level, using regression models that include such factors as temperature and gross state product. It also produces demand forecasts based on information provided by distribution businesses and direct customers of transmission businesses, which differ from the top down estimates. Given its responsibility for demand forecasting, AEMO is in the best position to undertake the technical analysis to understand why the two sets of forecasts diverge. However, given that the AER makes regulatory determinations that provide considerable weight to demand forecasts, and as that it will be expected to do so in a more sophisticated way in the future (as discussed in chapter 11), it should:

* act as an informed consumer of AEMO’s technical modelling, as this will assist the AER in its broader benchmarking analysis, and, through feedback and challenge, should assist AEMO to improve its own modelling
* take into account AEMO’s modelling in making its regulatory determinations.

The AER pointed out that such exercises are not without costs — particularly when comparing past and present estimates and results (sub. DR92, p. 6). The Commission considers that comparing benchmarking results with actual outcomes is a part of the process for improving and refining models. Of course, the realistic goal is not to have an ex-ante model to perfectly predict the future — rather, it is to be able to identify the specific causes of the discrepancies, and hence understand the limits of the model.

Recommendation 8.6

The Australian Energy Regulator should develop and maintain appropriate benchmarking databases and in-house expertise for the technical analysis required to undertake sophisticated benchmarking.

Recommendation 8.7

The Australian Energy Regulator should make all benchmarking input data publicly available (recognising that the businesses being benchmarked are regulated monopolies) except where the data can be demonstrated to be genuinely commercial-in-confidence.

Where the latter holds, the Australian Energy Regulator should still make the full datasets available to:

* independent researchers who are using the results for non-commercial purposes
* the consumer body involved in any negotiations described under recommendation 8.4.

Provision of data should be subject to statutory requirements for non-disclosure of information predetermined as commercial-in-confidence, drawing on existing models for data protection.

Recommendation 8.8

When making its revenue allowance determinations, the Australian Energy Regulator should make judgments about capital expenditure forecasts that take account of:

* any discrepancy between the Australian Energy Market Operator’s top-down demand forecasts and the aggregate of network businesses’ bottom-up demand forecasts
* any discrepancy between previous expenditure forecasts and actual outcomes by different parties.

### Collaboration

While there are several obstacles to international benchmarking, it may still provide useful information (chapter 4), and network businesses and governments already undertake such analysis. International collaboration between regulators, academic experts and global benchmarking specialists may improve the validity of international benchmarking. This has been long-recognised (Jamasb and Pollitt 2001, and Farrier Swier Consulting 2002), but the mechanisms for achieving it are still incomplete.

International collaboration would involve the consistent collection, auditing and reporting of data, and shared approaches to reporting results and the statistical testing of models.[[21]](#footnote-21) This would facilitate meta-studies, which help identify common variables that lead to robust benchmarking results. For example, if many (rigorous) individual country studies find that a limited set of consistently defined explanatory variables perform well in measuring industry costs — across diverse regulatory and operational environments — it suggests that such methods may be reasonably robust in any country.

Collaboration could also have other benefits, such as:

* increasing diffusion of best practice benchmarking techniques and data construction (such as meaningful measures of capital), which would build expertise and knowledge
* facilitating contacts between experts for solving technical and practical problems in benchmarking
* providing lessons and case studies on the pitfalls and unexpected benefits of benchmarking. For instance, the Commission’s own inquiry report has benefited from the experiences of Ofgem in the United Kingdom and FERC in the United States
* making it easier for secondments between agencies
* supporting external peer review of specific regulator’s benchmarking analysis.

#### Transparency, consultation and communication

Section 8.3 has already spelt out the need for stakeholder involvement in developing benchmarking models. But, equally, they (and experts) should have a role in commenting on results (which should all be public). They should also be encouraged to undertake their own analysis, and have the information required that would allow them to replicate any models.

From a scientific perspective, the AER should adopt processes that will increase the quality of its analysis. This would involve independent expert peer review of benchmark models to:

* establish their ongoing relevance, sensitivity to assumptions, and scientific validity (including the performance of any statistical models against accepted standards — including confidence intervals, parameter stability, and specification testing, as set out in chapter 4)
* assess the modelling strategy used to produce the results
* consider the policy implications of model and parameter misspecification
* assess whether the AER has adopted best practice models
* ensure that methodologies and results are presented and reported according to widely accepted standards.

Most stakeholders agreed that peer review was appropriate.[[22]](#footnote-22) Nevertheless, the Victorian Department of Primary Industries (sub. DR94, p. 8) expressed concern that external peer review processes may be unnecessarily expensive. However, the Commission considers that peer review is generally not a costly exercise, particularly in relation to the costs associated with the overall revenue determination process. It would not involve long-winded inquiries, but rather it would resemble the processes routinely practiced by research agencies and academics, including the Productivity Commission itself.

The benefits of peer review would be particularly strong given that the vast majority of benchmarking research on Australian electricity networks is prepared for industry participants (including regulators), and may not necessarily be scrutinised to the extent typical of academic research.

It is also important for the AER to disclose the impacts of factors outside the control of businesses, but that may be controllable by governments (since, as noted above, these are important for policy and regulatory development).

While the key role of benchmarking is the determination of efficient revenue allowances, the AER would also need to communicate the results to their disparate stakeholders in an accessible way. As Berg (2010, p. 56) has colourfully noted: avoid ‘sensational factoids’, but use clear presentation methods for experts and lay audiences.

Recommendation 8.9

The Australian Energy Regulator should collaborate with other leading regulators, academic experts and global commercial benchmarking specialists to enable robust meta-analysis of electricity network benchmarking results from individual country (and where credible, multi-country) studies. The collaboration should include cooperation in developing:

* the most meaningful measures of performance
* consistent data collection
* consistent reporting of results
* best-practice analytic frameworks.

Recommendation 8.10

The Australian Energy Regulator should submit its major benchmarking analyses of electricity networks for independent expert peer review to establish their ongoing relevance, scientific validity, adoption of best-practice, and to gauge the degree of uncertainty in the results.

Recommendation 8.11

The benchmarking analysis produced by the Australian Energy Regulator should include:

* accessible reporting of the results to inform consumer groups, network businesses, and others
* disclosure of the importance of factors outside the control of businesses, but that may be controllable by governments
* publication of the modelling strategy used to produce the results
* the sensitivity of the results to changes in key assumptions
* the performance of any statistical models against accepted scientific standards, including confidence intervals, parameter stability, and specification testing.

### Practicality and compliance costs

Data collection, consultation and modelling impose costs on the regulator and network businesses. Accordingly, the principle should be to only collect information or use processes likely to produce net benefits.

The AER considered that benchmarking methods might reduce compliance costs for businesses:

Unlike the current building-block approach, benchmarking does not unrealistically, burdensomely and intrusively aspire to duplicate or exceed the knowledge of business operation possessed by decentralised decision-makers in regulated entities. Rather, the informational and regulatory burden of a benchmarking program is limited to relative knowledge – a knowledge that only need pertain to relative performance. (ACCC/AER 2012a, p. 168)

This seems overly sanguine, since detailed bottom-up assessment is unlikely to be displaced over the medium term. Moreover, testing the validity of benchmarking models (using the learning process set out earlier) will still require bottom-up information. Consequently, the informational burdens are likely to rise, not fall (and this is one of the reasons why the Australian Government’s decision to allocate additional resources to the AER is justified, as discussed in chapter 21). That should change as the AER (and its international collaborators) develop better models and identify the most important data items.

In the meantime, the AER should routinely re-assess:

* its benchmarking approaches
* the resources it uses in benchmarking
* the compliance burdens on businesses.

These assessments should be publicly available and subject to independent refereeing.

Recommendation 8.12

The Australian Energy Regulator (AER) should periodically examine its benchmarking methodologies and processes — with input from an independent expert referee — to assess their usefulness in the determination process and the costs they impose on stakeholders. It should compare these costs with the likely benefits when determining the appropriate frequency and type of detailed benchmarking. In undertaking such assessments, the AER should consult closely with network businesses.

The AER should make all such assessments publicly available.

## 8.8 Conclusion

Although Australia has been relatively ‘unsophisticated’ in its use and application of regulatory benchmarking in the electricity sector, this is likely to change in coming years with improvement in the AER’s data collection and modelling capabilities. An increase in benchmarking for diagnostic and informational purposes is likely in the near term, given recent AEMC Rule changes. Over time, repeated use of benchmarking models (as well as ex-post analysis) will improve the reliability of the models’ estimation of network efficiencies, and increase the potential for them to have greater weight in regulatory decisions. Whilst there may be some shorter-term burdens for network businesses in providing additional data to the AER, improved confidence in benchmarking has the potential to simplify determinations and lower overall costs, leading to benefits for network businesses and consumers.

1. The term ‘building blocks’, refers to the procedure for determining the total revenue allowance (for example, AER 2012x) based on the capital assets of the business, total investment in capital (capex), operating expenditures (opex), depreciation, tax rates, the weighted average cost of capital and various rewards and penalties (such as an Efficiency Benefit Sharing Scheme — chapter 5). The building blocks model can be either simple or complex, depending on how its inputs are estimated. In that sense, benchmarking is not inconsistent with the use of a building blocks approach (ETSA Utilities et al., sub. 6, p. 10). [↑](#footnote-ref-1)
2. However, the term CPI-x can sometimes relate to a weighted average price cap, with x not related to productivity growth (chapter 4). However, this chapter refers to the CPI-x approach as necessarily capping the growth of average prices or revenues by inflation less TFP growth. [↑](#footnote-ref-2)
3. Neither Mountain and Littlechild (2010) nor Mountain (2011) intended their benchmarking analysis to be used to determine revenues directly. [↑](#footnote-ref-3)
4. Regulatory asset stranding would occur if the regulator were to (ex post) set an allowance that under-remunerated previously made, but still productive, investments — in effect, pricing at less than long run marginal cost. [↑](#footnote-ref-4)
5. Pacific Economics group (sub. 35, pp. 8‑9) argued that the AEMC was too pessimistic about data inadequacies, and that the Victorian use of CPI-x exemplified the practicality of the approach. Nevertheless, Pacific Economics Group acknowledged that in modelling Victorian TFP growth, it was necessary for the analysis to begin in 1998, rather than 1995, because of the atypical TFP growth occurring immediately after privatisation of electricity distribution businesses (p. 3). The Commission has proposed privatisation of all state-owned network businesses (chapter 7), which should create a similar burst in TFP growth, presenting exactly the problem identified by Pacific Economics Group. [↑](#footnote-ref-5)
6. For example, while Ofgem is often seen as one of the most prominent users of the CPI-x approach, the AEMC (2011b, p. 45) notes that in their case x is based on a productivity growth ‘assumption’. [↑](#footnote-ref-6)
7. An x above the industry-wide achievable productivity growth rate would not provide enough revenue *over time* for the businesses overall to fund efficient opex and capex. [↑](#footnote-ref-7)
8. Too dramatic a cut in revenue for any given business in the first year of a CPI-x regime would expose the business to immediate insolvency. [↑](#footnote-ref-8)
9. While methods that jointly use cross-sectional and time series (panel) data would help, given their better capacity to control for different operating environments, there has not been widespread testing of their reliability (Frontier Economics 2010b, p. 59). [↑](#footnote-ref-9)
10. With the inherent problems spelt out in detail by Rubinfeld (1985). [↑](#footnote-ref-10)
11. The AEMC (2011b) proposed a similar approach to pre-testing of TFP methods. In its review of TFP benchmarking, it recommended ‘paper trials’ of benchmarking before considering further Rule changes in the area. [↑](#footnote-ref-11)
12. For example, Aurora Energy (2012, p. 73) has indicated that it will increasingly decide whether to replace an asset when it has actually deteriorated (rather than basing replacement on asset ages, which was its past practice). This approach will reduce replacement investment, but would be likely to increase monitoring costs. [↑](#footnote-ref-12)
13. This was recognised by the ACCC/AER (2012a, p. 166) and several participants in the inquiry, such as Ergon Energy (sub. 8, p. 10). [↑](#footnote-ref-13)
14. These included the Total Environment Centre (sub. DR50, p. 2, p. 4); National Seniors Australia (DR62, p. 10), the ENA (sub. DR71, attachment A, p. 4), the AEMC (sub. DR89, p. 19) and CitiPower et al. (sub. DR90, p. 16). [↑](#footnote-ref-14)
15. All data collected by the AER through Regulatory Information Notices (or other compulsory processes) are confidential. In the Commission’s own experience, the sourcing of detailed data for the purposes of this inquiry has been met with various responses from networks, ranging from free unfettered use to complete anonymity. [↑](#footnote-ref-15)
16. The AER already routinely uses external consultants for economic and engineering advice, such as Nuttall Consulting (2010a) in the Victorian distribution network determinations, and Schweinsberg et al. (2011) on European benchmarking practices. But currently, it has limited current in-house capabilities in this area. [↑](#footnote-ref-16)
17. In the business world, an internal research capability is not only valuable in its own right, but because it raises the absorptive capacity of the business to others’ ideas (PC 2007). [↑](#footnote-ref-17)
18. The situation is different from many other tests — such as the accuracy of medical diagnostics (confirmed through assays) and clocks (confirmed by reference to the US National Institute of Standards and Technology’s highly accurate clock). [↑](#footnote-ref-18)
19. In theory, under a high-powered incentive regime (chapter 5), profit-motivated businesses will progressively reveal efficient costs, which could be an alternative estimate of ‘true’ costs. However, that process may be quite slow if lower-powered incentives apply or state-owned businesses (which may be constrained by a variety of non-commercial objectives imposed by their shareholders) continue their dominant role. [↑](#footnote-ref-19)
20. Defined by AEMO as the highest amount of electrical power delivered, or forecast to be delivered, over a defined period (day, week, month, season, or year) either at a connection point, or simultaneously at a defined set of connection points. [↑](#footnote-ref-20)
21. There appears to be greater efforts for international collaboration in water utility benchmarking than in electricity, with benchmarking consortia established for the Americas, Africa, and internationally (Berg 2010, p. 62). In utility regulation generally, the Public-Private Infrastructure Advisory Facility, the World Bank, and the Public Utility Research Center of the University of Florida have created a collaborative online repository of material on infrastructure regulation (http://www.regulationbodyofknowledge.org). However, this is more a collection of references and training tools, than a resource for benchmarking and data collection. [↑](#footnote-ref-21)
22. These included the Public Interest Advocacy Centre (sub. DR65, p. 24), Major Energy Users (sub. DR66, p. 33), the Energy Supply Association of Australia (sub. DR70, p. 5), and the Energy Networks Association (attachment A of sub. DR71, p. 6). [↑](#footnote-ref-22)