1 April 2026

Contingency FCAS Rule change proposals

Consideration of benefits

ACIL ALLEN

Context and background

- Grid Energy have proposed two rule changes concerning scheduling and payment for contingency FCAS:
  - The first proposal would require AEMO to co-optimise the size of the largest credible contingency during dispatch under most circumstances:
    - Co-optimisation means that the variable(s) are incorporated into NEMDE and the amount of the variables are determined (dispatched) to support the overall NEMDE objective of maximising the value of spot market trading
    - AEMO currently co-optimises energy and ancillary services (including contingency FCAS) but the amount of contingency FCAS is determined as an input and is not co-optimised in normal circumstances
    - Under some circumstances, usually for network contingencies, AEMO does co-optimise the amount of contingency FCAS.
  - The second proposal would require AEMO to allocate the costs of contingency FCAS using a ‘runway’ approach:
    - The runway approach is adapted from the way that airports charge for runways with charging determined by the length of runway that each plane type requires to take off and land.
    - Applied to the NEM, the runway approach would charge each generator (and some loads) based on their contribution to the size of the contingency.
      - For generators, the largest unit (as dispatched) would pay for all contingency FCAS dispatched to cover the difference between it and the next largest unit (as dispatched)
      - The largest and second would share payment for FCAS contingency to cover the difference between the second and third largest unit (as dispatched)
      - And so on ….

Context and background (2)

- The AEMC engaged ACIL Allen (including SW Advisory) to support its consideration of the two proposed rule changes.
- ACIL Allen's advice includes:
  - A backcasting assessment of the status quo and co-optimisation for financial year 2025 to assess benefits of co-optimisation
  - Simplified forward looking model to demonstrate the benefits of co-optimisation and the runway approach
  - Consideration of how the proposed rule changes would impact on participant behaviour including:
    - Significant change in risk for participants
    - secondary contract markets
    - Future investments in the NEM
- ***The analysis and findings set out in this presentation represent the views of ACIL Allen (and SW Advisory) and should not be construed as representing the AEMC's views. Our advice is one input to the AEMC's consideration of the proposed rule changes.***

Analysis and initial findings

- Our work suggests that there are material market benefits from co-optimising FCAS requirements with the dispatch of generating units, with the benefits expected to mostly occur when Energy and/or FCAS services are scarce (could be a few days per year):
  - Including network contingencies is most likely where the largest benefits will accrue.
- Co-optimisation of FCAS requirements with the dispatch of generating units, loads and network elements will improve market efficiency and ensure secure dispatches. It is a no regrets option:
  - AEMO currently relies on the SCADA data at the beginning of the dispatch interval:
    - Dispatch interval dispatch targets may result in a significant movement in output and the size of the largest contingency
    - Contingency and regulating FCAS services provided by the largest contingency unit are excluded in the calculation.
  - AEMO is enabling too much or too little FCAS at times compared with the optimal requirement: this results in suboptimal pricing of FCAS services which affects allocative and dynamic efficiency and is inconsistent with the NER requirement to maximise the value of spot trade.
    - When it enables too little FCAS the system is not usually in an insecure state because of the frequency response available from non-enabled providers.
    - The only way to avoid enabling too much or too little FCAS is to co-optimise the FCAS requirements and the dispatch of each unit's energy, regulation and FCAS.
- Co-optimisation of contingency FCAS will overcome reliance on non-enabled providers, recruit contingency FCAS efficiently, provide efficient price signals and ensure the power system operates in a secure state in respect of managing credible contingencies.

Analysis and initial findings (2)

- The more difficult issue is how to appropriately recover the costs of FCAS. NER 3.1.4 (a) 8 provides some useful guidance on this issue:
  - where arrangements require participants to pay a proportion of AEMO costs for ancillary services, charges should where possible be allocated to provide incentives to lower overall costs of the NEM. Costs unable to be reasonably allocated this way should be apportioned as broadly as possible whilst minimising distortions to production, consumption and investment decisions;
- For the largest contingencies, the ‘runway’ cost allocation satisfies the principles in NER but not for the 2nd, 3rd, 4th etc largest contingencies.
- A variation on the proposed runway approach using the shadow prices in the co-optimisation would overcome the inconsistencies of the runway approach and would satisfy the principles in NER.
- We don’t consider either the co-optimisation of contingency FCAS requirements or Runway pricing would change the level of contracts that participants would offer to the market:
  - Other factors are more significant drivers.
- Co-optimisation may marginally affect investor’s views of expected future revenues, leading to slightly less investment in larger units at the margin, in the long run (but this improves dynamic efficiency).
- Runway pricing is likely to impact how market entry is physically arranged to reduce the risk of being a one of the largest contingencies.

Determining FCAS raise requirements

- Our discussion of FCAS requirements will focus on generator contingencies but the approach can be extended to load and network contingencies.
- For simplification of the discussion, we will initially ignore load relief and just use one generic raise contingency service.
- To cover the loss of generator j, the total amount of FCAS enabled for all of the other generators must be greater than generator j's possible energy dispatch for the period [t-5,t]:

  the total amount of FCAS from units other than j >= unit j's energy dispatch over [t-5,t].
- To ensure that the power system has enough FCAS to cover all generator contingencies over [t-5,t] then:

  $\sum_{k \neq j} FCAS(k,t) \geq EnergyGen(j,t)$ for all j.
- If we add the FCAS raise services provided by generator j to both sides, we get set of equations (the unit can't provide its own contingency FCAS):

  $\sum_{k} FCAS(k,t) \geq EnergyGen(j,t) + FCAS(j,t)$ for all j.

This can be made into two sets of equations:

$\sum_{k} FCAS(k,t) \geq largest\_contingency$ and

$largest\_contingency \geq EnergyGen(j,t) + FCAS(j,t)$ for all j.

The approaches to determining the largest contingency size

Current approach

- AEMO's approach to determining the largest contingency size for dispatch interval t is:

  Largest_Contingency =
  Max[$SCADA\_Energy(j, t-5)$] for all units j.

- In a linear programming framework like NEMDE this is equivalent to the set of inequality constraints:

  Largest_Contingency ≥
  $SCADA\_Energy(j, t-5)$] for all units j.

Co-optimised approach

- The co-optimised approach assumes a unit follows a linear trajectory from its output at the start of a dispatch interval to its target along the way.
- If the unit is providing raise regulation it could deviate from this linear trajectory by the amount of raise regulation that it is providing.
- To cover for the loss of a unit over the dispatch interval [t-5,t] the co-optimisation approach ensures that there is enough FCAS at both the start and the end of the dispatch interval, this results in the two set of equations:
- Largest_Contingency $\geq SCADA\_Energy(j, t-5) + FCAS(j, t)$ for all units j
- Largest_Contingency $\geq Energy(j, t) + Reg(j, t) + FCAS(j, t)$ for all units j.

Current approach vs Co-optimisation

- If we look at the largest-contingency sizes as determined by the co-optimisation approach and the current approach we get the following for the start of the dispatch interval :
  - Co-optimisation

    Largest_Contingency $\geq SCADA\_Energy(j, t-5) + FCAS(j, t)$ for all units j
  - Current

    Largest_Contingency ≥ $SCADA\_Energy(j, t-5)$] for all units j.
- At the end of the dispatch interval we have:
  - Co-optimisation

    Largest_Contingency $\geq Energy(j, t) + Reg(j, t) + FCAS(j, t)$ for all units j
  - Current

    Largest_Contingency ≥ $SCADA\_Energy(j, t-5)$] for all units j.
- At the start of the dispatch interval the current approach has a shortfall based on the self provision of FCAS and at the end of the dispatch interval the current approach has a potential shortfall related to a unit's energy target being greater than its initial SCADA measurements and its provision of regulation and contingency FCAS.
- Further if the costs of FCAS are high the co-optimisation approach may back off units to reduce the amount of FCAS required to maximise the value of spot market trade.

General discussion of co-optimisation of requirements

- Co-optimisation would be relatively easy to implement in NEMDE optimisation. It would not be sensible to implement via generic constraints as this would involve hundreds of constraints, that would require ongoing and regular maintenance.
- In addition to changes to NEMDE some changes may be required to other AEMO systems.
- Even though AEMO is not enabling enough FCAS at times based on the mathematical analysis in the previous slides, this is not usually resulting in system security issues due to the number of units which are not enabled for FCAS but are still providing primary frequency response (non-enabled providers).
- However, one of the reasons AEMO does not include estimates of this non-enabled primary frequency response into the FCAS contingency requirements, like it does with load relief, is that AEMO still wants to provide clear price signals to the market so that, in the longer term, it is recruiting enough FCAS so that the power system could be operated without the non-enabled providers. The co-optimisation approach would reduce the reliance on these non-enabled providers to maintain power system security, and so there would be an economic benefit in doing this.
- If AEMO wanted to procure enough FCAS so that it wasn't dependent on the non-enabled provision of primary frequency control it could add a safety factor to its contingency requirement. However, this could be of a material size to cover all the shortfalls we identified in our analysis, with a high probability of in the order of 30-50 MW, depending on the service.
- Further, with the increasing penetration of fast responding technologies (e.g. BESS and pumped hydro) the current state of these unit based on their SCADA measurements at the start of the dispatch interval could be quite different to their energy targets at the end of the dispatch interval. Also, there is a potential for network contingencies to create bigger requirements. Therefore, the gap between what is currently done, and the co-optimisation approach could become too large in the future to be covered by the non-enabled providers.

Impact of co-optimisation

- Looking into the future, the NEM is likely to see:
  - Larger BESS installations that may affect both raise and lower contingency FCAS requirements
  - Increases in Network constraints impacting on contingency co-optimisation:
    - REZ’s or other local generation connecting to the meshed NEM and being run as credible network contingencies
    - Connection of Marinus Link
    - Networks operating during planned outages of key transmission elements.
- Given how local VRE generation can vary and how rapidly BESS targets for energy, regulation and contingency FCAS could change from one dispatch interval to the next, using current SCADA values to forecast the largest contingencies is likely to result in too much or too little contingency FCAS being purchased (noting that AEMO relies on non-enabled providers to avoid operating in an insecure state when too little is recruited).
- This is likely to result in contingency FCAS prices not reflecting the underlying requirement and not providing efficient incentives for providers to offer and invest in contingency FCAS capacity.
- Co-optimisation of requirements solves this problem and is thus well placed to ensure secure and economically efficient dispatches in a changing environment.

Changes in largest contingency over time

- We have considered how the largest contingency(ies) may change over time.
- As large coal units retire, the largest contingency is likely to be determined by:
  - Network contingencies
    - During maintenance on parts of the backbone network, especially during periods of high flow
    - Radial connections to large REZs – when production levels are high
    - Basslink or Marinus Link.
  - Large BESS installations connected radially.
  - Large open cycle gas turbines (in aggregate) connected through a single connection point.
  - We note that Kogan Creek (750 MW) is unlikely to be retired before the 2040s.
- While not part of the current rule change proposal, network contingencies should be part of any requirements for FCAS co-optimisation and cost recovery:
  - We recognise AEMO currently co-optimises some network contingencies.

Cost recovery

- In the NEM, there has been a guiding principle that FCAS costs be recovered on a 'causer pays' basis.
- Currently, contingency FCAS services costs are allocated to a class of participants that 'causes' the requirement. In the case of the NEM, generators are allocated the costs of raise services, and loads are allocated the costs of lower services. Both loads and generators are charged on a MWh basis not based on how they have affected the size of the FCAS requirement.
- The 'runway' methodology aims to charge more of the costs to those participants who affect the size of the requirement.
- When the runway approach is used to recover the costs of contingency FCAS, the largest contingency pays all of the costs of the extra requirement over and above the second largest contingency, the second largest contingency along with the largest contingency pays the extra costs of its requirement over and above the third largest contingency and so on, until some minimum size.
- The runway approach could be applied to generators, loads and networks based on how they affect a contingency requirement. When applied to networks the costs could be recovered from loads and generators using the relevant network elements on a MWh basis.
- However, the runway approach is likely to have significant detrimental impacts on economic efficiency when energy and FCAS is scarce, as the second, third, fourth and so on largest contingencies seek to reduce FCAS exposure without impacting the size of the largest contingency.

Modified runway cost recovery

- The runway approach would impose substantial incentives on the largest generator(s) to back off production to avoid high costs, especially when contingency FCAS costs are high, and energy prices are low. This could lead to a more or less optimised solution, depending on the relative pricing of the contingency FCAS and other services relative to the largest generator(s).
- To avoid units uneconomically backing off the runway cost allocation should always be done with co-optimisation of requirements. Under co-optimisation, when FCAS costs are high:
  - large units' output may be reduced, but they will still be able to provide inertia and system strength services.
  - However, substantial costs may be allocated to the second, third, and so on, largest generator(s). This would create incentives for these second- and third-largest generator contingencies to reduce output without affecting the size of the largest contingency, and therefore, the contingency FCAS requirement, dispatch, and prices.
- This would likely create significant inefficiencies in dispatch during periods when contingency FCAS supply was scarce and prices were high. A method to overcome this problem would be to apply runway pricing to the largest contingency for the difference between it and the next-largest contingency and then allocate costs proportionally on a MWh basis below that (including to the largest contingency).
- The above issues would be expected to be more significant should runway pricing be implemented without co-optimisation.

More efficient alternative to runway cost allocation

- An alternative method to the runway approach is based on the use of shadow prices associated with the FCAS constraints for each contingency FCAS service.
- The benefit of this approach is that only the generation units contributing to the largest contingency would face a non-zero price.
- Although this approach appears to be quite elegant, as all the shadow prices from the FCAS requirements constraints and largest contingency constraints are used to determine the FCAS price and the cost allocation to individual generators, there can be an arbitrariness to the cost allocation when there are many equally valid solutions for the shadow prices for the largest contingency constraints. In linear programming terms, this is known as dual degeneracy and can occur at times. There may be some linear programming methods that could remove or at least ameliorate this issue like what is done in dispatch with tie breaking.
- Based on each unit’s bids and offers this approach including the cost allocation does not end up with any ‘undesirable’ dispatches. That is there are no dispatches that units would have been better off with compared to the one where they are one of the largest contingencies and pay their FCAS costs.
- The approach does not end up with any of the uneconomic incentives that can occur with the ‘runway’ approach but may appear to excessively penalise the largest contingencies.
- If this was a concern to the AEMC the approach could be used just to allocate the extra FCAS required for the largest contingency like with the modified ‘runway’ approach and then recover the costs proportionally on a MWh basis for all the other generators, loads and network elements.

Network constraints impact on contingency co-optimisation and cost allocation to networks

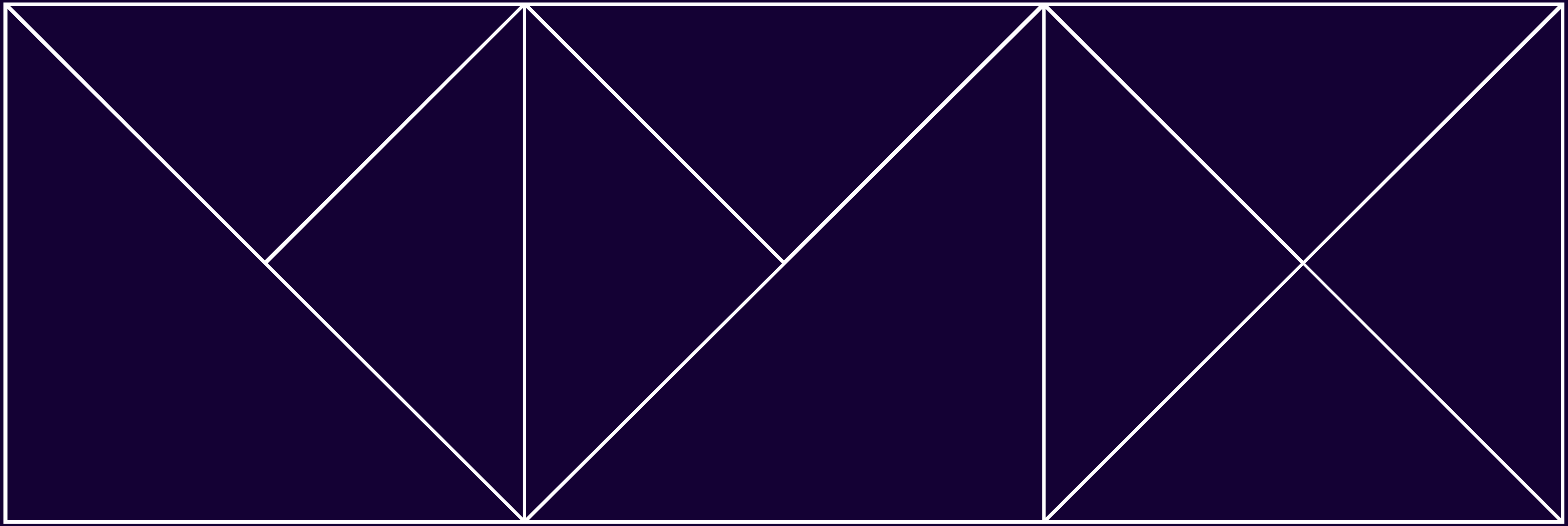
- Ideally the chosen cost recovery mechanism for network contingencies should use the same principles for intra-regional and inter-regional network flows and contingencies.
- When a network element is the largest contingency driving FCAS requirements then it should pick up costs like a generator or load. However, directly charging the network is unlikely to result in any efficient behavioural incentives. Some generators and loads are the beneficiaries of the increased flows:
  - Thus, a reasonable approach would be to charge the loads in the importing zone the costs of the raise FCAS (at least at the margin) on a $/MWh basis and the exporting generators the lower FCAS (at least at the margin) on a $/MWh basis
  - In the case of a REZ connecting radially to the meshed network, it could be considered as an aggregated generator and the REZ charged on the same basis as a generator and these charge in turn could be allocated on a $/MWh basis.

Applying proposed rule changes to loads

- Co-optimisation of loads and network elements as well as generation could be important to get efficient market outcomes if the increases in size and growth of data and AI centres continues.
- Before the NEM is faced with several large data centres, a prudent approach would be to have co-optimisation and appropriate charging in place.
- Larger loads driving FCAS costs could be identified (above a minimum size) with costs allocated similar to generators:
  - based on some variation on the runway or charging based the shadow prices in the co-optimisation could be made to satisfy the principles in NER 3.1.4 (a) 8.

Appendix 1

Largest Cost Allocation Using Shadow Prices: An alternative to the “runway” approach



Linear programming and marginal costs

- One of the nice features of linear programs (LPs) is that they determine the marginal costs of meeting each constraint.
- In the NEM, the RRPs and FCAS prices are determined from regional energy balance and FCAS requirements constraints.
- In the co-optimization of FCAS requirements the constraints can be split into two groups:
  a. The dispatch of FCAS to meet the FCAS requirement, and
  b. The determination of the FCAS requirement based on determination of the largest contingency
- In a mathematical programming co-optimization, these constraints look like the following for each raise contingency service x and units j and k:
  a. $\sum_k FCAS(k,x) \geq FCASRequirement(x) = \ LargestContingency(x) - load\ relief$
  b. $LargestContingency(x) \geq Energy(j) + Reg(j) + FCAS(j,x)$
- From equation *a* we use the constraint shadow price of the requirements constraint to get the FCAS price for service x, this is what is done in the NEM to get the FCAS price (marginal cost of FCAS).

Largest contingency shadow prices from co-optimisation

- The determination of the FCAS requirement based on determination of the largest contingency provides useful information for how to allocate the largest contingency cost
- The set of constraints

  b. $LargestContingency(x) \geq Energy(j) + Reg(j) + FCAS(j,x)$

  comprises one constraint for each unit and service x. For most units, their energy + regulation + FCAS will be less than the largest contingency for service x. Their constraints are not binding and will have zero shadow prices.

  On the other hand, one or more units, k, will have dispatches which are equal to the largest contingency

  $$LargestContingency(x) = Energy(k) + Reg(k) + FCAS(k,x)$$

- Each unit k with a combined dispatch target equal to the largest contingency will have a non-zero shadow price. This shadow price indicates how much the unit contributes to the largest contingency constraint. The sum of these shadow prices equals the shadow price of the FCAS requirement.

$$\sum_k ContingencyPrice(k,x) = FCASprice(x)$$

Shadow price cost allocation for largest contingency

- The shadow prices of the largest contingency (constraint marginal costs) provide an economic basis for allocating FCAS requirement costs, whereby each generator pays in proportion to its shadow price under its largest contingency constraint.
- The total cost of an FCAS service is:

$$Total\ cost\ of\ service = FCASrequirement\ x\ FCASprice$$

- Using the shadow prices from the largest contingency constraint the amount generator k pays is

$$CostFCAS(k) = \ FCASrequirement\ x\ ContingencyPrice(k)$$

- Note that

$$\begin{aligned}
&\textstyle\sum_k CostFCAS(k) \\
&= \textstyle\sum_k FCASrequirement\ \times\ ContingencyPrice(k) \\
&= \ FCASrequirement\ \times \textstyle\sum_k\ ContingencyPrice(k) \\
&= FCASrequirement\ x\ FCASprice
\end{aligned}$$

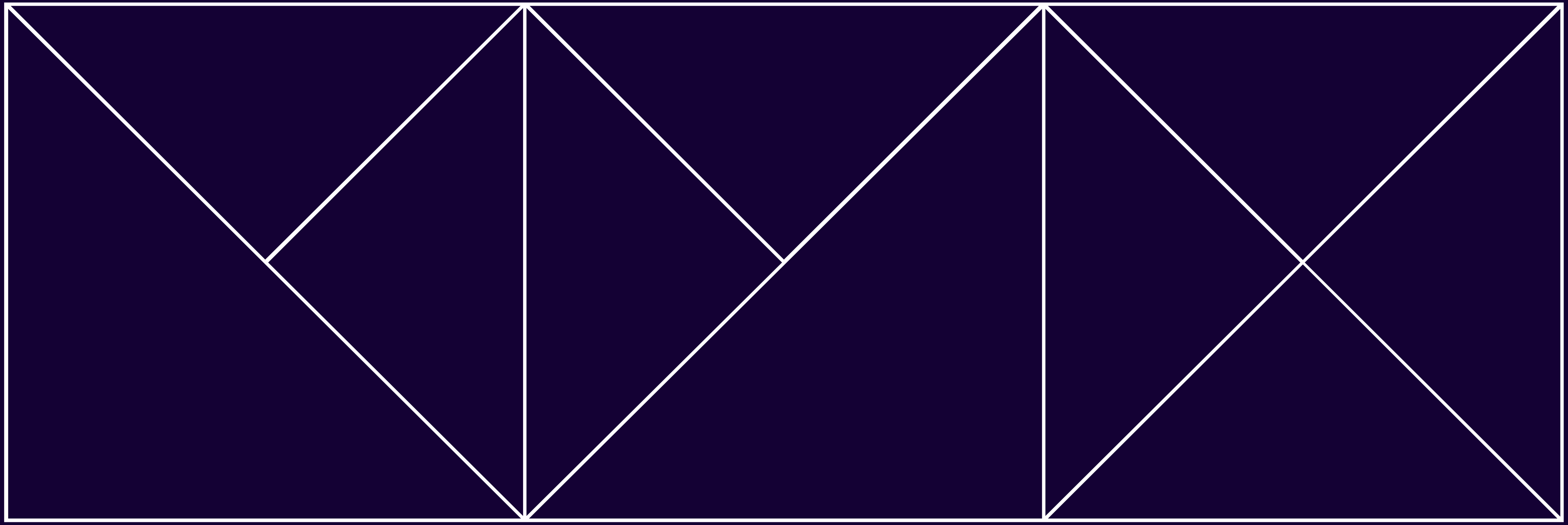
ACIL ALLEN

Conclusion

- Although this approach appears to be quite elegant, as all of the shadow prices from the FCAS requirements constraints and largest contingency constraints are used to determine the FCAS price and the cost allocation to individual generators, there can be an arbitrariness to the cost allocation when there are many equally valid solutions for the shadow prices for the largest contingency constraints. In linear programming terms, this is known as dual degeneracy and can occur at times. There may be some linear programming methods that could remove or at least ameliorate this issue like what is done in dispatch with tie breaking.
- Based on each unit's bids and offers this approach including the cost allocation does not end up with any 'undesirable' dispatches. That is dispatches that units would have been better off not being one of the largest contingencies and not paying their FCAS costs.
- The approach does not end up with any of the uneconomic incentives that can occur with the 'runway' approach but may appear to excessively penalise the largest contingencies.
- If this was a concern to the AEMC the approach could be used just to allocate the extra FCAS required for the largest contingency like with the modified 'runway' approach and then recover the costs proportionally on a MWh basis for all the other generators, loads and network elements.

Appendix 2

Some backcasting results – 2024-25



Overview

- Backcasting was undertaken using NEMPy, an open-source python-based NEM model
- Caution must be applied to the results as we undertook like-for-like modelling for the 2024-25 FY which did not include:
  - the incorporation of energy targets and FCAS and regulation enablement in determining the largest contingency
  - Network contingencies (these were excluded as our analysis was deliberately limited to generator-based contingencies)
- The modelling indicates that :
  - co-optimisation increases the value of spot market trade in most intervals and in aggregate
  - overall contingency FCAS costs (settlements) should be lower under co-optimisation.

Value of spot market trading – 1-17 August 2024

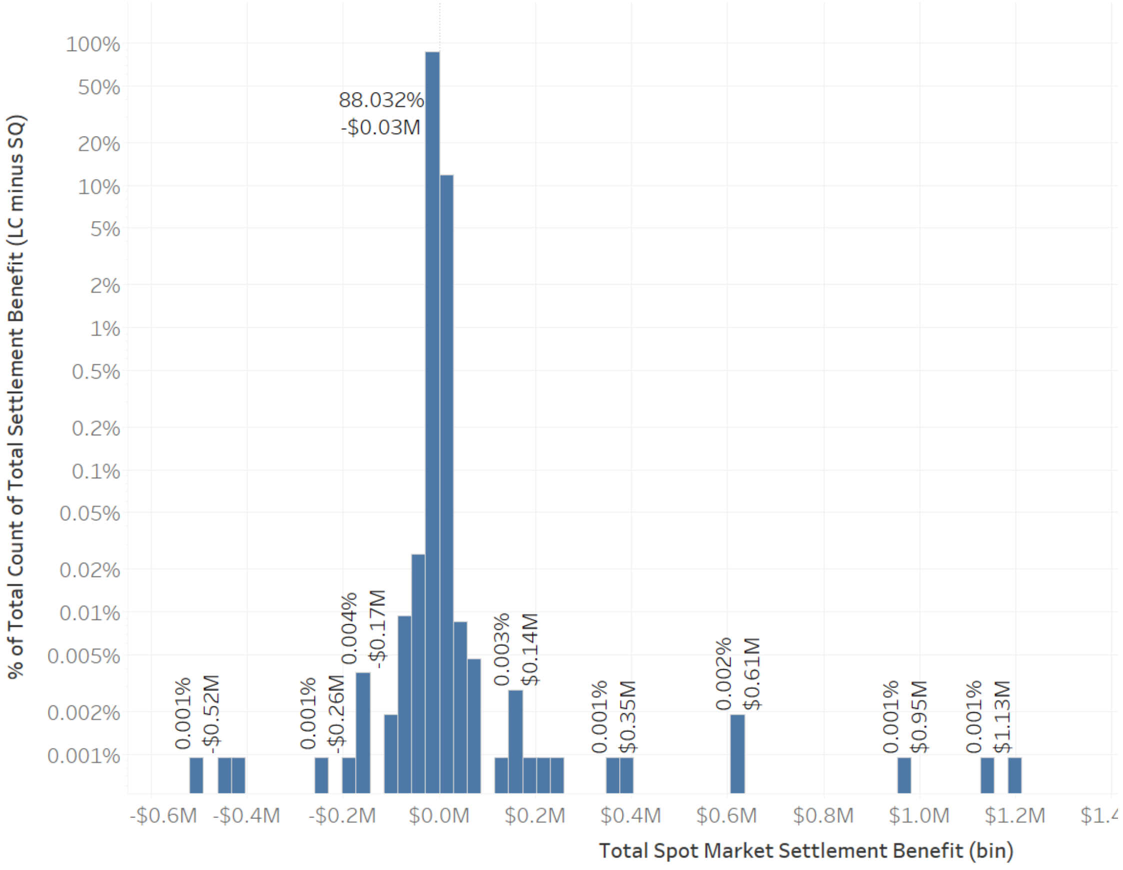
Day	Net welfare benefit
1-Aug	-$1,641
2-Aug	-$412
3-Aug	-$2,297
4-Aug	-$2,598
5-Aug	-$20,444
6-Aug	-$2,228
7-Aug	-$4,245
8-Aug	-$3,285
9-Aug	-$2,153
10-Aug	-$2,214
11-Aug	-$1,281
12-Aug	-$1,890
13-Aug	-$981
14-Aug	-$2,555
15-Aug	-$2,868
16-Aug	-$1,182
17-Aug	-$278
Total	-$52,551

- The table shows the estimated change in the value of spot market trading for the first 17 days of August 2024.
- Consistent with theory, co-optimising the largest contingency is expected to maximise the value of spot market trading (welfare).
- This was tested by comparing the welfare objective (objective function value) under SQ and LC during the first half of August 2024, including 5 August, when energy spot prices were materially higher under SQ than under LC.
- The value of spot market trading is estimated to be $52,551 higher under co-optimisation. If 5 August is excluded (an unusual day), the average daily increase in the value of spot market trade is around $2,030. Applying this average over 365 days of the year suggests that the annual improvement in the value of spot market trading would be around $ 740,000.

Changes in settlement costs

- The figure shows the frequency distribution of changes in costs to participants (total FCAS and energy settlement values), aggregated into bins of $1 million or greater.
- The results exclude the ten extreme single 5-minute dispatch intervals.
- When these extreme intervals are excluded, co-optimisation delivers a net benefit of around $3.7m.
- The extreme outcomes occur in a minimal number of tight intervals and are driven by discrete bid bands, typically from BESS operating at the top of energy bid stacks. As BESS capacity and competitive depth increase, bid band granularity is expected to improve, reducing sharp marginal price jumps and lowering the likelihood of these extreme cost outcomes.

Frequency distribution of Total Spot Market Settlement Benefit - FY25 (LC minus SQ)



For more information

Paul Hyslop
+61 417 392 079
p.hyslop@acilallen.com.au

Jonathan Ben-Tovim
+61 422 307 752
j.ben-tovim@acilallen.com.au

acilallen.com.au

Stephen Wallace
+61 419 234 353
stephen.swadvisory@gmail.com

SW Advisory

ACIL ALLEN