Opportunity Costing in the NZEM
Implications of Decarbonisation

Prepared for

MDAG

by

E Grant Read

Final Version

19 January, 2022

EGR Consulting Ltd

Executive Summary

1. The purpose of this report is to:
   - Lay out the basic theoretical framework of opportunity costing in hydro dominated electricity markets in a way that can hopefully form a common basis of understanding between those who may, or may not, have a background in electricity sector analysis of this particular type; and
   - Draw out the implications of that established theory for the valuation of hydro resources in the radically different sectoral environment that is now emerging.
2. We first outline the history of opportunity costing concepts, and of their application to the New Zealand electricity sector, noting that the theory described here was mainly developed in public sector environments, here and elsewhere, well before the NZEM was even conceived, and applies equally in a centrally optimised environment, or in a perfectly competitive market.
3. Accordingly, we explain the theoretical framework, as it would apply in a centrally optimised, or perfectly competitive context, and discuss the implications for operational, price, and water value patterns:
   - First, Chapter 2 focusses on the determination of the Marginal Water Values (MWV) for a single reservoir;
   - Next, Chapter 3 explains the complications that arise when considering the interaction between more reservoirs and system elements;
   - Then Chapter 4 examines the implications that removing thermal from the system, while also accommodating more renewable entry, would have on reservoir management, and MWV estimation.
4. Finally, in Chapter 5, we turn to discuss the extent to which all of the above extends to apply in a realistic market context, and the impact that market concepts like "risk aversion", "gaming", and "entry limit pricing" might have on observable outcomes.

Chapter 2: Single reservoir analysis

5. Starting from Chapter 2, we emphasise that much of what is commonly described as electricity price "volatility" is really predictable "cyclic variation". This leads us to examine reservoir management from a "deterministic" perspective, which aims to maximise the value of reservoir storage capacity to physically arbitrage by transferring water from low valued summer periods to high valued winter periods, and from low valued night-time periods to high-valued daytime periods.
6. From that deterministic perspective, we see that the common assertion that the marginal value of water must be zero when the reservoir is full, does not hold, if the reason the reservoir is full is because the manager has decided it should be full so as to carry as much water as possible forward from a low value period to a higher valued one. In fact, from this deterministic perspective:
   - Stock levels should be expected to cycle regularly up, during periods of relative over-supply, and down during periods of relative under-supply, with the key cycles being daily, for small hydro reservoirs, and annual for large ones.
   - Since reservoir capacity is costly, we do not expect reservoirs to be built so large as to be able to fully arbitrage away all marginal value differences, in either daily or annual cycles.
   - Thus, under deterministic assumptions, optimal reservoir management must involve holding the reservoir full (apart from minor daily cycling) for some time at the end of each filling cycle, and approximately empty (apart from minor daily cycling) for some time, at the end of each emptying cycle.

- For an annual reservoir, we should expect to see this pattern when a deterministic analysis is performed on the assumption of expected inflows, and probably for most individual hydrological years, with the length of time optimally spent in the full and empty states depending on how different the supply/demand balance is in summer vs winter, in each scenario.
- The day/night situation will be similar for most small reservoirs, on most days, but some may not need to reach their storage limits on some days.
- From this perspective the marginal value of water in storage (MVS) would be constant, at a lower level during the summer/night period, then rise over the period when the storage is full to a higher level, which is then maintained over the day/winter period, before falling again over the empty period.
- While at their bounds, reservoirs would effectively be operating in run-of-river mode, albeit with some daily cycling, but non-supply and spill are both very unlikely, with perfect foresight.

7. That perspective is clearly unrealistic, on its own, because there are many sources of true volatility in the sector, making a stochastic perspective important, too. From a "purely stochastic" perspective (i.e. assuming that flows and market requirements are purely random, with no predictable average patterns), the main function of any storage facility is to buffer the effects of random fluctuations in the supply/demand balance. From that stochastic perspective:
   - Since reservoir capacity is costly, we do not expect reservoirs to be built so large as to be able to fully absorb all such fluctuations, so stock levels should be expected to reach both full and empty bounds, so we may see spill, and also non-supply if the reservoir is critical for national supply security.
   - But storage should be managed to revert, when possible, to levels far enough above the lower limit to allow non-supply to be avoided when flows randomly drop off and, less importantly, far enough below the upper limit to allow high flows to be captured when they randomly occur.
   - The Marginal Cost of Release (MCR) will obviously be zero, when spill becomes inevitable, and that may happen even before the reservoir is full.
   - For a large annual reservoir "shortage" may mean national non-supply, and optimal reservoir management must favour high stock levels, with an inevitably increased probability of spill, over low stock levels, with an unacceptably high probability of expensive non-supply.
   - For a small daily reservoir, "shortage" may just mean inability to take full advantage of a (probably moderate) unexpected intra-day price spike. So, it may well be optimal to use the water when prices are known to provide a reasonably valuable use for the water, rather than holding water back in case a higher spike occurs, but eventually only finding less valuable uses.
8. We discuss how the interplay between these two perspectives can imply a wide variety of outcomes, depending on the balance between energy capture, storage capacity and utilisable release capacity, in different hydro systems. Thus, a wide variety of marginal water values should be expected, from different reservoirs, at the same time, and from the same reservoir at different times.
9. In general, the Expected Marginal Water Value (EMWV) can be estimated by simulating realistic management (i.e. without perfect foresight) of a large number of hydrology sequences, and determining the (conditional) MWV for each one, from some future "marginal economic opportunity" in that sequence.
10. The marginal opportunities available vary greatly, depending on the reservoir, storage level, time of year or day, and scenario, but:
    - There is no marginal economic opportunity available to release more water from a reservoir in periods when that reservoir should optimally be releasing at maximum. And there is usually no economic opportunity available in periods when that reservoir should optimally be releasing at minimum. So, increased price volatility, implying more extreme prices to levels in those periods will have no impact at all on the marginal water value of that particular reservoir.

- There is no marginal economic opportunity available to store water for release in periods beyond the next time when a reservoir is expected to be full in a particular scenario. And there is usually no economic opportunity available to store water for release in periods beyond the next time when a reservoir is expected to be empty in a particular scenario, either. So, we should expect to see significant cyclic variation in EMVW across seasons (for larger reservoirs), and across the day (for smaller storages).

11. As a result, water that can actually be held in storage might be given a high opportunity cost value (MVS) if it seems at all likely to be required to avoid non-supply at some later date, even when excess (unstorable) water is being spilled from the same reservoir, at an implied MCR of zero.
12. Importantly, the need to hold storage levels away from bounds to deal with volatility reduces the effective capacity available to arbitrage between low and high valued periods. So, it actually increases the expected cyclic variation between day and night-time prices, and between summer and winter prices, thus increasing the importance of the fundamental insight derived from the deterministic analysis: Namely that the underlying MVS must be rising while reservoirs are relatively full around autumn, and falling while reservoirs are relatively empty around spring.

Chapter 3: Complications

13. Chapter 2 discusses how, traditionally, the next time release from this reservoir would be "on the margin" has most often been the next time when storage reaches a "guideline" corresponding to the assumed SRMC of some thermal station. But Chapter 3 points out that:
    - The true SRMC of thermal generation should itself be determined by opportunity costing logic, in many cases, so it can actually vary substantially over time, and in any case is only inferred by hydro reservoir managers by observing thermal station offering behaviour.
    - Thus, the status quo might more properly described as finding an equilibrium balance between stock levels in all storages, both hydro and thermal, in which all marginal stock values are mutually determined by opportunity costing.
    - And the ultimate benchmarks are really the potential costs of non-supply, the internationally traded prices of some fuels, and (ultimately) the investment costs of the range of capacity expansion options defining the optimal mix of new capacity, as it appears to be at that point in time.
14. This same theory applies to smaller storages operating on a daily cycle, and to head ponds serving stations on river chains downstream from long term storage reservoirs. In this latter situation, complications arise once flow and/or storage limits are approached, and can imply that the MCR for water released from those storages is either higher or lower than that calculated for the top reservoir in the chain, thus implying a steeper aggregate offer curve than might be expected.
15. The new investment cost benchmark comes into play indirectly, by virtue of its influence on the balance of new plant entering the system, and hence on the opportunities available for existing plant to optimally contribute economic value by adjusting their production schedules. But that indirect influence is critical, whether in a centrally optimised or perfectly competitive setting, to optimising the economic value of a system whose costs are largely determined by capital investment.
16. In short, the setting of MWVs and prices under the status quo is a much more complex and fluid affair than many commentators seem to assume. But the influence of thermal is undeniably important, because:
    - Thermal generators (and guidelines) play a major role in moderating the effects of hydrological fluctuations, and that influence results in storage trajectories being actively steered away from extreme levels, and towards an idealised annual/daily cycle at more moderate storage levels, maintaining MWV at as consistent a level as possible;

- Thermal SRMC costs (along with spill and non-supply costs) have traditionally (although perhaps naively) been seen as setting the MWVs for most simulated hydrological sequences, and hence EMWV; and
- Thermal investment costs have traditionally been seen as the main (or at least most easily analysed) determinant of the optimal long run distribution of electricity prices, from which the opportunity costs of all storage options are indirectly derived.

Chapter 4: Impending Change

17. Increasing intermittent renewable generation while removing thermal will clearly put much more pressure on the remaining (hydro) storage capacity, as it tries to manage both short and long-term cyclic variation and volatility in supply/demand balances:
    - On one hand, the lack of thermal support will make it much more difficult for reservoirs to fulfill their "deterministic" function of transferring (the potential energy) stored in water from night to day, and from summer to winter. Long term storage will need to be held at much higher levels in autumn, if the risk of late winter non-supply is to be maintained at an acceptable level, but can also be expected to fall much more quickly to low levels when inflow and/or other intermittent generation is reduced.
    - On the other hand, the lack of thermal support to cope with increased volatility in the net demand/supply balance will also increase pressure on storages, both large and small, to fulfill their "stochastic" function of standing ready to surge generation quickly when generation from other sources falls, and to maintain that generation over a long enough period to cover extended shortfalls. And that suggests that water levels can not be allowed to fall too low, at any time.
18. The optimal overall balance between these two requirements is an empirical matter, but the combined impact of losing thermal system "storage", while needing to hold hydro storage at higher levels to manage stochasticity, suggests that there will be less effective capacity available to arbitrage between low- and high-priced periods. And that should logically suggest an increase in both intra-day and inter-seasonal price differentials.
19. The same may be true for EMWVs, but we may experience quite long summer periods in which major reservoirs are releasing at minimum rates, so as to maximise capture of inflows which will ultimately be valued at a high EMWV, reflecting a significant probability of being able to store incremental water for winter use.
20. And the system may evolve towards a situation in which EMWV has to be set quite high, right across the year, with hydro generation minimised over summer to maximise the capture of renewable generation for winter use.
21. Long-term storage in pumped hydro, or biofuel stored for use in a "green thermal peaker", would help to reduce both volatility, and cyclic variation.
22. But a significant contribution may need to come from demand-side options that can either reduce, or defer, electricity consumption when the supply/demand balance becomes critical. In particular, a large-scale development producing some internationally tradable commodity (e.g. Ammonia) might actually provide the nearest equivalent to the kind of "benchmark" that thermal SRMCs have been assumed to provide in the past.
23. Because solar generation peaks during the day, the overall impact on the deterministic component of cyclic intra-day price variation may be more mixed, but meeting winter evening peaks could be challenging, when wind generation is low.
24. Any technology that can flexibly adjust the supply/demand balance will be at a premium, and that includes short-term storage in batteries, which should be expected to play a major role in limiting both daily cyclic price variation and short-term price volatility.
25. Since the MVS set by each hydro/battery storage will depend on its own storage, and charge/ discharge capacity, there should still be a range of prices across each day.

26. Optimal price volatility will clearly increase, though, and so will the optimal volatility of hydro generation, and hence of river flows, with potentially significant socio-environmental impact.
27. In combination, these changes do suggest that system optimisation will become more challenging both physically and analytically. But that does not invalidate the overall opportunity costing framework described here, or suggest that there is any better way of understanding the optimal operation of storages in the emerging system.
28. In principle, entry cost economics actually determines the ideal equilibrium distribution of electricity prices which, in turn, drives the setting of optimal MWV. Of itself, removing thermal options from the expansion mix would clearly increase price levels. But that effect will be offset if renewable replacements are declining in cost. So, the combined impact on average price levels is an empirical question, depending on the balance of these effects.
29. Still, analysis of the long-run equilibrium price distribution implied by the costs of potential entrant technologies clearly reinforces an expectation that prices may be quite low for much of the (summer) time, but also very high in winter peak periods.

Chapter 5: Market Realities

30. All of the above discussion relates to an idealised situation in which both investment and operation of the electricity sector are either centrally optimised, or implicitly optimised by a perfectly competitive market. But the ways in which the real market may differ from this ideal are discussed in this final chapter, with a focus on how that might affect expectations about market behaviour in the emerging environment.
31. Our assumption that costs are convex and especially that capacity can be expanded continuously at linear cost, are obviously unrealistic, but they are much truer for the technologies now dominating expansion plans than they ever were for large scale thermal or hydro developments in the past. Still, we should expect to see significant fluctuation, from year to year, around the long run average trends derived from our optimisation-based analysis of the influence of entry costs on market prices.
32. Few real market participants will have a full understanding of the theory we have discussed, and none will have the analytical resources to fully implement it. So, we should expect individual policies to deviate significantly from the hypothetical ideal, but the broad shape of aggregate behaviour should still align with that theory.
33. Since the "snapshot" offer form employed by the NZEM does not allow participants to express physical constraints in the form required for inter-temporal optimisation, we should expect to see participants offering capacity affected by intra-day inter-temporal constraints at fairly extreme prices, in an effort to manage the physical situation.
34. And we suggest that further developments will be required in the DSM area, if that is to provide the kind of flexibility we anticipate to be necessary, once thermal generation capacity is withdrawn, or provide the kind of benchmark we assume in our discussion of hydro opportunity costing.
35. One key issue is that different optimisation models, each applying the same basic principles in its own way, are known to produce quite different looking "optimal" marginal water value curves. But the issue is not how high or low the MWV is, for a particular storage level, but what effect the overall curve has on the trade-off between system operational cost, spill, and shortage probability.
36. Simulation and experience suggest that equilibrium outcomes are actually not very sensitive to raising the MWV curve level. The system settles into a new equilibrium, with more storage, less shortage, and more spill, without necessarily raising system costs or prices by much at all.
37. In order to avoid physical risk (particularly in river chain management), and reduce financial risk, hydro operators can be expected to offer significantly steeper offer curves than a risk neutral analysis might suggest.
38. Those offers are likely to be targeted at stabilising output around contract levels and, for large generators, that effect may be exacerbated by "gaming" incentives to depress prices for power

effectively bought in from the spot market, when they face deficits, and to raise prices for power effectively sold to the spot market, when they have a surplus.

39. Removing thermal capacity is likely to strengthen the incentives for the remaining flexible capacity providers to make steeper offer curves, and that may become an issue, given the increasing need to provide flexible support to intermittent generator entrants.

40. A previous analysis suggested that market power may have been exercised to stabilise the industry price distribution around the shape corresponding to the optimal entry mix of conventional thermal capacity. But it will take some time for the industry price distribution to stabilise around the significantly different long run equilibrium price pattern, which Chapter 4 argues can be expected as a result of the radically different optimal entry mix in the new environment.

Contents

Opportunity Costing in the NZEM:
Implications of Decarbonisation

1. Introduction

1.1. Context

This report has been prepared for The Market Development Advisory Group (MDAG) to complement their recent work on price discovery with 100% renewable electricity supply, and their Issues Paper, to be released in February 2022.[1] The purpose of the report is neither to critique, nor to justify, the assumptions or conclusions of that work, which was entirely designed and performed by other analysts. Nor is it intended to propose solutions, or even directions, for the market development issues under consideration by MDAG. We focus instead on the basic theory of "opportunity costing" as it has been applied to the valuation of hydro resources in this country, and elsewhere, over many decades. Specifically, our purpose is to:

- Lay out the basic theoretical framework in a way that can hopefully form a common basis of understanding between those who may, or may not, have a background in electricity sector analysis of this particular type; and
- Draw out the implications of that established theory for valuation of hydro resources in the radically different sectoral environment that is now emerging.

The debate over the latter issue is obviously central to MDAG's considerations, and we understand that two contrasting viewpoints have emerged that might be crudely categorised as expecting "business as usual" vs a “bang-bang world".

In reality, we suspect that all analysts actually agree that business will not be quite "as usual", and that we will be seeing both prices and operating schedules alternating between extremes more often, over both daily and annual cycles, due to increasing reliance on inherently volatile energy sources such as wind and solar, and possibly reduced net storage capability (depending on investment decisions yet to be made).

So, the question is really one of how far the market might move in that direction, and how fast. We see that as an empirical question that should be addressed by the kind of modelling MDAG is doing, and not something on which we have been asked to express an opinion here. Our contribution is simply to explain how the theoretical framework discussed here could apply, even at the extremes of the spectrum. It could be that the market prices become so volatile that changes need to be made to market arrangements but, from a reservoir management perspective, we believe that the changes, e.g. to MWV estimation, are more changes of degree than of kind.

DISCLAIMER

The author has long history of applying the concepts and techniques discussed here, in both centrally planned and market contexts, in New Zealand and elsewhere, The author was also a significant contributor to the WEMS study which developed the NZEM market design, and has provided commentary on various aspects of market performance over the years, at the request of various parties. This included an overview of opportunity costing theory prepared for a broad consortium of New Zealand generators in 2018. Naturally there will be significant overlap, and hopefully consistency, between the views expressed here, and in those various other contexts. But the author has not worked closely with any market participant on

---

1 *Price discovery under 100% renewable electricity supply: Issues discussion paper*, MDAG report, to be released in February 2022

these issues, since the breakup of ECNZ. Thus, the views expressed here are solely the author's, not those of any other party.

1.2. History

Opportunity costing is a well-established concept, by no means unique to the NZEM, or to the electricity sector. The fundamental theory discussed here is in fact derived from optimisation theory, and applies to any kind of storable goods, feedstocks etc.[2] There is a long established and fundamental theorem stating that we can not determine the optimal allocation of any such stock, over time, without also determining, at least implicitly, a corresponding "shadow price" representing the "marginal value" of units in that stock.

As we will see, a stock management regime can not be strictly optimal unless the "marginal value" of a unit taken out of stock, at any time, equals the "opportunity cost" of saving it to use in some later period. Many stock managers may not be consciously aware of this theory at all. They may not manage stocks optimally, or they may use a stock management package without understanding its internal workings. But they still apply a heuristic form of the same reasoning whenever they think: "I'd better stop using that stock so fast, because otherwise I'll run out, and there will be hell to pay". Or: "I'd better start using that stock up faster, because otherwise I'll have nowhere to store it all". Or perhaps they just set "guidelines" or "trigger levels", or "rule curves" at which they start taking actions to build up, or run down, stock.

The theory has been consciously and consistently applied to manage the New Zealand hydro reservoir system, though, since at least 1979, via a number of reservoir optimisation models developed by the NZED, and MoE. Those developments were partly based on the original concepts articulated by Massé[3] in the 1940s, further developed by Adnet[4] and others in France, and applied to a deterministic model of the New Zealand hydro system in the author's own thesis.[5] Another major contribution came from Stage and Larsson[6] in Scandinavia, as implemented in the NZED's STAGE/VALWAT model.[7] The MoE's PRISM model[8] (later SPECTRA) then synthesised these ideas to implement a methodology later known as "Constructive" Dual Dynamic Programming (CDDP).[9]

But essentially the same theory holds in a market, such as the NZEM, even though it does not employ any kind of centralised reservoir optimisation. So, the MoE model was later adopted and adapted by ECNZ and its various successors.[10] And the same theory is central to alternative reservoir optimisation models and

---

2 This "duality" theorem is actually much more general, but we will only consider its application to stock management, over time.

3 P. Masșé Les Réserves et la Régulation de L'avenir dans la Vie Économique. Hermann & Cie. Paris, 1946

4 M Adnet et. Al, “Optimisation in the Use of the Means of Production and Transmission” *CIGRÉ* Paris, Vol 2, Paper 32-12, 1968.

5 E.G. Read: *Optimal Operation of Power Systems* Dept of Economics, University of Canterbury, Christchurch, New Zealand. August 1979. As summarised by:
E.G. Read: "Reservoir Release Scheduling for New Zealand Electricity - A Non-Linear Decomposition Algorithm", *New Zealand Operational Research*, vol. 11, no. 2, July 1983, p.125-14.

6 S. Stage and Y. Larsson. "Incremental Cost of Water Power", *AIEE W inter General Meeting* , 1961

7 J.F. Boshier, G.B. Manning & E.G. Read: "Scheduling Releases from New Zealand's Hydro Reservoirs" *Transactions of the Institute of Professional Engineers in New Zealand*, vol. 10, no. 2/EMCh, July 1983, p.33-41

8 E.G. Read, J.G. Culy, T.S. Halliburton & N.L. winter: "A Simulation Model for Long-term Planning of the New Zealand Power System", in G.K. Rand (ed.) *Operational Research 1987*, North Holland, p.493-507
Utilising the reservoir management module described by:
E.G. Read: "A Dual Approach to Stochastic Dynamic Programming for Reservoir Release Scheduling", in A.O. Esogbue (ed.) *Dynamic Programming for Optimal Water Resources System Management*, Prentice Hall NY, 1989, p.361-372

9 E. G. Read & M. Hindsberger “Constructive Dual DP for Reservoir Optimisation” in S. Rebennback, P.M. Pardalos, M.V.F. Pereira & N.A. Iliadis (eds) *Handbook on Power Systems Optimisation* springer, 2010, Vol I p3-32

10 J.Culy, V.Willis and M.Civil: Electricity Modelling in ECNZ Revisited *ORSNZ proceedings*, 1990

market simulation packages used in a great many countries, such as SDDP[11] and PLEXOS,[12] and those employed more locally, such as Emarket[13] and DOASA.[14]

The relevance of this history is that it underlines three major points:

- First, a modern observer might be inclined to think that the "market oriented" thinking described in this report has emerged to rationalise, and perhaps justify observed market behaviour. In fact, the opposite is the case. These ideas were developed and applied over many decades, in public sector contexts, to understand and manage reservoir economics with the goal of maximising national benefit. Familiarity with these concepts then lead New Zealand, in particular, to pioneer development of the NZEM market, in which those same principles now underpin market operations, and guide interactions between independent commercially motivated agents.
- Second, while other methodologies were experimented with in earlier years, there is now a very strong international consensus that the "market-like" concepts developed in this literature are the best, and perhaps only consistent way of understanding how hydro should, and generally does, operate, whether in a market environment or not. In fact, we believe that this "market-like" paradigm is also the best way to explain the inner workings of any of these optimisation models, whether they explicitly report marginal water values, or not. Hence, Chapters 2-4 all rely on this paradigm extensively, in developing insights into how reservoirs and water values "should" be expected to behave, in a perfectly competitive market, now and in future. It is only in Chapter 5, where we turn to consider the alignment between this idealised theory and real market behaviour that real-world issues such as contracting, risk management, and market power come into focus.
- And finally, the models referred to above have been collectively applied across electricity sectors with widely differing proportions of thermal vs hydro power, and widely differing hydro systems characteristics. SDDP, in particular, was developed in Brazil, for an electricity sector with very little thermal, and has been applied in many broadly similar Latin American systems. So, it should not be too surprising that we argue, in Chapter 4, that the same theoretical framework, and hydro valuation principles, can still be applied in New Zealand, as we move to eliminate thermal generation.

In the New Zealand context, the "opportunity cost of water" stored in major reservoirs has generally been referred to as the "Marginal Value of Water" (MVW), "Marginal Water Value" (MWV), or sometimes (more correctly) as the "Expected Marginal Water Value" (EMWV).[15] But confusion readily arises because hydro schemes differ widely in characteristics such as maximum release/generation rate, storage capacity, and flexibility.

So, misleading and/or confusing statements often start with something like "Because hydro is on the margin, we know that....". That begs the question: "Which precise hydro station is on the margin…... and what is its actual MWV at this moment?" As we will see, the same general theory applies to all storable hydro, and indeed to all storable inputs or outputs, but the MWV of a downstream station with limited head-pond storage may be quite different from the MWV of a larger reservoir, even in the same river chain, at the same time. So, the price implications of having "hydro on the margin" can be much more variable than often seems to be assumed.

---

11 M.V.F. Pereira, and L.M.V.G. Pinto "Multi-stage stochastic optimization applied to energy planning" *Mathematical Programming* v52, pp. 359–375, 1991

12 G. R. Drayton, and F. Valdebenito: *Stochastic Multi-stage Hydro Optimization: Making Better Choices Across Inflow Uncertainty* Energy Exemplar June 2019, Title of White Paper (hubspotusercontent00.net)

13 EMarket http://emk.energylink.co.nz/EMK:Multi_Dimensional_Water_Values

14 A.B. Philpott and Z. Guan. Models for estimating the performance of electricity markets with hydro-electric reservoir storage. Technical report, Electric Power Optimization Centre, 2013.

15 In most cases, the value is not calculated in water units either, but in terms of the potential generation available from that water, assuming it to be processed through all downstream power stations, at some assumed efficiency. Strictly speaking, then, it is not a marginal value of "water". Still, we will stick with the industry convention, and discuss the implications of varying downstream conversion rates later.

1.3. Overview

We have already outlined the history of opportunity costing concepts, and of their application to the New Zealand electricity sector, above. In the remainder of this paper, and then move on to discuss:

- The theory itself, as it would apply to a single reservoir, in a centrally optimised context (Chapter 2);
- The complications that arise, when considering the interaction between more reservoirs and system elements, also in a centrally optimised context (Chapter 3);
- The implications that removing thermal from the system, while accommodating more renewable entry would have on reservoir management, and MWV estimation, if we were operating under the same centrally optimised regime (Chapter 4); and finally
- The extent to which all of the above theory still applies in a more realistic market context, and the impact of market concepts like "gaming", and "entry limit pricing" might have on observable outcomes (Chapter 5).

The key conclusions may be summarised as follows:

1) The general theory for determining water values is long-established, and is the same whether applied in a centrally managed system or a perfectly competitive market.
2) The general theory provides that water values should be driven by the Expected Marginal Water Value (EMWV) of holding water for future use, where that EMWV represents the probability weighted average, over a wide range of simulated future scenarios, of the value obtained from releasing water to take advantage of the first marginal economic opportunity available in each scenario.
3) That same theory applies to all storages, both hydro and thermal, operating over both short- (e.g. daily) and long-term (e.g. annual) cycles, whether they are inked into river chains, or physically independent.
4) Much of what is commonly described as electricity price "volatility" is actually predictable "cyclic variation" (typically diurnal and seasonal) which optimal storage management will arbitrage away, as much as possible, implying a distinct cyclic EMWV pattern underlying the stochastic "noise".
5) For a system in equilibrium, the EMWVs will be determined by the costs implied by various forms of demand-side management, the prices of thermal fuel, spill probabilities, and (indirectly) new investment costs.
6) Moving to 100% renewable supply doesn’t change the general theory of MWVs but it will significantly alter the environment in which it is applied.
7) Outcomes will depend on empirical factors, but in broad terms we would expect:
   a) The loss of thermal system "storage", and the need to hold hydro storage at higher levels to manage random influences, reduces effective capacity available to arbitrage between low- and high-priced periods, implying higher intra-day and inter-seasonal price /MWV differentials.
   b) EMWVs may also exhibit stronger seasonality, but may eventually become high all year, thus reserving stored energy to cover winter loads, and summer peaks, despite a higher risk of spill,
   c) But the effect on average prices is not clear because it will be affected by new investment costs.
   d) Marginal generation expansion costs will continue to limit prices, and hence EMWVs, but imply markedly different equilibrium price patterns from those implied by thermal Entry Limit Pricing.
   e) Technologies that can flexibly arbitrage supply/demand imbalances will become more valuable, with the marginal expansion/operation costs of batteries limiting short-term price volatility and day/night differentials, while pumped storage and large-scale DSM options potentially do the same for summer/winter variations.

2. Opportunity Costing for a Single Reservoir

2.1. Introduction

This entire chapter describes theory that would apply in a centrally optimised environment. So, the "prices" discussed here could be "shadow prices" calculated within an optimisation model. Those prices might be reported by the model, and perhaps charged to consumers, or they might not. But, apart from the discussion of the real-world practicalities of managing demand response in Section 0, the theory developed here applies equally to a hypothetical perfectly competitive market environment.

Of course, the NZEM environment is not perfectly competitive, so the theory does not entirely hold. However, we will leave discussion of how market realities like gaming and risk aversion might affect the results and insights developed here until Chapter 5.

This section discusses the application of opportunity costing theory in a very simple situation:

- The load to be met in each hour of each day is defined by a "Load Duration Curve" (LDC);[16]
- With a shortage/non-supply cost applying if it is not met.
- These LDCs vary in a seasonal pattern, and must be met by;
- A set of thermal power stations, each able to freely buy fuel to generate at a known constant SRMC; [17]
- Plus, just one hydro power station, operating at constant efficiency;
- Fed by one water storage reservoir;
- Receiving random inflows, with some seasonal pattern; with
- All generation/release able to start/ramp/stop instantly and costlessly; and
- All reservoir/station capacities and costs constant over the year.

First, we lay out some basic assumptions, and explain the traditional representation of thermal generation, and demand response, in this kind of optimisation.

We then consider two distinct perspectives on this situation, each of which yields some distinct insights, and supports some commonly held beliefs about how MWVs should behave in a hydro dominated electricity sector. Those beliefs may often seem to be in conflict, but the reality is that actually they apply to all reservoirs, but that different perspectives may dominate when thinking about reservoirs with different characteristics, and perhaps at different times of day or year.

---

16 The LDC provides a simplified representation of the daily load pattern, in which load levels are ranked from highest to lowest. For simplicity we will assume it be hourly (rather than half-hourly). This representation is generally deemed sufficient for long term analyses such as this, with the more complex, and computationally challenging, "chronological" representation reserved for studies focussed on management of daily cycles, e.g. in river chains, as discussed in Section 3.3.

17 Traditionally, the "merit order" ranks thermal stations in order of increasing SRMC. In a purely thermal the cost of meeting loads has been minimised by "stacking" stations in merit order to "fill the LDC". That is, we have the station with the lowest SRMC operating as long as possible, to fill as much of the LDC as possible, then stack the station with the next lowest SRMC on top of it, to operate for as long as possible and fill as much of the remaining LDC as possible, and so on. Eventually, the entire LDC is filled, with the stations having the highest SRMC operating for as short a time as possible, if at all. So, although we may think of the station with lowest SRMC as having the "highest merit", we often refer to the stations with higher SRMC as appearing higher in this "merit order stack". The merit order of thermals is usually fixed, but hydro then appears at a variable position in the merit order, depending in its marginal water value, as discussed in this chapter.

Finally, we present a unified framework that brings those two perspectives together, and discuss how, together, they determine all the MWV patterns we should expect to see in reservoirs of different "size"., and discuss the role traditionally played by thermal generation, and demand response, in actually driving MWVs.

2.2. Underlying theory

Before diving into a discussion of reservoir management theory, we should outline some basic assumptions underlying that discuss, and explain some basic concepts that we will be using frequently.

2.2.1. Some basic assumptions

All of the discussion in this report makes certain assumptions that are basic to simplified models in both economics and optimisation theory:

- The quantities under discussion, like storage, release and generation, are always assumed to be continuously variable, between upper and lower bounds, with the lower bound being zero, unless otherwise stated.
- The cost, benefit, and efficiency functions we discuss are all assumed to be convex, so that increasing quantities yield gains at a decreasing rate, or costs at an increasing rate, as is assumed in basic expositions of both economics and optimisation, including the LP methodology used to clear the NZEM.
- So, we are assuming away "integer" factors such as the number of units committed, the cost of starting them, and also the possibility of "forbidden" generation ranges[18]. These factors can be significant, even in hydro systems, and the NZEM offer form is specifically designed to allow the best convex approximation to be presented to the market, but the details do not matter for the purposes of this discussion.[19]
- We also assume away any need to provide ancillary services. Ancillary services are actually quite important in the NZEM, and the offer form allows them to be explicitly represented to the market, and optimised, but the details do not matter for the purposes of this discussion.[20]
- For the most part we will also ignore the influence of "discounting" in the context of intra-day and intra-year reservoir management. In other words. we will treat cost and benefits as being equivalent whether they occur near the beginning or near the end of the planning horizons we consider.[21]

2.2.2. Marginal concepts

A great deal of this discussion focusses on marginal concepts. The "marginal value" of stock can be loosely thought of as either "the value of the last unit released" or the value of "the next unit that could be released". Those two values could be a little different, if stock units are "lumpy". For most discussions the distinction can be ignored, and the marginal value might be set at, say, the average of the last and next units' value. Thus, mathematically, we are assuming our cost/benefit/efficiency functions to be continuously differentiable.

---

18 These are not forbidden ranges in which storage is too high or too low, but particular ranges in which a particular machine can not operate, if committed to generate, e.g. below some minimum running level.

19 E.G. Read, G.R. Drayton-Bright & B.J. Ring: "An Integrated Energy/Reserve Market for New Zealand", in G. Zaccours (ed) *Deregulation of Electric Utilities*, Kluwer, Boston, 1998, p. 297-319

20 See above.

21 In principle, there should be a slight bias toward earning benefits earlier and incurring costs later, but that merely complicates the discussion without yielding much gain, in terms of insight.

This is not actually the case for NZEM offer curves, because they consist of distinct steps, with a distinct change of "marginal cost" when moving from one step to the next. In principle, that could lead us into a complex discussion about the handling of "special cases", in which a step quantity or price happens to coincide with another quantity or price in an optimal solution, or market equilibrium. A general discussion would not change the principles discussed here, though, and would yield more confusion than insight.

Some special cases worth understanding, though, and we will later discuss two of particular importance:

- The special case of a thermal generator that is (assumed to be) able to generate over a wide range, at constant marginal cost.
- The special case of a hydro generator, when release or storage are just at their limits, because then one more unit may be worth a lot more, or less, than one less unit.

One particularly important marginal concept is that of "opportunity cost". The term is often loosely thought of as referring to any profitable "opportunity" one might expect to have in the future. But it should be recognised that it is only "marginal" opportunities that actually set MWV. There will generally have been a great many more profitable opportunities that the manager has already decided to take, but those are no longer in play. What matters is the value to be derived from the next incremental opportunity, or perhaps the last. Importantly, those opportunities must be realistically attainable, too. The expectation of high prices in some future period may provide an opportunity for someone, but it does not present a <u>marginal</u> opportunity, for our reservoir manager, if the prices are expected to be so high that the reservoir is already expected to be releasing at maximum rate, in that period, irrespective of how much water they might decide to store or release today.

2.2.3. Optimisation and economic theory

The fundamental theory discussed here is derived from optimisation theory, and applies to any kind of storable goods, feedstocks etc. Optimisation models are used to find an "optimal solution", by adjusting many "decision variables" until the combination is found which maximises an "objective function", such as net benefit to the firm, or nation. But their ability to do that is limited by the presence of "constraints" set to preclude combinations that are physically impossible, socially unacceptable, etc.

Many optimisation models only report the "primal" solution, that is the optimal values of the decision variables, which in this case would include release, generation, stock level etc. But there is a long established and fundamental theorem stating that we can not determine the optimal solution to such a problem without also determining, at least implicitly, a corresponding "shadow price" representing the marginal cost that each constraint imposed, to the extent that it restricted the objective function from being even higher. The problem of finding these optimal prices is known as the "dual" problem, and this "duality" theory is very general.

We will only consider its application to stock management over time, though. In that context, it implies that we can not determine the optimal allocation of any stock, over time, without also determining, at least implicitly, a corresponding set of "shadow prices" representing the "marginal value" of that stock, in each period.

An optimal stock management schedule will then be characterised by:

- Keeping the marginal value of stock in the stockpile as constant as possible, over time; and[22]
- Keeping the marginal benefit from using the last unit released from the stockpile as close as possible to the marginal value of stock in the stockpile, in each period.[23]

---

[22] Strictly, this should be the discounted marginal value, but we will ignore the influence of discounting because, in this context, the MWV is re-set every time a bound is reached or, as we will see later, a thermal "guideline".

[23] This equality may not be exact, when "stock units" come in discrete sizes, but we can ignore that issue with water. Much larger inter-temporal differences may emerge when physical constraints prevent us from shifting as much stock as we would like from one period to another, but Section 2.3.1 discusses that issue in detail.

If that were not the case, there would be an unexploited "opportunity" to increase the benefits derived from the stockpile, by re-assigning one stock unit from a period in which it had a lower value in use, to some other period in which it had a higher value.

This is known as "arbitrage", so another way of characterising the optimum is to say that all profitable "arbitrage opportunities" have been exploited. From this perspective, we can also see that the marginal value of a unit taken out of stock, at any time, should equal the "opportunity cost" of saving it to use in some later period.

There are actually optimisation algorithms that apply this arbitrage concept explicitly but, irrespective of the actual mechanisms used to find the optimum, the result is the same. So, it can be helpful to imagine and explain the optimisation as being undertaken by a collection of hypothetical agents, within the optimisation engine, trading stocks over time and space, whenever and wherever there are gains to be made, until no more gains can be found.

The only difference, then, between a centralised optimisation and a perfectly competitive market is that:

- The hypothetical agents within the optimisation might be thought of as trying to maximise the global objective function, which in this case would be net national benefit; whereas
- The real agents in a perfectly competitive market, would each be trying to maximise their own benefit.

That makes no difference to either the primal (quantity) or dual (price) solutions, though, because, under perfectly competitive assumptions:

- The national benefit is just the sum of all individual benefits;
- Each agent is a perfectly rational benefit maximiser with access to the same information as a central planner would have;
- Each believes they can only increase their benefit by adjusting the "primal" (i.e. quantity) decision variables under their control, (and that they, individually, can not profitably shift prices);
- Neither the agents nor the central planner are risk averse; and
- While the true transaction costs would differ, we will ignore them in both cases.

Consequently, the arguments in this chapter, and the two that follow, will be presented entirely in terms of an intuitive understanding of how these hypothetical perfectly competitive agents would behave, rather than using formal mathematical proofs from optimisation theory. Note, though, that the understandings developed here are almost entirely based on a 1984 MoE report that does give mathematical proofs,[24] drawn from the author's 1979 thesis.

2.2.4. Thermal generation costs

The analysis in this chapter employs the traditional representation of thermal generation, having a well-defined SRMC, constant over the year. But that can really only be the case if the generator can "freely buy" as little or as much fuel as they may need, as and when needed. So, even if they do have a local stockpile, the " opportunity cost" of fuel in that stockpile is just the price at which more can be purchased which, when multiplied by heir heat rate, defines their SRMC, without any additional "scarcity rent" involved.

This assumption is analytically convenient, but it is debatable whether any generator ever did have such a contract, except perhaps in a public sector environment. Even there, a fixed SRMC calculated in that way could only be a true reflection of national costs under quite limited assumptions about the fuel extraction technology underlying the agreed price. Perhaps, in the MoE era, this assumption might have been valid for relatively small incremental purchases of Maui gas, if it had been available at the MoE's own internally

---

[24] E.G. Read Deterministic Reservoir Optimisation: An Application of the Economic Principles Ministry of Energy, Report ER4006 , 1984

calculated "Depletion Related Opportunity Cost" (DROC).[25] But gas never was available for purchase at that price, which bore no relation to the Maui contract price, and was actually negative for many years.[26]

The only stock that was traditionally replenishable, at a price that clearly represented a cost to the nation, was imported oil. But the benchmark set by oil purchases was always an international traded price that was notoriously volatile. Any local stocks should really have been re-priced regularly to reflect the opportunity cost of the gas, which, in this case, would be the cost of replenishment. So, oil should never really have been seen as providing a constant SRMC benchmark, either. More recently, the development of port infrastructure has allowed importation of coal, but that suffers from the same limitation as oil, when setting a benchmark.[27]

In summary, then, we are actually quite sceptical about the traditional assumption that water values can be clearly benchmarked against a set of stable thermal SRMCs, and Section o later discusses a more realistic analysis, in which, at least for New Zealand, most thermal fuels should actually be opportunity costed, much like hydro. Still, we will explain the traditional analysis, because it easy to understand, and still yields some useful insights.

2.2.5. Non-supply costs

The above discussion suggests that thermal SRMCs might not quite be the "god standard" they are sometimes assumed to be. In fact, "non-supply costs" arguably have a more valid claim to that title. The ultimate function of the entire power system is to supply consumer demand. And the goal of power system planning, or of electricity market design, is to ensure that consumer demand is met at minimum cost, and with acceptable reliability. There is obviously some trade-off between supply cost and reliability, with some consumers, and some societies, perhaps being prepared to pay more than others, for increased reliability. Always, though, reservoir management involves a trade-off between holding storage low enough to avoid spill, and holding it high enough to avoid what has traditionally been called "non-supply".

Section 0 will discuss how this trade-off may be generalised to include more subtle forms of "Demand-Side Management" (DSM) but, for now, we will just consider the possibility that "the lights may go off":

- In a centralised national cost benefit optimisation, we are, at least conceptually, trading supply costs off against an assumed welfare function measuring the aggregate benefit delivered to consumers by virtue of their having their electricity demand met, hour by hour. The incremental loss suffered by consumers when a unit they wished to consume is not supplied is called the "shortage cost", or "non-supply cost".
- In a perfectly competitive market situation, we assume the same thing. In that context, though, the marginal non-supply cost becomes an actual price paid to generators and, at least conceptually, paid by consumers.

There are a great many variables at play here, and a great many possible non-supply modes, implying a great variety of possible non-supply costs, depending on factors like the length and depth of the non-supply events, the notice given, etc. And this has led to some reluctance in adopting non-supply cost as a benchmark.

---

[25] As the name indicates, this was an opportunity cost just like the hydro MWVs we discuss here, but calculated over such a long planning horizon as to be virtually constant over a year.

[26] E.G. Read, J.G. Culy & S.J. Gale: "Operations Research in Energy Planning for a Small Country", *European Journal of Operational Research*, vol. 56, 1992, p. 237-248 describe how DROC was determined, and also how the MoE planned a supposedly optimal power station mix, assuming gas at that price, but then used the (very different) contract price to determine the actual merit order, and operating guidelines, as described here.

[27] And the logistics involved also suggest that immediately available local supplies might sometimes need to be valued at considerably higher levels while extra international supplies are being secured.

Abstracting away from that detail, though, traditional practice has just been to assume that each MW cut, in any period, would cost consumers a specified non-supply cost, in that period, up to a defined MW limit.[28] And then, perhaps, deeper cuts might be costed at a higher non-supply cost. Once simplified to that level, notional "non-supply stations" can be modelled in our optimisation and we will see that they play a very similar role to any real thermal station, in MWV setting. In fact, it was common practice to insert at least one such non-supply station below the highest price thermal station in the merit order, so that some degree of load reduction was called upon before fully loading al thermal stations.

2.3. Contrasting perspectives

We consider two distinct perspectives on the simple situation described above, one "stochastic", and the other "deterministic". Each yields a distinct and contrasting perspective on the situation, and those perspectives shed light on two frequently misunderstood aspects of the relationship between MWVs and observed physical operation, namely:

- It is often said that the MWV must be zero if the reservoir is full, but we will see that that depends heavily on why the reservoir is full, and also on whether we are talking about one more, or one less unit in storage.
- And this relates to another common confusion, between "Volatility", and what we will call "Cyclic Variation", in electricity prices, and MWVs.

We suspect that this latter confusion derives from many analysts with experience of other markets treating electricity price series as if they represented prices for the same product or instrument being traded at different times. And that perception leads to inferences being drawn on the implicit assumption that if there were any significant pattern, such as a systematic difference between day-time and night-time prices, it would quickly be arbitraged away, not only in financial markets, but in physical markets too. Thus, a trader who expected the night-time price of widgets to be higher than the day-time price would simply buy widgets during the day, and store them for night-time use, forcing the night price into line with the day price. And financial markets would follow suit.

So long as electricity remains difficult/costly to store, though, electricity delivered during the day is not the same product as electricity delivered during the night. Nor is electricity delivered in winter interchangeable with electricity delivered in summer, etc. This basic observation is important, because it drives the whole theory and practice of reservoir management and water valuation. Both large and small (i.e. long- and short-term) reservoir storages do cycle, to arbitrage as much price difference away as possible. Significant “deterministic” price cycles remain, though, because differences can not all be physically arbitrage. We will refer to these deterministic components of electricity price series as "cyclic price variation", as distinct from the "volatility" observed around those deterministic patterns.

Accordingly, we discuss "deterministic" and "stochastic" perspectives in separate sections below. Then we discuss a unified framework that brings those two perspectives together, with real world outcomes reflecting a balance between conflicting “deterministic” and “stochastic” effects.

2.3.1. Deterministic perspective

To understand this perspective, we need to think about how reservoirs would be managed in a world where we knew the precise pattern of future inflow and demand levels, and could optimise perfectly. In that case, it is very easy to characterise optimal reservoir management:

- The “marginal value” of water (MWV) would be determined by the value an incremental water unit would deliver, taking the most profitable “opportunity” (still) available, either now, or in future;

[28] The limits might vary from period to period, e.g. in proportion to load, and non-supply costs may vary, too, e.g. between peak and off-peak periods.

- The value delivered by the reservoir would be maximised if MWV was kept as constant as possible across time (i.e. All profitable arbitrage opportunities have been taken); and
- The power station fed by the reservoir should generate at its maximum/minimum possible rate whenever the electricity price is above/below MWV.

We need to tease out what some of those statements imply, though, and we will do so by building up the picture gradually.

Unlimited storage/release capacity case

First, imagine that we start out with a BOY stock level, and wish to end up with EOY storage at the same level.[29] And our reservoir is so large that, even when all seasonal arbitrage opportunities have been taken, the resultant "storage trajectory" never reaches its empty or full levels during the year. In that case:

- There is no impediment to inter-temporal arbitrage, except the limits implied by the LDCs specifying how much generation can be absorbed in each hour each period; so[30]
- Keeping the MWV "as constant as possible across time" means keeping it absolutely constant, in this case;
- As arbitrage continues, taking water from the hours with the greatest excess inflow to be utilised, and storing it to generate in the hours when it is most valuable:
  - The marginal value available by generating more in the periods where it is most valuable will be dropping, while
  - The marginal cost of taking it from the periods where it is least valuable will be rising; so
- Arbitrage will stop when:
  - No more water can be taken from the periods when it is least valuable;[31] and
  - Generation in all the higher valued periods has reached the LDC limit,
  - With our hydro generator setting the price in all of those periods at MWV.

Impact of release capacity limits

The above discussion ignores release/generation capacity limits, but they do play an important role. We have said that the power station fed by the reservoir should generate at its maximum possible rate whenever the electricity price is above MWV. But that means that the exact price level in those periods is completely irrelevant when it comes to setting the MWV, because there is no marginal economic opportunity (still) available to be exploited:

- If our MWV is set to 100, we will be generating at maximum whenever the price exceeds that level, and at minimum whenever price falls below that level.
- There could be thermal generating at a marginal cost of 101, and large enough to cap all prices to that level. That could severely impact the "profit", or "net benefit", calculated for this reservoir, and might well imply that it was uneconomic to build it in the first place. But none of that has any impact on the MWV, or on the optimal generation schedule.

---

[29] This assumption may be relaxed, as discussed below, but the basic concept of a requirement to carry water forward into the next year is critical, because we have assumed away both upper and lower storage limits, and would reach quite bizarre conclusions if we assumed that reservoir storage could just fall indefinitely.

[30] In the real world, and in the stochastic framework discussed later, the time order of decision-making obviously matters a great deal. With perfect foresight, though, we can think of all decisions being made simultaneously, so arbitrage can be thought of as "re-arranging" stock releases between two periods, irrespective of the order in which those periods occur.

[31] An explicit lower release bound is introduced below, but here we have implicitly assumed that negative release values can not occur.

- Or there could be no thermal station generating at any marginal cost below 1000, and prices could frequently rise to that level, or even higher. That could make our reservoir seem highly profitable, and may recover its construction costs many times over. But none of that has any impact on the MWV, or on the optimal generation schedule, either. Our generator is already generating as much as it can, in those high-priced periods, and has no marginal opportunities to exploit in those periods.
- The MWV can not set prices in those periods, either, because the hydro generator is not on the margin.[32]
- Conversely, for the periods where prices are below 100, it does not matter how far they fall below that level. If our reservoir is allowed to dial release right down to zero, it will be totally unaffected by prices in those periods, because it will not generate at all. If it must maintain a minimum release level, it will generate at that level, and whatever positive prices it does receive will contribute to its profitability. But, again, that will have absolutely no impact on MWV, and nor will the MWV have any impact on prices in those periods.

Even in this conventional setting, then, we can explore how MWV might be set in a "bang-bang" situation, with prices alternating between very high and very low levels, and nothing in between. But to reach sensible conclusions we must relax our assumptions slightly.

Impact of targets and efficiency variation

The discussion above is valid, and yields important insights, but it begs the important question: How do we know that 100 is the "right" MWV, particularly if the prices are alternating between the SRMCs of one generator, at say, 10, and of another at, say, 200?

- It definitely will be the right MWV if, when we simulate a storage trajectory, starting from the BOY storage level and generating at maximum in each high-priced period, and at minimum in each low priced period, storage exactly reaches the target at EOY.
- That is possible but, if it is true, it will also be true for any MWV between 10 and 200. If we set the MWV at 150, say, the simulation will give exactly the same result, because hydro will generate at maximum and minimum in exactly the same periods. By the same token, market prices will be unaffected, because hydro would never be on the margin in any of those periods.
- It is much more likely, though, that the simulated EOY storage level will be above or below the target.
- If the simulated EOY storage is below target, and we believe the target should be taken seriously, we are planning to release too much, and so must back off generation in some high-priced periods. If the price is going to be the same in all those periods, irrespective of how much we generate, it does not really matter which periods we choose, but hydro will then be marginal in those periods, and that will set MWV to 200.
- If the simulated EOY storage is above target, and we believe the target should be taken seriously, we are planning to release too little, and so must bring on generation in some low-priced periods. If the price is going to be the same in all those periods, irrespective of how much we generate, it does not really matter which periods we choose, but hydro will then be marginal in those periods, and that will set MWV to 10.

So, MWV remains well defined, and computable, but perhaps it, too, will become volatile? The problem with this picture is that the assumptions are unrealistically restrictive:

- First, market prices, or shadow prices in an optimisation are never going to be absolutely identical hour after hour, over a whole year. Even with perfect foresight, the wide range of supply/demand

[32] The fact that it is generating may put downward pressure on prices, but there will be no impact at all if the assumed thermal generator is large enough.

balances over each daily cycle, and the inability to arbitrage perfectly, means that there will be a range of prices over each daily cycle. So, our hypothetical reservoir operator will have a wide range of annual aggregate generation levels to choose from, each consistent with a particular MWV level corresponding to the marginal value of a marginal economic opportunity available in some hour of the year.[33]

- Second, in reality, generators do not operate with constant efficiency. The hydro manager will typically aim to operate at peak efficiency, and may set MWV accordingly. And it is actually possible to operate close to peak efficiency over a reasonably wide generation range, by committing more and more generating units, each operating at peak efficiency. Once all are loaded, though, it can only increase generation further by pushing all machines past their most efficient operating points, causing marginal efficiency to drop, and the SRMC of its generation to rise. If marginal efficiency is 90% at peak output,[34] then its SRMC will rise from MWV, at peak efficiency, to 11% above MWV, at peak output. So, as MWV is varied, there will actually be a whole range of hours in which hydro is on the margin, at some partial loading level. In this simplistic setting any or all of them could be used to "set" MWV, when adjusted for marginal release efficiency in that period. MWV will set the price in each of those periods too, when adjusted for marginal release efficiency in that period. In other words, they all simultaneously set MWV because there are no more profitable arbitrage opportunities between them (and no opportunities at all for arbitrage with periods in which generation is at its limits, as above).
- Finally, no sensible reservoir manager is going to set a hard EOY storage target, or stick rigidly to a target if it is set. If the simulated trajectory ends above the target, they will not simply spill water just to get down to the target. Instead, they will evaluate the marginal value of holding extra water over for use in the next operational year. They can not just conjure up water, if the simulated EOY storage is too low, either. Typically, they will not even set a "target", but estimate a complete marginal value curve, for water to be held in stock for the next operational year, then trade off the marginal value of using water in this year with that of using it next year.[35]

In summary, then "bang-bang" pricing in any one period, or even in a great many periods, does not imply a similarly volatile MWV. The MWV will generally be set to create a moderate storage trajectory, with releases set to use aggregate available inflows over the year. And that will typically involve setting MWV to match some set of moderately priced periods, somewhere in the planning horizon, while effectively ignoring more extreme prices, even if they are very extreme, and occur in many periods. We will discuss some special cases worth noting, though.

Impact of inflow/release capacity ratio

The MWV for any reservoir does not really depend on its absolute "size", but it will depend on the ratio between its inflows and utilisable release capacity. And that is especially true in the kind of extreme scenario discussed above:

- If a reservoir has low utilisable release capacity, relative to its inflow, then it will only be able to utilise all of its inflow by releasing at maximum in many periods. Obviously, it should release at maximum in all the high-priced periods, but the marginal economic opportunity setting its MWV will be in some lower priced period. Thus, its MWV will be low, irrespective of its storage capacity, and MWV could fall to zero, if the only marginal opportunities left are in periods where

---

33 That degree of precision may seem bizarre, from a real-world perspective. Under these simplistic assumptions, though, a detailed deterministic optimisation will use its assumed perfect foresight to set MWV from a particular (half) hour, some time within the planning horizon.

34 This is not the average efficiency across all units passed through the machine, but the extra power produced by the very last unit of water producing only 90% of what it might have done at peak efficiency.

35 That raises the question of how the MWV curve can be determined for EOY stock. In principle, if the reservoir is as large as we are assuming here, one could extend the perfect foresight horizon indefinitely, and a deterministic model could set MWV from prices in a far distant period. But the absurdity of that proposition really just underlines the fact that we need to move on to consider more realistic cases.

the electricity price is zero. Thus, it will tend towards a base-load operation, rather like that of geothermal.[36]

- If a reservoir has a high utilisable release capacity, relative to its inflow, then it will be able to utilise its inflows most effectively by only releasing in high priced periods. The marginal economic opportunity setting its MWV will be the lowest priced period it generates in, but that will still be relatively high. Thus, its MWV will be high, and it will effectively be acting as an energy-limited peaker, irrespective of its storage capacity.[37]

- If a reservoir has a high minimum release requirement, relative to its inflow, then it will be forced to utilise its inflows across a great many periods, irrespective of price. Once that requirement is met, though, it will need to become selective, and conserve any remining inflow to be released only in high-priced periods. The marginal economic opportunity setting its MWV will be in the lowest priced of those periods, but it could still be very high. Thus, its MWV may also be very high. The implication is that, irrespective of its storage capacity, this hydro must simultaneously operate as a base-load generator, and an energy-limited peaker. In market terms its offer would have a large component priced at zero, in order to see its minimum flow utilised for generation, if possible.[38] But the remainder of its capacity would need to be offered at a very high price, so that it is only called upon to manage extreme peaks. So, this is one of those "knife edge" situations, where the SRMC of one more unit of production may be much higher than the SRMC of one less.[39] It's not that the MWV is unstable or undefined. It's just that it only applies to part of the possible release.

In summary, then, a wide variety of (perfectly optimal) MWV settings and generator operation patterns can be inferred from the application of this basic theory, even in this highly simplified setting.

Impact of storage limits

All of the above discussion ignored storage limits. Such limits are an essential feature of reservoirs, though, and any optimally sized reservoir will have limits that materially constrain its ability to store water from lower priced periods for use in higher priced periods.[40] We will see that these limits imply "deterministic" MWV patterns that might be thought quite counter-intuitive, by those more used to focussing on stochastic aspects of reservoir management. Those insights are important, though, because they do carry forward into the real-world environment, despite the noise introduced by stochasticity. Specifically, they explain the underlying structural "cyclic variation" in expected electricity prices, whereas the stochastic perspective explains the observed "volatility" around that pattern.

The theory discussed here applies equally to a "small" reservoir arbitraging over a daily cycle, or a "large" reservoir arbitraging over an annual cycle. In New Zealand, our reservoirs are not large enough to reliably carry significant storage forward from wet years to dry years, so we will focus on an annual storage/valuation cycle, and discuss shorter term cycles in a later chapter.

---

[36] As above, though, it may still be very profitable, because it is generating at maximum in all of those higher priced periods, too. In fact, it must be making more profit than an otherwise identical generator that is forced to set a higher MWV to conserve a lower water supply, and so only generates in high priced periods.

[37] Just as in the previous case, though, increasing water supply must force the MWV down, so as to allow generation in more periods, but that extra generation is all added profit to the generator.

[38] In theory, a negative offer could be requited, but that would only be appropriate if the only way to release water was through the generator. If the reservoir operator can spill water past the generator, without any environmental penalty, the optimal perfectly competitive offer should logically be at zero, or more likely a little above zero, to cover avoidable costs of maintenance etc.

[39] In the limit, if the outflow requirement exceeds available inflows, the MWV theoretically becomes infinite, but it is irrelevant, because there is no capacity to offer at that price. So, some constraint would have to be relaxed, in order to find a feasible solution.

[40] In theory, a reservoir should have been sized so that the marginal aggregate expected lifetime benefit from being able to store one more unit of water, across a lifetime of storage cycles, matched the marginal cost of making a larger reservoir, at the time of construction. The optimisation may never be exact, but the marginal cost is positive, so we should never expect to see a reservoir whose limits have no impact in terms of limiting inter-temporal arbitrage.

The introduction to this Section (2.3.1) stated three key characteristics of optimal reservoir management in the deterministic perfectly competitive environment we are assuming here. One of those principles was that:

- The value delivered by the reservoir would be maximised if MWV was kept as constant as possible across time (i.e. all profitable arbitrage opportunities have been taken).

Put simply, the impact of storage limits is just that, unlike the theoretical cases discussed so far, the presence of storage limits means that:

- There will be otherwise profitable arbitrage opportunities that can not be taken, because the reservoir is not large enough to shift a large enough volume of water across time; and therefore
- Keeping MWV "as constant as possible across time" can no longer mean keeping MWV absolutely constant across the whole year.

Since the supply/demand balance is relatively tighter in winter (higher load, lower inflow) than it is in summer, the overall direction of desirable arbitrage will be to stock water in summer for winter use. So, storage will tend to rise over summer, and fall over winter, with the cycle repeated every year. To this point, the discussion has assumed that such a cycle can be repeated indefinitely, without violating any storage bounds. But we now need to be more realistic.

- Let's start by ignoring the storage limits, and imagine starting our arbitrage process from a "run of river" position, in which inflows are released immediately, so storage is constant at its beginning /ending level over the year.[41]
- The marginal value of water released in each period will be determined by the prices in that period, and we will refer to that price as the "natural price of water". While it may fluctuate a little from week to week, it will be at its highest some time in winter, and at its lowest some time in summer.[42]
- Now, let's find the highest priced period in which our reservoir could release more to increase generation, and the lowest priced period in which our reservoir could release less, and decrease generation. This gives us an opportunity for arbitrage.
- Then, let's switch some generation from the lower priced period to the higher, bank that gain, and then repeat, looking for the next highest, and next lowest periods between which we can arbitrage, and so on.
- On average, there will be more high-priced periods in winter than in summer, so this arbitrage process will cause hydro generation to fall in summer, and rise in winter. Consequently, our formerly flat storage trajectory will rise in summer, fall in winter, then rise again over the next summer to return to its original level.[43]
- Unlike in the previous case, though, this arbitration process will now reach a point where no more water can be carried forward from summer to winter, because there is not enough storage capacity to hold that much water.
- To be exact, we will reach a point when the difference between the highest and lowest levels of our arbitraged storage trajectory rise to the maximum feasible difference, as determined by the storage range/capacity between our maximum and minimum storage levels.[44]

---

41 The absolute EOY/BOY storage level does not actually matter, because we have not yet introduced any storage limits, and will drop these artificial EOY/BOY targets as soon as we do.

42 In a strict run-of-river situation the natural price would be different every hour, and exhibit a regular daily cycle, but that detail can be ignored for the purposes of this discussion about managing a large reservoir over an annual cycle.

43 In this deterministic world, ignoring the possibility of year-to-year changes in load, infrastructure etc, we can think of storage being in an infinite closed loop cycle.

44 For simplicity, we are assuming these to be constant over the year.

For the remainder of this report, we will call these times at which the maximum and minimum storage levels would be reached in a deterministic optimisation, (whether short or long), "autumn" and "spring". And the periods between them will be called "summer" and "winter". If we think of the available storage capacity as the "maximum stock carryover" level, the above logic tells us that, optimally:

- Releases over the winter must exactly equal winter inflows, plus the maximum stock carried forward from summer; and
- Releases over the summer must exactly equal summer inflows, minus the maximum stock to be carried forward to winter.

The arbitrage process does not stop there, though, because it is still possible to arbitrage between relatively lower priced and relatively higher priced periods in both summer, and winter. It's just that we can longer exploit any arbitrage opportunities between summer and winter. Focussing first on the winter period:

- We can continue arbitraging within that period, but notice that the highest priced period will tend to be in mid-winter, with lower natural prices for water towards the autumn and spring ends of the winter season. So, the general effect will be to shift releases from these two shoulder periods, towards the peak winter season.
- Shifting releases away from the autumn end of winter will tend to raise the storage trajectory over that early winter/autumn period. But we can not actually raise that trajectory above the maximum storage capacity limit. So, when the arbitraged trajectory is sitting right on the limit, we will have to stop reducing releases, up to that point in time, any further.[45]
- Shifting releases away from the spring end of winter will tend to raise the storage trajectory over that period, too. And we do not actually want to carry any more valuable winter water over into summer, so we will stop reducing releases over this late winter/spring any further when the arbitraged trajectory is sitting right on the lower limit. [46]
- Thus, as the arbitraging process continues, we will see the development of extended autumn and spring seasons, within which storage should theoretically be held at its upper or lower limit, thus allowing the trajectory to fall more rapidly, with higher releases, over the remaining (mid-) winter season.
- Eventually, there will be no more profitable arbitrage opportunities to exploit. The autumn and spring seasons will be as long as they can be, releases will be as high as they optimally can be across winter,[47] and we will have the same kind of result as before:
  - The MWV must be constant over the winter period, that is across the remaining period when the storage is falling freely and not at either limit, because there are no more profitable arbitrage opportunities left.

Exactly the same process can be applied in summer, with exactly the same kind of result:

- Arbitrage will act to shift releases away from mid-summer towards the spring and autumn shoulder periods.

---

[45] The simplified model of Read 1979 showed storage trajectories sitting continuously "right on the limit" over these autumn/spring periods. But that model used weekly release variables. In reality, a daily/weekly cycle will still be optimal. Thus, storage might typically be right on the maximum limit at 4am on Monday morning, but then fall away during the day, to return to the full level by 4am the next morning. Or it might only return to absolutely full after the next weekend. MVS will be constant during each of these cyclic excursions from the absolutely full level, only stepping up the next time the reservoir is absolutely full. This nuance makes no real difference to the theory discussed here, but it might be more accurate to say that storage will be held "approximately full", rather than "absolutely full" during this season.

[46] See note above.

[47] Noting that not all electricity price differences can be arbitraged away, because the release/generation capacity limit still applies.

- And that will eventually cause the storage trajectory to:
  - Lie (approximately) at its lower limit for an extended period in spring, then
  - Rise as steeply as it optimally can over summer, and then
  - Lie (approximately) at its upper limit for an extended period again in autumn.
- Again, there will be a constant MWV across summer, but that MWV must clearly be lower than the winter MWV, because it could not have been optimal to fill the reservoir unless we were carrying forward that maximum stock level into a higher priced season.

The arbitrage process described above may not be the most practical way to solve the primal reservoir management problem, or to find the dual prices and MWVs, but the conclusion is intuitively clear:

- Since water is more valuable in winter than it is in summer, MWV must rise as we pass from summer to winter.
- If the value of the available storage range is to be maximised, then storage must be full, carrying forward as much water as possible from the season when it has lower value to the season when it has higher value.
- So MWV must be <u>rising</u> when the reservoir is full.
- In fact, the optimal deterministic policy typically involves:
  - Holding the reservoir exactly full for several weeks,[48]
  - Which implies releasing inflows as they arrive (run-of-river operation), during that autumn period
  - With the MWV following the "natural price of water" curve up to the winter MWV.

Conversely:

- Since water is more valuable in winter than it is in summer, MWV must fall as we pass from summer to winter.
- If the value of the available storage range is to be maximised, then storage must be empty, carrying forward as little water as possible from the season when it has higher value to the season when it has lower value.
- So MWV must be <u>falling</u> when the reservoir is empty.
- In fact, the optimal deterministic policy typically involves:
  - Holding the reservoir exactly empty for several weeks, [49]
  - Which implies releasing inflows as they arrive, (run-of-river operation) during that spring period,
  - With the MWV following the "natural price of water" curve down to the summer MWV.

Also:

- MWV must be constant along each intra-season trajectory arc, i.e. whenever the reservoir is neither empty nor full, and
- This whole cycle must repeat, every year.

Obviously, no real reservoir manager is likely to follow this policy, particularly with respect to holding reservoirs even approximately full or empty for several weeks. The reason they will not do this, though, is because they are uncertain about what inflows will actually arrive, and rightfully wary of risking any

[48] Or, more exactly regularly cycling up to that level, over that period.
[49] Or, more exactly regularly cycling down to that level, over that period.

unnecessary spill, or non-supply. by definition, though, all such uncertainty has been assumed away when setting up this deterministic optimisation framework. So, in this deterministic framework:

- A reservoir will be filled at precisely the point in time when the manager, with perfect foresight, knows the situation is about to turn and, from then on, the problem will be avoiding shortage at the end of winter, not avoiding spill at the end of summer.
- And a reservoir will be emptied at precisely the point in time when the manager, with perfect foresight, knows the situation is about to turn and, from then on, the problem will be avoiding spill at the end of summer, not avoiding shortage at the end of winter.

While spill and shortage are technically possible in that framework, they are very unlikely:

- Spill will only occur if the reservoir manager, who knows the future situation perfectly at the beginning of summer, realises that net summer inflows will be so high that they can not all be utilised even with the maximum release the LDC can absorb, right across summer. In which case the summer MWV will fall to zero.
- Shortage will only occur if the reservoir manager, who knows the future situation perfectly at the beginning of winter, realises that winter inflows will be so low that they will not be able to sustain the minimum release necessary to meet the LDC, right across winter. In which case the winter MWV will rise to the assumed shortage cost.

The previous observations about the relationship between MWV and hourly prices still hold, within each season:

- Each seasonal MWV will be set by any one of the potentially many hours in which the hydro station is on the margin, in that season, and it will set the price in those hours, too.
- But MWV will be unaffected by how high, or how low, prices might be in the hours when hydro is generating at its maximum or minimum rates.
- In addition, though, the prices expected to pertain in winter will have no impact on setting the summer MWV, and vice versa. They are irrelevant, because the manager has no unexploited marginal economic opportunity to shift any more units from summer to winter (and no reason to shift any in the other direction), no matter what the price/value differential may be.
- Further, the prices in periods during which the reservoir is being kept full will not affect either seasonal price. The opportunity costing paradigm still applies to those periods, but, if the reservoir is being kept full, the best opportunity available to the manager is to release it within the week it arrives, because otherwise it will spill. Thus, its MWV will be set by some hour in that week, with hydro being marginal in that hour. [50]
- Likewise, the prices in periods during which the reservoir is being kept empty will not affect either seasonal price. The opportunity costing paradigm still applies to those periods, too, but, if the reservoir is being kept empty, the best opportunity available to the manager is to release it within the week it arrives, because prices will be falling as summer comes on. Thus, its MWV will again be set by some hour in that week, with hydro being marginal in that hour.

Impact of reservoir size

Many discussions talk about "large" and "small" reservoirs, but this can be misleading, because the issue is not about absolute size. If we double the absolute size of a hydro scheme, in all respects, the mathematical relationships we are discussing here just scale up proportionally. The scaled-up system may be twice as

[50] Strictly, if the reservoir really was to be absolutely full all week, this could come down to the only option being to release water in the hour it arrives, in which case each hour would have its own MWV, and that MWV would equal the price for that hour, unless hydro is also capacity-constrained, as above. As already noted, though, we expect storage to cycle away from the limits and back, on a daily /weekly basis, with MVS only updated at the end oi each such mini-cycle.

important, and the arbitraging it allows may have twice as much impact on market prices. But, at least under perfectly competitive assumptions (which ignore the latter effect), its optimal MWV/storage pattern will be just the same, i.e. with the quantities doubled, but MWVs the same. So, physically large and small reservoirs may have very similar operating characteristics.

The real issue here is the ratio between the capacity of the reservoir and the aggregate supply/demand imbalance it is ideally trying to arbitrage, with the latter being measured by the theoretical maximum storage capacity that would be required to arbitrage all price differences away:[51]

- If that storage capacity/supply imbalance ratio is greater than 1, the reservoir has enough capacity to handle the full range of variation presented to it, and MWV will be constant over the full annual cycle, as in the "no storage limits" discussion above.[52]
- As soon as the storage capacity/supply imbalance falls below 1, there will be distinct summer and winter prices, and intermediate seasons (autumn and spring) during which the reservoir should be held at one of its bounds, as in the "storage limits" discussion above.
- But notice that holding storage constant implies releasing inflows as they arrive, which is essentially how we define "run-of-river" operation. So, even a "large" reservoir will go through phases during which it operates in something like a run-of-river mode.

Under these deterministic assumptions, there can be no variation above/below expected flows but, before discussing the stochastic perspective, it is worth understanding the situation the reservoir would be in, during those times when its storage sat right at one of its limits. During those phases, its SRMC would be set by what we will call its "Marginal Cost of Release" (MCR), and this sits on a knife-edge:

- While the reservoir is being held at its maximum, any water that can be retained in storage should be assigned an +MWV that may be high, and is certainly rising, and which we will refer to as the Marginal Value of Storage (MVS).
  - So, the MCR of using that stored water would be set by that relatively high MVS, thus forcing our generator into the role of a high-priced energy-limited peaker.
  - But any extra water, above the inflow level that would just maintains storage at its limit, must be assigned an MCR of zero, because otherwise it would spill, thus forcing our generator to act simultaneously, in a base-load role.
- While the reservoir is being held at its minimum, any water that might be taken out of storage should be assigned an MVS that may be low, and is certainly falling.
  - So, the MCR of using that water, up to the received inflow level which just maintains storage at its limit, would be set by that relatively low forward-looking MVS.
  - But any extra release, above that level, would have to be assigned an MCR reflecting a higher MWV, possibly as high as the assumed shortage cost, because it is not physically feasible to release water that is not there.[53]

As the storage capacity/supply imbalance ratio falls further, the storage situation will get tighter and tighter, thus:

- Increasing differences between summer and winter MWVs; and
- Forcing the reservoir to operate in a run-of-river mode for longer periods; until

---

51 This can be determined by plotting the storage trajectory implied by following the arbitrage process though to its natural conclusion, as discussed in the "storage limits" section above.

52 Although still limited by its release/generation capacity.

53 The reservoir under discussion here need not be particularly vital to national supply security. But, even if it is, the SRMC/MWV does not necessarily rise as high as the shortage cost, because often it is quite possible to meet the load in such periods by using the hydro generation level implied by run-of-river operation, possibly topped up by thermal.

- Eventually storage will become so constrained that there are actually only very short periods over which storage is not at one bound, or the other; and
- In the limit, the reservoir would have no storage, and operate in run-of-river mode all year, in which case its MWV would just follow the constantly changing "natural price of water" profile.

2.3.2. Stochastic perspective

The whole discussion in the previous section may seem fantastical, and applicable only to a make-believe world with perfect foresight. We will see, though, that all of the intuitions developed in that section do, in fact, carry over to the real world, albeit in modified form. And they do significantly impact the price and MWV patterns we should expect to see in real world situations.

First, though, we will consider an equally simplistic and hypothetical stochastic scenario. Imagine a hydro reservoir operating in a world with significant inflow uncertainty, but no night/day or summer/winter cycles, in either inflow or demand, and no inter-temporal correlations, either. For simplicity, let's focus on the intra-year reservoir management problem, as above, and assume that the non-hydro resources available to meet load are identical in all periods but that each day, an inflow level and LDC are revealed, and we must decide how much of the load should be met by hydro.[54]

Under this regime, storage will rise or fall depending on how much inflow we receive, and how much we release. That process will not be entirely random because we will try to control storage to lie within the allowable storage range of the reservoir. Despite our best efforts, though, the reservoir could be filled. and possibly spill, dure to ongoing high inflows and/or low net demand. Or, it could be emptied because of ongoing low inflows and/or high net demand.

So, unlike the deterministic case discussed above, observing a reservoir to be full would not imply that MWV was rising as it carried maximal water forward from summer to winter, as part of a deliberate arbitrage plan, and with no risk of spill. Instead, it would be low to zero, due to a significant probability (and perhaps even certainty) of spill. To be exact, we will try to increase hydro generation to a level high enough to avoid immediate spill. If that attempt fails, we must spill, and set the MWV to zero. If it succeeds, MWV would be set to the SRMC of whatever alternative source would otherwise have supplied the last demand unit met by hydro.

Similarly, reservoirs would only be emptied because of ongoing low inflows and/or high net demand. So, observing a reservoir to be empty would not imply that MWV was falling as it carried minimal water forward from winter to summer, as part of a deliberate arbitrage plan, and with no risk of shortage. Instead, MWV could be high to very high, due to a significant probability (and perhaps certainty) of shortage. To be exact, we will physically have to reduce hydro generation to a level low enough to avoid running out. If load can not be met, we must set the MWV to the "non-supply cost", as discussed in Section 2.2.5. Otherwise, MWV would be set to the SRMC of whatever alternative source meets the first demand unit not met by hydro.[55]

At intermediate storage levels we have wider choices, though, and must decide how to manage storage optimally. It seems clear that a manager facing the same inflow level and load pattern will want to release more if the reservoir is nearly full than if it is nearly empty. And the chosen release level will determine the "marginal value of release", which optimally must equal MWV. So, there must be an MWV curve across the storage range, and that curve must be constant across all periods.

Unlike the deterministic case, though, we cannot hold MWV constant along each storage trajectory, because those trajectories are driven by events outside of our control. Thus, we may set a non-zero MWV, only to find that a series of high inflows forces us to spill. So, in retrospect, we will discover that the actual MWV for that scenario was zero, no matter what we may have thought at the beginning. Or, starting from the same storage level, and assigning the same initial MWV, we may find that a series of low inflows forces

54 The remainder traditionally being met by thermal, as detailed in Section 2.2.4.
55 As above, the SRMC/MWV does not necessarily rise as high as the shortage cost.

us into shortage. So, in retrospect, we will discover that the actual MWV for that scenario was very high, no matter what we may have thought at the beginning.

Setting an expected MWV

So, how do we set MWV? The above discussion helps us understand that, in fact, we can not "set" MWV at all, because we do not know which scenario will eventuate. But what we could do, in principle, is to simulate optimal storage management for each possible inflow/load scenario, and determine the eventual MWV for each one of them. Then, since we are ignoring risk aversion in this perfectly competitive discussion, we can base decisions on the Expected MWV (EMWV), as the probability weighted average MWV over all those scenarios.

That is essentially how the optimisation algorithms in most of the stochastic reservoir optimisation models cited above work, with two important caveats:

- We can not really simulate all possible scenarios, so these models either perform that simulation implicitly (as in the "Constructive" DDP method used by SPECTRA), or they simulate a limited sample of "important" scenarios (as in STAGE, SDDP, or DOASA).
- But, most importantly, we can not consistently say that we have "simulated optimal storage management" of any scenario, unless we have used the same model to determine each future release decision (and MWV), as we are using to determine this first decision (and EMWV), for the current storage level.

Simulating realistic future storage management[56]

This last "non-anticipativity restriction" is what makes stochastic reservoir optimisation difficult. In plain English, it's just a "no-peeking" rule, that says we can assume the hypothetical future managers, at any point in time, will have learned from what they will have observed in experiencing the particular scenario being simulated, up to that point in time, but will not know how that scenario will develop from that point in time forward.[57] This is easy to state, but it would take a very large amount of computing power to fully optimise decisions for each of the vast number of possible decision-making situations encountered across all of those scenarios. So, every optimisation model compromises on it in some way.[58]

Those compromises can cause different optimisation models to produce significantly different EMWVs for the same situation. In fact, many reservoir managers might only use simplified computational models, or just rely on heuristic rules, or experienced judgment. Fortunately, Section 5.3 will argue that the MWV estimated for any particular storage level makes much less difference than might be thought to long-run equilibrium outcomes. But what matters is that, throughout the decision-making process, managers will be thinking about how the future might pan out: What sequences of inflow/demand patterns might occur, how they might manage each of those sequences, and what impact that should have on the release decision they must make today, and its corresponding EMWV.

---

[56] The theory discussed here is developed in the author's thesis, and summarised in E.G. Read: "Stochastic Long Term Scheduling Models for a Power System", *ORSNZ Proceedings*, 1979, p.41-52.

[57] If we think of each scenario being defined as a complete beginning to end path through the ever-branching tree of future possibilities, then there will be a great many scenarios that all share the particular set of initial branches experienced up to any point in time. So, we can express the rule by saying that the future manager will not know which of the scenarios sharing the initial branches observed up to any point in time, will actually eventuate.

[58] Broadly, models like SDDP and DOASA assume a branching "probabilistic decision tree" of scenario possibilities, and assume that future managers will estimate EMWV by averaging MWVs over the possible futures for the time remaining, given the branch they find themselves on, due to the inflow/load levels that would have been experienced to that time, in all scenarios which start with that series of branches. So, the approximation in those models is partly that the number of future branches remaining, as possible continuations of the scenario represented by (the set of branches leading to) the branch they are assumed to be on, rapidly dwindles as the model simulates further into the future. As a result, the quality of modelled "optimal decision-making" falls off as we look further into the future.

Whether the "scenario simulation" is explicit or implicit, it should not be too surprising to discover that the correct MWV for the initial point is just the "expected marginal value" of water (EMWV) across all those scenarios.[59] And the relevant MWV for each scenario is determined by the value an incremental water unit would deliver, taking the first profitable marginal "opportunity" encountered in that scenario.

To be consistent, the future MWVs calculated in each scenario are actually, themselves, conditional EMWVs, and we will therefore refer to them as CMWVs. So, the CMWV for a future period in a simulated scenario is the EMVW a hypothetical future manager should calculate after observing how that scenario has played out, and the storage level reached, given the sequence of decisions they should optimally have made, up to that point in time.

It is not the whole sequence of hypothetical future CMWVs that ultimately set the current EMVW, though. What matters is that, after some periods in which there are not expected to be any marginal opportunities because (in the simulated scenario) hydro is expected to be releasing at either minimum or maximum rates, hydro will eventually be "on the margin".[60] That is, its CMWV is deemed to equal the marginal cost of some other generation source (typically thermal under the status quo), or load reduction option, or spill. So, if one more unit of water were to be held in storage now, and that scenario were to occur, that extra unit would be released in that period, in that scenario, and can be valued at the corresponding marginal cost.[61]

In particular, the CMWV for a simulated scenario will be set to zero, if the reservoir is expected to spill before hydro is "on the margin", in that scenario, and this will be reflected back into the EMWV calculated for the initial storage level from which that simulation started.[62] As storage approaches its upper limit, there will be more and more scenarios in which spill would be projected to occur, so the calculated EMWV will fall, and may reach zero. But that depends on the generation capacity of the hydro system, and any release limits. If there is enough storage capacity, and generation/release flexibility, to manage through any likely inflow scenario, the probability of actual spill may be very low, and the EMWV may still be set at a positive level, right up to the maximum storage limit. The CMWV for some scenarios may be quite low, though, reflecting the fact that, in order to avoid spill, the manager might have to dump water fairly aggressively, thus seeing hydro generation used to displace some fairly low value alternatives.

---

59 A risk averse manager might put more weight on high MWVs, from scenarios in which water ended up reducing the risk of shortage, to calculate a "Risk Adjusted EMWV" that would be higher than the risk neutral EMWV.

60 That is, generating between its upper and lower bounds.

61 Note that tine order matters in this stochastic environment, in a way that it did not under deterministic assumptions, because we only slowly discover what scenario is actually occurring. So, this first marginal economic opportunity will actually be shared by all the scenarios branching out from the initial scenario observed to that point in time, and they will all have the same CMWV, at that time. So, the point is that an extra water unit stored now will actually be released in that period, under all of those scenarios, irrespective of how any of them might develop subsequently, and what the "true" MWV for each of them might ultimately turn out to be, with perfect hindsight.
If an hourly simulation were to be performed, it might indicate hydro being on the margin, briefly, before very long. Realistic models use much longer simulation steps, and identify major crossovers with thermal/non-supply options as discussed in Section 2.3.3 below. So, the "first" identified marginal economic opportunity may be many weeks into the future.
We are not saying that the optimisation or EMWV setting processes can ignore the rest of the planning horizon, beyond this "first" marginal opportunity, though. The CMWV in a future period can not be determined without at least implicitly determining optimal management of all scenarios through to the end of the planning horizon. In fact, all "Dynamic Programming" methods explicitly work back through the scenario tree, from the end of the planning horizon. Thus, the "first" marginal economic opportunity can only be identified after the remainder of the planning horizon has been fully optimised.
When we simulate a relatively small sub-set of the (theoretically infinitely branching) scenario tree, using a weekly time step, a small variation in the scenario settings may determine whether the first period in which hydro is simulated as being on the margin occurs early in the simulation horizon, or later. That implies some degree of randomness. Computational models need to control for the error introduced in this way, but it does not really affect the principles under discussion here.

62 Hydro generation need not be on the margin during a spill incident, and the marginal value of electricity generation nay still be positive. But the marginal value of reservoir release is zero, nonetheless, once release exceeds the level that can be utilised to generate electricity.

EMWV behaviour

While this stochastic situation is computationally more complex, the same basic principles of arbitrage apply. While the arbitrageur can not know the future, value is still maximised if EMWV is kept as constant as possible across time, in each scenario, even though uncertainty makes that much harder to do.

All of the previous observations about the impact of release/generation capacity limits still apply, too. In each period of each scenario, the electricity price will equal CMWV when hydro generation is on the margin, and higher/lower than CMWV whenever generation is at maximum minimum, And the electricity prices in those periods still do not set CMVW, and are not set by CMVW, no matter how far they may be above or below it. Consequently, they do not affect the EMWV calculated for the initial storage level, either. But there would be no seasonal pattern to EMWV, under our "purely stochastic" assumptions. There are other possible limits that may have an impact, though.

Impact of flow rate limits and correlations[63]

So far, we have ignored the possibility of limits on release rates, and assumed that release is able to start/ramp/stop instantly and costlessly. Consequently, we have not considered the possibility that "precautionary release" might be desirable.[64] If no-one cares how high the spill rate gets, there is no reason to start spilling before it is physically forced upon us, even in a highly uncertain situation, with a high probability of future spill. Thus, EMWV can be positive, right up to the maximum storage level, implying that we should hold storage right at the level, if there is even a tiny probability that inflows will fall far enough to avoid spill, so that any incremental stored water might eventually be used to avoid some positive marginal cost. Thus, we could see hydro in the "knife edge" situation described previously, basing offers for generation from stored water at a positive (and possibly high) EMWV, while offering generation from extra inflows at (close to) zero.

The situation changes, though, if our reservoir manager faces explicit or implicit restrictions, or penalties, related to the rate at which water can be released, or the rate at which release changes.[65] In that case, it may be deemed prudent to start releasing at a moderate rate, some time before the reservoir is actually full, in order to reduce the likelihood that the required spill rate may become unacceptably high, as the situation develops.

Whether spill is penalised explicitly or implicitly, in a market situation, or some upper release rate limit is assigned a shadow price because it is projected to become binding in an optimisation, the marginal value of release above that level it will actually be negative in these situations. And, if the probability weighted cost of incurring a penalty due to spill outweighs the probability weighted benefit of later being able to avoid thermal generation or demand response, EMWV could theoretically become negative, too. But controlled spill should commence as soon as EMWV falls to zero, and that spill should generally be sufficient to hold EMWV at zero, until the risk of uncontrolled spill falls to an acceptable level.

In this context, inflow correlations become important, too. A manager who believes inflows will persist at high levels will obviously be inclined to ramp up release rates sooner than one who believes inflows will soon fall back to low levels. So, the EMWV in these situations really depends on forecasts, as well as storage level.[66] Technically, we should really be estimating EMWV for specified storage/forecast pairs, but a reasonable approximation can be to add some measure of forecast flow to the observed storage level,

---

[63] Here we only discuss constraints on release levels. Storage levels, and storage rates of change can also be subject to constraints other than simple upper and lower bounds, though. Manapouri, for example, is subject to a complex set of environmental limits, which are actually difficult to analyse accurately, because they are not convex. The general effect of such limits, though, is to restrict storage flexibility and force operators to behave as if they were approaching physical bounds, even when storage is at more moderate levels. And that must increase costs, reduce effective arbitrage capabilities, and increase price differences between periods.

[64] Theoretically, these limits apply in the deterministic case, too, but they are much less likely to have a significant impact in a situation where plans can be made to work around a perfectly foreseen situation, months in advance.

[65] Implicit penalties could include societal disapproval, and threats to future water rights.

[66] Actually, that is always true, but we will ignore that complication, except in discussing particularly critical situations.

when setting EMWV. In that case, the "effective storage level" could actually lie above the reservoir's maximum storage, some time before actual storage reaches that level, and spill should be expected to commence when the EMWV for the "effective storage level" falls to zero.

Correlation and concerns about river flow rates etc will be important at the opposite end of the storage range, too. If the reservoir is required to maintain a minimum release rate, it may become necessary to conserve water in order to meet that requirement some time before the storage is actually empty. Again, the EMWV in these situations really depends on forecasts, and a low inflow forecast could mean that the "effective storage" required to meet some acceptable probability of compliance already lies below the reservoir's minimum level, implying that release will have to be reduced to the minimum, well before actual storage reaches that level.

Major reservoirs can also face a different issue, in that the nation may rely upon their holding enough water to be able to maintain some minimum generation rate through a critical supply/demand balance period, typically in late winter, for New Zealand. This can be managed in much the same way, except that the implied "penalty" may be as high as the" shortage cost", or otherwise at least high enough to match the SRMC of whatever alternative generation source might need to be called upon in periods when the hydro generation level implied by run-of-river operation is insufficient to meet the load. But national security of supply arrangements may also kick in, forcing a generation pattern, and implied EMWV, that may not entirely accord with the manager's own estimates.

EMWV curves and surfaces.

It may not be obvious from the above discussions, but there is a fundamental, difference between the two perspectives discussed:

- A deterministic optimisation always starts from a specific storage level, and traces a specific storage trajectory across the planning horizon, to a specific ending lvel.
- As a result, it only calculates MWVs for points along that trajectory because, under deterministic assumptions, management has no reason to ever stray from that trajectory, or speculate about what MWVs might be at other storage levels.
- In a stochastic world, though, trajectories may wander far from anything "planned", and next year's starting storage may be very different from this.
- So, a stochastic model must determine CMWVs for all storage levels along any possible future trajectory, and will thus end up calculating CMVW over a wide range of storage.
- Accordingly, such models often report an "EMWV curve" across the entire feasible storage range.[67]

Obviously, we expect the EMWV curve to fall from a relatively high (possibly very high) level at the minimum storage level, to a relatively low (possibly zero) level at the maximum storage level. Since this section assumed away any seasonal patterns, the EMWV curve will be the same for all periods. More generally, though, it will be higher in winter and lower in summer, and we can talk about an EMWV "surface" consisting of all EMWV curves for the year, and draw contours on that surface, as discussed in Section 2.3.3.

Impact of reservoir size

Finally, before bringing together the insights derived from deterministic and stochastic perspectives, we should discuss the impact of reservoir size, under these simplistic assumptions. Other things being equal, a larger reservoir will obviously have lower probabilities of spilling, or running empty, than a smaller one.

[67] If we start from a particular storage level, the storage range of possible trajectories will gradually expand over time. But many models are set up to support simulations over many hydrological years, cycling around by setting the starting storage level in each simulation year to the ending storage in the previous simulation year. So, EMWV curves are ideally computed to cover ethe entire feasible storage range, in all periods, in which case they can also support simulations that were not included in the EMWV setting process.

But the appropriate measure of size is different, in this stochastic case, than it was in the deterministic discussion above. The issue here is not inter-seasonal carryover capacity, but the reservoir's ability to absorb the random fluctuations it is subject to. So:

- A reservoir might be considered "large" or "small" depending on the ratio of its feasible storage range, to the standard deviation of the probability distribution for its "inflow minus utilisable release requirements".[68]
- Generation capacity may limit a reservoir's utilisable release capacity, and hence limit a reservoir's ability to absorb inflow variations, and force it to spill more often, even if it has significant storage capacity.
- So, a "small" reservoir, in this context is one that has both limited storage capacity, relative to inflow/demand volatility, and limited utilisable release capacity, relative to inflows.[69]
- Offers from a small reservoir may become nearly vertical, since all it can do is to pass inflows through, as they arrive, in run-of-river fashion.
- In the limit, if utilisable release capacity is low enough, relative to inflows, the system may have to be constantly base-loaded, and still spill.

2.3.3. Unified framework

EMWV Estimation

To bring the stochastic and deterministic perspectives together, we just need to add seasonality to the assumptions underlying the stochastic section above. In this, much more realistic, world:

- EMWV is still determined by the average value an incremental water unit would deliver, taking the first marginal economic opportunity available, over many scenarios.
- Value is still maximised if CMWV is kept as constant as possible across time, in all scenarios (although that becomes even harder, now that storage limits constrain inter-seasonal arbitrage options).
- CMWV is still not set by electricity prices in periods when they are expected to be so high that release is maximised, or so low that release is minimised, no matter how high or low they may be.
- CMWV can still not be set by prices for periods beyond the next time storage is projected to be full/empty, under that scenario.

There are some changes, though, because there are now distinct times at which the storage is likely to be full/empty, under any scenario, and that implies a distinct seasonal pattern to the projected scenario CMWVs, and hence EMWV. To understand how this emerges, first imagine a world in which our uncertainty about the future reduced to us having to make a decision now, but knowing that we would discover exactly which scenario would occur, at the end of the week, then:

- Each of the scenario simulations discussed in the previous section would actually become deterministic, after the first week. [70]

---

68 The latter probability distribution should really account for correlations, so there may be some circularity, because a larger reservoir might be affected by correlations over a longer period, but this measure, and discussion, are only approximate.

69 Both mean and volatility may be significant factors here. But the ability of the load to absorb higher generation levels is also an issue, with the contribution of extra capacity falling off as we reach the point where the load in more and more hours can be met without it. So, it would not be easy to define an exact ratio measure.

70 In some models, the simulated trajectories are for historically observed inflow sequences, and the above process would reveal the best way in which each might be managed, if we had perfect foresight.

- So, each simulated trajectory would display all of the characteristics of the deterministic trajectories discussed earlier. That is, our hypothetical clairvoyant future manager would try to find a trajectory in which:
  - The full storage range was utilised, by pushing storage right up to the maximum in autumn, and right down to the minimum in spring; and
  - The storage level would generally sit at its maximum/minimum for an extended period in autumn/spring.[71]
- So, the calculated CMWV would be constant over summer, and again constant, but typically at some higher level, over winter.[72]
- And the calculated CMWV must generally be rising in autumn, and falling in spring, just as in the deterministic discussion above.

If we were to then average over all these trajectories, we would find a pattern quite a lot like the deterministic case, with storage trajectories tightly clustered around storage bounds in both autumn and spring, and CMWVs constant over the mid-summer and mid-winter periods, albeit differing by scenario, and possibly scattered over a fairly wide range.

Assuming clairvoyant future management is unrealistically optimistic, though.[73] As a next step, we could imagine adding a more realistic decision-making process in which observations of conditions in the first week only give us some idea as to which cluster of scenarios the future is likely to lie in. So, our hypothetical future manager must make a release decision on that basis, implement it, and discover which scenario will actually occur at the end of the second week. Following through this process, across the year, would yield a more and more realistic pattern of future decisions, and a more and more refined estimate of the initial EMWV. But that does not change the fundamental logic outlined above. Starting from mid-summer:

- There is clearly still a fundamental driver emerging from the stochastic discussion to avoid spill. So, if the simulated summer storage trajectory seems to be approaching the upper storage bound "too fast", our hypothetical future manager will be seeing the possibility of spill forecast in more and more scenarios.[74]
- Of itself, that will lower EMWV estimates, but they may still be quite high if there is still a significant probability that incremental water will not be spilled, but instead carried forward to reduce the need for high-cost thermal, and possibly shortage, before the end of winter.
- Thus, the simulation could project a knife-edge situation, in which excess water is eventually spilled (with an MCR of zero), even though the EMWV (or MVS) for water that can actually be stored remains positive, as in the deterministic discussion above.
- Or, penalties and limits on higher release rates could force EMWV all the way down to zero, not so much because storage is too high, but because it is deemed to be approaching the upper limit too fast, in some scenarios, thus indicating the desirability of "precautionary spill", as in the stochastic discussion above.
- Eventually, though, EMWV must start to rise, as it becomes more and more likely that any extra water can be carried though this critical period without spilling, and eventually used to reduce

---

[71] That is, unless the hypothetical future manager finds that the seasonal imbalance is so mild, in some scenarios, that all profitable arbitrage opportunities can be taken, without storage ever reaching either bound.

[72] Except in scenarios for which the hypothetical future manager can completely arbitrage MWV differences away, as above.

[73] Because the arbitrage process will be able to bring summer and winter MWVs much closer together in scenarios with a milder inter-seasonal imbalance, and might even manage to equalise them, in some cases.

[74] That is, in all the possible future projections of the scenario assumed to have been observed to that date.

thermal costs, or shortage, over the coming winter period, or possibly even carried through to the next winter.

- Once storage levels start falling, though, EMWV can be expected to stay at a relatively high winter level. How high that level eventually rises will be driven by developing expectations with respect to the late-winter/spring period.
- As noted earlier, electricity prices should not necessarily be expected to reach shortage cost levels, even if the reservoir is projected to be empty. In some scenarios, it may be possible to meet load requirements just by utilising inflows arriving across that period. (And of course, the generic reservoir we are discussing here may not be critical to the national supply system, anyway.)
- Still, if the reservoir is large enough to make a significant difference to national supply security, and simulated winter storage trajectory seems to be approaching the lower storage bound "too fast", our hypothetical future manager will be seeing the possibility of high prices forecast in more and more scenarios.
- That will raise EMWV estimates, but the effect will be moderated if there is also a significant probability that incremental water will be carried through into the summer, to be valued at lower summer prices, and possibly even spilled by the end of summer.
- Eventually, this effect will dominate, and EMWV must start to fall again to summer levels, as storage levels start rising, and the focus of concern eventually switches back to avoiding spill in the late-summer/autumn period.

To summarise, then, seasonal cycles still make it optimal to utilise as much of the storage range as possible to carry water over from summer to winter, and that still implies running storage as close to the full and empty levels, as is deemed to be acceptable. But likely penalties on extreme spill rates, and shortage probabilities, may force management to avoid holding storage too close to limits, particularly in major reservoirs of national significance. Thus:

- Rather than EMWV (or more exactly MCR) rising smoothly and predictably over autumn, it may be quite volatile over that time. EMWV ((or more exactly the underlying MVS) must eventually rise, though, because the whole reason for having a large reservoir is to be able to carry significant volumes of water forward from summer, where it has relatively low value, to winter, where it has relatively high value; and
- Rather than EMWV (or more exactly MCR) falling smoothly and predictably over spring, it may be quite volatile over that time, too. EMWV (or more exactly the underlying MVS) must eventually fall, though, to the relatively lower summer value, reflecting the possibility that spill (low-price dumping) may occur, before the value rises again for the next winter. .

Guidelines

All of the discussion above assumes that there is some kind of generation capacity available to meet the load not met by hydro, and thermal stations have traditionally been the main capacity in that role. We have assumed that each thermal station has a well-defined SRMC, which is constant over the year. Section 2.2.4 has argued that that is actually unlikely but, since the impact of removing thermal capacity from the New Zealand power system is our central focus, we need to understand the role traditionally played by thermal in setting MWVs.

It has been suggested that thermal SRMCs provide the "gold standard" against which hydro is opportunity costed. We suggest that the ultimate gold standard is actually non-supply cost, with the thermal SRMCs acting more like exchange rate pegs, or perhaps more accurately "resistance levels" in a modern market context.

That said, if the SRMC of a thermal station is believed to be fixed, across a year, we can draw the corresponding contour on the EMWV surface, and call that the "guideline" for that thermal station. In New Zealand, these guidelines have always peaked in autumn. It has commonly been stated that a thermal station should be "base-loaded" when storage falls to its guideline level, but that is not quite true. What really happens (under our simplistic assumptions) is that that thermal station swaps merit order positions with hydro, when storage falls to the guideline level. So:

- When storage is just above a particular guideline level, EMWV will be just below the SRMC of that thermal station, indicating that it is preferable, on an expected system cost basis, to meet loads with hydro generation, where possible, rather than with generation from that thermal station.
- When storage is just below a particular guideline level, EMWV will be just above the SRMC of that thermal station, indicating that it is preferable, on an expected system cost basis, to meet loads with generation from that thermal station, where possible, rather than with hydro generation.

The critical question, though, is what should happen when storage sits exactly on the guideline. In principle, hydro EMWV and thermal SRMC are identical, and the two could be used interchangeably, to meet loads across a broad band corresponding to their combined MW capacity.[75] If EMWV, and generation/release decisions, were reviewed continuously, then:

- Optimal operation would require that the thermal generation level be continuously varied so as to just keep hydro storage tracking along the guideline, thus keeping EMWV "as constant as possible", at the corresponding SRMC.
- If inflows are falling and/or loads rising, though, the downward pressure on storage levels will eventually become so great that the thermal station can no longer produce enough to stop storage falling[76], and it will then fall freely, until it reaches the next guideline (possibly a higher priced thermal, but possibly a non-supply level).[77]
- Similarly, if inflows are rising and/or loads falling, the upward pressure on storage levels will become so great, that the thermal station can no longer stop storage rising, even if it minimises its production[78]: So storage will then rise freely, until it reaches the next guideline (possibly a lower priced thermal, but possibly spill). [79]

Typically, though, EMWV for a major reservoir might only be reviewed weekly, in which case the focus of decision-making should really be on whether storage is expected to be above, below, or right on the guideline at the end of the week. And the answer to that question hinges on the thermal loading level chosen. At the beginning of the week:

- There will be a range of storage levels so high that end-of-week storage is expected to remain above the guideline, even if this thermal station remains above hydro in the merit order stack, thus making its minimum possible contribution to supporting storage.
- There will be a range of storage levels so low that end-of-week storage is expected to remain below the guideline, even if this thermal station remains below hydro in the merit order stack, thus making its maximum possible contribution to supporting storage.
- In between, though, there will also be a range of storage levels, from which end-of-week storage should optimally be expected to lie exactly on the guideline, with this thermal station sharing the same merit order position as hydro, and generating at a level intermediate between its minimum

---

75 In practice, both generation sources are likely to have a band of output levels across which efficiency, and hence marginal cost, are varying. So, there would actually be a range of EMWV levels, and hence storage levels over which partial loading of both sources would be optimal.

76 Noting that the thermal station's useful output is limited by the LDC, as well as its own MW capacity

77 If inflows are expected to be falling and/or loads rising, at this time of year, that imbalance will be reflected in a falling guideline. So, storage will not fall below the guideline unless loads are rising, or inflows falling, faster than expected.

78 Noting that the thermal station must still contribute to meeting the LDC, once hydro generation is at its maximum

79 Conversely, if inflows are expected to be rising and/or loads falling, at this time of year, that imbalance will be reflected in a rising guideline. So, storage will not rise above the guideline unless loads are falling, or inflows rising, faster than expected.

and maximum possible levels, on average across the week. (i.e. both hydro and this station are "on the margin"

- So, as a result, the beginning of week EMWV curve actually has a flat section, across which EMWV is constant at the corresponding SRMC, and we can think of there being two copies of the same "guideline", at the top and bottom of that flat "guideline band".

Control

This last observation was central to the workings of SPECTRA, where the upper and lower guidelines were called "augmented guidelines". The workings of that algorithm need not concern us here, but the width of the "flats" described above is important, because it determines the degree of control each thermal station can actually exercise over storage trajectories. This is not just determined by the MW capacity of the station but by the LDC, because the station will have no influence at all in periods when it can not generate because loads are too low. Thus:

- If the thermal station has such a low SRMC that it really is "base-loaded", i.e. operating 24/7 like geothermal in the New Zealand system, then swapping hydro above/below that level in the merit order will make very little difference to aggregate weekly hydro generation requirements. Thus, the corresponding "flat" band may be very narrow, implying that this station will have very little impact on any storage trajectory, real or simulated.
- If the thermal station has such a high SRMC that it seldom operates, like a diesel fuelled OCGT peaker in the New Zealand system, then swapping hydro above/below that level in the merit order might also be thought to make relatively little difference to aggregate weekly hydro generation requirements. And that would be true if it were swapping with a small hydro, because the generation difference only occurs over a few hours, for both of them.
- The issue here is not just the MW capacity of the thermal station, though, but that of hydro. If we assume minimum hydro releases to be base loaded, the remaining controllable hydro may be only operating at its absolute maximum for a few hours, but for much longer at some level above its minimum. So, having 1 MW of peaker capacity take over the task of meeting 1MW of load from hydro means swapping the last MW of hydro capacity, scheduled only for the extreme peak, with the first MW, scheduled for as long as hydro has the energy to support it. Thus, the difference can be quite substantial for a major hydro generator, implying that peakers might actually exert significant (per unit) control over storage trajectories.
- Moderately priced thermal stations may have a significant impact, too, depending on the shape of the LDC across the hours when they might operate.

In all cases, though, we may think of the thermal station as trying to achieve the basic goal of reservoir optimisation, which is to keep EMWV as constant as possible over time. And that goal corresponds to managing storage right along a guideline, for as long as possible.

It will not be possible to fully achieve that goal, beyond the point when the guideline reaches a storage limit, for all the reasons discussed in Section 2.3.1 on the deterministic perspective. But it will generally also prove impossible because the range of flow uncertainty (over a week, say) is likely to exceed the range of control available from each individual thermal station (that is, the aggregate difference between generation above/below hydro in the merit order, over a week, as discussed above). The greater the degree of control achievable by any generator, though, the wider the range of storage levels over which the EMWV curve will be relatively flat, at around its SRMC level.[80]

[80] If the "flat" corresponding to some generator was actually wider than the inflow uncertainty range, that generator could completely control storage trajectories over some range of storage levels, with EMWV exactly equal to that generator's SRMC over that range. That range would not expand indefinitely, though, because it would eventually be cut off by the upper/lower storage bounds.

So, returning to the explicit or implicit simulation process by which EMWV may be estimated, we can see that simulated trajectories are likely to soon be "captured" by a thermal guideline band, and then may be controlled to follow along that guideline for some time. So, the CMWV for that scenario will be the SRMC for that thermal station. So, if the goal was simply to determine EMWV from an established set of guidelines, we could just set MWV for each trajectory at the SRMC for the first guideline reached. But the guidelines themselves are contours on an EMWV surface, each point pf which can only be determined by averaging MWVs from multiple future simulations, all starting from that time and storage level. But those simulations could all be regarded as legitimate extensions of the primary simulation we are discussing here so, in principle, we should expect to get exactly the same EMWV if we carried our primary simulation on into the future, along each of those possible paths, with appropriate probability weights. The art and science of stochastic reservoir optimisation revolves, to a large extent, around determining how many scenarios, and of what type, should be sampled in order to give a good EMWV estimate.[81]

EMWV and non-supply costs

Traditionally non-supply costs have been set at high levels, and they have had a very significant impact on EMWV calculations, and hence on storage levels, both simulated and real. The difference between non-supply costs and normal price/marginal cost levels is generally agreed to be much greater than that between normal price levels and the zero marginal value assigned to spill. So, the optimal trade-off between spill and non-supply is very much weighted in favour of holding reservoir storage high enough to avoid non-supply costs, even if that does imply seemingly "wasteful" spill levels. In fact, optimal reservoir management is dominated by the requirement to manage the probability of non-supply events to an acceptable level.

If a non-supply cost is set at, say, 20 times the normal electricity price (or thermal SRMC) level, then non-supply occurring in one trajectory simulation will have as much weight in setting MWV as normal prices occurring in 20 other simulations. Particularly in models performing simulations on a limited set of historical inflow sequences, the calculated EMWV can be seen as roughly measuring of how many of those sequences are expected to fall to the empty level, or forbidden zone boundary (see below).[82] As a result, the assumed non-supply cost levels have a very significant impact on EMWV assessment. e should address three common mis-conceptions with respect to the role of non-supply costs, though:

- First, it is sometimes suggested that these "artificial costs" serve to inflate electricity prices and/or MWVs. It is quite the opposite, in an optimisation context, though. If the demand level were to be presented to such a model as a "hard" requirement, to be met at all costs, the internally calculated MWV could effectively become infinite.[83] At the very least, MWVs would have to be set higher, so as to push storage levels higher, in an attempt to meet the requirement. Thus, modelling "virtual stations" whose SRMCs are set by non-supply costs, is actually a way of moderating price projections and MWV estimates.[84] Accordingly, care must be taken to include enough non-supply

81 "Capture" is assumed to happen very quickly for the implicit simulations involved in SPECTRA's "constructive" DDP methodology, which works by explicitly tracing guidelines. "Sampling" DDP methods, like SDDP, do not explicitly calculate guidelines, so their simulated trajectories may pass through several guideline levels before being captured, and assigned an MWV. But this is not a difference in principle.

82 Thus, the difference between one simulated sequence "just touching" vs "just missing" that boundary can make an appreciable difference to the EMWV calculated for a point. The discrete nature of this effect is somewhat artificial, though, and results from the approximation of only simulating a limited set of trajectories. If enough simulations were performed, the EMWV should rise smoothy, as we look at lower storage levels, and higher non-supply probabilities.

83 In which case an optimisation model would report that the problem was "infeasible" or, in other words, it was impossible to find a solution that met the requirements.

84 In a perfectly competitive market, the issue is slightly different in that consumers are assumed to be facing these non-supply costs as prices, and so might reduce demand if those prices are higher than their "willingness to pay". Thus, the situation can not be manipulated as freely as it might in a central planning context, where the planner is just setting non-supply costs at a level assumed to represent what consumers "should" be willing to pay.

cost levels to ensure that enough flexibility is allowed to cut demand back far enough to make it possible to keep simulated storage within the allowable range, and so limit prices to finite levels.

- Second, it is often stated that the MWV associated with any (actual of simulated) storage trajectory should automatically be set to the non-supply cost, as soon as that trajectory reaches the empty level for a reservoir. As noted earlier, though, electricity prices (and hence MWVs) should not be set to that level in scenarios where it is possible to meet load requirements just by utilising inflows arriving, through that period.[85]
- Third, on the other hand, EMWV can actually rise to a non-supply cost level before storage reaches empty. If it looks likely that a period is coming up in which a second or third level of non-supply might be required to get load demand down to a level that can be meet with hydro generating at just the rate of incoming flows in enough forward simulations, the probability-weighted EMWV across all scenarios might actually reach the first non-supply cost level, implying that demand should be reduced to the corresponding non-supply level, so as to avoid the likelihood of deeper, and even more costly, non-supply a little later.

Buffer zones

If we treat virtual non-supply stations like thermal stations, we can also plot the contours of the EMWV surface corresponding to each non-supply cost level, and call this a "non-supply guideline". Then, when storage fell to that guideline level, the manager's optimal response would be to trigger the corresponding level of non-supply. In a centrally optimised system that would happen in a centralised way, but in a perfectly competitive market, the manager would raise offer prices to a level expected to induce that level of demand reduction.[86] Such a non-supply guideline may lie right on the empty level, for many periods. When it lies above the empty level, though, it may be referred to as "bottom guideline", or "basic rule curve", with the zone below it being referred to as a "forbidden zone".

Alternatively, such zones may be computed using a more physically oriented methodology, and perhaps imposed by an outside party, in which case they may be treated as defining the lower limit on acceptable storage. Managers may then compute EMWV curves for higher storage limits assuming that MWV rises to some stipulated non-supply cost when their own simulation shows storage reaching that level, irrespective of whether they agree that the forbidden zone has been computed correctly, or not. So, the top of the buffer zone starts to play a similar role to the lower storage limit discussed above. Management will try to avoid storage falling into that zone at all often, but might not want to carry much storage above that level either. So, EMWV may fall off quite quickly above the zone.

It is also common practice to introduce an upper "buffer zone", or "flood control zone" within which the focus is on spill management. And that zone could be quite wide, if respecting maximum flow rates, or avoiding flow rate penalties is a major concern. In that case, the bottom of the buffer zone starts to play a similar role to the upper storage limit discussed above. The EMWV curve for water stored below that level remains largely unaffected by possible events in the buffer zone, and could be assigned a high MVS value if it can be reliably carried over into the winter.

Using buffer zones does not change the basic principles we have discussed, though. They are really just a way of codifying the results from actual or hypothetical simulations of how storage trajectories should optimally be managed through critical periods, and they are imposed (sometimes by external parties) because they are simple to understand, and give more direct physical control than a full economic optimisation would. But notice that, whether the need to avoid holding storage too close to both upper and lower storage limits is expressed in the form of buffer zones, or inferred from a fully optimal EMWV

[85] So far, our discussion has assumed only one hydro reservoir in the system, or perhaps just one aggregate reservoir representing the national system storage in an optimisation. In which case, it is reasonable to think that this one reservoir could be critical to the national supply system. Most reservoirs will not be critical to the national supply system, though, and prices will not rise to non-supply cost levels unless a larger reservoir, or perhaps many reservoirs, are empty. Section 3.2 discusses water valuation in multi-reservoir systems.

[86] Noting, as above, that this discussion strictly applies only to a single national reservoir.

surface, the effect is to reduce the storage capacity that can be relied upon for inter-seasonal arbitrage. And that effect can only increase inter-seasonal price differences, on average.

In other words, the optimistic assumptions in a deterministic optimisation (with no buffer zones) lead us to under-estimate the volatility of EMWV generally, and particularly over the autumn and spring periods. But they also lead us to under-estimate the extent to which EMWV must rise over autumn, and fall over spring. Thus, the stochastic perspective not only explains the observed EMWV volatility, but actually also increases the importance of cyclic EMWV variation, as driven by the deterministic perspective.

Impact of reservoir size

Finally, if a reservoir is large enough, in the senses discussed above, the deterministic perspective will dominate, over most of the storage range, with the stochastic perspective coming into play primarily when storage is close to its limits. Conversely, if a reservoir is small enough, the stochastic perspective might be thought to dominate over its entire storage range, because such a reservoir has no appreciable inter-seasonal arbitrage capacity. And it is true that, in the limit, a reservoir nay be so small that it can only adopt a run-of-river operating regime, to which the deterministic arbitrage perspective does not apply at all.

That is not necessarily the case for all physically small reservoirs, though:

- Notice that the deterministic perspective discussion suggested that a reservoir should be considered "small" if its capacity is small, relative to the aggregate supply/demand balance variation it is ideally trying to arbitrage; but
- A reservoir with no appreciable inter-seasonal arbitrage capacity should not be trying to arbitrage over that planning horizon. Instead, it should be focussed on arbitraging over a daily cycle.

But the stochastic perspective discussion suggested that a reservoir should be considered "small" if the ratio of its feasible storage range, to the standard deviation of the probability distribution for its "inflow minus utilisable release requirement" was small. But that assessment should obviously apply to volatility within the planning horizon over which the reservoir is trying to arbitrage, and many reservoirs may actually have very predictable inflows over a day. We will discuss the management of smaller hydro reservoirs in Chapter 3, but note that the deterministic perspective may describe the operation of many of those reservoirs quite well.

3. Complications

3.1. Introduction

All of the discussion in the previous section related to a situation involving just one hydro reservoir, and focussed mainly on a situation where that one reservoir was large enough to play a vital role in maintaining a reliable national power supply, over a year. The real power system contains many more components, though, and many more storages, operating over both short- and long-term planning horizons. The treatment of non-storage renewables will be addressed in Section 4.2, when we discuss the impact of impending changes to the market plant mix. Here, though, we extend our discussion to consider the complexities introduced when considering multiple hydro, and/or thermal, storages, and then when storages are linked into downstream river chains. We also extend the previous theoretical discussion to include the important topic of plant expansion, and the relation between long run and short run marginal costs.

3.2. Multiple storages and regions

3.2.1. Multiple independent long-term hydro storages

Optimisation/arbitrage

In this theoretical context, "independent" does not refers to a reservoir being independently managed, but physically independent, in the sense that it is in a different catchment. "Long-term storage reservoirs" are reservoirs with a relatively high ratio of storage capacity to inflow/release capacity, irrespective of physical scale. And all of those reservoirs, jointly, will be trying to manage the national seasonal demand/supply imbalance.

The major implication, in a centralised optimisation context, is that the national optimisation problem now becomes multi-dimensional. That complicates the mathematics considerably but, in principle:

- There is now an EMWV to be estimated for each reservoir.
- For each reservoir, the EMWV curve discussed earlier for a single period now becomes an EMWV surface, in each period, still depending on that reservoir's storage, but now on every other reservoir's storage, too.[87]
- Some optimisation models will just determine a vector of EMWVs for a starting storage vector specified for a particular period. Others will estimate a complete set of EMWV surfaces for all reservoirs and periods, but it is not hard to understand why that can become very computationally demanding.
- Conceptually, the EMWV surface concept is important, though, and it is helpful to think about what it implies in the relatively simple two reservoir case which, in New Zealand, might be thought of as representing aggregate South Island and North Island storage.
- So, the North Island MWV surface depends on both North Island and South Island storage levels, because the North Island storage operator knows that the value of water stored in the North Island depends significantly on how much is stored in the South Island. And likewise for the South Island MWV surface.[88]

---

[87] In other words, a four reservoir problem has 4 EMWV surfaces (one for each reservoir) for reach week of the year, with each one of those surfaces being defined over a 4-dimensional grid of storage levels.

[88] SPECTRA only produced these two 2-dimensional surfaces, one for each aggregate island reservoir.

The arbitrage concept still applies, and we should think of optimisation being equivalent to arbitrage continuing until all profitable opportunities have been exploited. But arbitrage can now occur between reservoirs/locations, as well as over time.

- So optimal operation implies that trading would continue in each period, until the North Island and South Island MWVs were as close as possible. But they will not often reach equality, because:
    - There will be transmission losses between the regions, and
    - An inter-regional transmission limit may trap some potential generation in one or other island, or
    - One or other reservoir may reach its minimum, or maximum release/generation capacity limit.
- The effect of this arbitrage will be to balance storage levels between the reservoirs. If the North Island MWV is higher than the South Island MWV (after loss adjustment) then North Island storage must be too low, relative to South Island storage. So, optimal arbitrage will then favour South Island hydro generation over North Island hydro generation, thus tending to reduce South Island storage while allowing North Island storage to rise, until they eventually reach equilibrium, with EMWVs equal (+/- marginal losses).[89]

In summary, then:

- Week-to-week volatility of inflow levels is too great to be completely compensated by the thermal generation system. But optimal thermal operation will still tend to moderate storage fluctuations, and inter-temporal MWV differences, by attempting to manage storage along constant MWV guidelines, bringing it up toward a long-term sustainable seasonal trajectory if it falls too low, and letting it fall towards that level if it rises too high.
- Likewise, week-to-week volatility in the inflow balance will be too great to be completely compensated by the transmission system. But optimal transfer operation will still tend to moderate fluctuations in storage balance, and inter-regional MWV differences, by attempting to manage storage along MWV "balance guidelines", bringing storages back towards balance, whenever they deviate.

EMWV setting

The forward simulation concept still applies too, except that now we are conceptually simulating the joint evolution of linked storage trajectories across all reservoirs, and:

- EMWV should again be estimated by averaging the MWV's from all those simulations but, importantly, the opportunity cost setting MWV in one reservoir, under some scenario, might now be the opportunity to use a unit of water stored in that reservoir to back off generation from another reservoir, in some later period. So that opportunity might be valued at the CMWV for that other reservoir, which might, in turn, be determined by the opportunity to use the water saved to back off generation from the original reservoir, in some even later period, or perhaps some different reservoir in a multi-reservoir context. And so on.
- Eventually, though, an MWV will be set by the opportunity cost of spill, or non-supply, or thermal generation, in some location and period. So, a South Island EMWV might be determined by the opportunity cost of South Island spill, or of avoiding South Island non-supply. But it might equally be determined by the opportunity cost of North Island spill, or avoiding North Island non-supply, or indirectly displacing North Island thermal, at some later date.

[89] Simplistically, they actually swap in the merit order, just like hydro and thermal in Section 2.3.3 above. SPECTRA applied these concepts to determine " transmission guidelines" and "augmented transmission guidelines" which help to define the corresponding South Island and North Island EMWV surfaces.

The same optimal solution, and MWV surfaces, should arise in a perfectly competitive market situation, where we assume that each participant has the same information, analytical capabilities, and (lack of) transaction costs as the central optimiser would have had. Significant EMWV differences are still likely, though, due to transmission losses and limits, and the fact that each storage will have its own particular characteristics and limitations. Comparing each reservoir with a notional "national aggregate reservoir":

- A reservoir that has proportionately lower release capacity than the national aggregate reservoir, (which will be dominated by major long-term storages) will find itself having to release over a wider range of periods. So, its marginal opportunities must consistently be in lower priced periods, and that must make its EMW lower. The higher EMWVs enjoyed by other participants are simply irrelevant, because it has no spare capacity to take any more of the higher priced opportunities from which they are setting their EMWVs.[90]
- Conversely, a reservoir that has proportionately higher release capacity than the national aggregate reservoir, should optimally reserve its energy for generation only in higher-priced periods, and should set a higher EMWV than what might be calculated for the national aggregate reservoir.
- A reservoir that has inflows strongly correlated to those of the national aggregate reservoir, but with proportionately lower storage capacity, will find itself having to release over more periods in summer, and able to release over fewer periods in winter. So, its summer EMWV must be lower, and its winter EMWV higher, than what might be calculated for the national aggregate reservoir.
- But a reservoir with proportionately more storage than the national aggregate reservoir will find itself able to release over fewer periods in summer, and thus able to release over more periods in winter. So, its summer EMWV must be higher, and its winter EMWV lower, than what might be calculated for the national aggregate reservoir.
- In the limit, it could find that its MWV would actually be constant, over both summer and winter, for some scenarios, because neither upper nor lower storage bounds were threatened over the annual cycle. And the same could conceivably be true for a reservoir with inflows not strongly correlated to those of the national aggregate reservoir, but well correlated with load. EMWV is still likely to exhibit some cyclic seasonal variation, though, because storage limits will have impact on other scenarios

Further complications can result from limits on minimum or maximum flow rates, flow/storage rates of change, or complex environmental constraints such as those at Manapouri. But those complications only add to the general conclusion above that, despite the tendency of optimisation, or market arbitrage, to steer the system towards balanced storage management, with EMWVs differing as little as possible between reservoirs, there can still be quite a range of EMWVs across the national storage portfolio, at any particular time.

3.2.2. Independent short-term hydro storage

So far, we have mainly focused on major reservoirs that typically (threaten to) reach their upper and lower limits, around the same time each year, thus operating in an annual cycle in which MWV rises at one time of the year (prior to winter for most New Zealand reservoirs) and falls at another (after winter for most New Zealand reservoirs). But the general theory discussed above applies to all "independent" storages, of all sizes, over all planning horizons. The discussion at the end of the previous section implies that reservoir characteristics actually vary widely, thus creating a whole range of cases between "annual" and "daily" reservoirs, but we will shift to focus on storages at the other end of the spectrum.

[90] If it compares itself with other generators having the same inflow, it may consider that its profitability has been reduced due to low storage capacity. But we have already noted that its lower MWV actually indicates that it has enough inflow to generate in more periods than other generators, and so will be making a higher operating profit than other generators with the same storage/release capacity.

A really small reservoir may actually have two cycles in each day, one for each peak, while a station with no storage at all becomes a so-called “run-of-river” station, for which the MWV for each trading period is effectively determined by the market price in that trading period. But we will focus solely on the classic case of a hydro station supplied from a reservoir with insufficient storage to carry any significant stock through from summer to winter, but enough to carry a worthwhile volume through from night to day.

Looking first from a deterministic perspective, a reservoir like this should optimally exhibit exactly the same kind of behaviour as an annual reservoir, but over a daily cycle. So, it may typically reach its upper storage limit, or at least threaten to reach its upper limit, before the morning peak, then reach its lower limit, or at least threaten to reach its lower limit, after the evening peak. And this means that its MWV must also cycle daily, rising while the reservoir is being held as close to full as prudence allows before the morning peak, then falling while the reservoir is (close to) empty, after the evening peak.[91]

From a stochastic perspective, intra-day inflow volatility may be an issue for some of these reservoirs, but many should be able to forecast flows quite well over that horizon. On the other hand, while most such reservoirs are probably too small to worry about demand stochasticity, as such, price volatility may be a major issue. If prices are thought likely to spike to very high levels, at any time of day, the best use of reservoir storage capacity could be to hold stocks high so as to be ready to surge generation, and maintain output over the time when it appears to be most valuable. Section 4.2 suggests that that situation may well become more common but, traditionally, this kind of price spike has only been likely at fairly predictable peak times, when optimally managed short-term reservoirs should already be generating at maximum capacity. Thus, the deterministic strategy already discussed is also effectively building up storage to be available in such circumstances.

In other words, the deterministic perspective is probably most relevant at this point in time, and that perspective suggest a daily cycle with high daytime MWV until dusk, at which time storage may be held empty during a period of run-of-river operation until MWV falls to a low night-time MWV level. Then storage should build up overnight until the reservoir is full some time before dawn, and then held full for a period of run-of-river operation while MWV rises back to its high daytime level:

- In a "small" reservoir, that is one whose storage range is quite small relative to its desired daily cycle, night and day MWVs could be quite different:
  - If its utilisable release capacity is large relative to its storage plus inflow on that day, daytime MWV could be very high, and it might spend a long period in run-of-river mode, with storage at maximum, waiting to be released only when it is most valuable. After which it might spend another long period in run-of-river mode, with storage at minimum, waiting to back generation off and fill again, only when generation is least valuable.
  - If its utilisable release capacity is small relative to its storage plus inflow on that day, it might spend much of the day releasing at maximum.
  - If its minimum flow rate is high, it will also need to release at that rate throughout the night, while it slowly builds up storage.
- In a relatively "large" reservoir, that is, one whose storage range is large relative to its desired daily cycle, MWV could actually be constant all day:
  - If its utilisable release capacity is large relative to its inflow on that day, MWV will be at a high daytime level, and water reserved for release only when it is most valuable.
  - If its utilisable release capacity is small relative to its inflow on that day, MWV will be at a low night-time level, and water released at maximum rate for most of the day and night.

---

[91] There is no economic reason why a small reservoir should not be held empty for some time, because the nation is not depending on its output. At both limits it would still be running, but effectively in run-of-river mode.

- In the limit, MWV will only fall to zero if there is more water available than can be utilised, across the full day.[92]

3.2.3. Energy-limited thermal

Chapter 2 develops the conventional wisdom on MWV determination for hydro systems, on the assumption that SRMC of thermal generation is itself well-defined, and constant over the year. Unfortunately, that is seldom, if ever the case, in New Zealand, because local fuel markets are not liquid, and much of our thermal capacity is actually "energy-limited", and not in a very different situation from hydro.

We have already argued that thermal SRMCs really only provide a clear (national cost) benchmark, if whatever local stock they may hold can be freely replenished by buying in as much as might be required from a liquid market. We suggested that could be the case for oil, and imported coal, with the caveat that the implied "benchmarks" would be volatile international prices. Unfortunately, the local supply of other fuels seems to be subject to inter-temporal constraints that we may broadly described as "energy-limiting", to a greater or lesser degree.

The logistics of managing coal mining, delivery, and stockpiling locally were always quite complex. Traditionally there never was a market in which coal could be bought locally, in the kinds of quantities required by power stations, or with the flexibility required to meet demands that surged dramatically in dry years. Underground mining at Huntly could not be ramped up at all quickly, and flexibility was provided by "pre-stripping" some small open-cast mines in the Waikato, and later by stockpiling coal at Huntly. Either way, there was a limited stock available that had to be managed just like a hydro reservoir, with the added complication that the coal deteriorated quite rapidly, if exposed to air. In fact, a coal stockpiling module was added to SPECTRA, as an extra reservoir. Thus, the SRMC structure of local coal was anything but the simple constant assumed by most hydro reservoir modellers.

Nor does it seem likely that large quantities of gas are freely available for purchase at transparent prices in the New Zealand gas market. And we certainly expect some gas-fired peaking capacity to be dependent on limited line-pack storage in pipelines, meaning that the flexible supply component must be reserved for limited use in the highest priced periods of each day, by assigning a high opportunity cost. And, while we noted that, in principle, incremental gas might be drawn from a large field at a fairly stable "depletion related opportunity cost" the real situation w involves contracts.

Many studies assume that SRMC can be set by a fuel's per unit contract price. But that would really only be true for a fixed price, variable volume contract. Once the contract specifies a minimum or maximum take, though, the situation is really one of managing a scarce resource within those limits, and scarcity rents are involved. In fact, a strict "Take-or-Pay" contract really amounts to a lump sum having been paid for the right to draw supplies from an agreed stock, over some period, and subject to a variety of flow rate limits.[93]

Mathematically, this puts the gas-fired power station manager in very much the same situation as a hydro system manager, except that the annual volume is known in advance. As the season advances, the manager may have to offer generation at a progressively lower price in order to dispose of gas that would otherwise be unused at the end of the contract year, or the manager may have to offer generation at a progressively higher price, in order to avoid running out of gas before the end of the contract year. Either way, the per unit contract price is actually "sunk", and theoretically irrelevant, because the gas should really be opportunity costed in a similar way to hydro.

---

[92] Or perhaps even a later day, if this reservoir's storage limits are not expected to be reached within the day.

[93] The contracts we are referring to might more accurately be labelled "Pay for an Option to Take ". But such contracts actually reflect a physical and economic reality. The real cost of gas production largely arises from capital investments whose costs actually are sunk, and operating costs that are not really avoidable in the short run, either. From a national cost-benefit perspective, there is very little direct "SRMC cost" in actually extracting another unit of gas from the field, and the true cost of using it for power generation, today, is (once again) the lost "opportunity cost" of not using it for some other purpose (including generation on another day) later (and probably much later) in the field's life.

More flexible gas contract forms imply intermediate results, and coal production may have become more flexible with time. But there is little point discussing the details now that thermal generation and coal mining are being eliminated from New Zealand. The main point is that the traditional reliance on benchmarking hydro opportunity costs against constant-SRMC thermals has really been driven more by conceptual and computational convenience than national economic realities.

- With the exception of oil, and to some extent coal, New Zealand thermal generator SRMCs have never really provided the clear benchmark assumed in most optimisation models, and many industry discussions.
- Even for those tradeable fuels, the real SRMC benchmark would have been a fluctuating international price, thus adding another dimension of uncertainty to be factored into the EMWV calculation for hydro reservoirs.
- So, the status quo might really be characterised as finding an equilibrium balance between multiple storages, some of which are hydro and some thermal, as discussed in the sections on balancing multiple short- and long-term hydro storages above.
- In fact, unless New Zealand decides to draw on international markets for bio-diesel, or green hydrogen (in some form) to power thermal generators, the nearest analogy to internationally tradable fuels, in future, will probably be internationally tradable end products, as discussed in Section 0 below.

3.3. Downstream river chains

The theory discussed above applies to all "independent" storages, large and small. But that discussion assumed that each short- or long-term storage reservoir directly feeds a single generating station, which has zero flow delay, and a well-defined efficiency curve. So, when voluntary release from the reservoir is "on the margin", we could think of the MCR as being just the MVS of that reservoir, divided by the marginal generation station efficiency. Conversely, we have assumed that the MVS, for each hydrology scenario, will be set so as to match the electricity price expected in some future period when this station is projected to be on the margin, times the marginal generation station efficiency for the loading level expected in that period.

But the New Zealand power system has several hydro chains, typically fed from long-term storage reservoir, at or near the "top" of the chain, releasing water to allow generation in a series of downstream generating stations, each with its own small storage, or "head pond”.[94]

If they were not in a chain, each of these downstream storages would be operated in a simple daily cycle, as discussed above, but the fact that they receive water from a station upstream, and/or release water to another station downstream does complicate the situation. While it is common to talk, for example, about the MWV of a river chain, as if it was a single power station supplied directly from the top reservoir, this is not really a well-defined concept. Each storage in the chain has its own MWV, fluctuating in accordance with its own optimal operating cycle. And, while river chain optimisation seeks to keep all stations operating on synchronised cycles, this is often not possible, due to capacity imbalances and flow delays.

So, let’s build up the picture step by step:

- First, if we imagined that there were no tributary flows, flow limits, or delay times, we could imagine all of these stations would have been designed with exactly the same throughput capacity, so that water released from the top reservoir passed through them all, at the time of release, implying the same flow rate though each, probably subject to the same flow limits.
- In that case:

[94] The Waitaki system actually has two long-term reservoirs at the top of the chain, while the Waikato system has another chain of generating stations, with limited storage, above the main long-term reservoir. But these complications can be ignored for the purposes of our discussion.

- We can assume that none of these stations has any appreciable storage of its own.
- They would effectively respond like the single generation station assumed above, all shifting in synch up their generator efficiency curves, as release increased from minimum to maximum, beyond which point they would all spill the same amount of excess water, if any more were to be released from the top reservoir.
- The aggregate MCR for the whole system would either be zero, or the MVS for the top reservoir, divided by the combined marginal efficiency of the downstream stations.
- And, while each storage in the chain would theoretically have its own MVS, the MCR for each station would just be the aggregate system MCR, times the proportion of the potential energy of water stored in the top reservoir which is actually converted into electrical energy when passing the water through that station.[95]

If the top reservoir had ample spare storage range, it could adjust its discretionary release to optimally manage flow through all these downstream stations to match system prices. Consider the case, though, where the top reservoir has a high MVS, and should optimally only be releasing at its minimum rate, presumably in a constant stream. Under our simplistic assumptions, with no tributary flows, downstream stations may face the same flow limits as the top reservoir, and be forced to pass this release through in a constant stream, too. As soon as they do have some flexibility to vary flow rates, though, and some storage capacity to use, perhaps some extra flows from a downstream tributary:

- All storages in the chain should optimally be cycling, so as to manage whatever inflow is available to generate at the times of day when it has most value:
  - Perhaps the head pond in which the tributary arrives (let's say n) will be able to manage that inflow completely, so that it, and all stations downstream from it, can peak generation in a synchronised way, without reaching any storage limit, in which case, that head pond will have a single MVS for the day.
  - Or perhaps that head pond will find that its ability to do this is constrained by upper and lower storage limits, so it will have distinct night and day MVS levels, as discussed for independent reservoirs above.
  - Either way, though, the MVS of water stored in this downstream storage need not have any relationship to the top reservoir's MVS (which in this case is too high to justify any discretionary release, all day), or the top reservoir's MCR (which in this case is zero all day, because the top reservoir is releasing a constant stream, at the minimum rate).
  - So, there should optimally be generation offered from the chain, at a price that really has no connection to the MVS or MCR of water in the top reservoir.
- At this stage it becomes important to understand that the MCR for a storage in a river chain is NOT determined by its MVS. What matters is the DIFFERENCE between the MVS in the reservoir from which water is released, and that in the downstream reservoir where it will arrive:

---

[95] MCR for the top reservoir is either zero, or that reservoir's MVS, which is really defined as the value per unit of potential energy stored in that reservoir. The potential energy in each reservoir is proportional to its gross "head", that is the elevation of that reservoir above sea level. But the measure we are assuming here is basically the difference between that level and the elevation at which water will be released from the last downstream station in the chain, adjusted for efficiency losses in the generation process.
Roughly speaking, if the elevation between two successive storages in the chain is 10% of the total elevation difference down the chain, and release water flows directly into a downstream head pond (rather than losing head by running through any intervening river reaches) then the difference between the MVS of water stored in the upstream station's head pond and the same water stored in the next station's head pond must also be 10% of the total MVS calculated for the top reservoir. So, in this simplistic case, we can talk about the MWV of the chain, because all these MWVs are proportional to the MWV for the top reservoir.

  - This would not matter, if we could assume all MVS values in a river chain to be proportional to that in the top reservoir, as in the very simplest case, above.
  - But it does matter once the MVS in any storage is being set to manage incoming flows to allow generation in more valuable periods, and especially if storage bounds are limiting a storage's ability to arbitrage, and so equalise its MVS values over time.
- Suppose, for example, that the head pond for station *n* had a low night-time MVS and a high daytime MVS, because its storage limits are binding at some point during the daily cycle, but the reservoir below it in the chain *(n+1)* had a constant MVS. Then the difference between the MVS for *n* and that for *n+1*, which is what determines MCR for the intervening station *(n)*, must be higher during the day than it is at night.
- Similarly, if the station above *n* in the chain *(n-1)* has a constant MVS, then the difference between the MVS for *n-1* and that for *n*, which is what determines MCR for the intervening station *(n-1)*, must be lower during the day than it is at night.
- Following through that example, though, we now have one MCR potentially lower at night than in the day, and another MCR potentially higher at night than in the day, and that seems unlikely to be optimal. In fact, we need to consider arbitrage possibilities to improve the river chain release schedule as much as possible:
  - There are now arbitrage opportunities that should optimally be exploited by adjusting release in the upstream and downstream reservoirs so as to bring their MCRs into alignment, as far as possible.
  - So, it must be optimal for the upstream and downstream reservoirs to cycle, too, if they can.
  - But then new discrepancies will appear between MCRs for the stations served by those reservoirs and those further upstream, and further downstream.
  - In the end, upstream/downstream arbitrage must optimally continue until we find an optimal operational pattern; and
  - That pattern must involve all reservoirs in the chain, both upstream and downstream from the tributary inflow point, trying to cycle in synch, and thus use their combined storage capacity to manage the tributary flow in such a way as to maximise generation in the most valuable periods, as far as possible.
- Both flow and storage bounds will limit the extent to which this synchronised ideal can actually be achieved, though:
  - To start with, in this case, all stations upstream of the tributary inflow point are only processing minimum releases from the top reservoir, an they can not vary their release at all if they face the same minimum flow limit. As a result, their storage level will remain constant all day, irrespective of the economic value that might have been achieved by a more dynamic storage management strategy.
  - The stations downstream from the tributary inflow point have more flow to work with, but they are also likely to have higher minimum flow levels, reflecting the additional natural flow of the tributary.
  - The stations in a chain will often have quite different storage ranges, too, with some being more able to store flows received to generate in more valuable periods than others.
  - Upper release/generation limits also become important. Thus, a station with a relatively low generation limit may find itself generating at maximum right through the day, implying an MVS set from the most attractive marginal opportunity actually available to it, some time in the night.
  - That MVS, though will actually include the value the released water has when received downstream. And, if the next station in the chain has a relatively high release capacity it

might best be able to utilise that extra unit of water by releasing it some time during the day, and stations further downstream might be able to do so, too.

- If so, the MVS in the more capacity constrained upstream head pond will reflect the relatively high downstream daytime MVS of that head pond, plus a relatively low increment reflecting a night-time release down to that downstream head pond.
- The MCR for the constrained upstream station is driven by MVS differences though, and that difference will be low all day. [96]
- But that low MCR does not drive any more daytime release, because this station is already generating at maximum through the day.

The key conclusion from this discussion is just that MVS/MCR imbalances between different stations and times may well remain, even after all available arbitrage opportunities have been taken. Tributary flows will vary, though, and the optimal operational/MCR pattern will vary from day to day. On some days it will doubtless be possible to arbitrage all inconsistencies away, and support a moderate offer curve, representing balanced and efficient river chain operation across a wide range of aggregate output levels. But that is not likely to be possible, once conditions become more extreme. And we still need to consider the interaction between management of tributary flows and management of top reservoir releases:

- If its MVS is low enough, the top reservoir manager will consider releasing more than the minimum requirement, and must consider how that release will interact with the management of both tributary and minimum flows, downstream.
- If upstream flows were to arrive in a steady stream, downstream reservoirs might need to cycle harder to manage the increased flow over the day, but it should often be possible to schedule discretionary releases from the top reservoir, so as to make the downstream management problem easier, rather than harder.[97]
- A choice must be made, though, once the combined flows force some head pond storage to run up from its minimum to its maximum, over some period when it is generating at its maximum level.
  - It may be optimal to keep releases coming from the top reservoir, but spill water past this station, in order to allow more generation from stations further down the chain. In which case the MCR for this station, and any other stations spilling downstream, must be zero.[98]
  - So, it may actually be optimal to back off release from the top reservoir, thus storing more water there to be released in some later period, perhaps when tributary flows are lower.

Other complications can affect river chain management. The above discussion does not consider short term stochasticity, or flow delay times. Flow delays mean that storage cycles can not be synchronised in the simplistic way assumed above. Water released at peak time from one storage will not arrive in time to be used at peak time downstream, and compromises must be found. One storage will optimally be rising when another is optimally falling. and it may be impossible io deliver a surge of water downstream, and certainly to recall water already released, in time to deal with any change to the expected situation downstream.

The point of this discussion is not to develop an arbitraging optimisation algorithm to resolve these issues, but to note their complexity. Many hydro system operators may not even be aware of the theory discussed here, or conscious of the MVS and MCR patterns implicit in their dispatch solutions. Those patterns are

[96] In the limit, an MCR of zero will apply, if the upstream and downstream MVS are identical, implying tat there is no marginal opportunity to gain by releasing from one to the other.

[97] For example, by allowing some freedom for the head ponds above the tributary flow to cycle, in the example above.

[98] In an optimisation or simulation model that simplifies the representation of a river chain down to a single aggregate "power station", the aggregate marginal efficiency of that aggregate station will be falling off, as release rates rise, and the optimal aggregate MCR (MVS divided by marginal efficiency) for the top reservoir release rises accordingly.

potentially quite complex, though, and the MCR of generation can differ significantly between stations in the chain, and between periods of the day, particularly as flow levels rise.

It is quite possible, for example, that the MCR of generation from one station in the chain may be zero, even though MCRs from other stations are relatively high at the time, or that incremental generation is simply not available from some stations. Thus, the aggregate "SRMC curve" for a whole hydro chain may be much steeper than that implied by just dividing the MVS of the top reservoir, by an aggregate station efficiency curve. In fact, some sections of that aggregate curve may represent generation from stations whose MCR is not related to the top reservoir's MVS, at that time.

3.4. Optimal treatment of expansion costs

Historically, it was never possible to develop a comprehensive optimisation model that could simultaneously optimise expansion plans over a decades-long planning horizon, realistic stochastic reservoir management over both daily and annual cycles, and hour by hour planning of generator unit commitment etc. Even today, harnessing the power of cloud computing, many compromises must be made in order to render such models computationally tractable. But it is still worth understanding the optimality conditions such models would have to satisfy in order to optimally balance short- and long-run economics, because the insights derived from that theoretical understanding apply equally to a perfectly competitive market environment. And that, in turn, provides a basis for understanding the important concept of "Entry Limit Pricing" (ELP) in real electricity markets, as discussed in Section 5.5 below.

In fact, the sector's focus on the spot market, and the associated complexities like ancillary services, FTRs, and MWV, often seems to obscure a fundamental truth that actually, most of the sector’s costs are determined by investment decisions. That fact led the WEMS study, which determined the basic NZEM design, to see competition in the investment market as key to the long-run health and efficiency of the sector. Thus, while there are better and worse ways of organising short run interactions between participants, the fundamental roles of the spot market are to allow each asset owner to manage their own operations independently, while maintaining a reasonably efficient level of coordination, and creating an open environment to support competitive entry

The fundamental driver for long term cost minimisation and consumer protection was not supposed to be "spot market competition", as many commentators seem to think, but "entry competition". Specifically, it was believed that the investment costs of alternative technologies could, should, and hopefully would determine not only the optimal plant mix, but also the optimal long run equilibrium pattern of prices that should, in turn, drive operational consumption decisions. So, we consider it critical to understand how a top-down assessment of that optimal equilibrium price pattern relates to the theory we have developed so far, about the bottom-up processes involved in determining the MWVs which drive, and are driven by NZEM spot prices.

One of the factors that has made comprehensive models so difficult to solve is the realisation that expansion options can not realistically be represented by continuous variables. Due to economies of scale, thermal generation has always occurred in the form of discrete plant investments, each of which is reasonably large, at least in the context of the small New Zealand system. And hydro options are much harder to model, because each has its own unique characteristics, design options, and location. The realities of construction schedules, permitting delays, etc, are also difficult to model accurately.

SRMC vs LRMC in a simplified linear expansion model

Still, we can derive some useful insights by thinking about a highly simplified model in which all kinds of capacity are assumed to be continuously expandable, as required, at linear cost. We will also assume, at first, that all costs and technological options available for expansion are constant,[99] across the planning horizon, and that demand is growing, so that capacity of all kinds will eventually need expansion. A

[99] That includes hydro, so we are ignoring the reality that hydro characteristics are site specific, and that the most attractive sites are likely to be developed first.

complete discussion of the theory of optimal expansion lies beyond our scope, here, but some key observations may be made:

- If we assume that the system starts out with an optimal balance of all technologies, and load growth is not necessarily constant, but proportional, in the sense that peak and off-peak loads grow at the same percentage rates, then the technology balance will not change over time. Thus, the capacity of all technologies will expand proportionately across the planning horizon. [100]
- Under those assumptions the marginal cost of expanding all technologies would be constant, over time, and so would the Long Run Marginal Cost (LRMC) of meeting load growth.[101]

There seems to be a widespread belief that LRMC must be systematically higher than the SRMC prices we have discussed to this point, because it includes capital costs, and they (supposedly) do not. But this is a misunderstanding, perhaps resulting from differences in the way SRMC has been traditionally defined in New Zealand, vs overseas.[102]

- It is true that the SRMC of a particular thermal generation technology, for example, will be substantially lower than its LRMC. But the "SRMC price" calculated in an optimisation model, or determined by a perfectly competitive market equilibrium, is the system marginal cost of "meeting" load at a particular time. If load can actually be met, this will be the highest SRMC of any generator operating, and the resultant "operating profits" earned by plant with lower SRMCs will cover a substantial part of their capital costs.
- If load can not be met, though, different authors and jurisdictions may define SRMC differently:
  - Many have thought of SRMC as a purely supply-side concept, and assumed that it could never rise above the highest SRMC of any thermal generator in the system. Load-shedding may have occurred when demand could not be met, but those demand-side costs were not included in the definition of system SRMC.
  - In New Zealand, though, we have traditionally (at least since 1979) modelled load shedding options as a set of "virtual power stations", and included their non-supply costs in our SRMC definition.

One reason for that choice was that there really was no reason to distinguish, when the hypothetical prices involved were only being calculated within an optimisation run by a government department, and not used as a basis for charging, or regulating, any external party. Another is that, in an extended dry year crisis, Government policy included invoking at least one level of load reduction before "base-loading" the most expensive thermal station. So, the top end of our merit order included a mix of virtual and actual stations. Perhaps most importantly, though, most of our capacity was always hydro, the opportunity cost of which (as we have seen) has always represented a probability weighted mix of future marginal supply and non-supply costs. So, any time hydro is on the margin, the "system SRMC" has always included a significant implicit non-supply cost component.

This historical understanding probably underlies the willingness of the sector to adopt a market design in which prices are allowed to rise to whatever level might be required to equilibrate supply and demand. If that were not allowed, then a separate capacity component would have to appear, not just in payments to

---

[100] Including, for example, hydro energy capture, and storage capacity.

[101] Some energy sector discussions talk about "LRMC" as if it was determined by the marginal cost of expanding some particular type of capacity. It may be valid to talk about the LRMC of meeting base-load demand as being set by the LRMC, or Long-Run Cost of Energy (LCOE), of some base-load. technology, and perhaps align that with a Time Weighted Average Price (TWAP). But the LRMC of covering the whole range of loads must be determined by the appropriately weighted costs of all technologies involved, and that is what should align with the (always higher) Generation Weighted Average Price (GWAP), which equals the Load Weighted Average Price (LWAP), if we can ignore transmission losses and limits.

[102] It may also be a by-product of the fact that many markets were launched in situations where surplus capacity had been built under previous regimes, thus creating a situation in which SRMC prices were low, but expected to grow towards LRMC, over time.

generators for MW capacity, but in EMWV calculations that have traditionally been dominated by non-supply cost elements.

By way of contrast, if market prices had been capped at the maximum supply side SRMC, there would have been significant "missing money", not just for the generator with that highest SRMC (who would make no operating profit at all), but for all generation capacity available at that time (since all contributes equally to meeting the demand at that time). So, some form of capacity payment would have been required, with EMWV then perhaps reflecting the cost of meeting the implied requirement to have sufficient water in stock to meet capacity ticket obligations, should they be called upon.[103]

Either way, an optimisation model will simply calculate a shadow price on the constraint matching total supply to total demand, in each period, and that shadow price will automatically apply equally to all capacity types, and form an integral component of the economic justification for their expansion. So, adopting the generalised definition of SRMC that has been traditional in New Zealand:

- Under our simplified assumptions, such a model will continuously expand each capacity type at a rate that keeps its total marginal economic contribution to meeting demand across all periods, as valued at system SRMC prices, continuously equal to its marginal expansion cost.
- Thus, at a system level, the LRMC of balanced expansion to meet balanced load growth should exactly equal expected system SRMC prices, with no additional capacity component required. [104]

Optimal expansion and option values

The discussion above explains how expansion would be optimised by a centralised optimisation model. Such a model would achieve this by internally calculating the national economic benefits of expanding each capacity type, in terms of reducing the need to operate plant with higher SRMC, and reducing the ultimate incidence of non-supply costs, and weigh that against the cost of investment, to determine the optimal expansion of that capacity type.

Exactly the same result applies to a perfectly competitive market, though, except that the national economic benefits would be seen as "operating profits", and the computation described above would be seen as valuing a "call option" for each thermal capacity type, with strike price set at the SRMC for that type of capacity. In which case such plant should be expected to enter at a rate that keeps its call option value continuously equal to its marginal expansion cost.

Since (under our simplifying assumptions) the option value of thermal plant depends only on the cumulative distribution of prices above its SRMC level (and not its chronological detail) we often refer to entry cost as “disciplining", or "controlling" the shape of the annual "Price Distribution Curve" (PDC), formed by listing all hours in a year, in decreasing order of price.[105] To be clear, the details of that curve are not precisely controlled, between thermal SRMC levels, just the total above each thermal SRMC level, and hence the totals between each pair of thermal SRMC levels. The details of that analysis become more complex when we realise that thermal SRMCs are not as well defined as the traditional theory assumes, and that entry costs of intermittent renewable technologies discipline the curve, too. But the basic principles remain valid.

None of this is new to the market era, though. While the terminology has changed, the concept was first articulated in New Zealand during the MoE era. [106] It was then incorporated into the PRISM (later SPECTRA) model developed by the MoE, which applied this logic to value expansion of thermal generation capacity, and extended it to value expansion of hydro energy capture capacity, generation

---

[103] The original WEMS design proposed to synthesise these concepts by not capping market prices, but having capacity tickets defined as call options against those prices.

[104] Whereas an additional capacity component would be required if SRMC was seen to be a purely supply-side concept, and/or capped at the highest thermal SRMC, as is assumed in many international discussions.

[105] The PDC is clearly analogous with the LDC, and there is some degree of alignment, but it does not assume that the hour with the highest price will necessarily be the hour with the highest load.

[106] E.G. Read *Plant Factors for Oil-Fired Plant* Ministry of Energy, Report ER4007, 1984.

capacity, and storage capacity. The relationship to the economics of hydro expansion may be seen as follows:

- First, note that "continuous expansion" does not mean that plant capacity is expanded to match requirements in every hour, and then maybe contracted again when that hour is past. The expansion variables in an optimisation model will typically be annual, with expansion justified by trading off expansion cost with the expected average contribution made by that capacity over all hours of its planned life.[107]
- Likewise, it does not mean that capacity can be "expanded" to meet requirements in a dry year, and then maybe contracted again when that year is past. So, the expected value of the benefit delivered by expansion of any capacity type must be assessed over the full range of hydrological possibilities.[108]
- Ignoring risk aversion, we can simplify this by saying that the marginal cost of expanding any capacity type should equal the expected marginal value delivered by that expansion, over all hydrologies.[109]
- For hydro, this is quite a subtle concept, because expansion will involve simultaneously expanding some combination of energy capture capacity, MW generation capacity, and storage capacity.[110] If each aspect could be expanded independently, at linear cost, we would expect the marginal value delivered by expanding each aspect to align with the marginal cost of that expansion.
- We discuss the marginal value of expanding energy capture capacity below, but the expected marginal value of expanding MW capacity, in terms of being able to take more profitable opportunities in periods where hydro is operating at maximum capacity, should align with the marginal cost of expanding that capacity. And, the expected marginal value of expanding storage capacity, as determined by the average summer/winter or night/day EMWV differences, across all scenarios, should align with the marginal cost of expanding storage.

Energy capture costs and MWV

Energy capture can be increased by bringing more water into the system, to be utilised in new and/or existing stations. But it can also be increased by adding new upstream or downstream generation capacity to extract more of potential energy out of existing water flows. So, the marginal cost of expanding energy capture capacity may have more than one component. Assuming that marginal cost can be measured, though, we want to know how it relates to the setting of EMWVs via opportunity costing, There clearly must be some relationship because EMWV is really the marginal value of the potential energy stored by holding water in a reservoir, and we have argued that:

- The LRMCs of the mix of available entry technologies shapes the long-run equilibrium pattern of optimal system SRMC prices; while

[107] In principle, this evaluation should range over the whole project life cycle, but decisions can be based on comparison between the annualised cost and first year benefits, under our assumptions, because the value of expansion is constantly increasing.

[108] Ignoring the possibility that a planned project might be slightly advanced or delayed in response to conditions in its year of completion.

[109] The problem is still very difficult, in practice, because the management of all those hydrological scenarios still has to be optimised by a stochastic reservoir model, solved within the context of a long-term expansion model. But we are only concerned with concepts here.

[110] These components interact in subtle ways. Thus, adding a downstream station to a chain will obviously increase MW capacity. But if it adds 10% to the energy produced from each unit of water released from the top reservoir, it will add 10% to (energy) storage capacity and energy capture, too.

- EMWVs are all determined by a specific opportunity costing process that effectively samples from that same probability distribution of prices.[111]

Section 5.5 discusses the idea that managers might consciously move EMWVs towards levels approximating entry costs of particular technologies, or more subtly shaping the whole distribution of prices to align with the long-run equilibrium PDC discussed above. In a perfectly competitive market context, though, alignment must arise naturally, without conscious intent.

In fact, the theory we are discussing does not tell us what EMWV should be in any particular instance. Our previous discussions indicate that EMWV will effectively be set from electricity prices in a sample of hours, being specifically the first hours in which this particular hydro station is expected to be on the margin, under various scenarios.

- If stored energy is in short supply, it will be reserved to become marginal in a few hours where prices are expected to be relatively high. So, it may occasionally be competing with non-supply, but more often with generation from thermal stations with high SRMC, and from other storage-based generators, who have determined their own EMWVs from a similarly high-priced sample.

- If stored energy is in good supply, it will be used freely, and only become marginal in a few hours where prices are expected to be relatively low. So, it may occasionally be competing with spill, but more often with generation from thermal stations with low SRMCs, and from other storage-based generators, who have determined their own EMWVs from a similarly low-priced sample.

But, while the sample will be different, under different circumstances, and for different reservoirs, all of the hours in which profitable opportunities are identified will have been selected from a whole pdf of prices, the shape of which is strongly controlled by the marginal cost of expanding all capacity types in the optimal mix:

- If investment costs rise, that price curve will rise, and so will all EMWVs.

- If, say, the cost of expanding peaking capacity goes up, and the cost of expanding base-load capacity goes down, that change will be reflected in the equilibrium price pdf, thus raising EMWV for peaking hydro and lowering it for base-load hydro.

- And that will make a centralised optimisation, or agents in a perfectly competitive market, seek to build less base-load hydro capacity, going forward, and more peaking capacity, if it can be found.

In fact, in equilibrium, the storable inflow-weighted EMWV,[112] which measures the value of energy captured by a particular storage system across all periods in all hydrology sequences, should align with the marginal cost of expanding the energy capture capacity of that system, on average.

In summary, then:

- A comprehensive optimisation model would compute a probability distribution of system SRMC prices that aligns with the marginal expansion costs of every capacity type in the optimal plant mix.

- The precise valuation of hydro projects is complex, because each project is different, and the same project may play a different balance of roles (e.g. base-load vs peaking) under different hydrological scenarios, or in different phases of the same scenario, with that mix shifting over its lifetime.

- Still, a comprehensive optimisation model, or hypothetical perfectly competitive manager, would align EMWV, in every period of every simulated hydrology scenario, with those parts of that system SRMC price pdf with which the various capacity components of that reservoir are competitive; and

---

[111] Or, to the extent that EMWVs also drive the setting of those prices, we should really say that the EMWV and electricity price distributions are in equilibrium with each other.

[112] That is, we take the EMWV in each period when inflows arrive, weighted by the marginal increase in storable inflows arriving in that period.

- Finally, a truly comprehensive optimisation would determine that stochastic reservoir optimisation policy internally, account for it in setting the pdf, across all those scenarios and, if the average estimated EMWV is high enough, favour hydro expansion of a type that increases the capture of energy by that reservoir, and/or reservoirs of that type.

Deviation from LRMC equilibrium

Finally, Section 5.5 will discuss the implications of this theory in a more realistic market environment but, even in a centrally optimised for perfectly competitive environment, we should relax our assumptions, because development will never be as smooth or predictable, as assumed in the above discussion. Constant "surprises" will cause expansion to deviate from any optimal/equilibrium path, even in a centrally optimised environment. Thus:

- A surplus, or deficit, can be expected to develop every time load growth is lower, or higher than predicted;
- New technologies may also emerge, undercutting the economics of older plant types, so that they no longer recover their full investment cost, and fall off the list of options defining the optimal plant mix for future expansion; or
- Particularly for hydro, available expansion opportunities may get progressively more expensive, possibly eliminating that plant type from the list of options defining the optimal plant mix for future expansion, but boosting the economics of older plant of that type, so that they more than recover their full investment cost.[113]

System SRMC price levels should be expected to fall below, or rise above, LRMC whenever such events occur. And, since we can not really "undo" investments, or instantly create them, such discrepancies may persist for several years. Nonetheless, if we assume the likelihood of all those possibilities is properly factored in when assessing development options, we still should expect to see system SRMC price distribution anchored to a realistic assessment of LRMC entry costs, in the long run.

[113] Since that outcome is broadly foreseeable, for hydro, the expected boost to long term value should theoretically have been accounted for by the parties who "sold" the right to develop those earlier sites, in which case those projects should just break even, over their (very long) lifetimes, on average. But the historical development and valuation processes were obviously very different.

4. Impending Change

4.1. Introduction

The plant mix in the New Zealand electricity system is expected to change quite radically over the next few years, and that will force major changes to the way in which the system has traditionally been operated. Our goal here is not to discuss the feasibility, or economic viability, of these changes, but to develop an understanding of how the market would be expected to work with this new plant mix, and particularly how hydro reservoirs would be managed in the new environment. As in Chapters 2 and 3, we assume that operations in the sector are either being explicitly optimised by a central agency, or implicitly optimised by a perfectly competitive market. Thus, we defer discussion of market realities until the next chapter.

Basically, we argue that the conceptual framework developed to this point need not change, and we have already noted that essentially the same framework has routinely been applied to Latin American power sectors with little or no thermal generation capacity. While we do expect major changes in operational strategies, and in actual MWV curves, we believe those changes to be matters of degree rather than of principle, and fundamentally quantitative, rather than qualitative.

We do not attempt, or comment on, any quantitative analysis, though, or make any forecasts about the relative influence each of the effects discussed here might have on market outcomes. So, we make no judgments about whether the outcomes delivered by managing the emerging system under the current market regime will be "acceptable", let alone socially "optimal", or whether the changes might eventually become extreme enough to warrant any changes in market design.

Our discussion covers three broad areas:

- First, the increased penetration of intermittent renewables, and the withdrawal of thermal capacity, which we see as the two key factors driving change;
- Then the three key developments we expect to mitigate the impact of those changes, namely increased reliance on unconventional storage options, demand side response, and possibly green thermal capacity; and
- Finally, the general impact these changes to the entry mix can be expected to have on price patterns, and hence on EMWVs in the emerging environment.

4.2. Factors driving change

4.2.1. More intermittent renewables

Many discussions distinguish between renewable and non-renewable generation sources. For operational purposes, though, the real issues are:

- Whether the "fuel" powering the generation is storable, in the sense that, within limits, reducing generation in one period can allow more electricity to be generated in another period, and
- Whether that fuel is readily tradeable at some well-defined market price.

In the absence of any associated storage facility, neither of these applies to wind, solar, geothermal or run-of-river hydro. Accordingly, their impact is just to increase the variation and volatility of the net demand to be met by generation sources that do rely upon storable and/or tradeable fuel. Ignoring any correlation with load, the expected effects of each type of development would be:[114]

- Independent run-of-river hydro capacity (not associated with any hydro storage reservoir) may not expand by much, but any expansion probably increases predictable variation, over an annual cycle, and also volatility to the extent that runoff is unpredictable over the various time frames concerned.
- Wind capacity is expected to expand substantially, and that will increase short term volatility. It could either increase or decrease predictable seasonal/daily variation to a lesser or greater extent, depending on the extent to which it can be forecast over the various time frames concerned, and/or correlated with demand or other generation sources.
- Solar capacity is expected to expand substantially, and that will increase volatility to the extent that cloud cover is unpredictable over the various time frames concerned. It also has very predictable daily cycle and yearly cycles. The daily cycle is positively correlated with load, but not necessarily with peak load. So, while it has the potential to significantly change, and even reverse, the cycling of smaller hydro storages on particular days, the overall impact seems likely to be mixed. The yearly cycle is negatively correlated with load, and positively correlated with hydro inflows it must increase pressures on inter-seasonal arbitraging capacity.
- Geothermal may expand, too. This will be ignored here, because it does not introduce much new volatility of its own, except to the extent that it may be subject to breakdowns, just like any other generation capacity. It will have the general effect, though, of reducing the relative importance of both volatility and cyclic variation from other sources.

The overall impact on the daily cycle of demand/supply imbalances, and hence on predictable daily cyclic price variation seems unclear. It seems clear that there will be a significant increase in the annual cycle of demand/supply imbalances, though, and hence on the predictable summer/winter price/EMWV cycle. It also seems clear that volatility will increase, at both the daily and seasonal level. If volatility is strong enough, it will imply a significant increase in the number, and length of intervals in which the perfectly competitive market price could fall to (near) zero. On the other hand, in equilibrium, there should be an offsetting increase in the number or length of intervals in which prices are very high.[115] The impact on average prices is less clear, and we believe it will actually be driven by entry costs, including the entry cost of storage options, as discussed in Section 4.4 below.

Hydro system impact

Of themselves, increased volatility and/or cyclic variation make no difference to the principles discussed to this point. More extreme prices, in periods when hydro generation from any particular reservoir would not have been marginal, will not impact its EMWV at all. Of itself, increased volatility will not necessarily make a difference to average price or EMWV levels either. The quantitative impact will vary from system to system, though. There are obviously intermediate cases, but:

For long-term (annual) storage reservoirs:

- Changes to short term volatility, or variation, should have little impact on their optimal storage profile or their calculated EMWV (which is the MWV most commonly referred to in sectoral discussions).
- Inter-seasonal and intra-seasonal volatility only seem likely to increase, putting more weight on the "stochastic" objectives of avoiding spill and non-supply, thereby reducing the effective storage

---

[114] Here, we look at the impact of expanding capacity of each type, assuming all else equal, while ignoring the possibility that some of these technologies might expand more slowly than load, and thus actually form a decreasing proportion of total generation. The impact of mitigating technologies, like batteries, is discussed in Section 4.3 below.

[115] We have not discussed serial correlation, but that will obviously increase the likelihood of extended periods of high/low sun or wind generation, and hence of low/high electricity prices.

range that can utilised for inter-seasonal arbitrage, and increasing inter-seasonal price differences, on average.

- Increased inter-seasonal variation in the supply/demand balance can only increase summer/winter electricity price differences, too, and put more weight on the "deterministic" value of utilising the full storage range, at the expense of greater spill in autumn, and perhaps higher non-supply probabilities in spring.
- Higher summer/winter price differentials may also be expected to increase summer/winter EMWV differentials. but, the combination of these effects with thermal withdrawal may eventually lead to a situation in which EMWV would need to be held high all year, as discussed further below.

For short-term (daily) storages:

- Changes to inter-seasonal cyclic variation or volatility would have no real impact on operational patterns, or intra-day MWV setting.
- Increased/decreased intra-day variation would increase/decrease Day/Night MWV differences, and put more/less weight on the value of utilising the full storage range to deal with expected intra-day cycles, and comparatively less/more on the "stochastic" objectives of avoiding spill and non-supply.[116]
- Intra-day volatility seems likely to increase, putting more weight on the "stochastic" objectives of avoiding spill and non-supply.[117]
    - To the extent that volatility consists of broadly symmetrical up/down swings, operators can be expected to hold storage and flows further away from both bounds, preserving flexibility to swing generation suddenly towards one extreme or the other.
    - To the extent that volatility consists of sudden unpredictable downswings in generation from other sources, operators may need to hold storage levels high, and flows low, so as to ready to surge generation suddenly upwards, to compensate.
    - Either strategy will reduce the storage range available to manage intra-day arbitrage, and tend to increase day/night price/EMWV differences, on average.
    - Either strategy will also increase river flow and storage volatility, which may be of socio-environmental concern.

4.2.2. Less thermal capacity

The greatest change, indeed the driving force behind all of these changes, will be to eliminate fossil-fuelled conventional thermal capacity from the plant mix. This removes significant storage capacity, and all international tradability from the existing system, and consequently eliminates most of the traditional benchmarks against which hydro has been opportunity costed. That does not change any of the principles discussed, but it will materially affect the results of applying those principles, and system operation generally.

Impact on long term hydro storage management

First, for the annual planning horizon, the arbitrage logic remains the same, and EMWV can still be set by averaging MWVs found for individual scenario simulations. We can still draw contours on the resultant

[116] "Non-supply" may not really be an issue for any individual small storage, but a collective failure to supply would force prices to a very high level, so the issue is really about whether operators are incentivised to keep any storage in reserve to take advantage of such opportunities.

[117] ditto

EMWV surface, too.[118] But those contours will no longer act as “guidelines”, and no longer "capture” storage trajectories and guide them toward more moderate sustainable storage levels.

Low inflow sequences will cause storage to fall at a faster rate than it does now, so it is more likely to fall closer to the lower storage bound, earlier in winter, causing more frequent, and more intense, activation of non-supply options.[119] The only way to avoid (or at least reduce) more frequent non-supply around spring will be to hold storage at higher levels around autumn, thus raising the frequency and volume of spill.[120] This is all just the inevitable physical result of removing thermal capacity from the system, and it will impact on market prices and EMWV patterns, but it does not indicate any change to the theory, or failure in the market system.

Removing the inter-seasonal arbitrage capacity implicit in the thermal fuel storage/trading system should logically increase the expected differential between summer and winter price levels, and this effect will be exacerbated by the need to operate hydro storage cautiously, to deal with greater volatility. As noted earlier, that decrease in effective storage capacity actually increases the importance of the "deterministic" perspective of maximising utilisation of the remaining inter-seasonal transfer capacity. The effect on EMWV patterns is less clear, though.

Despite the high probability of spill, the EMWV (MVS) of water actually stored by the end of autumn will have to be higher, and perhaps much higher, reflecting the increased probability of prices reaching non-supply cost levels at some point during forward simulations of more and more hydro scenarios. The difference between this high MVS and the zero MCR value being received for spilled water should incentivise efforts to expand long-term storage capacity in various ways. The high autumn spill probability will moderate EMWV over summer, but summer EMWV must optimally remain positive, in order to ensure that long term storage hydro remains above the SRMC of non-storable sources, while the reservoir is filling.

That is, a centralised optimisation or perfectly competitive market should always ensure that as much intermittent renewable generation as possible is used to build up hydro storage, rather than being spilled. And storage should be built up until spill becomes (close to) inevitable, rather than water being released earlier in the season to displace whatever power might be available from those non-storable sources.

But Section 2.3.1 discussed a super-simplistic theoretical example where the only two competing technologies had SRMCs of 10 and 200, and noted that the MWV would not be set by (projected) prices in periods in which hydro was not (expected to be) on the margin. So, even if the summer was characterised by long periods of very low prices, those pries would not set EMWV for hydro reservoirs which should optimally be releasing only the minimum required to maintain river flows, over those periods. If (in this hypothetical world) we knew for certain that hydro would eventually be marginal at the 200 price, next winter, and that the reservoir would not become absolutely full in autumn, then that 200 price should set the EMWV right through summer, too.

More realistically, if simulation from the spring storage level showed that there was a 50% chance of that reservoir not being absolutely full by autumn, the EMWV must be at least 100 now, reflecting a 50% chance that a marginal unit of water will be carried through to winter, and then valued at 200. In the meantime, hydro could spend most of the summer releasing at minimum rates, and so maximising the collection of renewable energy for winter use, but occasionally ramping up to deal with short term situations where prices peaked above 100. While in that filling mode, it would largely be at the mercy of the weather, and specifically of its own inflows filling its reservoir at a rate it could not control, while releasing at minimum rates.

---

118 We can also still create multi-dimensional EMWV surfaces, and draw contours defining the stock balance levels at which inter-regional transfers will be stepped up to balance stock levels between storages. But that complication can be ignored for the purposes of this discussion.

119 Or some other DSM option, as discussed below.

120 With no thermal generation to be backed off, this may be manifested as extra water being spilled past hydro generation plant, or, if hydro does not spill, by geothermal, wind or solar generation potential being "wasted".

In reality, there would be a wider range of situations occurring over summer, in which our hydro reservoir could make contributions at intermediate prices. This could involve backing off release from some other hydro, and thus maintaining a balance between their storage levels, or some relatively low-valued DSM response. But there are also a large number of short-term hydro storages involved, and a potentially very large number of batteries in future, too. Even in this artificially extreme 10/200 world, each will be setting its own MVS, based on probability weighted values it can obtain from the opportunities it can access, over the day (or longer in some cases).

- Facilities with limited storage capacity, but high discharge/release capacity will save energy for use in only the highest priced periods, and may set an MVS near 200.
- Facilities with more storage capacity, and/or less discharge/release capacity will also take those opportunities, but find their marginal opportunities in lower priced periods, thus setting an MVS nearer to 10.

The interaction between all those short-term storages should thus imply a range of prices across each day, creating a range of opportunities, which our reservoir manager should plan to take, to a greater or lesser extent, particularly in managing the (hypothetical) 50% of scenarios in which the reservoir is expected to be full by autumn. So, those scenarios will typically each contribute some positive MWV value, and EMWV will be greater than 100, but:

- If it becomes increasingly clear, as summer progresses, that storage will not be completely filled, then EMWV will gradually rise towards 200, eventually reaching that level once it becomes clear that storage will not be filled before it starts falling to meet winter needs.

- If it becomes increasingly clear, as summer progresses, that storage will be completely filled, then EMWV will gradually fall below 100. Perhaps it will fall as low as 10, if it becomes clear that storage will definitely become full, even if all available opportunities above that price are taken, or even as low as zero, if spill becomes inevitable. It must gradually rise back up to 200, though, as the storage is held as close to the full level as prudence will allow, with increasing demand being met by decreasing inflows, and reach that level once storage starts falling to meet winter needs.

The winter/spring situation should be the mirror image of the above. The deterministic perspective argues for allowing the spring storage level to get very low, and keeping it low for some time, while the EMWV gradually falls to a low summer level. But, with no thermal support, the stochastic perspective suggests that reservoir\managers may need to hold storage significantly higher than at present, throughout that period, so as to avoid the possibility of shortages due to inflows dropping off in late winter or early summer. That suggests that only some simulated trajectories will actually reach very low levels and that, on average, more water may end up being carried through spring into summer.

As discussed above, that effectively reduces the storage range available for arbitrage, and must increase summer/winter electricity price differentials. But that does not necessarily carry through to summer/winter EMWV differentials:

- Those scenarios that do not experience a late winter/spring rundown will come into the summer with higher storage, making it easier to rebuild storage by the next autumn. And that must imply a lower probability that a marginal unit of water will be carried right though into the next winter, thus allowing EMWV to be set at a lower level, through the summer, possibly falling all the way to zero by autumn, then rebounding for the next winter, as above.

- Those scenarios that do experience a late winter/spring rundown will come into the summer with lower storage, making it harder to rebuild storage by the next autumn. And that must imply a higher probability that a marginal unit of water will be carried right though into the next winter, thus raising the possibility that EMWV may be high right through the summer, and perhaps rising all the way back to its winter level before autumn, as above.

The optimal balance between these two competing perspectives is an empirical matter that we are not in a position to comment on. It seems clear that the average summer/winter price differential will have to increase, and that there will be increased pressure on the storage system's ability to carry forward sufficient energy from summer to winter. Ultimately, the pdf of simulated storage trajectories and EMWVs must sit in cyclic equilibrium, with the probability of spill in autumn balanced against the probability of shortage in spring. And the latter must be given much more weight because it is so expensive.

The summer/winter EMWV differential could increase, too, but the situation could also evolve towards having a relatively high EMWV all year, with minimal discretionary hydro use over summer, so as to maximise renewable recharge ready for the next winter. The implied river flow pattern may differ significantly from the status quo but the socio-environmental implications of that change lie outside our scope, here.

Impact on short term hydro storage

Just as for the annual planning horizon, the intra-day arbitrage logic remains the same, but the absence of peaking thermal capacity will decrease effective storage capacity in the systems, and reinforce any tendency for increased in intra-day volatility driven by expanding renewable generation capacity to increase intra-day price variations.

Coping with that increase may require holding some hydro storage at high levels across the day, in order to be ready to mitigate the effects of unexpected reductions in solar and/or wind generation.[121] And, the more short-term storages move into a mode of opportunistic response, the less capacity they will have for regular daily cycling. Of itself, that means they will be less able to moderate day/night price differentials, which should consequentially increase.

On the other hand, solar will only produce during the day and, especially in summer, this may have a significant impact on the daily supply/demand imbalance cycle. Finding sufficient MW capacity to cover early evening peaks could be a challenge, but the overall need to shift energy from night to day, could actually reduce, thus at least partially offsetting the loss of intra-day thermal fuel storage capacity. So, the balance of effects is less clear, in this case, but the ultimate outcome will also depend on the extent to which mitigating technologies can be employed, as discussed in the next section.

What does seem clear, though, is that short term storage will become increasingly important, implying a need for greater managerial attention, focussed on extracting maximum value out of the physical capacity available. The socio-environmental implications of that change lie outside our scope, too, but the economic/market implications of smaller storages reaching their limits more often must be a greater variation in marginal water values, between storages, and across the day. And we might hope to see those values more explicitly calculated, and clearly reflected in market offers, thus partly taking over the current role of thermal in shaping an intra-day price profile, as discussed above.

4.3. Factors mitigating change

4.3.1. Possible development of “green thermal”

The possibility of "green thermal" has been floated, but the impact of any such development on hydro opportunity costing depends very much on the storability and tradability of its feedstock:

- At one extreme, a plant without any long-term fuel storage capacity would have to operate on a base-load, or short-term cycle basis, and will have no appreciable impact on EMWV setting for major reservoirs.
- At the other extreme, a large-scale plant burning internationally tradeable bio-diesel or green hydrogen (in some form), whether produced locally or not, could perform a very similar role to traditional thermal. The SRMC of that plant may be rather high, but a guideline could be drawn, below which that plant could be used to conserve hydro storage. And that would set an MWV benchmark for other storages, just like any conventional thermal SRMC.

121 Noting that solar generation reduction is only an issue during the day, and mostly during the middle of the day when it is generating most.

- In principle, if the same plant were fed only from a closed local hydrogen production/storage system, the situation would be more similar to that of current gas-fired plant. It might have more inter-annual flexibility, though, if utilising under-utilised old natural gas infrastructure. If so, that could introduce inter-annual arbitrage possibilities capable of reducing year-to-year price fluctuation driven largely by hydrological fluctuations. The principles for opportunity costing of that hydrogen reservoir would be essentially the same as for hydro, but the longer timeframe would tend to stabilise the optimal marginal value of its stock, thus providing somewhat more of a benchmark, than a purely intra-annual storage.
- Alternatively, a bio-fuel plant that depended on an inflexible local fuel source, such as wood, would face the issue of how much of that fuel could be stored, and for how long, without undue deterioration. Ideally, a stockpile could be built up over several years, then used to generate in a dry year, thus again creating inter-annual arbitrage opportunities, as discussed above. But, if the biofuel deteriorated at a rate similar to stockpiled coal, the overall flexibility might not be much better than for a plant dependent on local coal, as discussed in Section o.

4.3.2. Increased reliance on non-traditional storage

The combination of increasing reliance on intermittent renewables while losing thermal storage suggests that storage capacity will be at a premium in future, in both inter-seasonal and intra-day timeframes, and preferably inter-annual, too. To this point, our discussion has only considered the storage capacity of conventional hydro, once thermal generation is removed, but we expect there to be strong incentives to supplement that capacity with other storage options.

Batteries

Batteries seem unlikely to ever compete with hydro as a means of arbitraging over an annual planning horizon, but they are becoming increasingly competitive, as a means of counteracting volatility within and between dispatch intervals, and of moderating intra-day cycles. Battery storage could include a spectrum of options ranging from large purpose-built battery facilities at the transmission grid level, through to highly dispersed household installations, typically associated with rooftop solar, and mobile batteries in electric vehicles. Depending on their configuration, batteries can also be used for a wide variety of purposes, ranging from real-time regulation and contingency response to local emergency backup, to intra-day cycling.

It should be recognised that not all batteries are suited to all tasks, and adding versatility can significantly increase costs but, in principle, the same battery can serve several different functions, perhaps at different times, and perhaps even simultaneously. For example, Naidoo and Read analysed the optimal design and operation of an island power system in which the only non-thermal development prospects were solar, and very limited hydro.[122] They concluded that the same batteries should be used to supplement limited hydro storage by absorbing overnight excess hydro generation to cover the lack of solar when weather was projected to be dull and wet, and to store excess daytime solar for night-time use, when weather was projected to be fine and dry, with some batteries also responsive enough to supplement hydro's ancillary service capabilities, much of the time.

The details of operational optimisation obviously become complex when multiple alternative or simultaneous functions are in play, and particularly so if distributed resources have to be coordinated. Basically, though, batteries are storage devices, and the same general theory applies to their operation as to any other storage. Thus, a battery's daily operational cycle can be optimised in a very similar fashion to that any small reservoir, forming an MVS curve that would be entirely based on opportunity costs. The same deterministic/stochastic perspectives apply, and forecasting is likely to play a major role in

[122] R. Naidoo and E.G. Read "Analysing Renewable Power System Development for a Pacific Nation" *EPOC winter Workshop* , Auckland, 2017

determining optimal strategy. Guidelines (more likely referred to as "switching curves") can be formed in a similar way, and play a similar role. There is one important difference, though:

- Battery storage is not costless in energy terms, so batteries must be seen as adding load, to be met by adding further generation capacity to the system; and
- Even a free, and infinitely large battery could only arbitrage away day/night electricity price differences +/- the proportion of energy it loses in each charge/discharge cycle; and consequently
- The day- and night-time MVS for a daily cycling battery must differ by at least that loss factor; and
- That loss factor also sets an upper limit on the extent to which battery investment can compete away the night/day MVS differentials in any short-term hydro storages operating in a similar mode.

Of course. batteries are far from free, which means they will not be built with infinite capacity, either. Still, we expect they will exert a significant moderating influence on short term price volatility and cyclic variation, even after accounting for investment costs, as discussed in Section 4.4 below. Moderating day/night price cycles and short-term volatility will have minimal impact on the long run MWV pattern of major reservoirs, though.

Pumped storage

Unlike batteries, pumped storage hydro is a very site-specific technology. So, while pumped storage is, in principle, able to perform an even wider range of functions than batteries, all the way out to inter-annual arbitrage, that does not mean that any particular pumped storage development will actually be able to perform all those functions simultaneously, or even at different times.

Accordingly, while we will not discuss the possible impact of any particular pumped storage proposal, it is clearly very dependent on scale. While facilities with proportionately small storages play a critical role in managing daily cycles overseas, they would need to compete directly with (lossless) cyclic operation of conventional hydro storage here, and increasingly with batteries that may have lower losses per cycle. While some such developments might be profitable, the more likely role for pumped storage hydro in New Zealand is to provide extra long-term storage.

In principle, an inter-seasonal pumped storage facility would operate more like a large long-term battery than a conventional hydro system. That is, ignoring the (possibly significant) costs, losses and constraints around the switching process itself:

- pumped storage would not be costless in energy terms, and must be seen as adding load to be met by adding further generation capacity to the system; and
- There would be a significant 'dead band' within which prices could differ without triggering any pumped storage response; so
- Even an infinitely large pumped storage facility could only arbitrage away summer/winter electricity price differences +/- the proportion of energy it loses in each pump/generate cycle; and consequently
- Its ability to reduce inter-seasonal MVS differences in other storages would be similarly limited.

It is well known that a large facility of this kind is currently being investigated, and recognised that such a facility could, in principle, offset and even reverse most of the adverse impacts we have attributed to the replacement of thermal capacity with increasing development of intermittent renewables. A facility large enough to do that, though, would clearly have, and be intended to have, a major influence on market price patterns, raising questions about how it could be managed, regulated, or analysed under a "perfectly competitive", or even "workably competitive" paradigm. Thus, we understand that a variety of possible management regimes are being investigated. But the starting point for understanding how it might be operated, and how it might affect the sector, should still be to ask how its operation would be optimised from a national benefit perspective, or if it could operate flexibly in a perfectly competitive market, just like the other storage facilities discussed here.

In fact, we can determine an EMWV for the pumped storage facility, just like any other storage. It's just that, if we think of its MVS still being the value of what is actually stored, and setting its MCR when marginal, then we need to subtract the pumping losses when determining what its pumps should be prepared to pay as a load in the market. In that respect, the situation is essentially similar to that for a battery, but this is an upstream/downstream river chain situation, too, so we also need to account for the EMWV in the source from which it pumps, and to which it releases, at a different time.

The situation is further complicated if the same facility can also operate on a daily cycle. It seems technically possible that such an operational mode could be overlaid on a long-term empty/filling cycle, and it would be economically compatible with pump/generate triggers being set in relation to a slowly changing long-term MVS. But it remains to be seen whether that would be considered appropriate, given the overall goals of such a large-scale development, and its possible impact on competitive entry by other technologies.

If both batteries and pumped storage are built, and both are equally flexible, strict short run economic logic suggests that the technology offering the lowest daily cycling losses would be the first called upon to deal with intra-day price volatility, and to arbitrage between day and night-time prices. The ability of that technology to control price differentials will always be limited by its charge/discharge and storage capacity, though. So, those differentials could often end up being set at the wider limits implied by the "backstop" storage technology, or escaping those limits, too. That is as it should be, given optimally expanded storage capacity. And both technologies would be making a positive contribution in terms of limiting, or at least reducing, price differentials. There would be a significant impact on, and interaction between, the entry economics of both technologies, though, as discussed in Section 4.4 below.

4.3.3. Increased reliance on demand side response

The crude representation of non-supply costs in traditional optimisation models leaves much to be desired. "Non-supply" has never really just been a matter of the lights going out, without warning, and New Zealand has traditionally recognised the importance of other Demand-Side Management (DSM) options in managing the significant risks involved in harnessing volatile natural energy sources to meet variable demand in a, relatively small, isolated power system. And it is to be hoped that modern communication, computation, and control systems will allow more sophisticated DSM options, in future.

There are several distinctly different DSM categories to consider, though:

- First, elastic demand reduction means that consumers are just responding to the prices they see by not consuming electricity for a particular purpose that would otherwise be desirable, without any intention to use electricity for exactly the same purpose, at some later time.[123] In principle, this can be modelled just like a set of (probably small) "virtual thermal stations", with defined MW capacity and SRMC. A traditional EMWV estimation process would then form guidelines for those virtual stations, as for any other thermal station. The corresponding "stations" would never be explicitly asked to respond, but rather, assumed to respond to market price signals autonomously.
- Second, contracted demand reduction could involve consumers agreeing to reduce consumption when asked, in return for agreed compensation. If we can ignore the problem of distinguishing between "reduction" and "deferral" (as discussed below), this can also be modelled just like a set of (probably small) "virtual thermal stations", with defined MW capacity and SRMC. A traditional EMWV estimation process would then form guidelines for those virtual stations, as for any other thermal station. The corresponding stations would be explicitly "asked" to respond, though, rather than just responding to market price signals autonomously.
- Third, elastic demand deferral means that consumers are just responding to the prices they see by deferring the consumption of electricity for some purpose, with the intention of using electricity for

[123] That is, for a producer of goods, they would actually produce fewer goods, over an extended period, rather than making up any lost production in a later period.

exactly the same purpose, at some later time.[124] In principle, this should not be modelled as a "virtual thermal station", but as a "virtual battery" with defined charge, discharge and storage capacities, but a cycle cost, rather than a loss factor. Theoretically that virtual battery's behaviour might be modelled as responding optimally to actual price differences on, say, a day-to-day basis.[125] Pragmatically, though, many such responses will be pre-programmed, and just accounted for in load forecasting.

- Fourth, contracted demand deferral could involve consumers agreeing, in return for agreed compensation, just to defer consumption of electricity for some purpose until a later time.[126] In principle, that also might be modelled as a virtual battery or pumped storage station, with defined charge, discharge and storage capacities, and a fee instead of a loss factor.[127] Such a battery's behaviour might be modelled as responding optimally to expected price/EMWV differences, because its response will be explicitly called upon by some party, in response to some interpretation of market signals. But small-scale responses of this type will probably be pre-programmed, too.

Overall, there would seem to be significant potential for demand-side response to enhance effective system storage capacity, and/or replace thermal flexibility, but also much development still to be done in that area. Each possible type of development would behave somewhat differently, and could be modelled somewhat differently, with the principles we have discussed being applicable to manage the options involving "storage" of any kind.

The key thing is, though, that even the "reduction" options are implicitly utilising the storability and tradability of goods, feedstocks etc outside of the electricity sector, and perhaps the country, to substitute for the storability and tradability traditionally called upon within the electricity sector, or associated fuel sectors. Thus, when the New Zealand producer of widgets cuts back on production, the consumers of widgets, here and elsewhere, may defer their consumption, or draw down their own stocks, or buy from other manufacturers elsewhere, or some combination of all three. But the net effect is to draw on the inherent diversity and flexibility of the international production/storage/trading system, which is vastly greater than that of the local electricity sector.

Such demand-side flexibility could make a significant contribution, either from the aggregation of a great many small-scale (e.g. household) responses, or from a few large scale responses, such as proposals to establish a large-scale export plant (e.g. of ammonia) the export of which could be reduced in "dry years", or perhaps more exactly when hydro storage falls to a certain guideline level, or projected electricity market prices rise high enough to make New Zealand electricity sales more profitable than supply to the exporter over some planning horizon.

All of this DSM activity could help reduce price variation/volatility over various timeframes, but not all of it would necessarily provide the kind of well-defined benchmark that thermal SRMCs have traditionally played in centralised optimisation models. On the other hand, it should be recognised that the current market actually operates without clear benchmarks, too. Hydro managers do not know the actual SRMC of thermal plant, or even whether the thermal plant operator formally calculates an SRMC at all; they merely infer some estimate of the likely thermal response at projected future price levels, and they could arguably learn to infer estimates of the likely DSM responses in a very similar manner. In fact, something like an explicit arrangement with a major exporter could provide a clearer benchmark than any arrangement operating in the status quo.

---

[124] Households do this every day, when they set timers so that appliances are used in off-peak periods, and can be expected to do so on a much larger scale, once charging of electric vehicles becomes a routine household activity.

[125] We are describing this DSM option as distinct from the charging and discharging of actual batteries, including the batteries in electric vehicles, which could potentially provide power back to the grid. In practice, the two may be indistinguishable, though, particularly at the household level.

[126] Ripple-controlled water heating provides a classic example, in New Zealand, as does controlled heating of floor slabs, or heating/cooling of commercial buildings, etc

[127] That is, if compensation is arranged on a per unit deferred basis. Other compensation mechanisms might be treated as more akin to capital costs.

4.4. Impact of changing entry cost mix

Finally, we turn to an issue that may be less obvious, but possibly more important. In Section 3.4, we discussed the optimal treatment of investment costs, and noted that the optimal pattern of electricity prices in a perfectly competitive market (or of shadow prices in a centralised optimisation) is determined by the economics of the set of options that could optimally expand in the developing plant mix.[128] And that key insight follows through into real markets, as discussed in Section 5.5 below on Entry Limit Pricing.

In the past, that theory has primarily focused on the role of thermal generators in shaping (or "controlling", or "disciplining") the long run average PDC. That focus has partly been because the theory was developed in that context, and is very much easier to apply, and to understand, in that context, with thermal stations very naturally being valued as "cap options". But it also seems clear, for example in the results of MBIE's 2018 Pricing Review,[129] that those thermal expansion options probably have had a significant influence in shaping actual NZEM PDCs.[130] So, it seems important to note that ruling any technology out of the optimal expansion plant mix immediately makes that technology irrelevant as a shaper of the optimal equilibrium PDC, going forward, even if plant of that type are still operating in the system.

Pricing implications of expanding generation capacity

Specifically, the operating/fuel costs of existing plant are still relevant for day-to-day operations, and competition, but the capital cost of building more of that plant type no longer constrains how high prices should optimally be, over the periods when that existing plant should be optimally operating. To take a simple example:

- Let's assume that the last resort supply-side option in our current system is a Diesel-fuelled OCGT peaker, with only DSM options available to control the supply/demand balance if loads rise above total system capacity, up to and including that peaker. (Or, in other words, whenever optimal perfectly competitive prices rise above that peaker's SRMC.)
- If, in this hypothetical world, system extension has been occurring in an optimally balanced way, the expected value (over all hydrologies etc) of a call option corresponding to that OCGT peaker should always equal the current cost of building more capacity of that type.[131]
- Now, let's eliminate that plant type from our allowable expansion menu, and replace it with an option to install new plant that is identical in every way, but with twice the capital cost.
- Then, optimal expansion planning requires that there be no more OCGT capacity built until the corresponding call option has doubled in value.
- That is, the optimal centrally planned response would be to manage peak load growth by adopting more aggressive DSM measures, more often, across the periods when existing OCGTs are operating, and probably to have them operating over more periods, until optimal shadow prices (or perfectly competitive market prices) rise to the point where existing plant of that type is making twice the operating profit it was making under the status quo equilibrium.

---

128 We refer here to the "pattern" of prices, because the uncertainty introduced by natural elements such as hydrological fluctuation creates a whole probability distribution of prices, even for an otherwise specific period, like "Wednesday Morning 3am"

129 *Electricity Price Review: First Report,* Released by the New Zealand Ministry of Business Industry and Enterprise (MBIE), October 2018.

130 See discussion in our 2018 paper *An Economic Perspective on the New Zealand Electricity Market* Prepared by EGR Consulting ltd for a broad generator consortium, in response to MBIE's review paper.

131 Put another way, the economic contribution of that plant, as measured by the sum of optimal market clearing prices over all periods in the year when that plant should be operating because those prices exceed that plant's SRMC should equal the total cost of new capacity of that type, consisting of its annual capital cost contribution (and fixed overheads), plus its annual variable operating cost (i.e. units generated times per unit SRMC.)

- Those higher extreme peak prices would equally boost incentives to invest in other plant, further down the merit order, so the optimal plant mix would change to involve less OCGT capacity, and more DSM, but also more of every other type of plant capacity still available in the expansion menu.
- The increase in OCGT costs must optimally raise the LRMC of optimally meeting load, though, and also optimally increase reliance on DSM.
- So, the effect will clearly be greater if the OCGT is priced right off the market, or removed from the expansion menu entirely.

On its own, that last observation may be concerning, especially when we consider that essentially the same logic applies to all thermal capacity types, each of which has traditionally been thought of as capping the aggregate prices over the hours for which it can profitably operate, to its investment cost. So, the implications of simultaneously removing all thermal capacity options from the expansion menu, and hence from the optimal long run equilibrium plant mix has effectively already eliminated that traditional discipline on market prices.

Offsetting this, though, technological progress is making other options cheaper than they have been in the past, and the potential to expand cheaper options has exactly the opposite general effect. For example:

- Suppose that, instead of doubling the capital cost of OCGT, we left that alone, and halved the cost of new base-load plant generating at zero SRMC, with everything else exactly the same.
- Then, optimal expansion requires meeting load growth by, amongst other things, expanding this new cheap base-load capacity, until its call option value halves.
- The call option value for base-loaded plant with an SRMC of zero will just be the Time-Weighted Average Price (TWAP),[132] so that must eventually halve, and the Load-Weighted Average Price (LWAP) must fall, too, because the arrival of a new cheaper technology can only reduce costs for consumers.
- So, the perfectly competitive operating profits of all other existing plant could fall substantially, perhaps well below their historical or replacement costs, and there may be no more expansion of some of those plant types for some time.
- There will still be a requirement to meet all load levels, though. So, the existing plant, all the way up to the OCGT peaker discussed above, would still need to operate.
- Unless the capital cost of the new base-load technology actually falls below that of new/refurbished OCGT capacity (per reliable peak MW) that OCGT capacity would eventually have to be expanded, too, if load growth continues indefinitely.[133] So its call option value must eventually rise back up to the cost of OCGT expansion.
- But the optimal equilibrium plant mix resulting from the introduction of a new cheap base-load capacity will clearly involve more of that capacity, and proportionately less of every other plant type. And the path towards that optimal plant mix will probably involve allowing some existing pant to be retired earlier than might previously have been thought optimal, and not necessarily replaced.[134]

Each of the above discussions relates to one important factor in the current situation. We are simultaneously seeing the withdrawal of all thermal capacity types from the allowable expansion menu, and the addition of technologies with reducing cost profiles. The ultimate result, in terms of average price levels is hard to predict, but we certainly should expect future price profiles to be different. Specifically,

---

[132] Adjusted for its average availability.
[133] Or perhaps replaced, if existing plant wears out before expansion becomes necessary.
[134] No new capacity should be built that has a higher capital cost (per reliable peak MW) than the new base-load technology. But other plant types may be excluded from the optimal mix too, if their capital costs are not much less, and their operating costs significantly higher.

with the marginal expansion cost of "base-loaded" renewables like wind and geothermal setting TWAP, and thermal capacity expansion costs no longer capping price peaks, we should expect to see more periods with both very high and very low prices. At the same time, solar capacity expansion costs should limit summer day-time prices, but have much less impact on winter prices, and only indirect impact on night-time prices, via storage options, as discussed below. The overall effect may be to concentrate price spikes into evening peak periods, particularly in winter, and/or when wind generation happens to be low. And, if that pattern becomes predictable, it may actually be limited more by DSM responses, than by supply side options

The above discussion leaves it unclear whether we should expect to see the operating profits earned by existing thermal capacity increasing or decreasing over the transition period. Valuing existing hydro is even trickier, because "hydro" covers a wide range from extreme peaking capacity, all the way down to base-loaded capacity. And we have already suggested that the role of some hydro plant may need to change substantially. In fact, the value of hydro is driven by some factors not covered by the above discussion. Specifically, hydro systems include storage capacity that must surely become more valuable in the new environment, both because expansion of intermittent renewable capacity will be increasing volatility, and because withdrawal of thermal capacity is effectively removing much of the system's existing storage.

Pricing implications of expanding storage capacity

More generally the above discussion, and traditional theory, do not really cover the valuation of other storage options, such as batteries, either. The value of such options lies in their ability to arbitrage between low and high-priced periods. But, while the higher priced periods involved will obviously lie higher up the PDC than the lower-priced periods, we can not really analyse the value of any particular storage option without doing a more detailed "chronological" analysis. A facility that could only profit from one large arbitrage opportunity, from the lowest to the highest priced period in the PDC, would probably deliver much less value than one that could arbitrage between moderately low and moderately high-priced periods, every day of the year. So, we need to know basic things like how often prices alternate between high and low, in order to determine the number of storage cycles the proposed facility could actually make a profit from, each year.

Nonetheless, it should be clear that, other things being equal, expanding storage capacity will reduce the incidence of extreme prices, and increase the incidence of moderate prices, across the year, and hence across the PDC. So, the marginal cost of storage capacity will clearly be an important factor in determining the extent to which price volatility and cyclic inter-temporal variation actually do increase.

Ultimately, if storage capacity was perfectly efficient and absolutely free, the equilibrium PDC should be absolutely flat, at the LRMC of the cheapest expansion option, no matter how intermittent it might be. That is because, with zero capacity cost we could build an infinite battery, and so capture any generation from that base-load technology, at any time, in any hydrological year, and store it to be used at the (next) most valuable opportunity available. So, all inter-temporal price differences could be arbitraged away.

Battery capacity is far from free and, even if it were, it could only cap night/day price differences at its own charge/discharge cycle loss factor, as discussed in Section 4.3.2 above. Still, battery capacity may be added much more easily than hydro storage capacity, and the cost of that capacity will clearly limit the extent to which day/night price differentials, or short-term price volatility can increase, even if the aggregate capacity of smaller hydro storages proves unable to cope with this increasing challenge.

Theoretically, the cost of expanding pumped storage capacity should also limit day/might, summer/winter, and even wet/dry year price/MWV differences, too. Pumped storage is not cheap, though, and real facilities should always be sized so that limits on both pump/generation capacity, and storage capacity are binding from time to time, and probably quite often. So, it is the marginal economic opportunities that could be exploited by marginal expansion of each of those specific capacities that determine the marginal contribution of the facility, and should align with the marginal costs of expanding each type of capacity, under our simplistic assumption of continuous expansion at linear cost.

More exactly, if pumped storage facilities can operate across a combination of daily, seasonal and wet/dry cycles, their expansion costs should be seen as <u>jointly</u> limiting both summer/winter and day/night and wet/dry price differentials, with the same loss factor applying to both, but the capital cost somehow allocated between them. But the allocation of capacity costs between all of those modes must also account for the interaction between the economics of expanding and operating this pumped storage facility, and that of batteries and/or conventional hydro storage, performing similar roles in each of those various timeframes.

We will not pursue that discussion any further, because, while proceeding with the kind of large-scale developments currently being considered would clearly have a significant impact on the PDC, over many years, and a series of unique developments should optimally each have their own transient PDC impact, even in an centrally optimised or perfectly competitive environment, those deviations would not ultimately affect the long-run equilibrium PDC, which will always be driven by projected costs for generic entry options that are expected to be still available for future development, over the long-term planning horizon.

5. Optimisation Theory vs Market Reality

5.1. Introduction

Our discussion so far in this report has been based on standard optimization theory, as it was applied in the New Zealand Ministry of Energy, and as it might have been applied in an idealised centrally planned and operated environment, had the necessary computational capacity been available. We have argued, though, that the theory is equally applicable to a perfectly competitive market environment, and at times have used the terminology of markets interchangeably with that of optimization. Thus, we have talked about prices, and not always carefully distinguished between the prices that might be observed in a perfectly competitive market, and the shadow prices that might be reported by an optimization model, because we believe they are, in principle, identical. And we have treated generation in high-priced periods, as being equivalent to delivering high value to the national system.

The reason for developing our discussion in this way is that we believe a great many of the features that some may think have been recent by-products of the market era, and perhaps antithetical to the spirit of maximising national benefit, were, or should be, features of any centrally optimized regime given the same task of providing electricity to New Zealanders efficiently and effectively. For example, entry limit pricing is often described as a feature of the competitive market, and perhaps seen as an example of participants manipulating prices to achieve their own ends, in ways that some might consider undesirable. But we have seen that a centrally optimized regime should, and did, have its own version of that same mechanism.

The same can definitely be said of the concept of opportunity costing itself. Some critics may believe that hydropower should be offered to the market at what they believe to be its short run marginal cost of zero. But we have seen that opportunity costing is in fact integral to the way in which a centralized optimization would manage this resource, given the goal of maximising the national benefit delivered by electricity sector operations, and it thus should logically determine the prices set for consumers, too.

We are well aware, though, that the real market we are dealing with exhibits behaviour that does not entirely align with the theory discussed. Part of this discrepancy doubtless results from the much discussed "exercise of market power". But there are several other factors that should be noted before moving on to that discussion.

The limited goals of the section should be noted, though. We are not attempting to analyse market outcomes or behaviour, or to argue about whether they are good or bad, efficient or inefficient. We are not in a position to say what motivates particular participants to offer in particular ways, at particular times. The issue here is just whether the current market could plausibly continue to operate in much the same way as it does now, in the emerging environment. Ultimately, perhaps the question should be whether there is something about the discrepancies between real market behaviour and idealised theory that provides strong reasons to believe that the kind of behaviour observed today can not be taken as a guide to the kind of behaviour expected in the future. As a first step, though, we mainly focus on how and why observed outcomes may differ from theoretical predictions. Accordingly, we discuss:

- First, the validity of some of perfectly competitive market assumptions about the ability of participants to understand, analyse, and interact with the market situation when developing insights from the ideal;
- Second, the way in which opportunity costing is actually performed in the market, and the (surprisingly limited) impact we expect as a result of imperfections in that process;
- Then, the possible distortions involved in the offer formation process itself, including the contentious relationship between risk management and market power; and
- Finally, the implications of entry limit pricing, as it applies in the real market context.

5.2. Information processing and transaction costs

Discussions of market imperfections often focus on "lack of competition", but there are other important assumptions in our perfectly competitive market model that are really not applicable in the real world.

Imperfect information

We have assumed, for example, that the participants all have access to the same information set that an ideal central planner would have. But that will not be the case. To give one example Section 3.2.1 argued that the marginal value of water stored in one participant's reservoir should actually be seen as a function, not just of that reservoir's storage position, but that of all storages in the system, including the physical/contractual status of thermal fuel stocks. Only some of these "stock levels" are easily observed, though, and a hydro reservoir manager will presumably be inferring the status of thermal stocks from the offers made to the market.

The "imperfection" of each agent's information set does not necessarily imply that market outcomes are "inferior". The combined information set available to all participants will be greater than that available to a single central planner, and the information available to that planner may be simply wrong, in important respects. So, there is a broad argument in favour of outcomes generated by the interaction of multiple imperfectly informed participants being "superior" to, or at least more robust than, those produced by a central planner. They will be different, though.

Bounded rationality

Perhaps more importantly, we have also implicitly assumed that each agent has the same analytical capabilities, and is perhaps just as dedicated to pursuing optimization to its ultimate limit, as the central planner. In reality, a comprehensive integrated stochastic optimization planning/operational model of the whole system was not historically available to any central planner and, even today, the best models available to individual participants still struggle to meet that description. As discussed in Section 5.3.2, individual participants may have good models of their own systems, particularly for reservoir management, but with limited representation of any other system.

More generally, managers simply do not find it worthwhile to pursue optimisation to its theoretical limit. Most often, they "satisfice" to produce acceptably good decisions, with a reasonable expenditure of effort. Management practices are established as much for their convenience and familiarity, as for any claim to optimality. Individual managers develop ways of thinking, that may be unique to them, or may be well established business practices, but neither of those necessarily align with optimal economic theory. For example, we have suggested that historic accounting costs should theoretically be irrelevant when it comes to future decision making, but we are well aware that most businesses actually do use accounting costs, and markups on accounting costs, on a routine basis. Presumably, managers are familiar with these concepts, with performance possibly assessed against accounting metrics, have long experience of making good profits in that way, and see no reason to change.

As a result, individual decisions will be made that, from the perspective of an analytically minded observer, may not seem optimal, or even rational. But market outcomes depend on the aggregate interactions between a great many decisions made by diverse agents, and tend to be much more predictable than any particular decision might be. So, our assertions regarding the expected optimal behaviour of prices, marginal water values, and so on should be taken as expressing an expectation of broad alignment, on average, over the long term, with those aggregate market outcomes. We do not expect the situation to be materially different in the new environment, though.

Transaction costs

The above discussion relates to our assumption that transaction cost could be ignored. In reality that has never been the case. Transaction costs can be broadly defined to include all of the organisational, informational, social, personal, political and practical costs that limit society's ability to complete all transactions that might otherwise have been considered optimal by a perfectly informed rational decision-maker, intent on maximising national welfare. And, without detailing the precise reasons, it has been clearly documented that decision-making in the pre-market era fell well below that high standard. In other words, the centrally optimised ideal we have been using as our theoretical benchmark was no more real than the analogous perfectly competitive market ideal.

Transaction costs will always limit the extent to which any real-world market approaches that ideal, and the NZEM is no exception. We have already discussed the limitations implied by the practical barriers that keep participants from performing perfectly accurate MWV calculations. But two other limitations are worth noting here.

Transaction costs for DSM

One major concern relates to overcoming barriers that might limit the potential for DSM to partially replace thermal as a mechanism for managing potential supply shortfalls, and setting a benchmark against which stored water can be valued, as discussed in Section 0.

In the distant past, the New Zealand Government was in a position to impose operating practices, such as ripple control which deferred water heating loads from peak to off-peak periods and, when necessary, and impose measures to reduce electricity demand to a level which the power system could supply. These non-supply options were represented in optimization models as virtual stations, with SRMC prices set at estimated non-supply cost levels. The true cost of those measures was never really known, though, and our own experience was that changing that one uncertain parameter made a very large observable difference to the level of guidelines. So, the forbidden zone concept was often preferred to drawing guidelines drawn as contours of the EMVW surface matching such uncertain non-supply costs.

When the market was designed, it was hoped, and more-or-less expected that there would be quite rapid developments toward more flexible customer-focussed demand side arrangements. It was perhaps naively believed that a "demand curve" could be represented by actual bids in the market, and that brokers and aggregators would emerge to coordinate demand side management across multiple parties who had freely contracted to have their demand managed in this way, at some agreed price. So, it was expected that there would be significant consumer-driven price signals somehow appearing in the market, by now, and most likely in the spot market.

In reality, very short-term demand side responses were encouraged, and incorporated into the ancillary services market from the start, and demand side bidding by major consumers is now possible. But the general run of consumer demand remains inelastic, because most smaller consumers see no dynamic price signal to respond to. This may not imply too much loss of economic efficiency, if consumers actually do respond to typical cyclic price patterns by enrolling in pricing plans that provide day/night price signals and/or control particular loads. And the net effect of those arrangements must be reflected in load in forecasts. But, we believe that there is still considerable work to be done in reducing transaction costs if the opportunities discussed in Section 0 are to be systematically exploited, adequately represented in the market clearing process, and provide the kind of benchmark role previously given to thermals in setting MWVs.

Market design and offer form

Another aspect of this situation which is seldom considered, is the implicit assumption, in the theory of perfectly competitive markets, that mechanisms will be found to allow participants to interact optimally. But the optimal marginal cost structure emerging from some of our discussion is complex, and the New Zealand electricity market does not actually allow participants to present offers that embody all of that complexity. In particular, we have discussed the idea that a hydro station with a small reservoir or head pond, may have run-of-river flows to allow a base level of generation all day, but only enough stored energy to support extra generation for a few hours, in any particular day, say 1 for simplicity.

That situation would ideally be represented by the generator presenting an offer to the market for some of its capacity, at a price based on its MCR (which may be low in this situation), but with an explicit constraint limiting the number of hours to which the extra capacity could generate. Given an offer expressed in this way, an inter-temporal optimization would add its own optimal shadow price onto the participant's offer price, calculated to ensure that the capacity is actually called on for 1 hour, no less and no more.

The WEMS design process involved testing a more complex form of inter-temporal offer, and market-clearing optimisation, which would have directly incorporated the kind of river chain optimisation discussed in Section 3.3 into the market- clearing process. Instead, it was decided that participants should be expected to offer a strip of individual single period offers, each of which would have to include a section at a price set by the participant at a level which they expect to be only called for 1 hour a day. Then, if the offer is called in one hour, the participant will need to rapidly withdraw the corresponding offers for the rest of the day, or at least place such a high price on them that they will not be called again. The actual efficiency loss due to this particular market imperfection has not been seriously studied, to our knowledge. But it does not surprise us to see real market offers showing this kind of high priced "reserve capacity" offered right across the day. What cannot be deduced from those offers, though, is how tightly they might control the probability of being called to match the energy-limited capacity available, or how the participant might plan to adjust offers, once that energy-limited capacity has been called.

5.3. Opportunity costing and homeostatic equilibrium

5.3.1. Thermal opportunity costing

Section □o discuss opportunity costing of thermal fuels. While we believe the situation faced by thermal generators calls for some such practice, whether explicit or implicit, we are not in a position to say how closely internal practices might, or might not, match our theoretical description. It is safe to say, though, that, while we might expect significant volatility over the next few years, as thermal generators and thermal fuel supply systems are both wound down or re-purposed, the end result of that process will render the topic of thermal opportunity costing increasingly irrelevant.

5.3.2. Hydro opportunity costing

The central focus of our discussion is the opportunity costing of water stored in hydro reservoirs. In asserting the applicability of results from optimisation theory to an idealised perfectly competitive market in the previous sections, we implicitly assumed that participants in that market would have a very sophisticated understanding of some of the theory we have discussed, and be willing and able to construct complex models, not just of their own facilities, but of their competitors and even potential entrants. As noted above, that will not really happen, and even highly motivated managers of smaller systems will be simply unable to develop or employ the expertise to perform that kind of analysis well.

Instead, they may use simpler models, if they use any models at all.[135] If a manager just models their own reservoir's situation without any direct reference to other stock levels, the result will be a single dimensional marginal water value curve, for each period of the year, rather than the multi-dimensional MWV surfaces discussed in Section 3.2.1. A better model could involve a reservoir representing their own storage, plus another representing the remaining aggregate national storage. But a smaller reservoir might be better off focusing on the storage situation in the national storage system, and then setting its own MWV higher or lower, depending on the state of its storage, relative to the national average. Many may just use heuristics, rules of thumb, or experienced intuition, though.

There are more sophisticated models available, though, for optimisation of larger storage over an annual timeframe, and we expect that many managers of larger systems will be using them, or at least indirectly

[135] The discussion here focuses on optimisation over an annual horizon, that being the context in which the MWV concept is most often discussed. Stochasticity is less of an issue for intra-day river chain management so, historically, deterministic LP models were developed for each river system in New Zealand. But, while every solution of such a model implies a corresponding set of MWVs, and implied MCRs, the traditional focus of such models was more on the "primal" recommended release schedule, than on these "dual" prices. Optimisation of river chain management has become a very private affair, though, and we have not seen public discussion of the models actually used for that purpose, if any, for some time.

referring to their results to build up their own understanding of the national storage situation, and their best response to it. We referred to a number of those models in Section 1.2 and noted that they all, in one way or another implemented the basic opportunity costing theory described here. There has been ongoing debate, though, between the developers and proponents of those models as to which produces the best MWV estimates. We consider the diversity of viewpoints to be healthy, but it inevitably raises the question of what causes each model to be "biased" in particular ways, whether any of them is actually "correct".

We are certainly not in a position to comment on the "correctness" of any model here, but simply point out that all of these models need to incorporate a great many approximations in order to be computationally tractable. And our experience is that seemingly innocuous differences, like the frequency with which future managers are implicitly assumed to review decisions, can make surprisingly large differences in the results.[136] So, models are typically calibrated against some external simulation and/or market outcomes, and tuned to produce "plausible" results. In fact, some models simply develop a heuristic MWV surface, or guideline set, informed by an understanding of the theory, and calibrated against market results.[137] Accordingly, we think it pertinent to ask how much economic damage might be done, if the "wrong" MWV recommendation were to be accepted by one or more reservoir managers.

5.3.3. Homeostatic equilibrium

Section 3.2.1 has already discussed the interaction between two hypothetical reservoir managers: One managing an aggregate South Island reservoir, and the other an aggregate North Island reservoir. We discussed how arbitrage through the market would naturally act to bring those reservoirs into balance, in the sense that their MWVs were matched as closely as possible. As noted there:

> ".. if the North Island MWV is higher than the South Island MWV (after loss adjustment) then North Island storage must be too low, relative to South Island storage. But optimal arbitrage will favour South Island hydro generation over North Island hydro generation, thus tending to reduce South Island storage while allowing North Island storage to rise, until they eventually reach equilibrium, with EMWVs equal (+/- marginal losses). "

Now we should ask, what if the North Island MWV was only higher than the South Island MWV because the North Island manager used a model that tended to produce higher MWV estimates than the model used by the South Island manager? Clearly the same equilibration process will occur, and a balance will still be found. While randomly imbalanced inflows will constantly tend to force storage levels away from the balanced zone, market forces will constantly work to restore the equilibrium, as described above.

The balance found and maintained by the market will be different, though, than it would have been in the case where both managers used the same model. Because the North Island manager is consistently choosing higher imaginal water values, this balance will tend to hold more storage in the North Island, and correspondingly less in the South Island. As a result, we expect the North Island reservoir will run out less often, but also spill more often. And the opposite will be true in the South Island. The North Island manager may be very happy with that result. Or, if the North Island manager considers that the policy is resulting in too much spill, they will presumably tune their model to produce outcomes that tend to hold more water in the South Island, and less in the North.

Either way, the system will not collapse, and the two reservoirs will be managed effectively, with each manager achieving their goals, without having to resolve any debate over whose model is actually "correct". The net effect on the national storage position will be to hold somewhat more storage than the South Island

---

136 See E.G. Read & J.F. Boshier: "Biases in Stochastic Reservoir Scheduling Models", in A.O. Esogbue (ed.) *Dynamic Programming for Optimal Water Resources System Management*, *Prentice* Hall NY, 1989, p.386-398

137 For example, see J. Tipping & E. G. Read "Hybrid bottom-up/top-down modelling of prices in hydro-dominated power markets" in S. Rebennback, P.M. Pardalos, M.V.F. Pereira & N.A. Iliadis (eds) *Handbook on Power Systems Optimisation* springer, 2010, Vol II, p213-238 ,or the xxxwhat is JoC's model called ?????x model employed by MDAG for this study.

manager might have done, and somewhat more than the North Island might have done. In other words, the market solution is more moderate, and robustly maintains that natural balance over time.

Essentially the same thing will happen if there is only one reservoir, and one manager. This is because the reservoir system is fundamentally a natural system, which obeys a fundamental self-regulating (or "homeostatic") equilibrium relationship: Namely that outflows equal inflows, over time. If our reservoir manager adopts a new model that tends to estimate higher MWVs, that model will immediately start recommending offers that result in releasing less water. So, the storage level will rise until a new equilibrium is found, obeying the same basic relationship, that inflow must equal outflow, over time.

As discussed previously, the equilibrium formed will be a "stochastic equilibrium", with a different trajectory for each hydrology year, but cycling around to maintain generally higher storage levels than those maintained using the previous lower MWV surface. And we should expect higher spill probabilities, and lower shortage probabilities than were experienced in the previous equilibrium. Whether this solution is considered better or worse, will largely depend on the relative weighting society puts on the non-supply costs incurred during shortage events, versus the loss of energy implied by spill.

But how will the prices experienced by consumers compare? For simplicity, let's assume that the marginal water value feeds through directly to the market price. To understand the effect, we need to consider another basic equality in the electricity sector, which is that the sum of hydro and other generation must cover demand, minus a quite small volumes of "non-supply".

Since the release from this reservoir is the same, on average, as it was in the previous solution, its annual generation will be much the same too, apart from some extra energy lost due to increased spill.[138] In a traditional power system, the "other generation" is largely thermal, so the thermal generation level must be much the same, too, apart from a small adjustment upwards to generate the increased energy required to cover the extra spill, and the increased energy now supplied, but previously not supplied. So, this more secure power supply does cost something, but not necessarily very much, while the expected non-supply costs obviously fall.

If the marginal value of water is still being kept as constant as possible, the marginal cost of thermal generation must be, too. And since the level of thermal generation is very much the same, the marginal cost should also be very much the same. So, market prices should be very much the same when load is actually being met by hydro and thermal generation, and much lower in the periods for which non-supply prices no longer apply. If the MWV curves are too high, the extra spill will lose enough energy to push total costs up, and average prices with them. If the MWV curves are too low, though, the extra non-supply will cost enough to also push total costs up, and average prices too.[139] In between, the optimum is typically very flat, with very different looking MWV curves implying very similar simulated average costs and prices. [140]

The situation will be somewhat different in the emerging market though, with no thermal generation to call on. The trade-off then will be directly between volumes spilled by hydro and/or other renewable generation) and volumes "not supplied", as a result of some level of DSM activity. And (ignoring potential green thermal) the marginal water value will be set entirely from the marginal costs of those DSM options. The average volume of spill is likely to much higher, and so will the average volume of DSM, with the two now more intimately connected because (with no thermal) energy lost as spill may have to be made up by DSM at a later date. Thus, total cost may turn out to be more sensitive to differences in EMWV estimation methodology, in the new environment.

The same basic equilibration will occur, though, and we still expect average prices to change much less than the apparent change due to raising marginal water value curves across the board. And, as above, the change will not necessarily be in the same direction as that shift, either. Thus, setting MWVs too low could now much more easily set up a situation in which deep supply cuts had to be maintained for extended periods, leading to very high average prices.

---

[138] +/- an adjustment for any difference in average energy efficiency, which could go either way, depending on which model is better at maintaining hydro generation at a constant rate, over a year.

[139] In the limit, storages will always be empty, and hydro will have to operate in pure run-of-river mode.

[140] Hence, in part, the wide range of MWV curves that may all be considered "optimal" by one modeller or another.

Finally, we note that these conclusions align with the results of the simulation experiments run many years ago with the NZED STAGE model, demonstrating that very different looking MWV surfaces could deliver nearly identical simulated costs.[141] It also aligns with the conclusion of ECNZ, after the 1992 power crisis, that shifting MWV curves up could allow its reservoir operating policy to meet a 1:60 "Design Dry Year" criterion, rather than 1:20, without any appreciable cost increase.[142] Consequently, we conclude that, while marginal water value curves should be expected to move upward, as thermal capacity is removed from the system, that shift does not, of itself, indicate whether prices will increase or decrease. The overall effect on operational costs (including non-supply costs) and SRMC prices can really only be assessed by simulating market outcomes on one set of curves, versus another, and studying the total non-supply costs, plus thermal costs (for as long as they are relevant). But the long-run equilibrium outcome must still be driven by the entry mix, with price patterns ultimately driven by their entry costs, as discussed in Section 5.5 below.

5.4. Offer formation

Having determined marginal water values by opportunity costing any stored water, market participants must form offers. We've already noted that these offer curves (which must be increasing functions of generation level), are not the same as marginal water value curves (which are decreasing functions of storage). In fact, they are driven by the marginal value of release (MCR) which only equals the marginal value of stored water (MVS) under certain conditions. In particular, the marginal value of release will be (close to) zero, for any water that must be released in order to meet flow requirements, or to avoid violating storage limits, even though the MVS for water actually stored, or storable, in the reservoir may be quite high at the same time.

In Section 3.3 we discussed why the complexities of downstream river chain management mean that the marginal supply cost from any particular station on that chain may become quite dis-connected from the MVS of water stored in the top reservoir in the chain, or its MCR, particularly when tributary and/or release flows are high. And that would be equally true in a perfectly competitive environment, or a centralised optimization, where the downstream constraints could be reflected directly. But we have noted that the NZEM market design deviates from the optimisation analogue by requiring hydro generators to supply strips of single period offers. And the combined effect is that the best available offer curves may be steeper than is commonly thought, even in a perfectly competitive version of the NZEM.

Still, observation suggests that real market offer curves are often steeper than might be expected, just as a result of those factors alone. So, we now turn to consider what else is likely to drive offers to differ from those implied by the optimization theory described initially, and to ask whether those same factors are likely to continue exerting a similar influence on market outcomes as the plant mix changes.

First, we have already discussed the realities of "bounded rationality", and should never discount the likelihood that one of the major reasons why observed offers do not match what theory might recommend is simply that those making the offers are applying their own heuristic approaches, not necessarily driven by precise application of the theory we have discussed. Likewise, offers that seem inexplicable will often have been based on speculations or expectations that turn out to be false, or seem false to the observer. But there are also two key assumptions from our discussion of perfectly competitive markets, that need to be relaxed in order to understand real market behaviour, namely:

- The assumption of risk neutrality, and
- The assumption that market participants do not believe they can exercise "market power" by manipulating offers so as to profitably shift market prices either up, or down.

---

141 As belatedly reported by Read and Boshier [1989], above.
142 As recommended on p60 of: Davidson et al *Report of the Electricity Shortage Review Committee*, 1992

5.4.1. Risk management

Risk management is not a new concept, or unique to markets. Whereas government planners may consider national risk, though, market agents will be solely focused on their own risk, both physical and financial.

Physical risk

By physical risk we mean the risk of facing regulatory intervention, prosecution, social disapproval, or potentially destructive or dangerous events, if physical variables, such as flow/storage levels or rates of change, were to exceed limits that are physically safe, or set down by regulation, permits etc. The higher the potential penalties the more cautious we should expect management to be:

- That caution will be partly expressed by managers tightening up physical river chain management, in particular, by restricting variables to lie further away from bounds, thus increasing the likelihood that the kind of constraints discussed in Section 3.3 will come into play, and implying the likelihood that individual station offers and aggregate river chain offers will be steeper.
- But we might also expect hydro operators to more directly target physical release levels they believe to be safe, and desirable. And that can be done by setting very low, maybe zero offers up to the target dispatch zone, and very high offers for dispatch levels above that zone. In other words, by creating a steeper offer curve which, in the limit, could even be vertical at a target dispatch level.

That is understandable, and perhaps normal in other sectors with a strong focus on physical compliance, and targeted delivery. And this type of offering strategy could become more attractive to potentially flexible hydro operators if, with the loss of thermal support, hydro river chains are increasingly called upon to flex physical flow/storage levels in response to increasing wind/solar volatility, as in Section 4.2.2. On the other hand, flexible support for those highly variable technologies will become increasingly important. So, this may well become a contentious issue, in due time.

Financial risk

By financial risk, we mean the more frequently discussed risk of being "caught short", and having to pay a high price in the spot market to (implicitly) buy in power to meet contractual commitments to loads, or implied by hedges etc. [143] Or, on the other hand, of receiving a very low price for extra generation, above that contracted level:

- It is well known that the 'safest' policy here is to sell into the spot market exactly the quantity one expects to have to pay for, in meeting those commitments, in which case, the payments balance perfectly, and there is no risk.[144]
- And that goal could be achieved by submitting an offer curve with a vertical step, from a price of zero just below the target quantity, to a very high price just above it, or a set of offers in that form, from various stations that meet the target, in aggregate.
- Such an offer could be seen as equivalent to the normal business practice of a firm, having made arrangements to physically meet their own contractual commitments, not seeing any need, obligation or profit in offering flexibility to support their competitors.
- Thankfully, electricity sector offers do not normally take this extreme form, and we may speculate that this is partly because all participants realise that, if no-one offers any flexibility, then no-one will have any flexibility to call on, when they can not actually meet their own commitments.

[143] Load supply commitment may be physical, but the sector arrangements ensure that such physical commitments are met, by power from the spot market. The seller's exposure is financial, though, because that spot market power needs to be paid for.

[144] Ignoring price differentials due to transmission losses and congestion, for simplicity.

- And that is a very real prosect because load commitments do not depend on hydrology, whereas hydro generation capability does. So, if a hydro generator takes on commitments roughly matching its expected generation capability, it will generally find itself with either a surplus to dispose of, or a deficit to buy in.
- In other words, it will regularly need or want to generate at levels above or below its contractual commitment, and so it will need to open up a range of offer steps that it thinks will see it dispatched at a higher level, and effectively selling its surplus at a price at or above its calculated marginal cost of doing so, or at a lower level, and effectively buying its deficit at a price at or below its own calculated marginal cost of generation.
- In doing so, it is exposing itself to some risk, though, and the more risk averse it is, the more concerned it may be about being caught with production too far from its contract level, and the steeper its offer curve is likely to be.

Storage adequacy risk

There is also a longer-term aspect to this risk averse strategy. A risk neutral centralised optimisation would theoretically have no particular concerns about which reservoirs national stocks were held in, provided its internal simulations indicate those stocks will not be trapped in an appropriate location, by physical constraints on transmission etc. But a generator that relies on stored water to support its ability to meet future contractual commitments will want to manage that storage carefully, so as not to get caught short, or have to dump a surplus, at some future date. In doing so, we know from Section 3.2.1 that it should really also be taking careful note of the state of its competitors' storage levels:

- If others in the system all look to have surplus stocks, our generator can relax somewhat, expecting that there will be plenty of cheap power available on the spot market, if it might need to buy some in to meet its own commitments, when the time comes.
- Or, if all the others seem to be running short, it knows to put a high priority on conserving its own stock, with the goal of at least being able to cover its own contracts, and maybe selling some spare at a high future price.

This is all as it should be, and exactly the way the market is supposed to work in coordinating national storage across diverse participants, without imposing a centralised optimisation. But that theoretically optimal coordination process may be significantly impacted by risk aversion, in several related ways:

- First, our hydro generator can not directly observe all of the information required to make a confident judgment about the situation. The state of a competitor's hydro storage may be readily observable, but what about their own contractual commitments? And what about the state of thermal fuel stocks, and/or supplies contracted for delivery? What if the competitor's plant were to break down? And so on.
- And can it be absolutely sure the transmission system will be available to transfer generation from all of those reservoirs to its own customers locations, as and when required?
- We should expect any manager to be cautious about relying on a surplus about which very little definite can actually be said, stored in a facility managed and controlled by another party. Worse, the parties controlling those other stocks, and providing ancillary services to support inter-island transmission, are all competitors, and they could take advantage of our generator, if it ever were to end up in situation of having to depend on their stocks.
- Thus, even a mildly risk-averse manager will have strong incentives to focus on managing its own stock carefully, to meet its own commitments. rather than perhaps naïvely relying on an assumed ability to draw freely on the whole diverse national portfolio of storage options.
- And that may lead them to a policy of targeting storage levels, rather than openly offering whatever might be available on the dubious assumption that such openness will always be reciprocated and rewarded at any time of need.

- In other words, they will have incentives to submit relatively steep offer curves, focused on managing their own storage levels, rather than allowing the market to fully determine their storage management policy, as would be assumed in the perfectly competitive model.

Our goal, here, is not to assess the advantages and disadvantages of centralised optimisation, versus the more cautious, and perhaps more physically oriented, reservoir management style described above, but just to recognise that the pressures alluded to are likely to shape offers, and outcomes, in ways that deviate somewhat from the idealised theory discussed earlier. The issue is, though, whether we can expect those pressures, and that management style, to persist in the emerging market environment. And our judgment would be that, with thermal storage options gone, the remaining storage managers might tend to focus even more on maintaining physical control over reservoir levels, as well as over river chain flows as discussed above. And that may also become contentious, in a context where higher level policy goals are likely to favour providing a supportive environment for intermittent renewables.

5.4.2. Market power

Astute readers will recognise that the discussion on "financial risk" in the previous section almost completely mirrors a standard exposition of incentives to exercise market power. The only difference is that, as market forces move a participant's generation further away from their contract position, a large generator may realise that that they actually can exercise more control over their situation, and enhance profits, or reduce losses, by influencing the price they are being paid for their own surplus, or paying to buy power in. Specifically:

- A generator that would, in a perfectly competitive market, be generating at a level above their contractual commitments, may prefer to limit their excess generation, so as to get a higher price for the extra they do sell.
- Conversely, a generator that would, in a perfectly competitive market, be generating at a level below their contractual commitments, and so buying the balance, may prefer to generate more than the perfectly competitive ideal, and buy in less, so as to get a lower price for the deficit they do buy.[145]

This implies incentives for larger generators to move offer curves up, for quantities above their contract positions, and down for quantities below their contract positions, thus forming a steeper curve than mere risk aversion would imply, pivoted around the contract position, with an initial segment possibly priced at zero. The net impact on market efficiency and prices depends very much on the contract position, though. There should be no distortion at all, if the perfectly competitive generation happens to equal the contracted position, and very little, if it is close. But distortion can be significant if the discrepancy between perfectly competitive output and contractual commitments is large.

This kind of "gaming" will reduce the operational efficiency of the sector, but any assessment of its overall economic impact would need to analyse a much wider range of issues, including the implied incentives for vertical integration, in the form of business arrangements or contract purchases, and on entry incentives, as discussed in the next section.

The relevant issue for this discussion is mainly that, for large hydro generators, this effect should be expected to reinforce the incentives discussed above, to keep generation levels somewhat closer to contract levels than an idealised optimisation would recommend, and they can be expected to do that by making their offer curves steeper than perfectly competitive market theory would suggest. Whether that is considered to be undesirable, or not, the question is whether we should expect that behaviour to change in

---

[145] Most discussions focus on market power being exercised, or possibly “abused” to raise price, to the detriment of consumers. But, in principle, the situation is symmetrical, and the exercise of market power can also be about pushing the prices paid to suppliers down. And that will be to the advantage of all net buyers from the market, at that time. There is an asymmetry, though, in that electricity prices can be pushed much higher in times of relative shortage, than they can be pushed down in times of relative surplus.

the emerging market environment. And the answer will depend on the interplay of factors outside our current scope.

Naively, it might be thought that, in a situation of increasing price volatility, major generators would be forced to accept that they can not push prices any lower in periods where the perfectly competitive price is already zero, but might be keen to offset this by exploiting the opportunity to push prices even higher in non-supply situations. But they would surely realise that pursuing that kind of strategy aggressively would invite an ultimately detrimental regulatory response. In any case, their theoretical incentives would depend on their contractual position at that time.

Traditionally, supply crises have been associated with high load and low hydro availability, suggesting that a major hydro generator is actually likely to be over-contracted at such times, with strong incentives to push prices down, not up. That may be less the case in future, though, if supply crises become more associated with periods of low wind and solar generation. So, the situation could be that of an under-contracted hydro generator being asked to provide support for competing intermittent renewables.

It will be important to analyse this issue carefully, because both the ideal perfectly competitive PDC, and the circumstances under which market power could be exercised to distort that PDC, are likely to be quite different from the status quo. The incentives at play in the new environment will depend, though, on how one expects intermittent renewable entry to occur:

- If these new facilities are owned by independent investors reliant on spot market sales, large hydro generators may be over-contracted in this situation, too, and striving to keep prices down.
- But if the new facilities are themselves heavily contracted (including as self-suppliers), large generators may well be under-contracted in this situation, and wanting to get a good price for the support they are providing.

In the next section, though, we will turn to consider a rather different kind of gaming incentive, driven by perceived entry costs.

5.5. Entry limit pricing

Potential SRMC/ LRMC discrepancies

Sections 3.4 and 4.4 both discussed the theory of optimal expansion planning, in a centrally optimised or perfectly competitive environment, under the assumption that generation capacity could be continuously expanded, as required to meet load growth, at linear cost. We noted, though, that it never has been possible to expand generation capacity in a continuous fashion at linear cost, or easy to vary construction rates dynamically to keep pace with varying demand growth rate.[146] So, we should expect to see significant deviations both above and below the LRMC driven equilibrium PDC.

The potential impact of lumpy thermal investment seems obvious but, historically, there have been far greater problems with major hydro developments, because they have typically involved developing several stations and associated infrastructure, as part of a decades-long catchment development plan. Such development plans follow their own practical logic, which does not necessarily involve building the most economic stations first, and have historically been treated in a very political fashion. This has made plans very difficult to speed up or slow down, once started. Thus, there have been long periods where capacity development has fallen behind load growth or got well ahead of it. We have seen much less evidence of this kind of discrepancy in the market era, partly because no major catchment-scale hydro developments have actually proceeded.

[146] The "lumpiness" of generation units also means that operating costs are not exactly convex, either. But the discrepancy is much smaller in New Zealand than in most other markets, because hydro stations typically have several relatively small units that are easy to start and stop, and can be operated in a coordinated fashion to achieve a fairly smooth, and nearly convex aggregate output function. The issue is more significant for thermal generators, but we will ignore it because the distortions involved will gradually reduce, as those units are retired.

Significant volatility in the supply/demand balance has been observed recently, and should be expected over the next few years, given the potential for both major electricity-intensive loads and existing thermal generation to shut down suddenly. There could also be a major impact if one particular major pumped storage project currently being investigated, were to proceed. Apart from that, though, the future prospects for major hydro generation developments seem limited, while conventional thermal developments have been ruled out completely. Thus, there is probably little point exploring the potential discrepancies between real hydro project evaluation practices and the theoretical discussion in Sections 3.4 and 4.4.

Instead, we re-iterate the caution expressed there: That, unlike with some other technologies, we are not expecting a steady stream of very similar developments, all reinforcing the same predictable impact on the long-run equilibrium PDC. If potentially viable pumped storage projects do emerge, each would have its own unique impact, at least in theory. But we doubt that incumbents will find it worthwhile to analyse or specifically respond to, the potential impact of small proposals, while their responses to larger threats might focus more on lobbying than on offering strategies deliberately designed to influence the PDC shape in ways that make entry by a specific hydro project less attrative.

Conversely, though, newer technologies such as solar and wind, are much easier to develop in a fairly continuous fashion, and at a rate that can be varied to match load growth. So, in this respect, future market conditions should more nearly approximate the assumptions of our idealised model, as already discussed in Section 4.2.

Possible smoothing of LRMC/SRMC discrepancies

It is hard to get a clear picture of the situation because optimal SRMC price levels should vary greatly, from year to year, due to hydrological variation alone. But our (admittedly limited) analysis of the data presented in the report of MBIE's 2018 pricing review suggested that market PDCs over those years actually conformed remarkably well to theoretical predictions.[147] If anything, the issue we saw was that the PDCs matched the theory so well that it raised questions about the degree to which market power may have been exercised to achieve that outcome. If so, the incentives would not be directly related to contract positions, and thus quite different from those discussed in the previous section. Indeed, the two motivations are likely to be in conflict, at times.

It seems like a relatively complex analysis might be required to fully optimise a strategy, designed to influence a potential entrant's assessment of profitability across the full range of hydrologies. But very similar behaviour might actually be driven by a simple desire to give shareholders a stable rate of return, rather than just accepting the implications of strictly competitive responses to random volatility in the supply/demand balance. And it seems reasonable to expect that those participants who understand the issues, and believe they can influence prices in some relevant periods ,will probably exercise some market power to craft offers to move prices toward that level, irrespective of the optimal perfectly competitive MVS in any year.

Importantly, Chapter 4 argues that we should now expect prices to trend towards a very different pattern, corresponding to the very different optimal entry mix shaping the new environment. In particular, we should expect to start seeing more extreme prices, with stronger cyclic summer/winter variation, more low-priced periods, and stronger winter peak price spikes under certain conditions.

We have not discussed the speed with which that transition might happen, but, while existing thermal plant can be expected to still operate in ways that partially reflect the old realities, the entry mix supposedly shaping past PDCs has already been ruled out of contention, so the transition may already be underway. But it will take some time for the industry price distribution to stabilise around the significantly different long run equilibrium price pattern which Chapter 4 argues can be expected as a result of the radically different optimal entry mix in the new environment.

[147] See Appendix C of Read [2018] cited above.

Possible barriers to development of extreme peaking capacity

One way in which market outcomes may not be matching the basic theory discussed in Section 3.4, though, is in the possible under-provision of extreme peaking capacity. Capacity of that type would theoretically be reliant solely on income received during extreme price events, which are also likely to attract a threat of regulatory intervention. We have argued previously that such a threat should be expected to inhibit an optimal level of investment in peaking capacity.[148] And the price distributions observed in the MBIE review seemed consistent with that theory, too.

At one level the incentives to build new OCGT peaking capacity seem irrelevant, if new oil or gas fired capacity has been ruled out of contention. But the same issues would apply to new hydrogen or bio-Diesel fired OCGT capacity with even greater forces because these developments would also face a range of additional risks. The possibility that measures could be taken to reduce the risk of such investments lies beyond our scope, but it should be recognised that any assessment of the need for such measures is likely to depend significantly on assessment of the incentives operating on vertically integrated gentailers, rather than the stand-alone investors implicitly assumed in the simplistic discussion above.

[148] E.G. Read, M. Thomas & D. Chattopadhyay "The Impact of Risk on Capacity Investment in Electricity Markets" Keynote Address, *IAEE Proceedings*, Wellington, 2007