

# *Economic Behaviour in a Hydro-Dominated Electricity Market*

*Funded by*

**Mighty River Power**

*To inform Public Discussion*

*by*

**E Grant Read**

*2 March 2009*

## ***DISCLAIMER***

*This document is supplied for the purposes of facilitating discussions with the client, and within the industry. Neither the author(s) nor EGR Consulting Ltd., make any representation or warranty as to the accuracy or completeness of this document, or accept any liability for any omissions, or for statements, opinions, information or matters arising out of, contained in or derived from this document, or related communications, or for any actions taken on such a basis.*

***EGR Consulting Ltd.***

## **CONTENTS**

<b>Executive Summary .....</b>	<b>4</b>
<b>1 Introduction .....</b>	<b>5</b>
<b>2 Market Design Concepts.....</b>	<b>7</b>
2.1 Market Clearing Prices .....	7
2.2 The PDC and Entry Economics.....	9
2.3 Locational Issues .....	13
2.4 Risk Aversion and PDC Inflation.....	14
2.5 Contracts and Cost Recovery .....	17
2.6 An Integrated Transaction Cost Perspective .....	19
2.7 Conclusions .....	22
<b>3 NZEM Market Design Choices .....</b>	<b>23</b>
3.1 Background.....	23
3.2 Basic Design Philosophy .....	25
3.3 Why are offers not regulated? .....	28
3.4 Why are prices not capped?.....	31
3.5 Why is there no capacity component?.....	35
3.6 Conclusions .....	37
<b>4 SRMC for Hydro and Energy-Limited Thermal.....</b>	<b>39</b>
4.1 SRMC for Major Reservoirs.....	40
4.2 SRMC in River Chains .....	42
4.3 SRMC of Energy Limited Thermal .....	45
4.4 Locational Issues .....	50
4.5 Risk Aversion .....	51
4.6 Conclusions .....	52

**5 Final Perspective .....56**

**Appendix: Market Performance and Entry Barriers .....59**

## **Executive Summary**

1. This discussion paper considers some basic questions that need to be addressed before considering the kind, and extent, of market power that might be considered appropriate, or inappropriate, in the NZEM. It takes an “economic” perspective on market design, explaining the basic theory and comparing and contrasting the NZEM’s “energy-only” design with that of some other markets, in which both energy and “capacity” are traded.
2. We then consider both short and long run perspectives on market design, and the central role of contracting in reconciling those perspectives. Noting strong parallels with the design of other markets with similar cost structures, we explain why the NZEM market design places primary emphasis on long term efficiency, implicitly accepting that some deviation from short run efficiency, as defined by “SRMC pricing” will result.
3. In the context of the NZEM design, we also consider the implications of various forms of market regulation, and argue that capping market prices, or forcing spot prices to SRMC, for example would actually undermine the market design, and inappropriately deter entry.
4. We then turn to consider the question of how SRMC might be determined in a hydro-dominated market, where risk is a very significant factor, and storage capacity creates inter-temporal linkages which imply generation and pricing behaviour which looks very much like the exercise of market power, even under perfectly competitive, or centrally planned, assumptions. We also note that some New Zealand thermal producers actually have rather similar cost/production characteristics, and argue that it may be quite difficult, in practice, to determine what SRMC actually is, for any of these participants.
5. In summary, from a market design perspective, we argue that the real issue is long run market performance, and that this should be analysed in terms of the alignment of the market Price Duration Curve with the costs of entry, as calculated by risk averse investors. An appendix discusses this perspective, and notes that prices can not be said to have been “too high” in the NZEM, unless it can be shown that current capacity is, in fact, excessive, and/or it can be shown that capacity is not excessive only because there are artificial “barriers to entry”.

# *Economic Behaviour in a Hydro-Dominated Electricity Market*

## **1 Introduction**

Several questions have been raised, in recent years, about the design of the New Zealand Electricity Market (NZEM), and particularly about the potential exercise of market power within that market, and specifically in the wholesale spot market, on which we focus here. It seems evident, from some discussions, that there is a divergence of view about what the market design actually is, or was intended to be, and perhaps about what it should be. In particular, there is an evident tension between studies and concerns focussed on a narrow Short Run Marginal Cost (SRMC) perspective, and those derived from a broader Long Run Marginal Cost (LRMC) perspective.

This raises a particular risk that situations and behaviour may be assessed from a perspective derived from other markets, with different design philosophies, and that inappropriate and incompatible conclusions and “solutions” may then be imported from such markets. Accordingly Mighty River Power Ltd have requested EGR Consulting Ltd to assist in developing greater public understanding, and awareness, of some key issues by providing a background paper on the design of the New Zealand Electricity Market (NZEM).

This report ignores the issue of what might be considered legitimate, or illegitimate, exercise of market power in the context of that market design. Instead it focuses on the central market design issues defining the context within which that question might be considered, and on the kind of pricing and behaviour patterns that might be expected if market power were NOT being exercised.

- First, we discuss some basic market design concepts, noting the difference between the “one part” market design adopted in New Zealand, and several other jurisdictions, and the “multi-part” designs used in some other parts of the world. We particularly focus on the difference this may make to expectations about alignment of market prices with SRMC
- Then we outline the reasons behind the basic uncapped, LRMC focussed design philosophy of the NZEM, and discuss why various modifications to that design were not adopted, even though they may seem attractive in the short term.

- Switching tacks slightly, we next examine the inherent difficulty of defining SRMC in hydro dominated electricity markets, and explain the kind of behaviour and price patterns that may be expected to arise in such markets, assuming perfectly competitive, or centrally optimised, responses to varying hydrological conditions.
- Finally, we suggest an alternative, LRMC focussed, perspective on how the performance of the NZEM should be assessed.

Much has been written on some of the points touched on here, and much more could be written. Thus a comprehensive treatment is not possible in this context. Our aim has been to provide a reasonably accessible overview of the issues, rather than an in-depth development of any one of them. Thus we focus on the issues that are most pertinent, and perhaps most controversial, and do not attempt to “prove” any of the assertions made here. The perspective presented here is based on an extensive personal history of involvement with, and research on, electricity sector issues, particularly in New Zealand, before during and after the market reform process.

No “new” results are presented, and no systematic attempt is made to reference all the academic literature which might be brought to bear on these issues.<sup>1</sup> Some reference is made, though, to studies conducted by the current author and his colleagues, in the New Zealand context. In particular the reader is referred to Culy et al [1996]<sup>2</sup> for an economic perspective on the history of the New Zealand electricity sector prior to establishment of the current market. Read [1997]<sup>3</sup> provides a perspective on the goals of the current market design, and a commentary on initial experience with it.

---

<sup>1</sup> The formal academic literature is, in fact, much sparser than might be thought, partly because the study of hydro dominated electricity systems is only of significant interest in a few parts of the world, each of which has tended to develop and communicate its own understanding of the situation, in its own way.

<sup>2</sup> J.G. Culy, E.G. Read, and B. Wright: "Structure and Regulation of the New Zealand Electricity Sector", in R Gilbert and E Kahn (eds.) *International Comparison of Electricity Regulation*, Cambridge University Press, 1996, p. 312-365.

<sup>3</sup> E.G. Read: "Electricity Sector Reform in New Zealand: Lessons from the Last Decade" *Pacific Asia Journal of Energy* Vol 7, No 2, 1997, p. 175-191

## **2 Market Design Concepts**

### **2.1 Market Clearing Prices**

At the most basic level, the NZEM is an “energy-only”, or “one-part”, market<sup>4</sup>, in which all participants buy and sell at Market-Clearing Prices (MCPs). Thus it is based on the principles that:

- All participants providing (or purchasing) energy at the same location should be paid (or should pay) at the same rate, irrespective of their offers (or bids);
- The entire remuneration for generators should be provided by these spot market energy payments, or derived from them by way of financial contracts written against them; and
- Competitive discipline is relied upon to discipline offering behaviour, and hence to control prices.

This design has important implications for the pricing patterns we should expect to see arising from the market.

First, it is often loosely stated that, under “perfect competition” participant offers should be expected to reflect Short Run Marginal Cost (SRMC), and that MCP should thus reflect the marginal cost of generation at the industry level.<sup>5</sup> That expectation is examined in greater depth later, but here we note that in this kind of market MCP is not determined solely by the industry supply curve, but by the interaction of supply and demand curves. If demand is elastic, then MCP will sometimes reflect the marginal cost of (voluntary) demand reduction, rather than of generation.<sup>6</sup>

If there was no (voluntary) elasticity in the demand curve, though, MCP would (theoretically) equal the marginal cost of production until a capacity constraint was reached. But at that point the MCP should theoretically rise to very high levels, reflecting the cost to society of the involuntary load shedding needed to match

---

<sup>4</sup> We will use the former term, because “multi-part” offers may be employed in markets which are essentially “energy-only” in the sense that energy prices are capped only at very high levels, and there is no separate market for long term capacity investment.

<sup>5</sup> That is, the SRMC of the marginal producer.

<sup>6</sup> In fact most New Zealand discussions have extended the definition of “SRMC” to include demand reduction costs for several decades, thus making it natural, but potentially misleading, to refer to MCP as the “SRMC price”.

demand to the available supply. In general that cost would be much higher than for voluntary reductions of the type that might be expected in response to market price signals. This would be reflected in the “shadow prices” calculated by a centralised optimisation, and the same price pattern should be expected from a hypothetical perfectly competitive market.<sup>7</sup> In the real world, demand will be somewhat elastic, and prices may theoretically be set by load reduction at levels below the SRMC of the most expensive generator. But prices in an “energy-only” market should still be expected to exceed the SRMC of generation, perhaps by a very large margin, at least when supply is short.

Second, as an “energy-only” market, the NZEM differs significantly from markets, in which participants receive supplementary payments for “capacity” in various forms.<sup>8</sup> As such, it should be expected to produce patterns of “energy” prices, which differ from those arising in such markets, and also from many traditional forms of regulated electricity pricing, in which explicit capacity payments (or peak) charges often feature prominently.<sup>9</sup> What we might expect, though, is that the (optimal) pattern of energy prices in the NZEM should match the (optimal) pattern of energy/capacity prices, in combination, from a multi-part or traditionally regulated electricity market.<sup>10</sup>

Specifically, if the electricity system were still planned and operated using centralised optimisation models, those models would compute an SRMC-based price corresponding to the MCP, so long as the optimally planned system was able to meet demand.<sup>11</sup> There would be periods, though, in which an optimally planned system

---

<sup>7</sup> Technically, a centralised optimisation may report “infinite” shadow prices, but these over-state the situation.

<sup>8</sup> This discussion ignores ancillary service market(s). In New Zealand, these are “co-optimised” with the energy market, and ancillary service sales provide some additional revenue to support capacity investment. But this does not materially alter any of the discussion here. Ancillary service payments are only made to participants providing ancillary services, and only for the MW provided. This is not the same as providing a capacity payment to all capacity in the market, as occurs in markets that employ capacity pricing.

<sup>9</sup> At one stage, before there was any significant thermal generation in New Zealand, the wholesale “Bulk Supply Tariff” consisted entirely of such peak charges, meaning that “energy” was implicitly charged at a per unit price of zero. (This might be thought to be the SRMC of a pure hydro system, although that is generally not correct, as will be seen from later discussion.) A 50% energy component was introduced later. (See Culy et al [1996])

<sup>10</sup> But note that traditional capacity charges have at least partly reflected the cost of the transmission and/or distribution systems, and other overheads. For this purpose, though, it is only the “generation capacity” component of these charges which is relevant.

<sup>11</sup> This discussion assumes that load requirements are not just specified by a simple deterministic “hard” constraint. We assume that the range of possible loads is represented probabilistically (eg by a Load Duration Curve, or LDC), and that the optimization model is provided with an economic representation of the costs incurred when load is not supplied, either as a demand curve for voluntary reduction, or a cost curve for involuntary reduction.

would be unable to meet demand, forcing some form of rationing to occur.<sup>12</sup> During such periods there would be a shadow price limiting generation to be less than total installed capacity, and that would represent an extra payment required, over and above the SRMC-based prices, to all capacity available at that time. Ignoring economies of scale, it can be shown that the costs of all capacity in the optimal plan will be covered if this (notional) capacity payment is added to the (notional) payments calculated from SRMC-based prices. Conversely, the SRMC-based prices alone will always be insufficient to cover the investment cost of any plant, after fuel and variable maintenance costs are accounted for.

The same result would apply equally in a hypothetical perfectly competitive market. So, if the market is working well, we should expect the value of those price spikes to roughly correspond to the value which would be recovered from capacity charges under market design options which include such payments. In real life, this same price pattern may be approximated in two ways:

- By a combination of energy and capacity prices in a “two-part” market; or
- By energy prices alone in an “energy-only” market

In the first case, we might hope to see energy prices approximating the SRMC of supply at all times, complemented by capacity prices approximating the capacity constraint shadow prices in a hypothetical centralised optimisation. In the second case, we might hope to see energy prices approximating the SRMC of supply much of the time, but supplemented by moderately frequent (energy) price “spikes”. And the value implicit in those price spikes should approximate the value which would be recovered from capacity charges under a “two-part” energy/capacity market design. In other words, it should equal the value of the capacity constraint shadow prices in a hypothetical centralised optimisation.

The next section discusses the generalisation of this concept in a form that fits more naturally with discussions of market economics, rather than of centralised optimisation.

## **2.2 The PDC and Entry Economics**

While the precise chronological distribution of prices is important for some purposes, much can be learned by analysing the aggregate distribution of market prices, as expressed by the Price Duration Curve, or PDC, indicating the proportion of the time for which prices are observed above each price level. (Real) option values (OV) for

---

<sup>12</sup> Put another way, there will always be a probability level beyond which it would be more economic to risk the possibility of non-supply, rather than to incur the cost of building supply facilities which will almost certainly never be used. A sufficiently detailed probabilistic optimisation will reflect this by determining an optimal trade-off between supply and non-supply, implying a finite probability of non-supply.

any plant type can be determined from such curves, being the nett operating profit to be made by that plant type, assuming that it operates throughout the period for which the MCP exceeds its SRMC. In other words, it is the value of a call option, with a strike price set at that plant's SRMC.<sup>13</sup>

Generalising the definition of SRMC to include load reduction costs, as above, a notional PDC for an energy-only market can be produced under competitive SRMC assumptions. Prior to the market, in the latter days of the Ministry of Energy, the New Zealand electricity sector was planned using OVs determined from a PDC derived in exactly this way, using results from the PRISM/SPECTRA models. It is not difficult to show that more capacity of each plant type should be introduced if, and only if, its SRMC-derived OV exceeds its fixed (Capital +O&M) cost, FC.<sup>14</sup> This holds true under central planning, but also for a market. It is not too difficult to show that, subject to some caveats discussed below, investors should have commercial incentives to introduce new capacity of each type when its MCP-derived OV exceeds its FC.

Accordingly, the alignment of the SRMC/MCP-based PDC with entry costs provides an important test of market performance. Details are discussed below, as they apply to a hydro-dominated market. But one basic observation should be made here. It is easy to show that, ignoring scale economies, technological progress and resource depletion effects, and adopting the extended definition of "SRMC pricing" discussed above, OV/FC alignment implies that:

$$OV=FC$$

That is:

$$\textit{Expected revenue assuming "SRMC-based" MCP prices MINUS Fuel Costs = FC}$$

This can then be re-arranged to form:

$$\begin{aligned} \textit{Expected revenue assuming "SRMC-based" MCP prices} &= \textit{FC PLUS Fuel Costs} \\ &= \textit{LRMC} \end{aligned}$$

This relationship holds on average, over the long run, and applies to each plant type, and hence to each load class. That is:

---

<sup>13</sup> We will ignore variations in operating efficiency across the output range and note that under competitive assumptions, plant will make no nett operating profit when it is itself marginal, because the price will be set by its own SRMC.

<sup>14</sup> Strictly this is the FC for bringing forward a small increment of capacity to the year concerned. In reality investment is lumpy, and this matching is not exact. But this does not really affect the principles discussed here.

- The (generation weighted) average SRMC-based MCP received by base-load plant (and hence the average price paid by base loads) must match the LRMC of such plant;
- The (generation weighted) average SRMC-based MCP received by “shoulder” plant must match the LRMC of such plant;
- The (generation weighted) average SRMC-based MCP received by “peaking” plant must match the LRMC of such plant; and, as a result
- The (load weighted) average SRMC-based MCP paid for any pattern of peak/shoulder/base load must match the LRMC of meeting such load.

Oddly, though, many discussions treat the Long Run Marginal Cost (LRMC) as if it were an upper bound on SRMC pricing. Thus it is common to see market price projections tracing a rising curve of “SRMC prices” up to the point where they equal LRMC, after which it is assumed that entry will occur and limit prices to LRMC thereafter. This may be a reasonable picture of the actual performance of many markets, including the NZEM in its early years, but it should be recognised that it can only be a valid picture of a market in disequilibrium, starting with excess capacity. If this pattern is exhibited by many markets worldwide, that presumably suggests that most of these markets were established in situations where excess capacity already existed. From an economic perspective, though, it does not represent a typical, or sustainable, long term pricing pattern, but a reflection of a pervasive tendency toward excess investment under prior regulated regimes, worldwide.<sup>15</sup>

What we should really expect to see, in the long term, is that average annual SRMC-based prices sometimes lie above LRMC, and sometimes below it, equalling LRMC on average. A very similar pricing pattern might be expected from an optimised centrally planned system, too. But this has often been obscured by political factors. In New Zealand, for example prior to the establishment of ECNZ, the prices charged to the generality of loads were held constant for long periods of time, in nominal terms, and then sometimes increased very abruptly.<sup>16</sup> But this pattern had more to do with politics, and government revenue requirements than with the underlying economics of a system which veered between really quite significant under- and over-supply. Hopefully markets will succeed, where centralised planning obviously failed, in minimising fluctuations around the LRMC benchmark, but such fluctuations will still occur, particularly when the market experiences any kind of shock. Thus it seems reasonable to expect that “shortage” situations, such as that which faced California, will also occur in other markets, from time to time, and that prices will exceed SRMC for significant periods during such periods of relative shortage. If they do not, and if this means that average prices fall below LRMC, the market design will

---

<sup>15</sup> Prices may also rise, over time, because LRMC itself is rising, perhaps due to resource depletion and increasing environmental pressures. But this is a different issue.

<sup>16</sup> Discounted price offers made to attract major electricity intensive industry were another matter.

have failed in one of its primary objectives; that of setting prices to sustainable levels, on average.

This discussion has important implications from a regulatory perspective. It seems that regulators are generally comfortable with periods during which SRMC prices lie below LRMC. Indeed some regulators seem concerned to force prices down in situations where they are arguably above SRMC, even when those prices are still well below LRMC: That is below a long term sustainable level. But it is by no means clear that regulators will look so benignly on periods when SRMC prices rise above LRMC, as they must do if entry is to be economic in the long run, or when the “SRMC” price is explicitly or implicitly set by demand reduction. It seems distinctly possible that prices could be capped below their natural level in such situations. If this is so prices will have to rise above SRMC during surpluses, if a sustainable long run equilibrium is to be maintained, on average.<sup>17</sup>

For similar reasons, New Zealand electricity prices were not raised to “SRMC” levels during times of relative shortage during the historic centralised planning era. In particular, they were not raised high enough to reduce demand during dry years, and physical restrictions were not infrequent. But prices were not forced down to SRMC levels during times of relative surplus either, and certainly not to zero when water was being spilled in wet years. This may partly have been due to administrative inertia, but it also reflected a widespread, if not universal, view that the assets built to meet public electricity demand must be paid for, preferably by electricity consumers, irrespective of the way in which SRMC varies from time to time, due to hydro variation, planning mistakes, or whatever. In other words, in order to recover costs, electricity prices have historically been set well below SRMC much of the time, and well above SRMC much of the time, in New Zealand and around the world.

This can be seen as reflecting a fundamental conflict between the forward-looking perspective of economics, with its emphasis on finding the best use of resources irrespective of what they may have cost to develop; and the backward-looking perspective of accounting, with its emphasis on paying for resources already committed, whether or not they were economically justified in retrospect. Alternatively, though, we may characterise the conflict as being between short and long run perspectives on economic efficiency; or broadly between achieving productive and allocative efficiency in the short run, versus dynamic efficiency in the longer run. Later, in Section 2.5 - *Contracts and Cost Recovery*, we explore the kind of contractual mechanisms that can, theoretically, allow both to be achieved

---

<sup>17</sup> More exactly, the issue is not whether prices will actually be capped, which may be unknowable in advance, but whether participants now think there is a possibility they will be capped, and account for that possibility in making their operational/investment plans. The greater the probability of capping, and the tighter the caps, during shortage periods, the greater the deviation from “SRMC” required to balance the books during surplus periods. But that topic is examined in more detail later, in Section 3.4 - *Why are prices not capped?*

simultaneously. First, though, we should consider how locational issues, and risk attitudes might impact on the framework outlined above.

## **2.3 Locational Issues**

All of the above discussion has implicitly assumed that “the market” exists at a single location, so that a single PDC applies to all capacity. But the NZEM is a locational market, and we should consider locational issues, too.

Basically, all of the logic above applies at each location, with the caveat that these locations are not independent. Thus there is a PDC at each location, and in long term equilibrium the OVs implied by that PDC must align with the FCs of entry at that location. Now, though, the prices at a location are set partly, and perhaps predominately, by power imported from, or exported to, other locations. Also, the “entry” options for a location include not only expansion of local generation, but combinations of generation/transmission expansion that allow more power to be imported and/or exported.

The discussion that follows, in relation to issues such as risk and contracting also applies at a regional level, with the extension that the set of potential contracts to be considered includes combinations of energy contracts at that location and at other locations, with “FTRs” or other transmission hedging arrangements that can potentially allow commercial competition to occur wherever physical transmission is possible.

Basically the smaller, and more “isolated”, a location is, the greater the importance of:

- Risk as a factor for local investment in both generation and retailing;
- Contracts, and/or capacity tickets, and/or FTR type arrangements<sup>18</sup> as a means of mitigating that risk; and
- Transmission under/over investment as a likely factor in local market disequilibrium.

Clearly it is quite possible for the market to be significantly out of equilibrium in particular locations, even if it is broadly in equilibrium at a national level. Thus we may expect deviations from long run equilibrium to persist for much longer at the local level, where transmission and generation investments may occur very infrequently, than at the national level.

---

<sup>18</sup> Including location swaps, for example.

All of this may imply some differences in emphasis, but does not really change the conclusions reached above, or below. Accordingly, we will continue to ignore regional considerations for the remainder of our discussions on market entry and equilibrium, and on market design issues.

## **2.4 Risk Aversion and PDC Inflation**

While the theory discussed in previous sections relates to an uncertain PDC, and implies some degree of “risk”, we have not considered “risk aversion” at all. It should be recognised that risk is a rather more significant issue in hydro dominated markets, than it is for typical electricity markets, and that it may not be easy to provide risk averse investors with sufficient assurance that they will be able to obtain an adequate return for the risk involved. The risks of strategic response from other participants, regulatory intervention, technical failure, and changes to load growth, technology or fuel prices affect these markets, just as they do any other market. But participants in, and potential entrants to, a hydro dominated market face additional risks.

For a start, they must try to assess the true underlying supply/ demand balance, and the whole price probability distribution, from observation of prices in a relatively small sample of hydro years, which may have been significantly wetter, or dryer than average.<sup>19</sup> Since extremes play a major role, in determining expected values and risk, it might take several decades to collect an adequate sample, from a hydrological perspective. But then, observations of market conditions, as opposed to hydrological conditions, will be rapidly outdated by changes to the system, fuel prices, and political conditions.

Then, once built, generation designed to provide the last increment of capacity to meet the 1:20 security standard used in traditional capacity planning, let alone the 1:60 standard implied by more recent announcements, can be expected to generate significant power in only one year out of 20, or perhaps 60. In other words there is a significant probability that it will not be called upon ever, during its entire technically viable life-time.<sup>20</sup> Thus the need to provide contractual guarantees, and/or to ensure that generators can recover costs by pricing above SRMC when loads do not fully contract, becomes much more pressing in such a market.

Accordingly, risk and risk aversion are important factors here. All of the above discussion may be thought of as assuming risk-neutrality, though. Thus, when we say that peaking plant will enter if its “OV exceeds its FC”, this may be interpreted in

---

<sup>19</sup> For one such analysis, see: J Tipping, E G Read and D C McNickle “The Analysis of Spot Price Stochasticity in New Zealand’s Electricity Market” *IAEE Proceedings*, Zurich 2004.

<sup>20</sup> This is actually not far from the experience of stations such as Marsden B, or the old Whirinaki station.

NPV terms. But we have not actually said what risk premium should be applied, or what discount rate is to be used when determining OV or, for that matter, FC. In reality, any potential entrant realises that investment in a peaking plant is naturally risky, and that any threat of intervention is likely to increase that risk. So the potential entrant will presumably apply a risk adjustment to the discount rate used for project evaluation, thus raising FC by a potentially significant amount. Read et al [2007]<sup>21</sup> argue that this could have a significant impact on the effective PDC expected in long run equilibrium, particularly in a energy-only electricity market:

- In a market which provides guaranteed payments for “capacity”, the providers of that capacity should be expected to determine FC at a moderate discount rate, and this should be equivalent to the OV for such capacity determined from an optimal SRMC based PDC;<sup>22</sup> but
- Since participants in an energy-only market receive no such guarantee, they must determine FC at a “risk-adjusted” discount rate, and that FC should be equivalent to the OV for such capacity determined from a PDC with higher prices occurring with a higher probability.

One may argue about how significant this effect actually is, particularly in a market dominated by vertically integrated “gentailers” who are exposed to the risk of not being able to meet customer obligations if they can not access sufficient capacity in extreme conditions. In principle, though, an energy-only market can not be in long term equilibrium unless this latter condition holds, with a PDC which may be inflated to significantly higher levels than might be calculated on a risk-neutral perfectly competitive basis. The issue is whether those higher prices occur more often because market prices exceed SRMC, or whether it is that SRMC itself must be higher, more often. In principle, either or both may occur.

PDC inflation of this type is not necessarily inconsistent with “SRMC pricing”. If market rules were to enforce SRMC pricing, but prices were not capped, potential entrants would simply refuse to enter, thus “withholding capacity” in the ultimate sense, until the PDC rose high enough to support entry, with an appropriate risk premium. In part this may occur because the lack of investment forces less efficient, and hence more expensive, plant onto the margin, more often. Thus peaking plant may be required to generate more as “shoulder” or “peak support” plant, and shoulder plant as base-load plant. In part it may occur because shortages eventually

---

<sup>21</sup> E. G. Read, M Thomas and D Chattopadhyay “The Impact of Risk on Capacity Investment in Electricity Markets” keynote presentation, *IAEE Proceedings*, Wellington, 2007

<sup>22</sup> Such a guarantee does not eliminate supply-side risk, but probably reduces overall risk below normal commercial levels.

become acute enough to force SRMC/shortage cost prices up to an acceptable level, given the risks involved.<sup>23</sup>

The result will be a plant mix with less capital investment, higher running costs, and more shortage than may seem optimal, when assessed from a traditional central planning perspective, where aversion to commercial risk is not normally considered to be a significant factor. Alternatively, though, greater entry, a lower incidence of shortage and a more optimal plant mix, can be sustained if potential entrants believe that their risks can be reduced by pricing above SRMC at peak times, and/or during surplus periods.

Specifically, if energy prices were always limited to (supply-side) SRMC then the plant with the highest SRMC in the system could never make an operating profit at all from spot market sales. In order for the plant mix to be optimal, though, the OV of the equivalent call option (as determined by that part of the PDC where prices exceed this maximum supply-side SRMC) must still equal the FC of that plant. But the market can not be in long term equilibrium, with sustainable entry of peaking plant, unless that plant can cover its FC. If capacity payments are not possible, then the equivalent must be recovered in the form of a markup on SRMC prices.

In fact, to sustain entry of peaking plant that operates for PEAK hours per annum, assuming  $FC_{\text{peak}}$  expressed as an annuity per MW of capacity, then the premium over SRMC pricing during those PEAK hours would have to average  $FC_{\text{peak}}/\text{PEAK}$  \$/MWh. Further, it can be shown that every MW of supply side capacity should receive this capacity price which, in combination with the remainder of the PDC, is just sufficient to support optimal entry for each plant type in the optimal capacity mix. Thus no further markups should be required, for that purpose. In reality, capacity provision in hydro-dominated systems relates to much more than simple provision of MW capacity at peak times.

As discussed in Section 4.1 – *SRMC for Major Reservoirs*, the prospect of high prices due to a possible future energy shortage, for example, feeds back into high opportunity cost-based offers, and prices, for many periods before that event is projected to occur. Strictly speaking, though, this does not cause the PDC to inflate above SRMC levels. What it does is to cause the SRMC of hydro, as determined by these opportunity cost calculations, to rise.<sup>24</sup> We have already discussed the particular problems of hydro generators facing sustained low prices, though, and the

---

<sup>23</sup> Price capping merely removes this second option forcing more reliance on the first, as discussed in Section 3.4 - *Why are prices not capped?*

<sup>24</sup> In theory, these opportunity cost calculations should not be much affected by premiums applying only in extreme peak periods, because opportunity costs relate only to opportunities available, on the margin, and most of the hydro system will already be fully loaded in such periods. Opportunity costs may be more affected by prices expected in future “shoulder” periods, and hence by the SRMC of “peaking” plant used to provide energy, in a sustained “dry year support” role.

need for the “bottom end” of the PDC to inflate, in order to support entry of such capital intensive plant. Thus the impact on the PDC may be much more pervasive than this simplified discussion might suggest.

Relying on ill-defined markups like this may be considered a less than perfect way to run a market, but Section 2.6 – *An Integrated Transaction Cost Perspective* points out that this is, in fact, the normal way in which most other markets operate, with considerable success. In reality, neither option is perfect, and the choice between them is discussed further in Section 3.5 – *Why is there no Capacity Cost Component*. But the point here is not to debate what the NZEM design should be, just to note that an energy-only market of this type can only work if the PDC is allowed to inflate in the general way described. In particular, prices may need to settle above SRMC on average, and in particular situations, in order to sustain a reasonably efficient plant mix, in long term equilibrium.

## **2.5 Contracts and Cost Recovery**

Contracts have a direct impact on risk, and on the effective price distribution, and hence on the economics of entry. Obviously, if participants buy or sell 90% of their output/demand on contract, then the effective average price is largely determined by the contract price, rather than the spot price. This, in turn, has significant implications for behaviour.

Provided contracts trade at a freely determined price, rather than being imposed on one side of the market or the other, the logic described above remains valid, with one major difference: Contracts re-define and re-assign risk, and consequently allow the backward and forward looking perspectives, described above in Section 2.2 – *The PDC and Entry Economics*, to be reconciled.

A potential entrant should be able to sell a contract for the expected output pattern of a unit at around the expected value of that output pattern in the spot market. Thus, looking forward, it will try to time its entry so that the OV of such a contract corresponds to the FC of its proposed plant.<sup>25</sup> So the economic optimality conditions, described above in Section 2.2 should be expected to hold, in prospect, at the point

---

<sup>25</sup> Strictly, the optimal contract form for a stand-alone thermal station would be a call option, with a strike price set at that station’s SRMC, and the value of such a contract should be OV, as discussed above. In theory, such contracts should be attractive to loads looking for peak power, and to hydro producers looking for dry-year backup power. In reality, the market for such contracts may be thin, and generator output may need to be sold in some bundled form more suited to consumer requirements, and backed by a portfolio of plant and/or contracts. If so, potential entrants may find it difficult to compete with incumbents, and this may create barriers to entry. But that does not affect the theoretical discussion here. If the contract takes the form of selling power, eg via a standard “contract for differences” (cfd), rather than selling a call option at OV, the entrant would need to sell the output of the plant at its entry cost, including both FC and fuel cost components.

when a participant commits to building a new plant, at least when evaluated from that participant's perspective.

Looking back, though, participants may have quite different views, both about what their costs actually were, and about market prospects. Thus they may find a significant discrepancy between their cost recovery "requirements", and SRMC-based prices. If they sold a contract for all the new capacity, though, the average price they receive will be the contract price, not the MCP. If that contract price met their expected cost recovery requirements at the time it was agreed, it will continue to meet those expected requirements as conditions change. Contracted entrants would then only have to deal with any discrepancy between the actual and expected cost and/or performance of their own plant.

A perfect match between contracted and actual capacity is unlikely, in reality, and virtually impossible for a hydro generator. But the perfectly contracted case describes the opposite end of the spectrum from that at which many analyses start; that is with a pure spot market with no contracts at all. Reality will lie in between, but note the implication that the closer the real world lies to the perfectly contracted case, the closer revenues will lie to cost recovery requirements, and to LRMC, as it was expected to be when the contract was signed. This is true, irrespective of spot market prices, making participants less concerned about whether market prices lie above SRMC, or below it. Thus a perfectly contracted generator is indifferent to the level of market prices, and has no incentives to influence them, either up or down, so long as it is able to generate if, and only if, market prices exceed short run operating costs. As their optimal perfectly competitive generation levels deviate from their contract levels, though, generators will find themselves more exposed to spot prices.

If their production is below contracted levels, generators are actually net buyers in the spot market, and will wish to see prices lower, if possible. If production is above contracted levels, they are net sellers in the spot market, and will wish to see prices higher, if possible. Totally uncontracted generators, having no guaranteed income and relying solely on spot market prices to recover costs, represent an extreme case. But, although hydro generators may find themselves over-contacted in dry years, and thermal generators may find themselves over-contacted in wet years, most generators will be contracted for something between 0 and 100% of their expected output, on average. And generators must not only recover their costs, on average in the long run, but generate cash flow to cover many of their costs, including the funding requirements of sunk capital costs, on a year by year basis. This will not be possible if prices fall too low, and particularly if they fall to zero, which may be the SRMC when hydro is in surplus supply.

So, whatever other motivation may apply, most generators will need to receive prices above SRMC in order to cover their costs, during periods of relative surplus. If they can not do this via contracting, they will need to see prices above SRMC during those surplus supply periods. Given the strong correlations between inflows to one catchment and another, this will also be the time when individual hydro producers will normally have an uncontracted surplus to sell, making them reluctant to push

prices down further by releasing all of that surplus on to the market. The extent to which generators can actually influence prices relates to the degree of “market power” they may be able to exercise, and that is not our focus here. The point here is simply that, in a world of imperfect contracting, generators may not be financially viable unless prices exceed SRMC during extended periods of relative surplus.<sup>26</sup>

## **2.6 An Integrated Transaction Cost Perspective**

One possible conclusion from the above discussion is that all generation, and hence all load, should be contracted for its expected output at all times, and over all time scales. In theory, this should minimise risks, remove the need for prices to deviate from SRMC, and allow both long term and short term markets to operate with minimum distortion, and maximum efficiency. In reality, this does not happen. So we may ask: Is this a distortion of the market, and something that should possibly be “corrected”? Or is this perhaps optimal, when seen from the context of some broader theoretical framework?

In this case the appropriate theoretical framework is provided by the theory of transaction costs. But a helpful perspective may also be provided by comparing electricity with other sectors whose cost structures are similar. There are, in fact, many industries with basically similar, capital intensive, cost structures to electricity. So the critical question to be considered is this: If regulation to force a high level of contracting, and/or SRMC pricing is the right answer for electricity, why is it not adopted more widely throughout the economy?

The cost structure of the electricity sector is actually little different from that in a myriad other industries, where prices routinely exceed SRMC, because prices need to be maintained at such levels in order to provide an adequate return on investment, given the risks involved. But we will focus on two sectors with which we are all very familiar: Airlines; and Hotels. In both cases SRMC is very low, except on rare occasions when all capacity is fully booked. In both cases, though, SRMC pricing is a rare exception.

In the case of airlines, the incremental cost of taking another passenger on a scheduled flight is very near zero, and “SRMC pricing” actually does occur, in the form of frequent flyer miles. But this is only in the context of what is effectively a long term contractual relationship, where it can equally be interpreted as a (relatively small) discount on regular prices. Not only are prices in that industry quite volatile, in both the short and long term, but (unlike electricity) consumers receiving apparently

---

<sup>26</sup> In light of previous discussions, these periods when price exceeds SRMC must be offset by other periods during which price is less than SRMC (broadly defined to include load reduction costs) if long term market equilibrium is to be maintained. But note that if hydro generators are over-contracted during dry periods they will find themselves net buyers from the market, and thus motivated to put downward pressure on prices by supplying more volume (to the extent that they have discretionary volume and remain a net buyer).

identical services, at the same time, may be paying very different prices. It is entirely possible that, sitting across the same row of the same aircraft, there will be passengers who paid full fares, half fares, quarter fares, and nothing at all.<sup>27</sup> Even “worse”, from an economic perspective, there will often be empty seats which could easily have been filled by grateful passengers, had they been offered at the “SRMC price”, of close to zero. Very similar observations may be made about the hotel industry. Hotels routinely charge positive prices, even quite high prices, on nights when there are actually spare beds, making SRMC, again, very close to zero.

But deviation from SRMC pricing is not limited to these industries. In fact it is ubiquitous in the everyday world of business, even in sectors which can be reasonably considered “competitive”. Prices deviate from SRMC wherever a “markup” is added to cover risk or overheads, for example, or royalties are charged on intellectual property rights of any kind. So why do regulators not intervene to force SRMC pricing in all of these industries? Clearly, regulators are taking a wider and longer term perspective. They understand that what is really most important, particularly in capital intensive industries, is that the market facilitates efficient investment over time, in the form of new aircraft, hotels, software packages etc. Accordingly, they rightly focus on the existence of possible barriers to entry, and accept that short run pricing will be routinely distorted, with consequential impacts on short run efficiency.

Let us be clear: This policy does impose real costs on the economy. There really are people sleeping under bridges when beds are free at the Hilton. There really are empty seats in planes and movie theatres that would have been filled if tickets were free. And consumers, every day, go without all kinds of goods that they find too expensive, at retail prices, but which they would readily buy and use if available at the SRMC of production and distribution. The aggregate cost of all these distortions must be very great indeed. Our point here is not to criticise such practices, though, but to note that they have long been considered legitimate, indeed desirable. In fact, it can be shown that many, if not most, desirable economic outcomes require investments, the fixed costs of which can not realistically be recovered without “distorting” prices away from, and above, SRMC.

This point is often made, and generally accepted, where industries are dominated by scale economies, so that the SRMC of an optimal capacity investment and/or output level is consistently less than its average cost. This would apply to investment in transmission lines, for example, or software development. But, while scale economies do apply at the scale of individual hotels, planes, or generation units, they are less relevant at the more macro scale of overall industry capacity. Thus the need to price above SRMC is less obvious for these industries. There is another common factor at work here, though. What these industries have in common is that perfect forward contracting is not possible, or more exactly that the transaction costs of such

---

<sup>27</sup> But see further discussion below, about the way in which the “services” being delivered here may actually differ.

contracting would exceed the economic costs of living with the distortions implicit in the current regime.

We have referred to the hotel sector, above, but it should be recognised that this is actually just a small subset of a much broader accommodation sector. In that broader sector, it is actually possible to contract forward, and most people actually do manage it, with respect to the vast majority of their person-night requirements. The most common form of forward contracting is called “home ownership”, and home-owners spend most of their person-nights in their own homes, paying only SRMC. Those with term rental agreements are in much the same situation. The hotel sector just represents an extreme outlier in the distribution of accommodation contracting arrangements, from ownership through rentals, time shares etc, to hotels.<sup>28</sup> It is really unrealistic to expect customers to enter into contracts to book hotel beds over anything like the term over which the fixed costs of building a hotel must be recovered. So the hotel supplier must recover all of their costs, with a considerable risk element, from “spot sales”.

Similar comments apply to the “transportation” sector, in which airlines and taxis also lie at one extreme of a wide spectrum of arrangements, covering that part of the demand which customers can not reasonably foresee, or arrange for themselves, and hence can not make long term arrangements for, by buying a car, for instance. In these cases, what we should expect to see, and in fact do see, is a whole spectrum of arrangements being offered; ranging from arrangements in which the customer takes full responsibility for the fixed costs, eg by outright purchase of a vehicle, and then obtains “service” at SRMC; through to arrangements in which the customer takes no responsibility for the fixed costs, and can only obtain “service” at prices that recover LRMC, with a suitable risk-adjusted rate of return. In fact any one customer, at different times and for different reasons, is likely to access “accommodation”, or “transport” via a mix of any and all of these arrangements.

Thus the airline passengers mentioned above, paying different prices to sit in the same row, are actually receiving different “services”, because they are travelling under differing contractual arrangements, entered into at different times, and with different terms and conditions. The more flexible those terms are, the less firm the contract, and the more nearly the price approaches LRMC, rather than SRMC.

Similarly, too for the ideal electricity market, and the theory lying behind an energy-only electricity market like the NZEM is that loads should be sufficiently motivated to contract forward to ensure that a reasonable balance is struck, with the majority of load covered by contracts, and thus hedged against spot market risk, but with suppliers also free to extract a reasonable risk-adjusted rate of return from that part of the load that opts not to contract forward, thus forcing suppliers to take all the risk of providing for a load which may not even eventuate.

---

<sup>28</sup> “Loyalty card” schemes also form part of this spectrum, by offering prices closer to SRMC, or occasionally at SRMC, in return for a longer term, if implicit, commitment by the customer.

This is the way effective markets generally work, and customers in most markets know and understand that, if they leave their bookings to the last minute, they might get a bargain, but they may equally be left out, or end up paying a premium price for the last bed in town.

This point may not be well understood by consumers in the electricity sector, though. Understandably they compare contract prices, retrospectively, with the spot prices that actually eventuated. Much of the time, it turns out that spot prices were low, and they may feel that they paid “too much” for the contract. But this retrospective assessment ignores two important effects. First, a cfd on electricity prices, particularly in a hydro-dominated system, includes a significant component of “insurance” against the possibility of very high prices. And the very nature of insurance contracts is to provide a negative return, when assessed retrospectively, in most periods. Second, the proper comparison is not against the prices that actually did occur, but against those that would have occurred, had the contract not been in place. Collectively, consumers should recognise that the less they are prepared to contract forward, the higher the risks faced by generators, and the higher spot prices will have to be to provide an adequate rate of return. For most individual consumers, the effect of their contracting will be insignificant, but some major electricity users will be large enough to have a noticeable impact in a small market like New Zealand, and particularly in transmission constrained regional sub-markets.

## **2.7 Conclusions**

The main point to emphasise, from this overview of electricity market design, is that a tension exists between a theoretically optimal market design paradigm that ignores real-life transaction cost issues, and the paradigm applied in all other sectors, where transaction costs are implicitly accepted as a fact of life.

The theoretical optimum involves a high degree of forward contracting, presumably at prices matching the LRMC of entry; and minimal deviations from SRMC pricing, in the short run. Theoretically, this could allow both short and long run efficiency to be maximised, simultaneously.

The transaction costs of imposing such a regime would be significant, though, as would the efficiency loss due to intrusive regulation. Thus it is not clearly superior to the more laissez faire alternative adopted in other sectors, where forward contracting is generally much less than 100%, and prices routinely deviate from SRMC, often by a large margin. So there is a trade-off to consider, and compromises must be made. The next section discusses that trade-off, and the choices made, in defining the NZEM design.

## 3 NZEM Market Design Choices

### 3.1 Background

The history of the New Zealand electricity sector prior to establishment of the current market is surveyed by Culy et al [1996], while Read [1997] provides an update, with commentary on initial experience with the current market design. That design evolved in several stages, starting with corporatisation of the Government's electricity sector assets as the Electricity Corporation of New Zealand (ECNZ). The key electricity market design options, including much of the theory discussed in the previous section, were then debated extensively during the late eighties and early nineties, with the current author being heavily involved in those debates. The detail of those debates, or of subsequent history, is not important, but the following summary may be helpful in trying to understand the reasons why the current design was adopted. In particular, it is important to understand that these design choices were made consciously, after careful consideration, and based on a reasonably complete grasp of the theoretical options, and the consequences likely to follow from the design choices available.

From a wholesale electricity market design perspective, the first major step was establishment of a simulated SRMC-based market pricing framework by the Electricity Corporation of New Zealand (ECNZ). That pseudo-market could be described as an exercise in self-regulation by what was then a (near) monopoly. As described by Read and Sell[1987],<sup>29</sup> the development introduced the key elements of the market pricing framework described in previous sections, including half-hourly spot pricing combined with longer term contracts defined as financial “contracts for differences” (cfd). The key difference was that the half-hourly “spot prices” were not determined by competing market offers, in real time, but by running ECNZ's optimisation models, a week in advance. Thus this was an early attempt to simulate the operation of a perfectly competitive market, with strict SRMC pricing. But the “market” also operated within quite tight limits, because there was a requirement for the distribution companies, who bought ECNZ's output at that time, to be contracted for a very high proportion of their load. Importantly, cost recovery required the addition of an “up-lift” payment, called the Pool Price Margin, which effectively played the role of the “capacity payments” discussed here.

---

<sup>29</sup> E.G. Read and D.P.M Sell: *A Framework for Electricity Pricing*. Arthur Young report, released by the Electricity Corporation of New Zealand, November 1987.

The current market design was basically established by the Wholesale Electricity Market Study (WEMS) of 1992, in which the current author played a major role.<sup>30</sup> So far as the wholesale market is concerned, it recommended three major changes to the ECNZ pseudo-market. First, the ECNZ assets were to be broken up, and strict model-based SRMC pricing was to be replaced by a more normal market arrangement, in which prices would be determined by competing offers, in an uncapped market. Second, requirements to contract for a high proportion of load, via cfd “energy” contracts were to be relaxed. But, third, a requirement was to be imposed that load serving entities cover a high proportion of their load with “capacity tickets” defined as call options, and providing protection against extreme price spikes. Thus this would effectively have been a “two part” market.

WEMS was then followed by the Wholesale Electricity Market Development Group, WEMDG [1994]<sup>31</sup>. The WEMDG group included extensive representation from the industry, as for WEMS, but also from consumer groups, and it deliberately employed different consultants, so as to benefit from a wider perspective. Still, it basically endorsed the WEMS design, with one key difference. Whereas WEMS had advocated what was basically a two-part energy/capacity market, WEMDG rejected the capacity ticket proposal, thus creating the energy-only NZEM design, which we have described here. That design was then implemented in 1996, following separation of TransPower and partial divestiture of ECNZ generation assets to form Contact Energy, as a competitor to ECNZ.

The WEMDG wholesale market design remained basically unchanged when the remaining ECNZ generation assets were divided between competing SOEs, and full retail competition establishment, with vertical integration, in 1999. Since that date the most significant events have been Government intervention to build dry year backup capacity at Whirinaki, and the establishment of the Electricity Commission. But neither change affected the fundamental structure of the market.

---

<sup>30</sup> See, in particular, J.G. Culy, E.G. Read, and F.T. Baird *A Managed Transition Toward a Facilitated Market: Rationale*, New Zealand Wholesale Electricity Market Study Report, WEMS/4 October 1992 and *Towards a Competitive Wholesale Electricity Market*, New Zealand Wholesale Electricity Market Study Report, WEMS/5 October 1992.

<sup>31</sup> WEMDG *New Zealand Wholesale Electricity Market, Wholesale Electricity Market Development Group, Final Report*. 1994

## 3.2 Basic Design Philosophy

Much of the literature on electricity market behaviour uses SRMC pricing as a reference point. The designers of the NZEM also had a thorough understanding of that SRMC perspective, having previously been involved with, and advocated, a market design based on strict SRMC pricing. Nonetheless, WEMS and WEMDG placed greater emphasis on a long run perspective, to the likely detriment of SRMC pricing, and thus short run efficiency. So it seems pertinent to ask why.

In part, the decision was motivated by the difficulty of objectively determining what SRMC might actually be, in a hydro dominated system, as discussed in Section 4 – *SRMC for Hydro and Energy Limited Thermal*. In part, it reflected an aversion to intrusive regulatory intervention, as discussed in Section 3.3 – *Why are offers not regulated?* But the WEMS/WEMDG/NZEM market design also emphasised an LRMC perspective, primarily because it was believed that what really mattered most in the electricity industry, like any other capital intensive industry, was to get the long run signals right. And this decision was made despite a realisation that it could imply sometimes significant deviations from SRMC pricing, with consequent economic distortions:

- It was never expected, at least by the designers, that the market would be seen to produce optimal short run operational outcomes, for the capacity mix actually available. If one thinks one has full knowledge of the costs involved, it should always be possible to show that a theoretically superior outcome could have been produced, particularly in hindsight. But the point is that such “knowledge” is essentially an illusion, because the costs are not necessarily even well defined, let alone agreed. Thus the market outcome should never appear optimal, from any one perspective, but should hopefully be more robust, being produced by the interaction between a variety of participants, with different perspectives, each informed by intimate knowledge of their own situations, at least.
- Nor was it expected, at least by the designers, that the market would produce spot prices that were particularly “low”, for the capacity actually available. As we have seen spot prices must be high enough, on average, to cover the full fixed and variable costs of whatever investments are actually made. But the point is that competition and innovation in a de-regulated investment market was expected to provide a better national portfolio of investment options, implemented at lower development costs, and this was believed to be the key factor in keeping average price levels, including spot prices, lower than they

would otherwise need to be to cover the cost of the required level of capacity investment.<sup>32</sup>

This long term emphasis seemed particularly important in New Zealand, where costs have traditionally been dominated by the investment costs of transmission, and of renewable generation, none of which have any significant operating cost.<sup>33</sup> On the supply side, only the thermal system has significant real operating costs, making these a much smaller proportion of total system costs than for pre-dominantly thermal systems. This means that the impact of price/dispatch distortions on total supply costs is also proportionately smaller.

But it was also well understood that the actual degree of any short run distortion, and also of any short run wealth transfer, would be heavily dependent on the level of contracting. In a perfect world, all loads might be contracted for 100% of their expected requirements, using option contracts which ensured that they were fully hedged for that expected requirement, but also 100% exposed to spot prices for any deviation. Under those conditions it can be shown that the incentives of producers to deviate from SRMC offers is actually minimal, so distortion becomes a non-issue, and risk is also minimised.<sup>34</sup>

On the demand side, it was also believed that, while low demand elasticity may imply significant short run volatility of SRMC prices, and possibly allow significant deviations from SRMC prices, it also suggests the actual economic impact of such deviations will be small.<sup>35</sup>

In any case, even if spot prices are highly distorted, heavily contracted loads will have minimal real risk exposure. Theoretically, their marginal decisions should still be affected, but this is only true if spot prices are actually passed through to them. In reality, the vast majority of retail customers in New Zealand, accounting for a significant proportion of the load, do not actually face spot prices in real time. In fact

---

<sup>32</sup> There was never any reason to expect that prices would be lower than they had been historically, either. As explained by Culy et al, electricity pricing in New Zealand, particularly for domestic consumers, had been historically driven as much by politics as by a requirement to recover costs, and it was not considered desirable that that situation should continue. And of course development costs were expected to rise, because cheap accessible hydro development options had either been exploited or protected from development, while the introduction of the Resource Management Act meant that environmental concerns would have significant cost impacts on new projects, rather than being over-ridden by statutory declarations, as had often happened in the past.

<sup>33</sup> Section 4 discusses the difficulty in defining SRMC for such sources, but the point here is that, while these “opportunity costs” do define a valid SRMC, at the margin, they are not, of themselves, real costs to the economy.

<sup>34</sup> See T.J. Scott and E.G. Read: "Modelling Hydro Reservoir Operation in a Deregulated Electricity Sector", *International Transactions in Operations Research*, vol.3, no.3-4, 1996, p. 209-221.

<sup>35</sup> That is, consumers will not curtail their normal activities by much when prices rise, at least in the short term. They may suffer a personal or commercial loss as a result of paying higher power bills, but that is not a welfare loss to the nation, merely a wealth transfer.

they may not see any change to price signals at all, even when spot prices are elevated for several months. This has obvious implications for any consideration of the supposed wealth transfer effects of these prices. But it also has significant implications for market design choices. To the extent that the economic rationale for enforcing SRMC pricing in the spot market rests on the belief that this will enhance allocative efficiency by reducing distortion to consumption patterns, that rationale is undermined by the observation that the prices charged to decision-makers controlling consumption do not reflect the dynamic structure of spot prices anyway. Accordingly, it was considered that the inefficiency due to deviation from SRMC in the spot market, while still significant, would probably be less than that arising from other distortions in the sector.<sup>36</sup>

In summary, then, it was considered that if a compromise had to be achieved between short and long run efficiency, it was better to err on the side of fostering long run efficiency. Thus the key issue was believed to be reducing barriers to entry, and avoiding intrusive regulation, not just because of the direct expense involved, but also because of its likely negative impact on productive efficiency. The implication is that, in this market design, fostering allocative efficiency by aligning prices with SRMC was, at best, to be a secondary consideration. In fact, we have argued that the design actually relies upon prices deviating significantly from SRMC, on a regular basis, to provide a sustainable environment for long run capacity investment.

Our goal here is not to argue for a particular market design, or to explore options for what the NZEM design could, or should be. But the extent to which prices should be allowed, or expected, to deviate from SRMC depends partly on market design choices. Thus the next few sections briefly consider the rationale behind design choices made in three key areas, and examine the implications of those choices, in terms of their expected impact on behaviour in the market, and performance of the market. Those choices relate to three key questions, namely:

- Why are offers not regulated?
- Why are prices not capped?
- Why is there no capacity component?

---

<sup>36</sup> Historically, a much greater distortion resulted from the fact that, under central planning, electricity prices were not varied in response to changing hydrological conditions at all. And, in the current context, one would also think that a much greater distortion arises, at least for domestic customers, as a result of limits being placed on fixed charges, thus forcing costs to be recovered by adding substantial mark-ups onto the energy price component, whether or not it reflects SRMC.

### **3.3 Why are offers not regulated?**

Despite the emphasis on achieving long run efficiency, consideration obviously could, and was, given to mechanisms designed to achieve maximum short run efficiency as well, thus providing the best of both worlds. One obvious option would be to try to force offers to match SRMC. But that option was rejected, for two main reasons.

First, the decision was partly motivated by the difficulty of objectively determining what SRMC might actually be, in a hydro dominated system. Those difficulties, which are elaborated in Section 4 – *SRMC for Hydro and Energy Limited Thermal*, were very much appreciated by the designers of the NZEM, who had extensive experience with the development of computer models to perform such assessments. Thus it was thought wise to avoid a market design in which alignment with SRMC was a primary goal, implying a requirement to make, justify, and debate such assessments a major focus of activity.

More generally, the idea that intrusive regulatory intervention might lower costs was considered to fly in the face of conventional regulatory wisdom, at least as understood in most sectors other than electricity. Productive efficiency gains seemed most unlikely at the organisational level, where the transaction costs involved in that whole process, including the de-motivating and distracting impact of intrusive investigations and interventions, would most likely outweigh any benefits. Efficiency gains would be conceivable at the sectoral level, though, if it could be shown that the loss in coordination (allocative) efficiency, due to distortions away from SRMC pricing under the status quo, were greater than the increased transaction costs, plus losses in productive efficiency within firms, and dynamic (investment) efficiency, due to regulatory intervention. After much debate, though, the WEMS study concluded that this was not likely, partly because the actual impact of SRMC pricing at the wholesale level would often not be passed through to the retail level (because of contracted prices), as discussed in Section 3.2 – *Basic Design Philosophy* above. The overheads of establishing such a function in the small New Zealand market were also considered to be a significant issue.

Second, though, it was considered that forcing offers down to SRMC levels would actually not be desirable, in terms of maintaining a long run equilibrium, with acceptable capacity margins, for the reasons already discussed in Section 2.4 – *Risk Aversion and PDC Inflation*. As discussed in Section 2.6 – *An Integrated Transaction Cost Perspective*, it was believed that the electricity sector should evolve toward a paradigm which has proved successful in other sectors, under which prices might deviate significantly from SRMC. In the absence of a perfect contract market, this was thought necessary in order to support sufficient entry by risk averse investors, and also to provide discipline to that contract market, and encourage consumer contracting, as discussed in Section 2.6.

In most sectors, it is also clearly understood that the market simply will not work if the supplier is restrained from charging premium prices to customers who refuse to book ahead. What incentive would anyone have to book ahead if they knew that a regulator would force suppliers to make seats/rooms available at a near zero SRMC to last minute purchasers?<sup>37</sup> And what incentive would a potential hotelier have to invest, if they suspected that a regulator might intervene in this way? The overall effect would surely be to delay investment until accommodation shortages became common enough that hoteliers could reasonably expect to make an acceptable profit, given the risks, by charging premium prices when all accommodation was fully booked.

Similarly, a requirement to force electricity suppliers to offer SRMC prices in the spot market could reasonably be expected to kill the contract market for electricity, and thus to make entry riskier, and less attractive. The overall effect would again be to delay and distort investment, raise prices, and increase the frequency of shortages, as discussed in Section 2.4 – *Risk Aversion and PDC Inflation above*. Consumers may find such measures attractive, in the short term, because they depress prices temporarily, and have the appearance of “controlling market power” by banning “capacity withholding”. But, while one may be able to force incumbents to make existing capacity available, forcing potential entrants to create new capacity is another matter. Such measures will not really serve consumer interests, in the long run, if their effect is merely to ensure that the capacity needed to meet consumer requirements is “withheld” from the investment market.<sup>38</sup>

Of course another option would be to regulate contract prices, rather than spot market offers or prices. Simply regulating prices would not suffice, though, unless contracts were actually available. Thus consideration was given to requiring generators to offer contracts at regulated prices. If the entry market is reasonably competitive, this kind of intervention seems unnecessary, since contract prices should ultimately be disciplined by the contracts offered by competitive entrants. Still, the prospect of driving prices down will always seem attractive in the short run. There would be limited value in pursuing such a policy, though, unless it could be effective in depressing prices over the long run. And we have already argued, in Section 2.4 - *Risk Aversion and PDC Inflation*, that forcing prices down below a level capable of supporting risk averse entry will distort investment patterns and imply a greater likelihood of shortage than would be considered optimal under central planning paradigm. In fact Section 2.4 argues that, without an explicit capacity payment, entry

---

<sup>37</sup> In reality there would still be some incentive, but only at peak times when customers may fear there will not be enough capacity, in aggregate. But that motivation would also be largely removed if a regulatory authority were to impose “capacity standards” and “supply obligations” on these sectors, as is not uncommon in the electricity sector.

<sup>38</sup> Ironically, though, consumers may continue to support such short sighted policies, even in the long run, because their reference point is the capacity that has actually been built, and the efficient utilisation of that capacity. Unfortunately, they have no way of knowing what investment opportunities have been deterred, and how much this has driven prices up.

of peaking plant could never be supported at all, if prices could never rise above the SRMC of such plant.<sup>39</sup>

Still, it might be thought that at least such intervention could produce a sustainable long run equilibrium with lower prices, so long as the reduced security standard was considered “adequate” by the regulator and/or entry of sufficient peaking plant could be subsidised. Unfortunately, this is not true, though, unless demand is actually declining faster than the rate at which existing capacity fails. If any new, or replacement, capacity is to be built at all, prices must eventually rise to the level where that capacity becomes economic. Thus, for example, the regulator could insist on contracts being available at expected SRMC prices, and this may depress prices temporarily. But the long run impact must be to delay entry until the capacity situation is tight enough that the SRMC based PDC, including shortage components, is high enough to finance that new plant. In other words, the long run PDC may be distorted, but average price levels must be essentially the same, despite the intervention. In fact, we should expect the PDC to be higher, if regulatory action increases perceived risks for potential investors.<sup>40</sup>

Overall, routine regulation of spot or contract offer prices did not, and does not, seem a particularly attractive option, and was rejected by both WEMS and WEMDG. Given the emphasis, on long run efficiency, it was felt that regulatory attention would be better directed to reducing entry barriers, for example.<sup>41</sup>

---

<sup>39</sup> Technically, entry might not be deterred if the potential entrant could be assured that the allocation of discounted contracts was a one-off event, never to be repeated. But, once such intervention has occurred, it is hard to see how anyone could be certain it would never happen again. And the prospect of such intervention poses a two-fold threat for an entrant, who must consider the probability of later finding their own position being undercut by new discounted contracts issued by other parties, or being forced to issue discounted contracts of their own.

<sup>40</sup> Although the PDC could be lower, if the intervention was implemented in a way that reduced perceived risk, eg by guaranteeing contracts for potential entrants.

<sup>41</sup> The offering pattern of incumbents is not irrelevant in that context, since it may form part of an entry deterrence strategy. Such gaming strategies relate to market power issues that are not considered here, but the plausibility and likely effectiveness of that type of strategy in the New Zealand context was considered speculative, and pervasive regulation did not seem justified simply as a precautionary measure.

### 3.4 Why are prices not capped?

Most other electricity markets impose some kind of cap on prices, although that cap may be set at very high levels in other energy-only markets, such as Australia. Obviously, this is an option that could be, and was, considered for implementation in New Zealand, too. In part it was rejected because of a general aversion to regulatory intervention. But it is also not consistent with the general market design paradigm, and theoretical framework described in previous sections.

Market price or offer caps obviously have a direct impact on the PDC, and hence on the economics of entry. Any (actual or prospective) capping of market prices implies a diminution of (actual or prospective) revenue to both incumbents and potential entrants, and thus implies a *prima facie* risk of deterring entry, leading to under-supply of capacity in the long run.<sup>42</sup> In theory, the optimal plant mix, under perfectly competitive or centrally optimised assumptions, must imply a finite probability that generation capacity will be fully utilised. And that implies a finite probability that prices will have to rise high enough to reduce demand, without any form of physical intervention, in those situations of full capacity utilisation.

Accordingly, if price caps were to be imposed, or if potential entrants think that there is any possibility of such caps being imposed in future, capacity adequacy could only be assured by one of two mechanisms. Either:

- Some means must be found to reward capacity by payments additional to those received from the energy market; or
- Participants must be allowed recover the deficit by pushing prices above their perfectly competitive SRMC levels when capacity is less than fully utilised.

As discussed in Section 3.5 – *Why is there no capacity component?*, WEMS actually proposed a two-part energy/capacity market in which participants would have received payment for capacity, as well as for energy. But that proposal was not implemented, and the point here is that, if no capacity payment is provided, the value taken out of the market by capping the price must be replaced by some other means, if optimal entry is to be supported. The only way this can occur, in an energy-only market, is by allowing the sub-cap PDC of energy prices to inflate, as discussed in Section 2.4 – *Risk Aversion and PDC Inflation*. That is, the lower the price cap, the greater the extent to which prices must be allowed to settle above SRMC at other times. The alternative, if SRMC pricing is enforced right across the PDC is to accept greater distortion of the long run entry profile.

---

<sup>42</sup> Here we interpret “capacity adequacy” in terms of the optimal economic level of capacity, that is the level at which the marginal benefit of extra capacity equals its marginal cost. Of course this may differ significantly from public/political perceptions of capacity adequacy.

Thus no explicit price cap was imposed in the WEMS design. It was widely felt, though, that the industry was subject to an implicit “threat of regulation”, and that this threat would inevitably impose limits on how high prices could rise, and how long high prices could be sustained, during any real crisis. In other words it was thought likely that if a serious and prolonged crisis occurred, most likely in a dry year, the Government of the day would not sit idly by and let the industry raise prices to their theoretically optimal level: That is to the level at which price alone was sufficient to reduce demand back to a level that could be met by available capacity. Instead measures would most likely be introduced to subsidise entry of alternative supplies, and/or to force prices down, while rationing demand by other means.

An implicit price cap of this nature has much the same impact on the top end of the PDC as an explicit cap, and thus implies a similar requirement to inflate prices above SRMC over the lower part of the PDC. In other words, if market participants believe there may be limits on their ability to charge what the market will bear during periods of extreme short supply, they must compensate by charging more than SRMC during other times.

As it happens, the scenario that has unfolded is that, rather than introduce a capacity market, entry of one particular peak-opping plant has been subsidised, without making equivalent capacity payments available to other market participants. This seems like an attractive short term expedient, but it should be recognised that using a subsidised peak-opping plant (ie Whirinaki) to effectively cap the top end of the PDC creates similar issues to imposing a price cap. The price capping potential of Whirinaki was demonstrated by the Electricity Commission during the winter of 2008 when market prices were affected by the offering of Whirinaki below SRMC, with unrecovered costs in the market being recovered by the EC levy.

Such “subsidised” entry could actually be economically optimal, in the sense that the plant might be expected to operate profitably, at a reasonable commercial discount rate, on the basis of receipts from spot market sales.<sup>43</sup> If so, the resultant PDC could also be optimal, and entry by other plant types would not have been unduly discouraged. But if a genuine subsidy does need to be paid, in expected value terms, it must be that SRMC prices, at the top end of the PDC, are not enough to cover the FC of this entry. In other words  $OV < FC$ , for the subsidised plant. But this is also true for all other plant.

---

<sup>43</sup> This does seem possible if entry is otherwise being deterred by factors that made it too risky. It seems likely if loads are reluctant to contract, perhaps because they believe that they may not secure the benefits of contracted capacity for their own exclusive use, physically or commercially, if a real crisis occurs, and/or prefer to rely on the political process for protection. Thus the “subsidy” required may be more in the form of guaranteeing expected revenues so as to reduce the risk premium, than increasing expected revenues. Effectively, the regulator would be contracting on behalf of all consumers, collectively, because the transaction costs of doing so are lower than the transaction costs of each individual trying, and probably failing, to negotiate an acceptable contract on their own.

Conceptually, imposing a cap on the energy market price may be thought of as equivalent to allowing the market price to find its natural level, above that cap, but then automatically issuing every MW of load with a (retrospective) 1 MW call option, the strike price of which is set at the cap. If that cap/strike price were to be set at the SRMC of the most expensive plant in the system, then that plant could never make an operating profit from spot market sales. In order for the plant mix to be optimal, the OV of the equivalent call option (as determined by that part of the PDC where prices exceed this maximum supply-side SRMC) must still equal the FC of that plant. But the market can now only be in long term equilibrium, with sustainable entry of peaking plant, if that plant, at least, receives a capacity payment to cover its FC.<sup>44</sup>

Recall, though, that the OV for any plant is just the value of a call option applying in all periods where the MCP exceeds its SRMC, including those periods when it also exceeds the market price cap. In other words the optimal (uncapped) OV for plant with lower SRMC equals the optimal uncapped OV for peaking plant, plus the value of a call option based on capped market prices, and applying all the time when MCP exceeds that SRMC. So, a market price cap that reduces the OV of peaking plant will reduce the OV of all capacity by exactly the same amount. So, whatever subsidy is required to make investment in peaking plant profitable, the market must also pay the same amount, per MW, to all other capacity, if an optimal plant mix is to be maintained.

In particular, capping prices at the SRMC of peaking plant would reduce its OV to zero, thus requiring a subsidy equal to the full investment cost of peaking plant,  $FC_{\text{peak}}$ , in order to maintain an optimal investment level for such plant. And the same will be true for all other plant types in the optimal plant mix. In the absence of a capacity market, these cost recovery requirements can only be met by allowing a markup on SRMC prices, set so as to restore the PDC to a level that is just sufficient to support optimal entry for each plant type in the optimal capacity mix.<sup>45</sup>

In other words, capping the PDC carries with it the implication that the remainder of the PDC must somehow be inflated in a similar manner to that discussed in Section

---

<sup>44</sup> This follows because no markup is possible on its SRMC price, which forms the market price cap. More generally, the cap could be set to some higher level, so that some operating profit is made, and only a partial subsidy is required. The same will be true if the market price is only capped by a subsidised entrant, because the market prices can then be expected to rise above the “cap” on some occasions.

<sup>45</sup> Specifically, assuming that  $FC_{\text{peak}}$  is expressed in terms of an annuity, a plant operating for H hours per year, on average, must receive an average price premium of  $FC_{\text{peak}}/H$  \$/MWh, over and above the SRMC price which might be expected assuming an optimal plant mix, under perfect competition. The logic of Section 2.4 suggests that, in order to sustain entry of plant near the top of the merit order, these markups would have to be concentrated near the peak period, when that plant operates, and this may not be possible if the SRMC of that plant is close to the market price cap. Thus it may be necessary to set the price cap well above the highest SRMC. Otherwise, a range of high SRMC plant may still need to be partially subsidised, even if they are able to price right up to the market cap, when operating.

2.4 – *Risk Aversion and PDC Inflation.* This is not to suggest that price capping, or subsidised entry, will allow, or facilitate incumbents to raise prices in the short term. If incumbents have insufficient market power they may well have to accept a loss in value in the short to medium term. Thus consumers may benefit from lower prices over that period, too. But the point is that, in the long term, entry will be deferred until it can be supported by the capped market PDC: That is, until the uncapped portion of the PDC rises high enough above its optimal level to offset the loss in value from capping at the top end.<sup>46</sup>

In summary, market price caps are employed in many markets, for fairly obvious reasons, but they seem theoretically problematic, and were not favoured by WEMS or WEMDG. In principle, capping market prices distorts the PDC, and leaves us with the option of subsidising plant to operate during the time when the market price binds, or perhaps accepting a sub-optimal plant mix. Thus even the Australian market, which sets its market price cap to a very high level, also retains a “reserve trader” concept, under which some plant is contracted to operate only when the price cap binds.

---

<sup>46</sup> It has been sometimes been suggested that intervention of this form risks starting the market down a “slippery slope” scenario, under which more and more capacity, of all types, must be subsidised to enter. Provided enough peaking plant continues to be subsidised, though, it should be possible to keep the probability of shortage to an optimal level, or less, and to keep prices below their optimal level, if not down to the SRMC of peaking plant, at the top end of the PDC: That is in those periods when the peaking plant operates. But, even with SRMC pricing, the PDC can still inflate by deferring entry, and shifting investment from more to less capital intensive plant, including the subsidised peaking plant. Indeed it must inflate in this way if long run equilibrium is to be maintained. The result is a sub-optimal plant mix, with higher costs, and of course higher prices are required to cover those costs but, although no formal proof has been attempted, it does seem possible that a sustainable equilibrium could exist. There is still an incompatibility, though, between maintaining SRMC pricing and maintaining an optimal plant investment pattern, in an energy-only market:

### **3.5 Why is there no capacity component?**

Finally, many jurisdictions have adopted some form of market, or centralised contracting, for capacity. One way to do this would be to require loads, or load serving entities, to buy “compulsory insurance”, in the form of “capacity tickets”, or “cap contracts”,<sup>47</sup> as proposed by WEMS. The implication would be to force the price of such contracts up until it is high enough to underpin entry of whatever capacity is needed to meet a security standard considered “acceptable” by the regulator.

This was the original WEMS design and, under that proposal, it was hoped that the price of capacity tickets could eventually be set entirely by market forces, both on the supply and demand side. Thus it was hoped that, ultimately, the provision of such an instrument would allow trading to reach an economic equilibrium, in which purchasers of capacity tickets were satisfied that they had bought an adequate level of “insurance”, at a price which allowed capacity ticket suppliers to recover their costs.

It was expected, though, that the market would have to be “managed”, at least initially, by setting a capacity ticket coverage requirement to be met by load serving entities. Thus the level of security, and corresponding demand for capacity tickets, would be set by some non-market process and could, in principle, be made arbitrarily high. But the market could still reach a sustainable equilibrium to supply that amount of capacity, even if the capacity standard was actually excessive in economic terms.<sup>48</sup>

This kind of market design imposes some overheads, but reduces the risk for potential entrants, and particularly for peaking plant. So it may be expected to lead to greater competition, lower risk premiums, and lower prices, in the long run. Many North American markets include some form of capacity payment mechanism, and some academics have recently recommended designs very much like the original WEMS design.<sup>49</sup>

---

<sup>47</sup> That is call options, with a relatively high strike price, effectively creating a market price cap, from a load perspective.

<sup>48</sup> Despite the hopes expressed by WEMS in this regard, it was clearly felt that the capacity level which any regulatory authority might set was likely to be higher, even in the long run, than the “economic: capacity level: That is, the capacity level that customers would freely choose, if faced with the true cost of meeting the standard, assuming they had sufficient understanding of the situation, could contract robustly enough to secure the benefits of contracted capacity, and had no incentive to game the political process. If so, that would exacerbate many of the problems discussed here, but does not really change the nature of those problems, or of those conclusions.

<sup>49</sup> P Cramton and S Stoft: *The Convergence of Market Designs for Adequate Generating Capacity with Special Attention to the CAISO’s Resource Adequacy Problem* A White Paper for the Electricity Oversight Board 25 April 2006

There is no universal agreement on this issue, though. It would be fair to say that, while WEMS concluded in favour of a two-part energy/capacity market design, opinion within the WEMS study group was actually fairly evenly balanced. After further consideration, the more broadly representative WEMDG group clearly favoured the energy-only design. And it should be said that, when the effect of dry years, ad hoc intervention, and supply side shocks is stripped away, we are not aware of any convincing evidence that this market design actually has produced a capacity short fall in New Zealand.<sup>50</sup> It should be recognised, too, that this market design is by no means unique to New Zealand. A number of other markets, including Australia and Singapore, have adopted energy-only designs with apparent success. Most recently, Texas has adopted a design very similar to the New Zealand market, after many years of experience with alternative market paradigms, and extensive observation of alternative market designs operating elsewhere in North America.

Of itself, neither option is really ideal. Theoretically, customers in an energy-only market may expect to face higher prices, greater price volatility, and more frequent outages than might be considered “ideal”, and than they would face in a market with capacity payments, or traditional regulation. But society may prefer to opt for this market design if the transaction costs of contracting, or establishing more elaborate and/or intrusive market regulation to avoid this situation, exceed the benefits from doing so. Thus the energy-only market design may be optimal if the costs imposed by the obvious dis-benefits are less than the transaction costs of adding a capacity component to the market design, and/or imposing more rigorous regulation.

In particular, concern may be expressed that this arrangement gives the body setting capacity requirements considerable power to set requirements in excess of what market participants would willingly pay, if contracting on their own behalf. The resultant distortion to the plant mix could well be greater than that implied by not having a capacity market in the first place, and the cost would ultimately have to be borne by consumers. WEMDG, which included significant consumer representation, obviously found these arguments persuasive and, while alternative proposals have been raised from time to time, consensus in the industry probably still supports that position, on the grounds that:

- The transaction costs of imposing a contracting regime, or establishing more elaborate and/or intrusive two-part market arrangements, may well be more than the costs of persisting with an energy-only market design; and

---

H-P Chao and R Wilson : *Resource Adequacy and Market Power Mitigation via Option Contracts*  
Electric Power Research Institute 03/18/2004

S Oren “Generation Adequacy via Call Options Obligations: Safe Passage to the Promised Land”  
*The Electricity Journal* Volume 18, Issue 9, November 2005, Pages 28-42t

<sup>50</sup> An appendix outlines how we believe this question could be answered, but we are not aware of any authoritative study along these lines actually having been done in New Zealand.

- Despite public perceptions many analysts believe that the energy-only market design is actually performing well enough, in terms of providing sufficient capacity, and reasonable prices.

The point here, though, is not to debate whether WEMDG's judgement was, or is, correct, or to promote any alternative design. The point is merely that the NZEM now operates according to an energy-only design. As such, it relies upon prices deviating significantly from SRMC, on a regular basis, so as to provide a sustainable environment for long run capacity investment.

### **3.6 Conclusions**

It should be clear, from our discussions that we consider the NZEM design to be predicated on the assumption that significant deviation from SMRC pricing is not only acceptable, but necessary, at least in some situations. Without that freedom, we consider it unlikely that participants would be able to obtain sufficiently high spot or contract prices to underpin the economics of sustained new entry, particularly for peaking plant. This is particularly so when one considers the inherent risk involved in investing in such plant. Thus, in opting for an energy-only market design, WEMDG acted consistently by not placing limits on offers or market prices.

If prices or offers were to be limited, the likely intent would be to force prices down and/or capacity provision up, in the short run. In the long run, though, these two goals seem incompatible. If a sustainable equilibrium is to be maintained in the market, intervention must be accompanied by, or expected to induce, a balancing reaction:

- Intervention to force market prices down, on any occasion, must be offset by an expectation that, in the long run, prices will rise on other occasions, either because participants withdraw existing capacity when it becomes uneconomic to maintain that capacity in the spot market, or because potential entrants withhold potential capacity from the investment market, until average prices cover entry costs, with sufficient certainty.
- Intervention to force capacity provision up is really only possible if the regulatory authority itself enters the market as a buyer of capacity, or requires market participants to do so. If that occurs, though, prices must ultimately rise to induce, or at least cover the costs of, extra capacity provision.

In particular, capping prices, or requiring electricity suppliers to offer SRMC prices in the spot market, or the threat that this could happen in future, can be expected to have a negative impact on the contract market for electricity, and to delay competitive entry until the expected PDC, and the risk of shortage, rise higher than is likely to be considered desirable. Selectively subsidising entry may serve to keep the probability of shortage down to an acceptable level, and this has actually occurred in the NZEM.

Theoretically, though, it implies the likelihood of distortion to the remainder of the plant mix, and raises issues which, in our opinion, remain unresolved in the NZEM at this time.

One possible market design would force spot prices down to SRMC, while guaranteeing capacity payments in some way. The regulator could run a competitive tender for capacity contracts, or require loads to do so. WEMS proposed the latter, using “capacity ticket” contracts for peaking (or more exactly in the NZEM context, “dry year backup”) capacity. This kind of market design might improve both long run and short run efficiency, but it might not perform significantly better than the current design. It does impose overheads, and create problems of its own and thus it was ultimately rejected by the NZEM designers.

As a result, the NZEM became a simple unconstrained energy-only market. Theoretically, if risk were not an issue, and/or contracting perfect, such a market might produced a perfect alignment between both short and long run economics, with spot prices at SRMC and contract prices at LRMC. But risk is an issue, and contracting imperfect, in this market as in any other. Thus adequate entry is expected to be partially supported by allowing spot prices to exceed SRMC, particularly at peak times, but also in other circumstances. The threat that prices will significantly exceed SRMC is also a fundamental part of the market design, since that threat is supposed to motivate forward contracting by loads, and hence entry by alternative suppliers.

In other words, this market has been designed to operate just like the vast majority of successful markets operating outside the electricity sector, and with similar cost structures, where pricing above SRMC has always been considered absolutely normal.

## 4 SRMC for Hydro and Energy-Limited Thermal

Discussion of economic behaviour in electricity markets often focuses on the extent to which prices are considered to deviate from SRMC. The previous sections have suggested that prices may actually have to deviate from SRMC, perhaps significantly, in order to produce a sustainable equilibrium, particularly in an energy-only market design such as the NZEM. But this section focuses on the other side of that question, namely determining what SRMC might actually be in a system dominated by hydro and energy-limited thermal plant.

It will be seen that this is actually quite a complex question, and that the SRMCs of hydro, gas, and coal plant can be expected to exhibit quite complex patterns, correlations, and connections, over daily, weekly, monthly and annual time scales.<sup>51</sup> We should make it clear, though, that none of the discussion in this section relates to deviations from SRMC pricing, let alone market power. In this section we assume SRMC pricing, and merely:

- Explain the kind of price patterns, correlations and connections that would arise, internally, within any sufficiently detailed centralised optimisation;
- Note that exactly the same patterns, correlations and connections should be expected in a hypothetical perfectly competitive market; and
- Argue that the same general conclusions should (hopefully) apply in real markets, if they are working properly, even though participants may not be able to clearly articulate or analyse how all of these factors interact.

---

<sup>51</sup> Strictly speaking, SRMC is actually difficult to define in thermal systems too, even without consideration of energy limits. This is because unit commitment decisions must be made, perhaps on a daily or weekly basis and, once committed, plant may not wish to shut off even when market prices are below their fuel costs for a few hours. Similarly, once de-committed, plant may not wish to start up, even when market prices are above their fuel costs for a few hours. This implies variation in the effective SRMC over a daily or weekly cycle. This observation applies to some plant in the New Zealand system, too, but it will be ignored here because it is a relatively less important feature, and is also relatively well understood from studies in other markets.

## 4.1 SRMC for Major Reservoirs

Although our discussion has already referred to conditions in the hydro-dominated NZEM market, none of that discussion is actually specific to hydro systems. Electricity markets are inherently risky, and all markets face the central problems of coordinating short run supply and demand side activities, while incentivising efficient entry by risk averse investors in the long run. Thus the basic market design considerations are the same. But some of the problems discussed above are exacerbated in a hydro market setting.

Obviously, hydrological risk is a major factor in such markets. In the absence of any storage capacity, hydro generation could only utilise flows as they arrived, and load would have to be curtailed to match those flows. It may be argued that the SRMC of hydro generation would be zero in such a market, but this would only be true when flows exceeded what was required to meet the load level which might be induced by a zero price. The rest of the time, the effective SRMC of hydro would effectively be infinite, or at least indeterminate, and the “SRMC” market price would actually be set by the marginal cost of curtailing load to the level which could be generated, given the real-time inflows. So the market could be expected to experience a volatile bimodal price distribution, alternating between zero, during times of surplus, and shortage cost levels, during times of relative shortage.<sup>52</sup>

This price pattern may be thought of as similar, on average, to that in a variant of the traditional regulated pricing regime that relied entirely on capacity/peak payments, with no “energy” charge at all.<sup>53</sup> All the above theory still applies, though, and the expected long run average price level should still be that required to induce new entry, given the market risk.

Introducing thermal generation, still with no hydro (or fuel) storage, would reduce price volatility by introducing intermediate price steps, corresponding to the SRMC of each thermal unit, with a significant probability that the price would lie at one of those levels. But this does not fundamentally change the situation and, again, the above theory still holds.

Introducing hydro storage has a more radical impact, though. Clearly it will mitigate uncertainty by allowing flows to be stored for use in the most needy periods, and this will reduce short run price volatility. But this means that the SRMC of hydro is no longer zero (or infinity), but is given by a “marginal water value” (MWV), which is the “opportunity cost” or “option value” of a marginal unit of water stored for future

---

<sup>52</sup> The frequency of shortages would probably be quite high, in such a system, although the price implications of modest shortages may also be modest.

<sup>53</sup> And that is exactly the way the “Bulk Supply Tariff” was structured in New Zealand, for many years, while the system was purely hydro.

use. In a pure hydro system, that marginal unit of stored water may ultimately be spilled, in which case the opportunity cost of using it today will turn out to be zero. Or it may ultimately help to reduce a future shortage, in which case the opportunity cost of using it today will be some kind of load reduction, or shortage cost, and may be very high. But the effective SRMC for hydro generation is given by the expected marginal water value at any time, and this will not normally lie at either extreme, but vary continuously as storage levels, and expectations change.

The calculation of these expected MWVs lies at the heart of reservoir management optimisation, in a centrally planned system.<sup>54</sup> In New Zealand, the old NZED STAGE model, and the PRISM/SPECTRA models developed by the MoE, both made that calculation explicit, and this is also true of newer models such as SDDP. In other models the MWV calculation is implicit, but mathematically equivalent. Thus the “SRMC” of releasing water from a storage lake is almost always the “expected opportunity cost” of not having that water available for use in some future period.<sup>55</sup>

If there is no thermal generation in the system that expected opportunity cost will be a weighted average of the spill value (zero) and the shortage cost arising when demand can not be fully met. Thus it will vary as a function of the calculated probability of spill occurring before the next time the reservoir is empty, or conversely of the reservoir being empty before the next time it is full. And that varies as the state of the reservoirs varies, over time, but will clearly be lower if the reservoirs are relatively full, for the time of the year, thus reducing the probability of future shortage.

The addition of thermal generation to the system does not fundamentally change this, but tends to mitigate the effect. As the proportion of thermal increases, so does the probability that prices will be set directly by the SRMC of some thermal generator, rather than by the expected MWV of some hydro generator. Thus the weight given to expected spill and shortage events may actually be quite small in determining MWVs, most of the time. Instead, expected MWVs will normally be set by the likelihood that a unit of water saved now will eventually be used to displace generation from some

---

<sup>54</sup> Extensive discussion of the theory of MWV calculation for the deterministic case may be found in E.G. Read [1982a]: *Economic Principles of Reservoir Operation I: Perfect Foresight*, International Short Course on Reservoir Scheduling, University of Tennessee, Knoxville, (CBA Working Paper No. 151. E.G. Read [1982b] *Economic Principles of Reservoir Operation II: Uncertain Future*, International Short Course on Reservoir Scheduling, University of Tennessee, Knoxville, (CBA Working Paper No. 152) extends it to the stochastic case.

<sup>55</sup> The only exceptions are when the reservoir is either empty or full, in which case, the “expected opportunity cost” relates to the marginal cost of the thermal generation, or load shortfall, in the current period, because the optimal policy is to release whatever inflows arrive, for a while, and the MWV equals avoided by passing through those inflows. Where water must be released, or spilled, because is not possible to store any more water for the future, this MWV may be zero.

thermal generator at some future date, the value of which will be determined by the SRMC of operating that thermal generator.<sup>56</sup>

## 4.2 SRMC in River Chains

It should be recognised that the theory discussed above applies to all reservoirs and “headponds”<sup>57</sup>, at all levels in the system, and over all time periods. In all cases the relevant opportunity cost is calculated over the period until the reservoir storage bounds are next expected to be reached.<sup>58</sup> And in all cases the MWV must change when such a bound is reached. Specifically, Read [1982a] explains why, somewhat counter-intuitively, the MWV must rise whenever an upper storage limit is reached, and must fall whenever a lower storage limit is reached.<sup>59</sup>

If the reservoir is large, then it will typically reach its upper and lower limits, or at least threaten to reach its limits, around the same time each year, thus operating under 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 a small reservoir will exhibit exactly the same kind of behaviour over a shorter period, often operating on a weekly or daily cycle. So it may typically reach its upper 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

---

<sup>56</sup> Or at least that is the conventional wisdom, derived from markets in which it can be assumed that the SRMC of operating a thermal generation is itself well-defined. As discussed in Section 4.3, though, that is not necessarily the case where thermal plant is “energy limited”, as may often be the case in New Zealand.

<sup>57</sup> These are small storages, immediately above a hydro station, often with only a few hours storage capacity.

<sup>58</sup> It is easiest to think about a deterministic problem here, where we know the inflows, and can determine the optimal time at which storage should next reach on or other limit. The stochastic version of this theory, as described by Read [1982b], is quite complex. The principles discussed here carry through to that case, though, except that changes occur more subtly and continuously, as expectations change over time.

<sup>59</sup> This result, and the timing of the change, is clear-cut in a deterministic optimization model. In reality, because of uncertainty, operators try to avoid having reservoirs actually reach their storage bounds, and the MV change occurs a little more gradually, over several periods, as the threat of reaching the bound builds up, and then recedes. But this requirement to try to avoid actually reaching the limits means that the effective bounds on storage range are actually tighter than a deterministic model would imply, and managing storage to those tighter effective limits means that the total MWV change over the periods involved must actually be greater than for a deterministic model. Thus consideration of a deterministic model still provides a reasonable guide to real-world behaviour.

peak. And this means that its MWV must also cycle daily, rising before the morning peak, then falling after the evening peak.<sup>60</sup>

When reservoirs are linked into river chains the situation becomes much more complex. While it is common to talk, for example, about the MWV of “the Waikato river chain”, this is not a well defined concept. Each reservoir, or headpond, 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 delay times. In such a chain, the SRMC of release is not determined by the MWV of the releasing reservoir, at the time of release, either. It is determined by the difference between the MWV of the releasing reservoir, at the time of release, and the MWV of the downstream reservoir, at the time that incremental release is expected to arrive there, which may be several hours later.

Conversely, the MWV of the upstream reservoir must be determined by a trade-off between the opportunity cost of not keeping water in that reservoir, for later release, and the opportunity cost of not having that water arrive at the downstream reservoir, for release there, after some delay. In each case, though, the opportunity cost must be calculated on the basis of the opportunities available before that reservoir next reaches a storage bound. And the periods involved may be very different for the two reservoirs because they may be of very different size, and (given the delays) at very different stages of their daily cycle.<sup>61</sup>

The point of this discussion is not to develop an optimisation algorithm to resolve these issues, but to note their complexity. That complexity becomes much greater once it is realised that one can not resolve the issue by considering just two stations. One would have to iterate both up and down a whole chain of stations, to find a generation dispatch solution, and MWV pattern, that was simultaneously optimal for all stations in the chain. Overlaying uncertainty about both inflows and market prices

---

<sup>60</sup> A really small reservoir may have two cycles in each day, one for each peak. In the limit, a station with no storage 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. (As noted earlier, the SRMC of increasing supply from such a station is not zero, as is sometimes asserted, but indeterminate, because it can not produce any more than the minimum of its capacity or the inflows it receives.)

<sup>61</sup> Suppose, for example, that the delay time is 2 hours, peak load is at 6pm, and both reservoirs actually hit their storage minimum soon after, say at 7pm. Then the MV in both reservoirs will be high until 7pm, but then drop suddenly. But the SRMC of release from the upstream reservoir is not determined by the MWV, but by delayed MWV difference. And that difference will actually rise suddenly at 5pm, because water released after 5pm will arrive too late to also be released to meet the evening peak from the downstream reservoir. After 7pm the SRMC will drop abruptly, though, because water in the upstream reservoir is then too late to be released to meet the evening peak from either reservoir. This example is over-simplified, though. The optimal solution may well avoid having SRMC rise so high at the peak time by having the upper reservoir reach its minimum at 5pm. But that would mean that the daily output cycle of the two reservoirs was offset by the delay time, which means that they can not both be in synch with the load cycle.

does not make the situation any less complex, either. Read [1979]<sup>62</sup> illustrates the kind of operational patterns that may emerge, using the Waikato river chain as an example. That thesis developed MWV-based methods for optimisation of major long term storage reservoirs, but found the river chain optimisation problem too complex to tackle in this way. Other attempts in the literature have been similarly abortive, and river chain optimisation packages generally solve a “primal” version of the problem, in which the MWV is only implicit, and generally not reported.

Thus hydro system operators may not even be aware of the theory discussed here, or conscious of the MWV patterns implicit in their dispatch solutions. Those MWV patterns are potentially quite complex, though, and the SRMC of generation will generally differ between stations in the chain, and between periods of the day. It is quite possible, for example, that the SRMC of generation from one station in the chain may be zero at exactly the same time as the SRMC of release from another is very high.<sup>63</sup> If generation can be drawn from anywhere in the chain, the SRMC of generation from the chain as a whole will be less volatile, but it will rise as increasing requirements must be met by release schedules of decreasing efficiency, and should be expected to vary over the daily cycle, perhaps significantly.

Many discussions assume that we can think of the entire chain as having a single piece-wise linear SRMC “supply curve”. Conceptually, and ignoring uncertainty, such an SRMC “curve” could be derived by a technique known as “parametric programming”, in which an optimization model representing the river chain, with all of its downstream storage, generation, flow, and delay time restrictions is asked to produce more and more output. Conceptually, we could expect that such an SRMC curve might start out fairly constant, while the output requirements can be met without fully utilizing the chain’s capacity in any respect, and then rise in progressively steeper steps as various constraints start to bind. But the situation is actually much more complex than this, because the inter-temporal linkages implied by the storage and delay terms mean that the SRMC curve for any period depends directly on the output requirements in all other periods. Thus we can not actually derive a piece-wise linear SRMC supply “curve” for any one period. Instead we must determine a multi-dimensional piece-wise linear “surface” for all periods simultaneously. Uncertainty about future demand and supply conditions also means that this surface will evolve continuously, as expectations change over daily, weekly and longer cycles.

---

<sup>62</sup> E.G.Read *Optimal Operation of Power Systems*, Phd Thesis, University of Canterbury, 1979.

<sup>63</sup> If there are limits on spill, or on river flows, MWV can actually be negative at some points in the river chain, particularly during flood conditions. And SRMC, which is a difference between successive MWVs can be zero, or even negative, if water must be released to meet minimum flow requirements at some point in the chain.

### **4.3 SRMC of Energy Limited Thermal**

Section 4.1 – *SRMC for Major Reservoirs* discusses the conventional wisdom on MWV determination for hydro systems, on the assumption that SRMC of thermal generation is itself well-defined. Unfortunately, in New Zealand, that is not always the case either, because much of the thermal capacity is actually “energy-limited”, and not in a very different situation from hydro. Conceptually, gas “reservoirs” actually have similar characteristics to hydro reservoirs, except that they are not replenished, and drawdown occurs monotonically over many years, rather than in daily, weekly, or annual cycles. Thus the same general theory applies, except that the MWV is replaced by a “Depletion Related Opportunity Cost” (DROC), and (under deterministic assumptions) this rises steadily over the years at the discount rate, until the reservoir is empty.<sup>64</sup>

Of itself, this physical analysis does not imply any significant extra constraints on the power system. Nor does it imply any difficulty in determining SRMC for gas-fired generation, because DROC changes over such a much longer time horizon than MWV, and is not much affected by year to year variations in the demand for gas-fired generation, eg due to hydro fluctuations. In reality, though, gas producers also have cash flow requirements, and often sell gas via “take-or-pay” contracts that require purchasers to make annual contract payments, and then impose restrictions on the extent to which gas “purchased” in one year can be rolled over for later use, and the extent to which gas to be “purchased” in later years can be used earlier. Maximum and/or minimum restrictions may also be placed on daily, weekly or monthly quantities.

The problem is that when any of these restrictions bind, or threaten to bind, optimal utilisation of this (perhaps artificially) limited resource implies the need to adopt an opportunity costing methodology that is conceptually very similar to that for hydro. And the true opportunity cost-based SRMC of gas-fired generation will then cycle on a daily, weekly, monthly or annual basis, just as for hydro.

To see this, first consider a very simple hypothetical case, in which there is only one gas-fired generator, in a hydro-dominated system, and that generator is supplied under an annual “take-or-pay” contract with no provision at all for roll-over, or purchase of extra gas, and no opportunity to trade. So this generator’s annual gas purchase must be used, or lost, within a year, and the generator faces very much the same situation as a hydro generator with a stock of water than must be used, or lost, within a year. The supply, in this case, is not (normally) at risk, but the demand, being the residual not supplied by hydro, certainly is. And the “per unit cost” is, in principle, irrelevant in determining the SRMC for gas supplied under a contract that effectively involves

---

<sup>64</sup> DROC can never fall, because the reservoir is never full after the first period, whereas MWV does fall, every time an upper storage limit is approached, . Strictly speaking, NWV should rise, like DROC, at the discount rate, over the hours, days or months when storage is not approaching either limit. But this effect is generally ignored on those relatively short time scales.

payment of a lump sum, agreed in advance, for a fixed quantity of gas. Once agreed, this becomes a fixed cost, just like the capital cost of hydro plant.<sup>65</sup>

Such a generator is “energy limited”, and must ration its use of its limited gas resource entirely on the basis of opportunity costs, calculated so as to just use up that resource over the annual time horizon. A centralised optimisation model, optimising the dispatch of such a plant in the context of a hydro-dominated system, would ignore the purchase cost of the gas, and endogenously determine an opportunity cost, and hence an SRMC for gas generation, so as to achieve that goal. But that same model would also have to determine opportunity cost based MWVs for each hydro reservoir in the system. Thus the SRMCs for hydro and gas would be jointly determined, and very closely related, and neither would be determined by the purchase price of gas. In a wet year, the opportunity cost SRMC of gas would have to fall low enough to ensure that the annual gas quantity was used, despite the hydro surplus, and the correspondingly low MWV. In a dry year, the opportunity cost SRMC of gas would have to rise high enough to ensure that only the annual gas quantity was used gas was used, despite the hydro shortfall, , and the correspondingly high MWV.

In reality, such a system is unlikely to exist, because such inflexible gas-fired generation actually does nothing to complement annual fluctuations in hydro output. If the system were that inflexible, the SRMC of both gas and hydro would probably fall to zero in wet years, leaving water to spill and/or gas unused. And the SRMC of both gas and hydro would have to rise high enough to produce electricity prices high enough to choke off demand in dry years. In reality, gas-fired generation would have to provide greater flexibility than this, in order to play a swing producer role in a hydro-dominated system. This flexibility could be provided by contract provisions to purchase, anticipate, or defer, the supply of incremental gas. Or flexibility could be provided by arrangements to trade gas with other users, as (expectations with respect to) hydro inflows vary, on a short to mid term basis.

Either way, the per-unit costs relating to such incremental trading, purchase, anticipation or deferral, will become relevant to the opportunity cost calculation. They do not render that calculation irrelevant, though.

First, in the limit, if the gas market is flexible enough, and this generator is physically and commercially unrestricted in trading its gas in that market, the opportunity cost of using gas purchased under its take-or-pay contract for generation will still not depend

---

<sup>65</sup> There are logical connections, in the longer run, because participants will not enter into contracts to purchase gas that they think is over-priced, on average, relative to the prices they can obtain for gas-fired generation in the electricity market. This impacts on the LRMC of gas-fired generation, but the discussion here relates to determination of SRMC, as hydro output varies, in a time-frame where contract provisions will already have been agreed.

on the price it paid for that gas at all. Instead it will be the market traded price for gas at that time.<sup>66</sup>

Also, in the limit, if variations in electricity generation account for a sufficiently small proportion of the gas market, the market traded price for gas will not fluctuate much as a function of inflow conditions in the hydro sector. This may well be the situation in the US or Europe, say, where there are many alternative uses for gas, and a relatively liquid market will be able to absorb the fluctuations induced by hydrological variations in their, comparatively very small, hydro generation systems. And under those circumstances, the daily, weekly, monthly and annual quantity provisions normal in gas contracts may turn out to have very little influence at all on the calculation of gas opportunity costs, and hence SRMC.

We have not investigated current conditions in the New Zealand gas market but, at least historically, the situation in this small isolated market has been rather different from that in the US.<sup>67</sup> Any trading flexibility will serve to mitigate the effects discussed here, but only unlimited trading would eliminate them entirely. And the gas market has not been liquid or flexible enough to allow unrestricted trading within daily, weekly, monthly or annual time frames.<sup>68</sup> Gas-fired generators have faced physical and/or commercial restrictions on their trading; and the range of variation in electricity generation required to fully match fluctuations in hydro generation has been a significant proportion of the total gas market. Unless the daily, weekly, monthly and annual quantity restrictions in gas supply contracts were to become so relaxed that they could be ignored, we consider that the opportunity costing of gas for electricity generation must remain a significant issue.

Those opportunity cost calculations may be relatively more complex than for hydro, because the opportunity cost of using gas now might be determined by the implied need to trade more or less gas on the market, purchase incremental gas from the supplier now or later, bring forward gas usage planned for a future day, week, month or year, or defer gas usage to a future day, week, month or year. Just as for hydro, though, the calculations must be continuously revised, as expectations change with respect to the future requirements for gas-fired generation, due to variations in load, hydro inflows, or plant availability. And the calculations should really consider a wide range of possible future scenarios, looking forward.

---

<sup>66</sup> As above, the contract price for gas may align closely, on average over the long run, with traded prices, but that is because the traded price determines the contract price, not vice versa. Thus one price may be substituted for the other, for the purposes of long term studies. But it would be a mistake to use (historical) contract prices as a proxy for (forward looking) market prices, when determining the SRMC for gas generation on a time scale of months or shorter.

<sup>67</sup> Isolation is not complete, because gas can be indirectly exported as methanol, for example, and imports remain a long term options. But the New Zealand situation is still very different from that in the US, for example.

<sup>68</sup> That is of hourly quantities within a day, daily quantities within a week, weekly quantities within a month, or monthly quantities within a year.

In fact the theoretically correct opportunity cost calculation may be so complex that it is not actually performed, explicitly, by the managers of gas-fired generators. But that is not the point. The point is that the implied SRMC of gas-fired generation must actually be changing whenever the manager adjusts output so as to avoid violating any kind of daily, weekly, monthly or annual quantity restriction, irrespective of how that manager may conceptualise, or rationalise, that decision.<sup>69</sup> In particular, if the manager (wrongly) thinks of the per quantity price in the contract as setting “SRMC” then he or she may think that what they are doing is adjusting offers to reflect something other than SRMC. But that is not actually the case. What is really happening is that the effective SRMC itself is varying, in accordance with daily, weekly, monthly and annual cycles. Since those variations are strongly linked with the dynamics of, and fluctuations within, the hydro sector, the SRMC of gas is also strongly linked to the SRMC of hydro, and vice versa, and both vary jointly, but not identically, in all of those time scales.

The above discussion suggests that gas-fired generation is significantly less flexible, and its SRMC correspondingly less obvious, than it is assumed to be in many studies and models of the New Zealand electricity system.<sup>70</sup> This in turn, means that greater flexibility must be found from other sources, and that further complicates the SRMC calculation, unless those sources themselves are fully flexible. Neither geothermal nor wind add any significant controllable flexibility to the system, and their SRMC might best be described, like that of run-of-river hydro, as indeterminate. Perhaps oil-fired generation, which is seldom used, might be considered flexible enough that a well defined (if considerably uncertain) SRMC can be determined from the world traded price of the relevant grade of oil, adjusted for transport. Shortage costs may also be considered to provide a clear SRMC component, free of any opportunity cost considerations, although the level of that component is a matter of debate.

The situation faced by coal may be a little different from that of gas, though. In this case, there will be a genuine SRMC element in the calculation, if there are options to increase supplies at extra cost, and unrestricted by daily, weekly, monthly or annual limits. There may also be a genuine SRMC element if there are options to trade coal between electricity generation and alternative uses, either in New Zealand or

---

<sup>69</sup> Hourly restrictions are different, and can be properly accounted for in electricity market offers without any inter-temporal opportunity cost calculations.

<sup>70</sup> Including SPECTRA for example, where these restriction are ignored for algorithmic convenience.

overseas.<sup>71</sup> Historically, though, those options have been limited and, while we have not investigated current market conditions, we suspect that the bulk of the coal supplied to generation plant is still supplied from relatively inflexible sources, on contracts with significant take-or-pay elements. And we suspect that there will not be a sufficiently liquid market for the grade of coal used in electricity generation, or significant alternative users outside the electricity sector, with enough flexibility to absorb the wet/dry year swing. At least, we think it unlikely that the market will be so liquid that generators can assume no limits to their trading. So a very similar opportunity costing logic applies to coal, too, particularly if there is a coal stockpile involved.

Again, we suspect that the opportunity cost calculations theoretically required here may not be performed, explicitly, by the managers of coal-fired generators. But, again, that is not the point. Conceptually, and analytically, the effective SRMC of coal-fired generation is changing whenever the manager adjusts output so as to avoid violating any kind of daily, weekly, monthly or annual quantity restriction, irrespective of how that manager may conceptualise, or rationalise, that decision. The manager may even think that they are adjusting offers to reflect something other than the “SRMC” implied by the per quantity price in the contract. But that is not actually the case. What is really happening is that the effective SRMC is varying, in accordance with daily, weekly, monthly and annual cycles. Once more those variations are strongly linked with the dynamics of, and fluctuations within, the hydro and gas sectors. So the SRMC of coal is also strongly linked to the SRMC of hydro and gas, and all probably vary jointly, but not identically, in all of those time scales.

---

<sup>71</sup> Unlike gas, nearly half of New Zealand’s coal is exported, suggesting that fluctuations in the requirements of coal-fired generators could possibly be accommodated by varying export quantities. This option is only relevant, though, to the extent that coal intended for these two uses is actually substitutable, in the time frame necessary to deal with variations in hydro availability. If coal is being diverted from export, the export coal would have to be an acceptable input, chemically and physically, for generation purposes, and the physical infrastructure would have to be in place to transport it from the export mine to the station. Export contracts would also have to be flexible enough to allow variation in quantity, or substitution of alternative coals, sourced internationally. Similarly, if coal is being diverted to export, it would have to be an acceptable input, chemically and physically, for its intended purpose in the export market, and the physical infrastructure would have to be in place to transport it from the mine supplying the station to the export port. But, while we have not investigated this market, we understand that much of New Zealand’s coal exports consist of metallurgical coking coal, from mines in the South Island, whereas coal-fired generation capacity is situated inland, in the North Island. In any case, the opportunity cost of diverting metallurgical coking coal to be burned in power stations is likely to be very different from the SRMC of their normal fuel.

## 4.4 Locational Issues

While Section 4.2 – *SRMC in River Chains* talks about several generation stations in a river chain, and flow delays between them, it has nothing directly to do with locational issues. In fact it assumes that generation from all stations in a chain is interchangeable, in the sense that it can be sold at the same price, in each period. Similarly, the discussion of SRMC/price interactions between hydro and gas/coal “reservoirs” implicitly assumes that “the market” exists at a single location, so that a single price applies to all capacity.

This is a reasonable approximation if there are no transmission limits between the locations involved, as will often be the case for stations in a single river chain, for example. But the NZEM is a locational market, and we should consider locational issues, too.

Basically, all of the logic above applies at each location, and the prices, and implied opportunity costs at those locations are not independent. Thus the state of South Island hydro storage will still affect the assessment of SRMC for North Island hydro, and for North Island coal/gas stations, too. But the effect may be attenuated by marginal losses and/or transmission limits, on intervening lines. In the absence of losses, it can be shown that the MWV in similar reservoirs will often be exactly equal, in equivalent energy terms. But the presence of losses means that equality will only be maintained within an error bound given by +/- the marginal losses. Transmission limits may further limit the system/ market’s ability to trade-off storage in one reservoir against that in the other, thus leading to a greater divergence in MWVs.

In a perfectly competitive market, the prices in various regions will be set by the opportunity cost-based SRMC of the generation capacity that can meet incremental loads in that region, in the period concerned. These prices will be equal across the system, +/- marginal losses, if no transmission limits apply, but may diverge more strongly when transmission limits apply. MWVs will also tend to diverge more strongly where transmission limits apply, but this effect will be significantly less extreme, because the MWV of a reservoir reflects the opportunity cost of being able to use water to meet load requirements at any time up until that storage is next expected to reach, or at least threaten, its bounds. Even if transmission is limiting, many of those opportunities may relate to future periods in which free trade will be possible, and that will tend to align MWVs through all earlier periods, back to the present. But, if MWVs are tending to diverge, due to surplus of inflow in one region, say, one reservoir will tend to release at its maximum, and the other at its minimum, during periods when inter-regional trading is possible. In that case, the SRMC of hydro generation is no longer set by the MWV, and will be indeterminate.<sup>72</sup> The

---

<sup>72</sup> Theoretically, it will be +/-infinity, depending on the release bound involved.

implication is that local prices are not set by this hydro station, and may be much higher (if release is at its maximum), or much lower (if release is at its minimum).<sup>73</sup>

None of this really changes the conclusions reached above, though. While the transmission system may limit the strength of some interactions and linkages, in absolute terms, the complexity of those linkages, and the corresponding SRMC patterns, is increased. In the limit, MWVs may diverge significantly, as one region becomes increasingly isolated from the other, in terms of electricity market trading, on the margin. At other times, though they will be closely linked, as above. As noted earlier, then, opportunity costing means that each participant's SRMC, in any period, may change as a result of changes in variety of factors other than that participant's own supply position, in that period. But now the factors which might affect SRMC estimates in this way include (expected) changes in the status of transmission constraints. And, so long as a generation station is not marginal, its marginal production cost may consistently be above or below the nodal MCP there.

## **4.5 Risk Aversion**

Finally, while the theory discussed in previous sections relates to uncertainty, and hence may be thought to imply some consideration of "risk", we have not actually considered "risk aversion" at all. It should be recognised that risk is a rather more significant issue in hydro dominated markets, than it is for typical electricity markets. We have noted that it may not be easy to provide risk averse investors with sufficient assurance that they will be able to obtain an adequate return for the risk involved, but risk aversion also plays a major role at the operational level. The impact of risk aversion on the management of energy-limited generation seems to have been almost entirely ignored in the international literature, but bears further examination.<sup>74</sup> It should be obvious that a contracted hydro generator can be expected to err on the side of caution, by setting water aside so as to be available to meet future contractual and/or retail commitments under a wide range of possible hydrological and/or market outcomes. Similarly, for an energy-limited coal or gas generator.

This is not wrong, and it is not new. Indeed, while public sector reservoir management was always based on a balanced sample of historical hydrological years, it was also biased in the direction of caution by adding "buffer zones", "safety factors" etc. Realistically, the general public expects a high degree of reliability in its electricity supply, which means that it will almost always be found, in retrospect, that

---

<sup>73</sup> Actually the same will often be true for reservoirs with differing storage/inflow/release characteristics, even without transmission limits. So long as a generation station is not marginal, its marginal production cost may consistently be above or below the nodal MCP there. As discussed earlier, this will often be the case for stations in a river chain, even when no transmission limits apply.

<sup>74</sup> One exception is: A L Kerr, E.G. Read and R.J. Kaye "Reservoir Management with Risk Aversion", *ORSNZ Proceedings* 1998, p167-176

“too much” water was retained early in the season, only to be released later when, on average, the market supply situation will actually be less tight. This is optimal, and it occurs because there is considerable asymmetry between the costs and risks involved in under-supply, and those involved in over-supply.

The key point here is to note that this pattern of behaviour will involve setting a price on hydro (or coal/gas) generation that differs significantly from the SRMC that would be calculated by a centralised optimisation model that assumes risk neutrality.<sup>75</sup> Specifically, generators will typically be retaining supplies (with upward price pressure), prior to the winter season, only to release more later, when, on average, market prices may actually be lower. This may seem irrational, or perhaps manipulative, from a risk neutral perspective.<sup>76</sup> But the need to behave in this way is dictated by the desire, indeed the effective requirement, to operate cautiously. Risk aversion may also be expected to amplify the response to events that, in themselves, may not seem major, but that might be considered as indicators of an increased likelihood that more severe problems will arise in future periods.

## 4.6 Conclusions

The above discussion provides a reasonably comprehensive guide to the difficulties of determining hydro dominated market environment, with energy-limited thermal plant. In practice there are a great many “reservoirs” involved, operating over a wide variety of time scales. Effectively, coal and gas stocks form extra “reservoirs”, and this further complicates the assessment of opportunity costs, and hence of “SRMC”.

In principle, and to a large extent in practice, the SRMC of production from any one of these sources can not be determined independently from that of any other. But the theoretical linkages described above are not created by the market, and have nothing to do with “market power”. A sufficiently detailed centralised optimisation model would account for them all endogenously, and internally compute SRMC “shadow prices” for all of these resources, jointly, in the course of determining its optimal dispatch solution. Those shadow prices may not be reported but, if examined, they are likely to exhibit quite complex patterns of variation, and connection, on daily, weekly, monthly and annual time scales. Most importantly, whether reported or not,

---

<sup>75</sup> That is for virtually all centralized optimization models of which we are aware.

<sup>76</sup> Consideration may usefully be given as to whether profits could actually be increased, on average, by such a practice, but that is not our concern here. As noted above, the participants withholding water early in the season may actually make a loss on the marginal unit withheld, when assessed at market prices. But this does not actually tell us whether such withholding increases or decreases profits, once price impacts are accounted for, because profit maximising oligopolists generally appear to make a loss (or forgo a profit) on marginal units provided (or withheld) in a gaming situation.

these internally calculated SRMC prices would determine and explain the dispatch solution produced by the centralised optimisation model.<sup>77</sup>

Theoretically, a perfectly competitive market should be able to perfectly account for all of these linkages, and reproduce all of these subtle SRMC interactions. In reality markets are not that perfect, and nor are the information sets or models available to market participants. So they must account for many of these effects subjectively, using their best judgement, rather than through formal analysis. This makes it rather hard to say, objectively, what SRMC actually is, for any participant, in this kind of market situation. One would expect individual judgements to differ, for a great variety of reasons.

Broadly, though, opportunity costs will depend on future market prices, which depend on the offers expected to be made by other participants, which depend on the SRMCs assessed by those other participants which, in the case of hydro or energy-limited thermal, are also opportunity costs. Of course, these opportunity costs will be assessed, internally and privately, by competing generators. Thus each participant must determine the MWV in their own reservoir(s)/stockpile(s), taking account not only of the probability distribution of their own inflows, or supply contract provisions, but also the assessments which they expect each of their competitors to make with respect to the opportunity costs of operating their own hydro/coal/gas resources, given their own private data and probability assessments. And each of those assessments is, itself, equally complex, and also dependent on each of those other parties' assessments of one's own situation.

The general effect is that changes in (perceptions about) the likely supply situation of any plant, whether energy-limited or not, must cause all energy-limited participants to re-assess their SRMCs, in ways which tend to reinforce one another. Thus a developing dry year, or a major plant failure will, and should, immediately cause all hydro/gas/coal participants to raise their SRMC assessments. All of this means that SRMC values for hydro, and for energy-limited fuels, will regularly rise to quite high levels, possibly every autumn, in anticipation of a possible crisis, long before the period in which the crisis is predicted to (potentially) occur, and while storage is still relatively high. And they can be expected to remain at such levels, often for several months. Occasionally they will continue to rise until an actual shortage occurs. But most often they will just fall back to more normal levels because the looming crisis

---

<sup>77</sup> Once stochasticity is accounted for, the true complexity actually becomes too great for many optimisation models to handle. For example, although its PRISM predecessor once had a coal stockpiling module, SPECTRA does not model annual energy limits on either gas or coal. So it can not capture the kinds of interaction described here, and must assume an exogenously determined SRMC for both fuels. This approximation is not correct, but it is required because the optimization methodology employed in that model can not readily be generalized to handle a larger number of reservoirs. If managers employ such limited models, and take them as a guide to "SRMC", they will then need to "adjust" SRMC for opportunity cost effects, outside of the model, and perhaps iteratively adjust model inputs, to achieve an acceptable outcome. But that does not change the fact that it is really the SRMC itself that is changing, as above.

dissipates and/or is averted. Thus, in a hydro dominated power system, and particularly one with energy-limited thermal plant, a properly operating market can be expected to exhibit quite significant variations in average price from year to year, depending on hydrological conditions. And prolonged periods of elevated prices can routinely be expected, even in years which, in retrospect, turn out not to have been particularly dry.<sup>78</sup>

Since it is the change in expectations, rather than any change observed in the current period, that typically drives MWV, the impact on behaviour can also be counter-intuitive. It might be thought, for example, that a tightening of the demand/supply balance causing prices to rise in the current period would always induce an increase in output, or at least no decrease. That would be the case for plant that is not energy-limited, and also in cases where a single, short, one-off, event, such as short generator outage or dry period, has no discernible impact on longer term MWVs, and thus on the SRMC supply curve. The situation may be radically different for energy limited plant, though, if such an event creates the expectation of an ongoing trend, or extended situation. As soon as the likelihood of an extended outage or drought becomes apparent, hydro and energy limited gas/coal generators should re-assess their opportunity costs, and raise offer prices, so as to reduce output, and conserve water/fuel to be used in later periods when the or drought may create an even tighter supply/demand balance. That is their SRMC curves should rise to such an extent that their output actually reduces, even though the demand for output, and the prices obtainable, have increased in the market. This may occur over a period of months, for a prolonged outage or anticipated drought. Or it may occur over a period of hours, for relatively short outages, or load increases, for example.<sup>79</sup>

Finally, it might be thought that all of these correlations and connections between the effective SRMCs of energy-limited plant have something to do with “market power”, or even “collusion”. But, while, both market power and collusion could certainly arise in such an environment, this discussion of correlations and connections actually has nothing inherently to do with either. Nor is there is anything particularly unusual about what is going on here. The revaluation of contracted gas, or stockpiled coal, due to a change in the expected availability or price of hydro power in the local electricity market is no different from the revaluation of those same resources due to a change in the expected availability or price of oil on world markets. And nor does it differ from the routine revaluation of shares, or of hotel rooms or airline seats due to changes in perceptions about likely supply or demand.

In all cases the opportunity cost, and hence the SRMC, do actually change, at least when these concepts are properly defined in economic, rather than accounting terms.

---

<sup>78</sup> Noting that New Zealand has relatively small reservoirs, which do not allow much storage carryover from year to year.

<sup>79</sup> See P Stewart, E.G. Read and R James: “Intertemporal Considerations for Supply Offer Development in Deregulated Electricity Markets” *IAEE Proceedings*, Zurich 2004

And that re-assessment may imply an initial reduction in output as part of the optimal response to a developing crisis. This is merely an optimal reaction to changing expectations, such as would occur in a sufficiently sophisticated centralised optimisation, or in a perfectly competitive market. All that is different here is that the interactions are more obvious and more explicit, when considered in the context of a relatively small and inflexible system, and applied to commodities such as water, electricity, gas, and coal, for which liquid international markets may not be readily accessible from New Zealand, in the required time frame. In this context many feedback loops which might normally be considered “open”, in many analyses elsewhere, must be treated as “closed”, implying a need for something more like a general equilibrium type analysis to calculate opportunity costs jointly, and simultaneously, rather than applying partial equilibrium analyses sequentially and/or independently.

To repeat, then, we are merely explaining the kind of price patterns, correlations and connections that would arise, internally, within any sufficiently detailed centralised optimisation, and arguing that exactly the same patterns, correlations and connections should be expected in a hypothetical perfectly competitive market. Real markets may not exhibit all of these patterns, correlations and connections quite so explicitly, and participants may not even be able to clearly articulate or analyse how all of these factors interact. But a market in which such patterns, correlations and connections were not evident should be judged to have fallen short of the ideal, perhaps significantly so, and that should be a matter of concern.

Real markets may also provide opportunities for the exercise, and perhaps abuse, of market power, and this may distort pricing patterns away from the perfectly competitive ideal discussed here. That is another matter, and not our concern here. We would say, though, that the complexity of the underlying situation must make it difficult to determine whether, and to what extent, market outcomes might actually have deviated from the perfectly competitive ideal, on average, or in any instance.

## 5 Final Perspective

This paper has not discussed market power issues in any detail, but has been prepared in a context where the identification and assessment of market power is expected to become a significant issue. One way of thinking about market power would be to examine the extent to which market offers, and hence market prices, deviate from what might be expected in a perfectly competitive market. If SRMC pricing is expected to characterise a perfectly competitive market, one may eventually need to ask:

- What is SRMC, and
- Should prices deviate from SRMC, and
- If so, to what extent?

Accordingly, this paper has not attempted the kind of analysis that might yield a definitive conclusion as to how or where a line might be drawn between legitimate deviations from SRMC pricing, such as occur pervasively throughout the economic system, and the illegitimate “abuse” of market power. Nor has it examined any evidence relating to actual behaviour. Instead it merely sets the scene, by considering two broad areas, in relation to the questions above.

- First, Section 2 – *Market Design Concepts*, and Section 3 – *NZEM Market Design Choices*, have argued that some degree of deviation from SRMC is an expected, if not necessary, aspect of the NZEM market design;
- Second, Section 4 – *SRMC for Hydro and Energy Limited Thermal* has argued that SRMC is actually quite difficult to define and determine, in a hydro-dominated power system, where energy limits also apply to some of the thermal generators.

The major implication of the discussion is that a wide range of market behaviours, and outcomes, which might be thought to have something to do with market power are also quite likely to arise in a perfectly competitive market, or in a centrally planned environment. Persistently high prices, for example, do not provide conclusive evidence that market power is being exercised, let alone abused, in a hydro dominated market. And the same applies to highly correlated changes in market offers. This kind of offer or price pattern should be expected, and accepted as a necessary and appropriate feature of such markets, even if perfectly competitive. Risk aversion also plays a major role at the operational level, and can be expected to produce patterns of behaviour very similar to those which might be expected from the exercise of market power.

Thus, while assessment of deviations from SRMC pricing is certainly an interesting research area, it may be quite difficult to achieve any certainty about what participants actually believe SRMC to be, let alone identify the motives behind any deviation from it. Nor is the task made any easier if it is accepted that some deviation from SRMC is required, as part of the market design, so that the issue becomes one of determining whether the observed degree of deviation from SRMC is in fact optimal.

Partly because of these difficulties, the designers of the NZEM did not make alignment with SRMC a primary goal. So the question now arises as to whether, given that market design, this is really the most useful direction of investigation, either to determine the effectiveness of the market design, or to examine whether market power, in any form, is being inappropriately exercised in the New Zealand electricity sector.

We have argued that the NZEM has been designed to be like other markets, throughout the economy, rather than like some other electricity markets, where prices are regulated to lie close to SRMC, while capacity is at least partly paid for by explicit capacity payments. So the basic focus of the NZEM design is really on ensuring that the PDC rises high enough, often enough, to induce sustainable entry, in the belief that the overall cost structure of the industry is dominated by fixed costs, and that a competitive entry market will serve to discipline those costs.

This leads to a focus on LRMC, rather than SRMC, as the reference point with respect to which competitive pricing should be measured, and the primary indicator of market performance. This is, after all, the price at which a competitive entrant should be able to provide power, and competitive entry will not occur until it is possible to sell contracts at such prices. Even LRMC is not a “limit”, though, because prices must equal LRMC, on average, being above that level for long enough to balance periods of excess supply, when competitive pressure may force prices below LRMC.

Many electricity markets have started in a situation of over-supply, with SRMC prices well below LRMC, but this can not be sustainable in the long run. On the supply side, incumbent generators have found themselves thrust into a surplus situation with no long term commercial arrangements in place. But this is an artificial situation. If the market had evolved naturally over time, rather than being created by break-up of public assets, it is debatable whether competitive investors would ever have entered an inherently risky market in such quantity without long term contracts, or a large scale retail base, but absolutely certain that they would not have if they had also thought that their spot and/or contract pricing would be forced down to SRMC during a surplus.

Recognising this, many markets have started with “vesting contracts” in place, to protect both sides of the market, with prices often well above SRMC, and then allowed spot and new contract prices to trend toward LRMC at whatever pace participants have been able to achieve. Other markets have left incumbent generators to negotiate their own contractual arrangements, with the expected result that contract

prices have been below LRMC, and/or uncontracted generators have struggled to achieve prices which might be considered sustainable, in the long run.

On the demand side, this can give consumers unrealistic expectations about the present and future cost of supplying electricity, and help create a perception that the market is “failing”, once prices start moving towards more realistic sustainable levels. But, a history of rising prices spot or contract prices is not, of itself, evidence of any problem with the market, especially if the market starts off in surplus and/or LRMC itself is rising.

Eventually, though prices must at least reach LRMC and, at that point, the current market design assumes that competitive entry will be sufficient to discipline prices, on average, to that level. So the critical question is really whether that is actually the case, or whether some intervention is necessary to impose regulation to force prices down to LRMC. Thus our own preference, from a market design perspective, would be to focus first on that critical question; that is on the alignment between the market PDC and entry costs, as perceived by risk averse investors.

Since markets are always in disequilibrium, perfect alignment should not be expected. Careful consideration must be given to the recent history of the market, in terms of hydrology, and of “shocks” due to resource, technological or policy changes, for example. An appendix briefly outlines how such issues might be accounted for, and the effect they might be expected to have. But it is only once such a study has been conducted that we could reasonably reach a judgment with respect to market performance, in terms of what we would characterise as its primary design goal of aligning the PDC with entry costs.

Then, if it turns out that they do not align, we would want to ask what barriers there may be to entry, and if those might be due to incumbents exercising some kind of market power in the market for entry (eg by blocking access to attractive development sites, or threatening to undercut entrants by “predatory pricing”), or due to other factors such as the threat of regulatory intervention, or to failures in the contracting chain. Spot market behaviour would be a relevant issue in any such investigation, but primarily inasmuch as it might provide evidence with respect to the possibility that a broader long terms strategy was being pursued, eg to deter entry by increasing the risks faced by potential entrants.

## **Appendix:**

# **Market Performance and Entry Barriers**

We have argued that attempting to enforce SRMC pricing would be inappropriate in the NZEM context, and that makes assessment of deviations from SRMC of limited relevance. Instead, we suggest that the emphasis should be on whether the market is fulfilling its intended function which is, primarily, to provide appropriate long run signals, while facilitating short run coordination between alternative suppliers, and between them and consumers. Thus we should really be asking

- Does the PDC align with LRMC of relevant plant?
- Are there any barriers to competitive entry by alternative suppliers which might allow prices to persist above LRMC?
- Are there problems in the contract market, and/or the wholesale/retail “contracting chain” which are leading to sub-optimal results, such as excessive risk for entrants, leading to increased risk of non-supply?

In principle, ignoring the possibility that incumbents may raise barriers to entry, it is actually very easy to determine whether NZEM prices are, or have been, “too high” in recent years:

- If entry has been excessive, and we now have “too much” capacity, then we might conclude that prices (or price expectations), if anything, have been too high over the period when that excessive entry was occurring.<sup>80</sup>
- If entry has been inadequate, and we now have “too little” capacity, then we might conclude that prices (or price expectations), if anything, have been too low over the period when that inadequate entry was occurring.<sup>81</sup>

This simple test indicates the basic direction which investigations of market pricing should be directed, either to discover why prices have been too high, or why they have been too low. Thus it is pertinent to ask where the current NZEM PDC lies with respect to long term FCs. We note that entry has been occurring over recent years, which suggests that the PDC should be matching FC for at least some plant types. There are three complications to consider, though.

---

<sup>80</sup> With market forces perhaps now pushing prices below LRMC level, as a result of excessive entry.

<sup>81</sup> With market forces perhaps now pushing prices above LRMC level, as a result of inadequate entry.

- First, there are other reasons why investment might have been too low, if it has. For example, the new SOEs formed by the breakup of ECNZ, had a limited set of feasible development options and the lead times to develop options have extended significantly. Also, there has been a significant learning curve with respect to technologies like wind generation over that period. So investment could lag market demand, for those reasons.
- Second, a rational investor, or central planner, should be asking what the PDC looks like over the whole range of hydrological conditions, not just what the PDC has looked like over the very small sample of hydrological conditions that actually occurred in the past few years. One does not, or should not, build new capacity in response to high prices driven by dry year conditions, but to a shift in the underlying probability distribution from which that price sample was drawn. Thus it is quite possible to observe (temporary) high prices and (underlying) excess capacity in the same year, in a perfectly competitive or centrally planned system. The study by Tipping et al suggests that the NZEM seems to have experienced a higher than average number of dry years recently, with correspondingly higher prices. This can be expected to have raised public awareness, and concern, to levels which are probably not justified by the underlying supply/demand balance.
- Third, it is the expectation of future prices that should drive investment, and that expectation may turn out to be significantly in error, if the market experiences some kind of shock. In this case, the NZEM has recently experienced a series of shocks, all in the same direction. Apart from the ongoing impact of local “environmental” resistance to developments, the system has seen an unexpected reduction in gas availability, rising world fuel prices, the sudden imposition of policy driven restrictions on capacity investment, and possibly inflow reductions due to climate change. These will all have raised expected LRMC levels, and we should expect to see prices rising now, reflecting an upward shift in the (expected) long run equilibrium PDC, with a higher probability of shortfall, to account for these factors.

Further, because these changes were not expected, we should expect to see price overshoot, with prices lying above even the new (higher) LRMC levels for a few years while the market adjusts to the new situation. Basically, if the market was (thought to be) in long run equilibrium prior to these effects becoming evident, we should now expect to see the market out of equilibrium, and experiencing relative shortage.<sup>82</sup> Conversely, if the market now appears to be in equilibrium, and if the PDC is not now lying above the new LRMC levels, we should really be concerned to explain why there was excess capacity investment, relative to expectations, in prior years.

---

<sup>82</sup> The issue here is not whether any of the factors have really changed, but whether market analysts today employ more, or less, optimistic cost/availability assumptions than they did a few years ago, when performing their FC/OV comparisons.

Leaving aside the weight of public opinion and political concern, we are not trying to express any opinion here as to whether there actually is “too much”, or “too little” capacity in the NZEM, or whether prices have been “too high”, or “too low”. But, if there is, a regulatory response might be envisaged:

- In the first case, regulatory intervention may be justified to place downward pressure on prices, perhaps by tightening offer rules if, but probably only if, it can be shown that this is not just a temporary situation resulting from a “shock”, but a long term structural problem, presumably arising out of a lack of competition in the market for entry.
- In the second case, regulatory intervention may be justified to place upward pressure on prices, perhaps by adding capacity payments or loosening offer rules if, but probably only if, it can be shown that this is not just a temporary situation resulting from a “shock”, but a long term structural problem, arising perhaps out of fear of a political intervention in response to higher prices.

These are not the only possible conclusions, though. While this market design perspective does suggest a lesser degree of concern about deviations from SRMC pricing, it does suggest that the market for entry may be a legitimate focus for concern. At the national level, we are concerned, primarily, about whether entry into the generation market is competitive, or whether incumbents might raise barriers to competitive entry:

- One possible way of deterring entry would be for incumbents to refuse to provide necessary supporting “ancillary services”, but the NZEM market design makes this quite difficult.
- Another obvious strategy would be for incumbents to block access to desirable development sites, or resources, and this possibility may be worthy of examination in the NZEM context. This would not produce a PDC which was “too high” relative to actual entry costs, but would imply that entry costs, and hence the PDC, were too high.
- But there is another, less obvious, way in which incumbents might raise barriers to entry. A central concern in the literature about “entry deterrence” is that incumbents could deter entry by building too much plant, then pricing high, but threatening to price low for long enough to drive out any competitor which might be tempted to enter.

This hypothesis was advanced in an early NZIER study of NZEM design issues,<sup>83</sup> but its relevance to current conditions is debateable. If the strategy were being played effectively, it seems possible that there could be too much capacity, and prices which

---

<sup>83</sup> See SJ Gale & AE Bollard *A Theoretical Approach To Electricity Generation Restructuring* NZIER Report to the Officials Working Group, July 1990

are also too high, on a sustained basis. Or, if the strategy failed, and did not actually deter entry, we could see entry followed by a period in which there could be too much capacity, but with prices which are too low to sustain further entry, possibly falling to SRMC levels.

Again, we come back to the very basic question, though: “Is there too much capacity in the NZEM?” If not, it seems unlikely that this entry deterrence game is being played. And, if the entry market is deemed to be reasonably competitive, we must then ask whether market power is really a major problem in the NZEM, given its design goals.

But this discussion has been focussed on entry to the generation market, at the national level. All of these issues become more critical at a regional level, and for retail markets. At that level, entry may require being able to:

- Build generation capacity in the right place, and/or
- Gain physical access to generation elsewhere via transmission system enhancement, and/or
- Gain commercial access via some form of transmission capacity right.

The first will obviously be difficult, in many instances, while the last is not possible under current market arrangements, and opinions vary with respect to the effectiveness of current transmission planning processes. Thus, barriers to competitive entry into regional retail markets could possibly be a legitimate focus of concern. Still, while many of those barriers may create an environment in which the exercise, and potential abuse, of market power is more likely, most are not likely to have been created by market power, or for the purpose of enhancing market power. Nor does the existence of potential barriers prove that market power exists, or has been abused.