Models for estimating the performance of electricity markets with hydroelectric reservoir storage*

Andy Philpott[†] Ziming Guan[‡]

May 12, 2013

Abstract

We describe some new results of an empirical study of the New Zealand wholesale electricity market that attempts to quantify efficiency losses by comparing market outcomes with a counterfactual central plan that accounts for hydroelectric reservoir shortages. The study extends previous work by studying differences in welfare transfers between generators, consumers and transmission owners, as well as studying the effects of risk aversion on an optimal central plan.
Keywords: electricity market, hydroelectricity, stochastic inflows, risk-aversion, market power.

1 Introduction

In this paper we describe an empirical study of the performance of the New Zealand wholesale electricity market (NZEM) over the years 2005 to 2009, which extends the study on production inefficiency reported in [19] to encompass two more years. The study also looks at welfare transfers between participants as

*The work has benefited from modelling and theoretical contributions from many people. We wish to acknowledge in particular the contributions of Geoffrey Pritchard, Golbon Zakeri, Tony Downward, and Vitor de Matos to this work. The views and opinions expressed in this paper are, however, solely attributed to the authors.

†Electric Power Optimization Centre (EPOC), Department of Engineering Science, University of Auckland: a.philpott@auckland.ac.nz.

‡University of British Columbia and BC Hydro: Ziming.Guan@bchydro.com. The work described in section 4 of this paper was carried out while Dr Guan was working as a postdoctoral fellow in EPOC.

well as investigating the effects of risk aversion on our benchmark policies. This work has been developed in our research group over the past few years under the acronym EMBER (Electricity Market Benchmarking Exploring Risk).

The NZEM is a nodal electricity pool market with a high proportion of hydroelectric generation. Unlike markets consisting solely of thermal plant, markets with hydroelectricity have an inter-temporal aspect arising from the fact that energy (water) can be stored for later delivery. This complicates the decision making of hydro-generators as their computation of the marginal cost of releasing water must involve some modelling of opportunity cost and possible shortage costs. Since these models depend on uncertain inflows, the computation of benchmark policies involves optimization under uncertainty as outlined in [19].

Our benchmarks are the result of simulating a system optimal policy. If electricity market participants maximize productive and allocative efficiency, then they will maximize social welfare for the system as a whole. Maximizing system social welfare over an uncertain future is a challenging problem to define, let alone solve. For the purposes of defining a risk-neutral benchmark, we define the objective of this problem to be the expectation of operating benefits of participants (defined to be revenue minus fuel cost for producers and the integral below the demand curve minus payment for consumers), taken with respect to a stochastic process of inflows that is common knowledge. In this setting it is possible to solve an approximation of the welfare maximizing problem and simulate the policy over historical inflow sequences. This provides a benchmark against which historical outcomes can be assessed.

In previous work [19], we focused on estimating the productive inefficiencies that result from market dispatches over the years 2005 to 2007. The benchmark central planning policies that we computed were risk neutral and as expected were less expensive on average than the market outcomes, at least in the three year sample that we simulated. The year 2008, which was not included in the previous study, experienced very low reservoir inflows, and so one purpose of the current work is to include this year in our study. The results are of some interest in the sense that the benchmark policies computed using the technology of [19] incur significant shortage costs in 2008, a recent dry year. We might expect this, as the benchmark policies we compute are risk neutral and so make savings on average, even if every now and then they incur large costs. We also show in this paper how the inclusion of risk aversion in our model can lead to a less extreme outcome in the simulation without a significant increase in productive inefficiency.

We also extend our previous analysis to study the marginal prices of energy emerging from the model. These can be compared with observed prices in the wholesale market. The analysis of prices is interesting in several respects. It enables us to estimate welfare transfers between agents (i.e. consumers, suppliers, and the grid owner) in each trading period, and compare these in the centrally-planned counterfactual and the market. Our study can be compared with the counterfactual model applied to the New Zealand market by Wolak [27]. The Wolak model uses a benchmark (called Counterfactual 1) in which hydro generation is set to its historical levels, and thermal plant is then offered in each period at its short-run marginal cost. It is argued (incorrectly) in [27] that the demand-averaged marginal cost computed from this experiment will be biased above the truly competitive marginal cost, so using this benchmark as a counterfactual will give conservative estimates of price markups. In our previous work [19] we discuss the shortcomings of this counterfactual[1], as motivation for developing a benchmark based on multistage stochastic programming.

Given our previous criticisms of Counterfactual 1, it is interesting that our benchmark value for 2005 gives a similar figure to that published in [27]. Our results, however, come from a different benchmark policy. In our counterfactual policy, the opportunity value of water is influenced by the possibility of shortages, so the resulting marginal water value estimates (in contrast to Counterfactual 1 in [27]), can exceed the cost of the highest dispatched thermal plant. Indeed, the highest weekly average energy price we observe in our counterfactual (in 2008 when a shortage occurs) exceeds the historical average price, whereas in most other weeks, the estimate of weekly average prices in the counterfactual solution is lower than that observed in the historical solution.

The aim of benchmarking market performance is to identify inefficiencies, diagnose the causes of these and then devise institutional arrangements and instruments that might be used to reduce them. The causes of inefficiency identified by our work are not clear. Wolak's paper [27] attributes price differences to unilateral exercise of market power, but it is difficult to discriminate between this and the competitive behaviour of risk-averse agents. Indeed, the correspondence between a competitive equilibrium in hydro-dominated markets with risk averse agents, and a socially optimal plan is not well understood. In theory, under very special circumstances (no strategic play and risk neutral players), it is possible for a Walrasian equilibrium to give a stochastic process of prices with respect to which every agent optimizes its own expected benefit (revenue minus variable cost) with the outcome of maximizing total expected welfare. However, as shown by the examples in [3], the stochastic process of prices that yields an equilibrium might be very complicated with none of the stagewise independence properties that make computing optimal policies easy for generators.

When agents are risk averse or do not have a common view of the (random) future then a competitive equilibrium is harder to identify. Agents in such an equilibrium (if it exists) will estimate marginal water values based on their risk-adjusted view of the future, and their actions in aggregate will yield equilibrium

[1] There have been several similar critiques of Counterfactual 1 in [27], see e.g. [6].)

prices that are then used to form these views. Determining these prices for a multistage equilibrium would be very difficult. Furthermore, if we were to seek an equivalent socially optimal plan then it is necessary to integrate individual risk measures into a system risk measure to be optimized. It is well known that this is not possible for general nonlinear utility functions. On the other hand in the special case where agents optimize coherent risk measures [2], then it is possible to obtain a system risk measure in complete markets, i.e. where all risk can be traded by market participants (see [7]). Although we cannot expect such complete markets in practice, the formulation and solution of risk-averse centrally planned hydrothermal models in this paper provides a first step towards a competitive equilibrium benchmark of a complete market with risk-averse agents.

A similar problem of market incompleteness in hydroelectricity systems can occur in a risk-neutral setting, as shown by the paper by Lino et al [9]. They demonstrate using a computational example where all agents are located on different river systems that a risk-neutral centrally planned solution gives rise to system marginal prices that will clear the market with an optimal dispatch if each agent optimizes its own objective using these prices. Lino et al [9] then show how inefficiency can result from different agents operating hydro stations on the same cascaded river system. In this case, a market instrument pricing the transfer of water between generating stations is needed to recover the economically efficient solution. For the duration of this study, each of the main hydroelectric river systems in New Zealand was operated by a unique generation company, so the inefficiencies identified by [9] do not arise.

The results in this paper show that the production inefficiencies in moving to a market model from a central plan are not that great in relative terms. The differences in prices between these solutions on the other hand are quite large, leading to very different distributions of benefits. As mentioned above it is hard to say whether the price differences we observe are due to unilateral exercise of market power. Many of the conditions for Walrasian equilibrium are missing, and the stochastic price process that one would like to use to clear the market as a stochastic optimization problem is never revealed to agents, but defined by a single realization of prices appearing every half hour out of the system operator's dispatch software. It could be that the market structure we are working with has some way to go to providing a better approximation to the prices we would obtain from a Walrasian auctioneer. Nevertheless, the presence of the differences between the benchmark and the historical outcomes give some grounds for deeper investigation.

The layout of the paper is as follows. In the next section we describe the New Zealand wholesale electricity market. We then outline the features of a suite of optimization models that we use in our study. To make the computation times reasonable these models are approximations of the system models that are used to dispatch and price the market. For this reason, the data for these models must be aggregated, and a lot of attention is paid in this section to how this aggregation is performed. In section 4 we repeat the analysis of [19] with updated data, and compare results. Section 5 compares the benchmark policies computed in section 4 with those constructed with some more recent stochastic optimization models. These are able to incorporate risk aversion, and being faster to solve, can produce more accurate approximations to optimal centrally planned policies. We show using the new models that historical prices are generally higher than the counterfactual solution and estimate welfare distributions in the market as compared with the counterfactual. The final section makes some conclusions.

2 The wholesale electricity market

Since 2004, New Zealand has operated a compulsory pool market, in which the grid owner Transpower plays the role of Independent System Operator (ISO). In this market all generated and consumed electricity is traded[2]. Unlike most electricity markets in other parts of the world, the NZEM has no day-ahead power exchange. Bilateral and other hedge arrangements are still possible, but function as separate financial contracts. Trading develops by bids (purchaser/demand) and offers (generator/supply) for 48 half hour periods (called *trading periods*) over 244 pricing nodes on the national grid. (Although demand side bids are included in the official description of the ISO dispatch model, there is currently very little demand-side bidding in the NZEM, so we will omit them from further discussion.)

The offers of generation made by generators to the ISO take the form of *offer stacks*. These are piecewise constant functions defining the amount of power offered at up to five different prices that may be chosen by the generator m. We can represent the offer stack for generator m by the (step) function $C_m(x)$. In the New Zealand market the generator offer functions C_m are not publicly known at the time of dispatch, but are published the following day. These data are made available as part of a Centralized Data Set (CDS) distributed by the New Zealand Electricity Authority [11].

All the prices in the wholesale electricity market in New Zealand are computed by the ISO using a linear programming model called "Schedule Price and Dispatch" or SPD. This represents the New Zealand transmission network by a DC-load flow model. The full version of SPD includes constraints that ensure voltage support, $N-1$ security for line failures, and meet requirements for

[2] Small generating stations with capacity of 10 MW or less are not required to make offers. From 1996-2004 a voluntary wholesale market existed, where approximately 80% of electricity was traded; the remaining 20% by bilateral contracts.

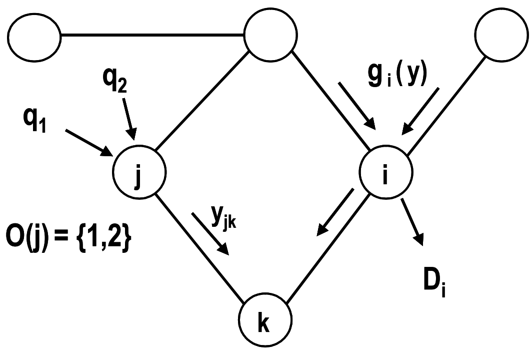

Figure 1: Generic network model illustrating notation

spinning reserve that are dispatched at the same time (see [1]).

If we ignore these additional features then the problem solved every trading period by SPD can be described mathematically using the generic network model shown in Figure 1. For each node i the set $\mathcal{O}(i)$ defines all the generators at node i, where generator m can supply any quantity $q_m \in Q_m$. The demand at node i is denoted D_i. This gives the following market dispatch model:

$$\begin{array}{llll} \text{MP1:} & \text{minimize} & \sum_i \sum_{m\in\mathcal{O}(i)} \int_0^{q_m} C_m(x)dx & \\ & \text{s.t.} & g_i(y) + \sum_{m\in\mathcal{O}(i)} q_m \quad = \quad D_i, & i \in \mathcal{N}, \\ & & q_m \in Q_m, \quad m \in \mathcal{O}(i), & i \in \mathcal{N}, \\ & & y \in Y. & \end{array}$$

Here the components of the vector y measure the flow of power in each transmission line. We denote the flow in the directed line from i to k by y_{ik}, where by convention we assume $i < k$. (A negative value of y_{ik} denotes flow in the direction from k to i.) We require that this vector lies in the convex set Y, which means that each component satisfies the thermal limits on each line, and satisfies loop flow constraints that are required by Kirchhoff's Law. The function $g_i(y)$ defines the amount of power arriving at node i for a given choice of y. This notation enables different loss functions to be modelled. For example, if there are no line losses then we obtain

$$g_i(y) = \sum_{k<i} y_{ki} - \sum_{k>i} y_{ik}.$$

With quadratic losses we obtain

$$g_i(y) = \sum_{k<i} y_{ki} - \sum_{k>i} y_{ik} - \sum_{k<i} \frac{1}{2} r_{ki} y_{ki}^2 - \sum_{k>i} \frac{1}{2} r_{ik} y_{ik}^2.$$

In our model the quadratic losses are modelled as piecewise linear functions of arc flow which enables MP1 to be solved as a linear program (at least when losses are minimized by the optimal solution).

Bids and offers start 36 hours before the actual trading period. Up to 4 hours (pre-dispatch) before the trading period starts, a forecast price is calculated to guide participants in the market. From 4 hours to the start of the trading period every half hour a *dispatch price* is calculated (and communicated). Two hours before the start of the trading period, bids and offers for the period in question are locked in. From that point onwards any new prices reflect the ISO's adjustments in load forecasts and system availability.

During the half hour period the ISO publishes a new real-time price every 5 minutes and a time-weighted 30-minute average price. The real-time prices are used by some large direct-connect consumers to adapt their demand. The above prices are a guide only, as the final prices are calculated ex-post (normally noon the following day, unless there are irregularities or disputes) using the offer prices as established 2 hours before the trading period, and volumes metered during the trading period.

3 The models

Our study will make use of the same suite of models defined in [19]. For completeness, we repeat the description of these models here. In the Appendix we outline the changes that have been made to these models to bring them up to date with the latest public data set. The models we use are:

CP1:	A determininstic dispatch model solved over one trading period;
CP3:	A determininstic dispatch model solved over one week with 3 load blocks
CP48:	A determininstic dispatch model solved over one day;
CP336:	A determininstic dispatch model solved over one week;
YEAR:	A stochastic planning model solved over one year;
INTER:	A model that is used for estimating water release and spill over a day or week;
EP:	A model that is used to calibrate nodal demand from historical dispatch and prices.

We examine a counterfactual proposal that supposes that the national electricity system is controlled centrally and is dispatched sequentially by the yearly and weekly models. This is compared with the actual dispatch in the wholesale market.

3.1 The dispatch models

Ideally, the dispatch model that should be used in our study is SPD, the full-scale version of MP1 with 244 nodes and reserve and security constraints (or its publicly available version vSPD distributed in the Electricity Authority's EMI system [10]). However we need to solve such a model many times in simulation, and so we have chosen to approximate this system with a model having only 18 nodes. The representation that we use is shown in Figure 2.

This approximation ignores constraints in the full model that arise from voltage support, $N-1$ security, spinning reserve and frequency keeping, and so it is likely to underestimate prices. Moreover since we aggregate electricity load into regions around each of the 18 nodes, we ignore in our approximation the thermal line losses that occur in lines that join the points within each region. This means that regional totals of historical demand will underestimate the true demand that we should use at each of the 18 nodes in the simplified network. One option is to apply a uniform scaling to demand, but this does not reflect the fact that demand is concentrated in some nodes (e.g. TIW which contains a large aluminium smelter) and not in others (e.g. WKM that meets the needs of a dispersed region in our model.) To overcome this, we estimate demand values for each node by solving the following model for each trading period being studied:

$$\begin{array}{llllll} \text{EP:} & \text{maximize} & \sum_i \bar{\pi}_i g_i(y) & & & \\ & \text{s.t.} & g_i(y) + \sum_{m\in\mathcal{O}(i)} \bar{q}_m & = & D_i, & i \in \mathcal{N}, \\ & & -\alpha \bar{D}_i \leq D_i - \bar{D}_i \leq \alpha \bar{D}_i, & & & i \in \mathcal{N}, \\ & & y \in Y. & & & \end{array}$$

As in the full-scale dispatch model SPD, the loss functions g_i in EP are modelled as piecewise linear functions of line flow y. Here $\bar{\pi}_i$, $\bar{D}_i$, and $\bar{q}_m$ are respectively the historical nodal price at the node i representing the region, aggregated demand for the region, and aggregated dispatch for the region, all summed from data recorded in the Centralized Data Set. It is known (see e.g. [20]) that the line flows from any given optimal dispatch maximize $\sum_i \bar{\pi}_i g_i(y)$ where $\bar{\pi}_i$ are the nodal prices. The problem EP seeks to scale aggregated demand $\bar{D}_i$ for each node i (by at most $1 \pm \alpha$) so that the demand estimates D_i obtained are consistent with the historical dispatch and historical prices according to EP. Here α is chosen to be 0.1[3].

In a centrally planned market, the offer curves C_m would be determined by the marginal cost of supply. For simplicity we assume that generator m can supply any quantity $q_m \in Q_m$ at cost ϕ_m per MWh. The single-period centrally planned

[3] Details of this estimation process, and all data used in this paper are available in the online companion downloadable from the data repository at http://www.epoc.org.nz/ember.html.

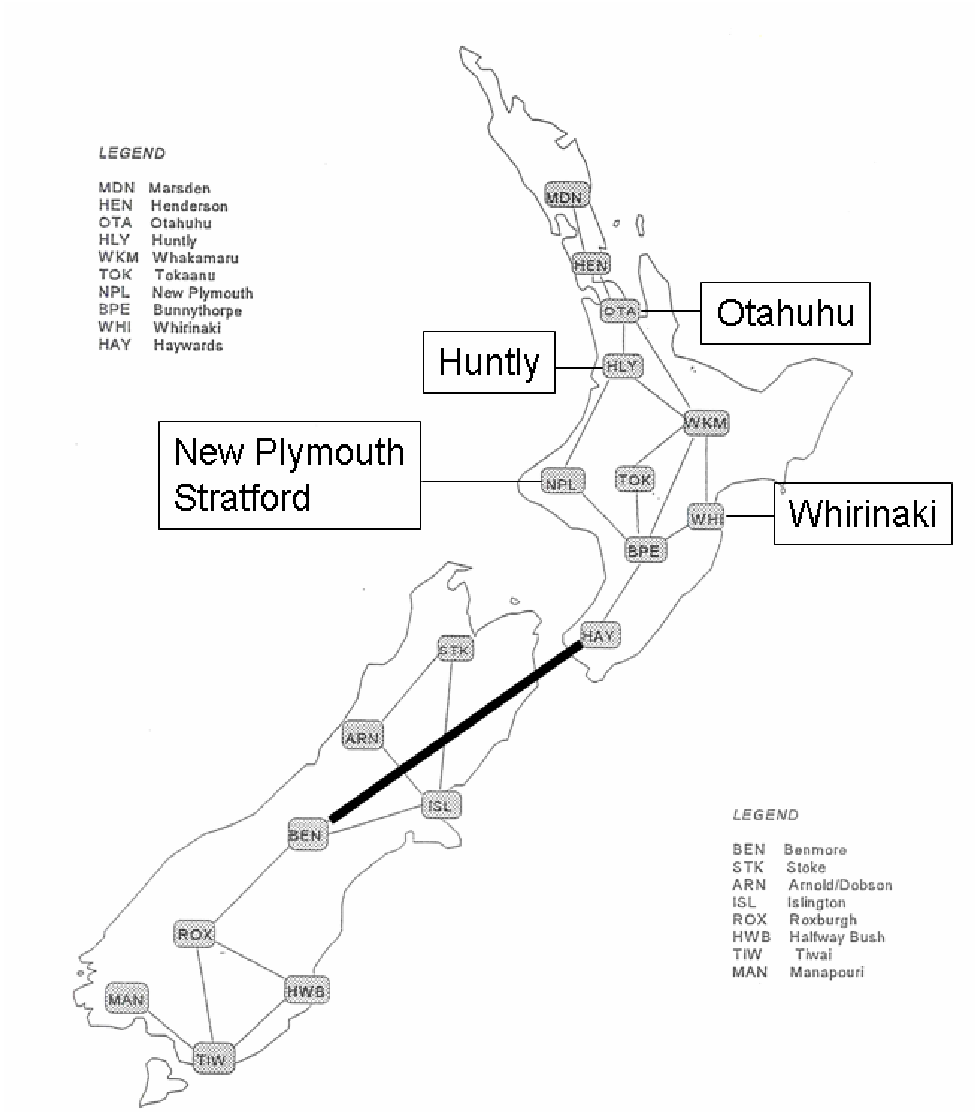

Figure 2: Approximation of New Zealand transmission network showing location of major thermal generators. The bold line represents a HVDC cable connecting the South and North islands.

economic dispatch model (CP1) then minimizes cost in a single trading period.

$$
\begin{array}{llll}
\text{CP1:} & \text{minimize} & \sum_i \sum_{m\in\mathcal{O}(i)} \phi_m q_m & \\
 & \text{s.t.} & g_i(y) + \sum_{m\in\mathcal{O}(i)} q_m \quad = \quad D_i, & i \in \mathcal{N}, \\
 & & q_m \in Q_m, \quad m \in \mathcal{O}(i), & i \in \mathcal{N}, \\
 & & y \in Y. &
\end{array}
$$

3.2 The short-term hydro model

To investigate the dispatch of hydroelectricity over the course of a day a national river-chain dispatch and nodal pricing model (CP48) combines offers with river scheduling constraints over 48 half-hour trading periods, $p = 1, 2, ..., 48$. A diagram of the physical system for this model is shown in Figure 3.

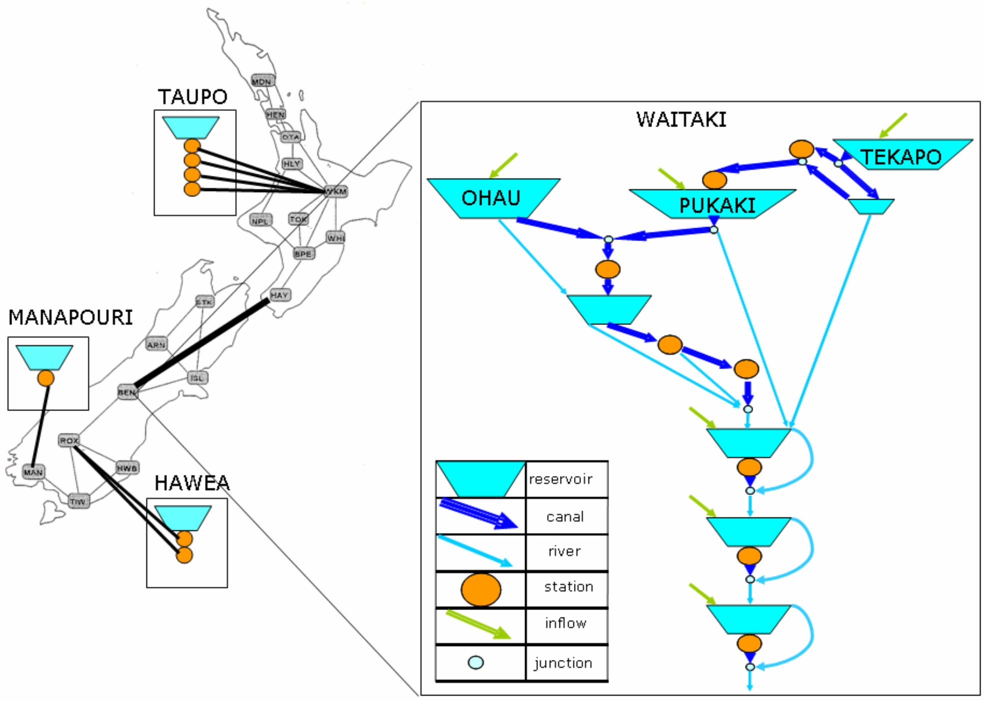

Figure 3: Approximate network representation of New Zealand electricity network showing main hydro-electricity generators

In the model we discriminate between thermal generation f_m, $m \in \mathcal{F}(i) \subseteq \mathcal{O}(i)$, and hydro generation $\gamma_m h_m$, $m \in \mathcal{H}(i) \subseteq \mathcal{O}(i)$. The parameter γ_m, which varies by generating station m, converts flows of water $h_m(p)$ into electric power.

This gives the following formulation:

$$
\begin{aligned}
\text{CP48: min} \quad & \textstyle\sum_{p=1}^{48} \sum_{i\in\mathcal{N}} \sum_{m\in\mathcal{F}(i)} \phi_m f_m(p) \\
\text{s.t.} \quad & g_i(y(p)) + \textstyle\sum_{m\in\mathcal{F}(i)} f_m(p) + \sum_{m\in\mathcal{H}(i)} \gamma_m h_m(p) = D_i(p), \quad i \in \mathcal{N}, \\
& x(p+1) = x(p) - A(h(p) + s(p)) + \omega(p), \;\; p = 1, 2, \ldots, 48, \\
& x(1) = \bar{x}(1), \qquad x(49) = \bar{x}(49), \\
& 0 \le f_m(p) \le a_m, \quad m \in \mathcal{F}(i), i \in \mathcal{N}, \\
& 0 \le h_m(p) \le b_m, \quad 0 \le s_m(t) \le c_m, \quad m \in \mathcal{H}(i), \\
& 0 \le x_m(p) \le r_m, \quad m \in \mathcal{H}(i), i \in \mathcal{N}, \quad y \in Y.
\end{aligned}
$$

Here the water balance constraints are represented by

$$x(p+1) = x(p) - A(h(p) + s(p)) + \omega(p)$$

where $x(p)$ is the reservoir storage at the start of period p, $s(p)$ denotes spill in period p, and $\omega(p)$ is the uncontrolled inflow into the reservoir in period p. All these are subject to capacity constraints. (In some cases we also have minimum flow constraints that are imposed by environmental resource consents.)

The node-arc incidence matrix A represents the river-valley network, and subtracts controlled flows that enter a reservoir from upstream from those that leave a reservoir by spilling or generating electricity. In other words row i of $A(h(p)+s(p))$ gives the total controlled flow out of the reservoir (or river junction) represented by row i, this being the release and spill of reservoir i minus the sum of any immediately upstream releases and spill. With small modifications to the water balance constraints, our model can also represent transit times for flows in parts of the river system. These can be as long as 22 periods (11 hours). In this setting the boundary conditions are augmented to account for in-transit flows at the beginning and end of the day.

3.3 The medium-term hydro model

To investigate the dispatch of hydroelectricity over the course of a year, a hydro-thermal release policy must be determined. This involves the solution of a large-scale stochastic dynamic programming model which is defined as follows. Let $x\,(t)$ denote the reservoir storage at the beginning of week t, and let $C_t(x, \omega(t))$ be the minimum expected fuel cost to meet electricity demand in weeks $t, t+1, \ldots, T$, when reservoir storage $x(t) = x$ and week t's inflow is known to be $\omega(t)$. Here $C_t(x, \omega(t))$ is the optimal solution value of the mathematical program:

$$
\begin{aligned}
\mathrm{P}_t(x,\omega):\ \ \min\ \ & \textstyle\sum_{i\in\mathcal{N}}\sum_{m\in\mathcal{F}(i)}\phi_m f_m(t) + \mathbb{E}[C_{t+1}(x(t+1),\omega(t+1))] \\
\text{s.t.}\ \ & g_i(y(t)) + \textstyle\sum_{m\in\mathcal{F}(i)} f_m(t) + \sum_{m\in\mathcal{H}(i)}\gamma_m h_m(t) = D_i(t), \quad i\in\mathcal{N}, \\
& x(t+1) = x - A(h(t)+s(t)) + \omega(t), \\
& 0 \le f_m(t) \le a_m, \quad m\in\mathcal{F}(i),\, i\in\mathcal{N}, \\
& 0 \le h_m(t) \le b_m, \quad 0 \le s_m(t) \le c_m, \quad m\in\mathcal{H}(i), \\
& 0 \le x_m(t) \le r_m, \quad m\in\mathcal{H}(i),\, i\in\mathcal{N}, \quad y\in Y.
\end{aligned}
$$

To solve this we have used the DOASA code [17] which is based on the SDDP technique of Pereira and Pinto [14]. This approximates $\mathbb{E}[C_{t+1}(x(t+1), \omega(t+1))]$ using a polyhedral function defined by cutting planes that is updated using samples of the inflow process. Weekly demand is represented by a load duration curve with three blocks, and $\omega(t)$ is sampled from historical inflow observations. We use a simplified transmission network comprising three nodes: one for the South Island (SI), one for the lower North Island (LNI) and one for the upper North island (UNI). The hydro system assumes that six reservoirs, Manapouri, Hawea, Ohau, Pukaki, Tekapo and Taupo, can store water from week to week. The remaining reservoirs are treated as run-of-river plant with limited intra-week flexibility. It is important to note that we assume inflows to the main catchments are stagewise independent. These are sampled from the historical weekly inflow series distributed with the CDS [11][4]. Full details of the DOASA model for this study can be found at the online companion [18] to this paper.

The solution to $\mathrm{P}_1(x_1, \omega(1))$ defines a set of thermal plants to run and a set of linear functions (or *cuts*) whose pointwise maximum approximates $\mathbb{E}[C_2(x(2), \omega(2))]$. Indeed the DOASA code yields an outer approximation to $\mathbb{E}[C_{t+1}(x(t+1), \omega(t+1))]$ at each stage t, and so this defines a policy at this stage by solving the

[4] In any year y we select inflows for each catchment in the years from 1970 to $y-1$ as equally likely random outcomes in each week. Thus for 2005 we have 35 (vector) outcomes per stage giving a stagewise independent scenario tree for DOASA of 35^{51} scenarios.

single-stage approximating problem:

$$
\begin{aligned}
\mathrm{AP}_t(x,\omega):\quad \min \quad & \sum_{i\in\mathcal{N}}\sum_{m\in\mathcal{F}(i)}\phi_m f_m(t)+\theta_{t+1} \\
\text{s.t.} \quad & g_i(y(t))+\sum_{m\in\mathcal{F}(i)} f_m(t)+\sum_{m\in\mathcal{H}(i)}\gamma_m h_m(t)=D_i(t), \quad i\in\mathcal{N}, \\
& x(t+1)=x-A(h(t)+s(t))+\omega(t), \\
& 0\le f_m(t)\le a_m, \quad m\in\mathcal{F}(i),\, i\in\mathcal{N}, \\
& 0\le h_m(t)\le b_m, \quad 0\le s_m(t)\le c_m, \quad m\in\mathcal{H}(i), \\
& 0\le x_m(t)\le r_m, \quad m\in\mathcal{H}(i),\, i\in\mathcal{N}, \quad y\in Y, \\
& \alpha_{t+1}^k+\beta_{t=1}^k x(t+1)\le\theta_{t+1},\, k\in\mathcal{C}(t+1).
\end{aligned}
$$

The policy obtained by DOASA depends on the number of cuts used. The experiments that we carry out below use both 200 cuts and 300 cuts per solve.

3.4 The weekly simulation model

The DOASA model defines a policy by the cuts that give the outer approximation of $\mathbb{E}[C_{t+1}(x(t+1),\omega(t+1))]$. In other words we can represent the expected future cost of meeting demand when $x(t+1)$ remains in the reservoirs at the end of week t by θ_{t+1} where

$$\theta_{t+1}=\max\{\alpha_{t+1}^k+\beta_{t+1}^k x(t+1),\, k\in\mathcal{C}(t+1)\}.$$

The policy is then to release water to meet demand at least cost over the week while accounting for the future cost. Given a set of cuts this policy can be simulated over a given period using the observed sequence of inflows ω to the hydro reservoirs. In each week we solve a model that includes the 18-node transmission system. This involves the solution of the following problem with $P=336$ trading periods:

$$\begin{aligned}
\mathrm{SP}_t(x,\omega): \quad \min \quad & \textstyle\sum_{p=1}^{P}\sum_{i\in\mathcal{N}}\sum_{m\in\mathcal{F}(i)}\phi_m f_m(p)+\theta_{t+1} \\
\text{s.t.} \quad & g_i(y(p))+\textstyle\sum_{m\in\mathcal{F}(i)} f_m(p)+\sum_{m\in\mathcal{H}(i)}\gamma_m h_m(p)=D_i(p), \quad i\in\mathcal{N}, \\
& x(p+1)=x(p)-A(h(p)+s(p))+\omega(p),\ p=1,2,...,P, \\
& 0\le f_m(p)\le a_m, \quad m\in\mathcal{F}(i),\ i\in\mathcal{N},\ p=1,2,...,P, \\
& 0\le h_m(p)\le b_m, \quad 0\le s_m(p)\le c_m, \quad m\in\mathcal{H}(i),\ p=1,2,...,P, \\
& 0\le x_m(p)\le r_m, \quad m\in\mathcal{H}(i),\ i\in\mathcal{N}, \quad y\in Y,\ p=1,2,...,P, \\
& \alpha_{t+1}^k+\beta_{t+1}^k x(P+1)\le\theta_{t+1},\ k\in\mathcal{C}(t+1).
\end{aligned}$$

Observe that $\mathrm{SP}_t(x,\omega)$ assumes perfect information about demand and inflow over the week being simulated, but does not anticipate inflows beyond that. The simulation is therefore likely to give an optimistic estimate of the fuel cost from implementing this policy for the week.

4 Market benchmarks over 2005-2009

We now decribe the first set of experiments that were carried out. All of these assume a counterfactual with a risk-neutral central planner and are solved over the planning horizon 2005-2009[5]. The Centralized Data Set (CDS) maintained by the New Zealand Electricity Authority records the offer curves for every generator in the wholesale market. It also records the historical dispatch level of each generator and daily reservoir inflows. Given costs per MWh of gas and coal generation it is therefore possible to compute the cost of fuel required to generate the electricity dispatched by the wholesale market in each half hour. This cost can be compared with the same cost as optimized by a central plan.

There are several difficulties with such an approach. The first of these concerns dispatch that has limited control. Examples of such dispatch is that from cogeneration, geothermal plant, run-of-river hydro and wind. Although these have low marginal cost, their availability is subject to the vagaries of inflows and wind, and so we cannot centrally dispatch these in a counterfactual. We choose to fix all cogeneration, geothermal generation wind generation, embedded generation, run-of-river generation and small hydro plant at their historical levels.

[5] At the time of writing, system hydrology data were available only until the middle of 2010, which restricts our study to the years 2005-2009.

This leaves the large hydro systems (Manapouri, Clutha, Waitaki and Waikato) available for control along with the major thermal plants (Huntly (4 units plus e3p and P40), Otahuhu, New Plymouth, Stratford, and Whirinaki). These are the only generators that we allow to offer energy within our model. In reporting costs, our measure will be the cost of fuel burned by these five plants as evaluated with the table of estimated fuel costs as shown in Table 1 and Table 2 (obtained from [12]). All costs are measured in 2008 dollars. Coal costs are assumed to be constant at \$4/GJ ([4]). The short-run marginal cost for any plant can be obtained by multiplying the heat rate by the fuel cost.

Station	**Fuel**	**Heat rate**
		(GJ/MWh)
Huntly	Coal	10.50
e3p	Gas	6.80
P40	Gas	9.50
Stratford	Gas	7.30
New Plymouth	Gas	11.00
Otahuhu	Gas	7.05
Whirinaki	Oil	11.00

Table 1: Heat rates and fuel types of thermal plant in New Zealand

		Diesel	Gas
2005	Mar	23.53	4.49
	Jun	25.42	4.21
	Sep	27.31	4.13
	Dec	26.53	5.14
2006	Mar	27.62	5.12
	Jun	32.69	5.07
	Sep	31.07	5.18
	Dec	26.07	5.67
2007	Mar	24.87	6
	Jun	26.23	5.97
	Sep	26.67	6.01
	Dec	29.71	5.57
2008	Mar	31.68	4.11
	Jun	38.45	5.13
	Sep	38.64	5.36
	Dec	29.00	5.77
2009	Mar	25.40	6.92
	Jun	23.26	7.11
	Sep	24.39	7.46
	Dec	24.33	7.44
2010	Mar	25.36	7.75
	Jun	27.52	8.41
	Sep	25.69	7.86

Table 2: Quarterly real gas and diesel wholesale prices in (2008 NZ)\$/GJ.

We note that the fuel prices here are estimated average values assuming fuel can be purchased on demand. Natural gas is typically acquired under a take-or-pay contract that gives a different operating imperative from that faced by a purchaser with more flexibility. Similarly coal is typically used from a stockpile that is periodically restocked; in this setting, supply shortages can lead to high opportunity costs. We argue, however, that a central planner might avoid many of the contractual problems in obtaining thermal fuel that a number of competing generators might face, which would make our assumption less important. In any case including these effects leads to a more complicated optimization problem than we would want to study here, so we ignore them.

To enable a fair comparison with the market outcomes, we have de-rated stations at which plant have been removed for planned maintenance, as outlined in the POCP database [21]. The schedule in POCP defines the starting and end time of scheduled maintenance for generators, which includes the offering generators and three others (Tokaanu, Rangipo and Waikaremoana) that we treat as fixed. We have observed that in some declared maintenance periods, the generators still offered or were dispatched energy. For the offering generators, we define the capacity loss due to maintenance to be the nominal capacity minus the maximum of their total offer and their dispatch (which can sometimes exceed the nominal capacity by a small amount) in the period of interest. For Tokaanu, Rangipo and Waikaremoana, for which only dispatch data are available, the capacity loss is defined to be the nominal capacity minus their dispatch. Note that since a generator may not offer or dispatch at its maximum available capacity in any period, our model overestimates the capacity loss due to maintenance.

We also assume that transmission outages are known at the time of dispatch. The time periods of HDVC line outage, and the HVDC flow are available in the CDS. We define the HVDC capacity loss in each such trading period to be the nominal capacity minus the HVDC flow. Other line outages can be detected by examining historical nodal prices. For each line, the ratio of historical nodal prices at the ends of the line and the ratio of power sent and received along this line are computed. If the former exceeds the latter, then this line is deemed to be constrained by some contingency. Some care is needed in treating lines in loops, as a contingency in one line can affect price differences around the loop. If this is the case, then the line in the loop with the highest ratio of nodal prices between its endpoints is assumed to be the one with the contingency. The capacity loss is then defined to be the difference between the nominal capacity and the power sent in the EP model.

In order to compare the dispatch of a central plan with that of the market we need to ensure that they both have the same boundary conditions. In other words, a market solution may burn more fuel than a central plan, while leaving all reservoirs with more water in them at the end of the day. In our daily experiments we wish to impose the same boundary conditions on both models in order to compare the efficiency of the dispatch.

The first task in doing this is to estimate what these boundary conditions should be in our models, since they do not represent all details of the river systems involved. For example the CDS contains only daily averages of tributary inflows, many of which vary over the course of a day. Even the averages are sometimes approximations of the true values, and reservoir levels at the beginning and end of a day fail to capture water flows that may be in transit between days. Moreover the conversion factors γ can vary with reservoir head level, and so nominal values of these might not correspond to the water releases associated with a given dispatch level.

The two largest river chains (Waikato at node WKM and Waitaki at node BEN) are block dispatched by the market. This means that the actual dispatches at nodes WKM and BEN respectively were allocated over the stations on their respective river chains so as to give a desirable hydrological flow pattern. We do not have information on how this reallocation was done. We use the historical dispatches $d_m(p)$ obtained from CDS data and solve the following model (called INTER) to find a block dispatch of each hydro chain i where $\mathcal{H}(i) \neq \emptyset$, that is consistent with $d_m(p)$ while meeting hydrology constraints and boundary conditions.

$$\begin{aligned}
\text{INTER: min} \quad & \textstyle\sum_{p=1}^{T} \sum_{m\in\mathcal{H}(i)} \lambda s_m(p) + \mu \left\| x(T+1) - \bar{x}(T+1) \right\|_1 \\
\text{s.t.} \quad & \textstyle\sum_{m\in\mathcal{H}(i)} u_m(p) = \sum_{m\in\mathcal{H}(i)} d_m(p), \quad i \in \mathcal{N}, \\
& (1-\alpha)\gamma_m h_m(p) \leq u_m(p) \leq_m (1+\alpha)\gamma_m h_m(p), \\
& x(p+1) = x(p) - A(u(p) - s(p)) + \omega(p), \\
& x(1) = \bar{x}(1), \\
& 0 \leq f_m(p) \leq a_m, \quad m \in \mathcal{F}(i), i \in \mathcal{N}, \\
& 0 \leq h_m(p) \leq b_m, \quad 0 \leq s_m(t) \leq c_m, \quad m \in \mathcal{H}(i), \\
& 0 \leq x_m(p) \leq r_m, \quad m \in \mathcal{H}(i), i \in \mathcal{N}, \quad y \in Y.
\end{aligned}$$

The objective of INTER is to minimize spill of water and absolute deviations from historical end storage levels using penalties λ and μ respectively. To minimize these deviations the water duty conversion factor for each station m is allowed to vary between $(1-\alpha)\gamma_m$ and $(1+\alpha)\gamma_m$ (where $\alpha = 0.1$ is a typical choice and γ_m denotes the nominal conversion factor with average headpond levels). The solution from INTER gives a (possibly new) set of hydrological boundary conditions for which the dispatches computed from MP1 for this day are feasible, while allowing for block dispatch. These final boundary conditions are the ones used in our comparisons of the central plan with the market. In other words, the solution from the INTER model is assumed to give the actual releases of water through each hydro station, since the historical dispatch from the CDS will give the total dispatch from each river chain in each trading period, that has possibly been redistributed by the operators. Because we do not have access to detailed information about each river chain, the end-of-day storage levels in INTER will be slightly different from historically observed storage, and so we penalize these differences to try and match history as much as possible. If we consider it more important to match historical reservoir levels then we should penalize deviations from these levels more severely than spill. The difficulty in doing this is that the historical solution might appear to spill large volumes of water to match what might be erroneous daily storage or inflow observations. Recall that the lake inflow values are daily averages, and tributary inflows are estimates, and so variations in historical generation might be accommodating changes in these inflow values that we do not have recorded.

Since spill appears to be relatively rare in practice, our approach for most stations is to penalize spill more heavily than matching the end conditions. (The exception is Manapouri, for which we do the opposite as very high inflows and resulting spill is a lot more common in this catchment). Penalizing spill means that the market solution in each trading period is not forced to spill past generating stations so as to match an historical end-of-day boundary condition. Instead we replace the boundary condition by a synthetic one, i.e. a set of storage levels that the market would have attained with average inflows, minimal spill and its historical dispatch. As a comparison the central plan is then computed using the synthetic boundary conditions. We observe that with more detailed, accurate and timely public information about historical river chain operation, there would be no need for the INTER model.

The decision horizon for INTER can be either a single day or a week. Since reach flows have delays, we augment the time horizon T with extra days to avoid missing in-transit flows which will contribute to the next stage's available water.

4.1 Experiment 1: Daily model

We now describe the results of repeating the experiments carried out in [19]. The first experiment recomputes the values that were listed in Table 3 of [19].

YEAR	MARKET	CENTRAL	DIFFERENCE	PERCENTAGE
2005	$451,896,771	$426,720,979	$25,175,792	5.6%
2006	$491,720,636	$471,071,279	$20,649,358	4.2%
2007	$487,837,719	$473,346,254	$14,491,466	3.0%

Results from [19, Table 3]. Differences in annual total fuel cost (in $NZ) between the market dispatch (MARKET) and central planning models

(CENTRAL) solved each day.

This shows the total of daily fuel cost for the market dispatch as compared with the central dispatch (solved each day with matching boundary conditions). Repeating the experiment with new CDS data gives an updated Table 3.

	MARKET	CENTRAL	DIFFERENCE	PERCENTAGE
2005	451,927,646	411,521,364	40,406,282	8.9%
2006	491,931,598	457,058,522	34,873,076	7.1%
2007	487,840,928	463,921,900	23,919,028	4.9%
2008	508,434,796	425,477,457	82,957,339	16.3%
2009	456,681,573	398,891,865	57,789,708	12.7%

Table 3: Differences in annual total fuel cost (in $NZ) between the market dispatch (MARKET) and central planning models (CENTRAL) solved each day.

The values in the MARKET column are slightly different in the updated table owing to some revised data in the CDS and differences in the estimated demand values obtained using EP (see Appendix 1). One can observe a substantial annual saving in fuel cost over the five years. Since CENTRAL uses the same amount of water each day in each reservoir, the difference in fuel cost cannot be attributed to conservatism in retaining water by risk-averse hydro generators. On the other hand, the half-hourly rolling horizon dispatch mechanism described above requires hydro generators to control their own water usage over a rolling horizon by choosing offer prices. When other agents are doing the same, the resulting uncertainty in future dispatch levels could produce inefficiencies when compared with a daily dispatch mechanism (like CENTRAL) that computes the dispatch for all 48 trading periods at once.

4.2 Experiment 2: Weekly model

In the daily model, the boundary conditions effectively constrain the total hydro generation to be the same under the market assumptions and the central plan. We now examine a central planning model (CP336) of 336 trading periods over one week (and extra days beyond this to deal with in-transit flows). Here we compare a central dispatch for all river chains assuming perfect information about demand and inflows in all periods in the week with the cost of the dispatch in the market dispatch. To produce a set of target reservoir levels for the end of the week that both models share, we solve an instance of INTER with 336 trading periods.

The experiment was carried out in [19] and the results were given in Table 5 of that paper reproduced below.

YEAR	MARKET	CENTRAL	DIFFERENCE	PERCENTAGE
2005	$451,896,771	$407,369,223	$44,527,548	9.9%
2006	$491,720,636	$450,223,740	$41,496,896	8.4%
2007	$487,837,719	$455,319,555	$32,518,164	6.7%

Results from [19, Table 5] showing total fuel cost savings (in $NZ) from weekly central plan dispatch compared with market dispatch

Repeating the experiment with new CDS data gives Table 4. As in Table 3 the cost savings are larger.

	MARKET	CENTRAL	DIFFERENCE	PERCENTAGE
2005	451,927,646	392,539,663	59,387,983	13.1%
2006	491,931,598	437,130,668	54,800,930	11.1%
2007	487,840,928	448,052,399	39,788,529	8.2%
2008	508,434,796	409,119,753	99,315,043	19.5%
2009	456,681,573	384,849,589	71,831,984	15.7%

Table 4: Total fuel cost savings (in $NZ) from weekly central plan dispatch compared with market dispatch

4.3 Experiment 3: Yearly model

In this section we investigate inefficiencies that might arise in the market from generators moving water from week to week in a suboptimal fashion. Studying this requires some care as perfect foresight here can generate large gains for a central plan that would not be realizable in practice. As observed by [8], the appropriate benchmark is the solution to a stochastic dynamic program that seeks to minimize expected fuel cost. As discussed above, we use a sampling-based model with six reservoirs. This gives water-value surfaces that can be used in a weekly dispatch model $\mathrm{SP}_t(x, \omega)$ to determine the optimal dispatch in each trading period of the week to minimize the thermal fuel cost in that week plus the expected future cost of using fuel from the end of the week. Since fuel costs are measured in 2008 dollars we do not discount future costs in our stochastic model.

As discussed above we also make use of costs for unserved load. These depend on the type of customer and the amount of load reduction.

	Up to 5%	Up to 10%	VOLL	North Is	South Is
Industrial	$1,000	$2,000	$10,000	0.34	0.58
Commercial	$2,000	$4,000	$10,000	0.27	0.15
Residential	$2,000	$4,000	$10,000	0.39	0.27

Table 5: Load reduction costs ($/MWh) and proportions of each load that is industrial, commercial, and residential load.

The last two columns of Table 5 show the proportion of load of each type in each island. This shows that 58% of South Island load is industrial, 15% commercial, and 27% is residential. In simulations we assume for simplicity that these proportions are the same at every node of the transmission grid. The costs (in NZ$/MWh) of shedding load are also shown in Table 5. We assume that up to 10% reduction in load can be achieved at a relatively low cost, but the value of unplanned interruption (or reduction above this level) is very high ($10,000/MWh). Therefore if, for example, load in the South Island was 1000MW, we could shed up to 5% of 580 MW at $1000/MWh and at $2000/MWh, we could shed 5% of 420MW (150MW commercial and 270 MW residential) plus a further 5% of 580MW industrial load.

The DOASA model minimizes expected fuel cost and so it is risk neutral. This means that shortages might be more frequent than is considered desirable by planners. In practice, when reservoir levels become very low, voluntary load shedding is expected to occur through a public electricity savings campaign. In the years of our study, load shedding campaigns could be triggered when total national reservoir levels fell below a critical level called a *minzone*. We have implemented this in DOASA and any violation of the minzone incurs a penalty of $9000/MWh. This means that up to 10% of load reduction (that has penalty costs of at most $4000/MWh) will occur in preference to minzone violation. This means that a market simulation might shed load even though lake storage volumes are not at their minimum levels. We have also added two minzone constraints that reflect reservoir levels in each island. The South Island minzone is 250 GWh less than the national minzone and has the same violation penalty. The North Island minzone is computed to enable environmental minimum flow constraints at Lake Karapiro to be met with high probability. Violations of this minzone are penalized at a lower value of $500/MWh, as breaches in these flow constraints would typically be allowed in preference to load shedding.

The solution to DOASA yields a set of 200 cuts for each week that define the future cost of meeting demand. Using the cuts from DOASA we simulate a central-plan policy by solving $\mathrm{SP}_t(x, \omega)$ for each of 13 weeks, using starting reservoir levels obtained from the end of the previous week. At the end of 13 weeks, we re-solve DOASA with the computed reservoir levels and a new 12 month planning horizon to obtain new cuts. This process then repeats.

The results of our experiment are shown in Figure 4 which compares the historically observed storage with that computed by the simulation of the central plan.

Observe that the trajectory of water storage for the central plan in Figure 4 is more extreme than the market trajectory. In the winters of 2005, 2006 and 2007 the central plan does not use the more expensive thermal plant as much as the market, and then recovers reservoir storage in the spring by augmenting hydro

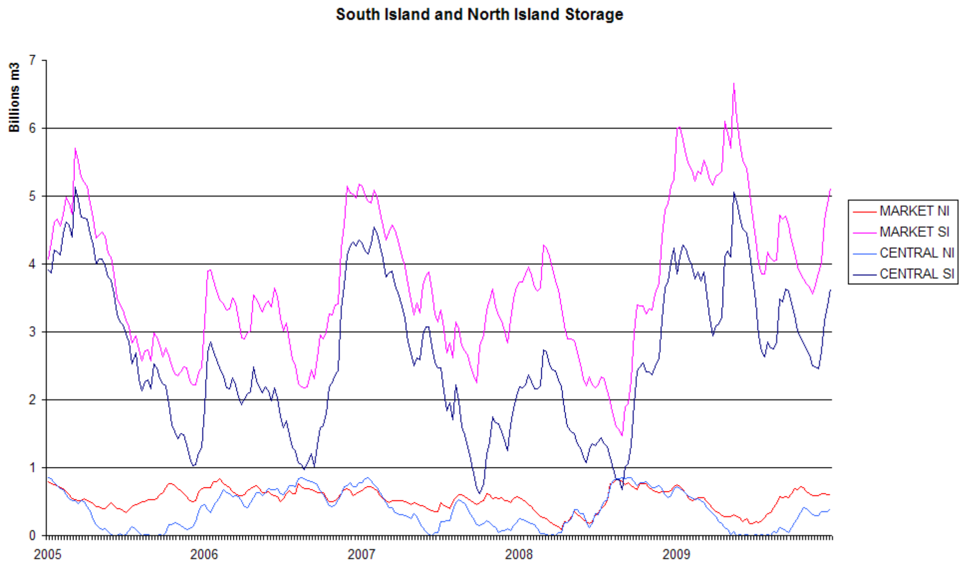

Figure 4: New Zealand reservoir storage levels (m^3) from the central planning simulation (CENTRAL) compared with historical (MARKET) levels over calendar years 2005-2009. Here NI and SI refer to North Island reservoir levels and South Island reservoir levels respectively. The MARKET solution retains approximately 700 GWh more storage at the end of 2009.

with base-load coal generation. The MARKET solution ends 2009 with higher reservoir levels than the central plan. This amounts to approximately 700GWh of energy.

	MARKET	CENTRAL	DIFFERENCE	PERCENTAGE	LOST LOAD
2005	451,927,646	364,867,582	87,060,064	19.3%	
2006	491,931,598	427,947,922	63,983,676	13.0%	12,999
2007	487,840,928	405,242,828	82,598,100	16.9%	
2008	508,434,796	438,479,906	69,954,890	13.8%	69,702,472
2009	456,681,573	329,628,389	127,053,184	27.8%	

Table 6: Annual fuel costs (in $NZ) incurred by a central plan that uses DOASA to plan water releases.

Table 6 shows how much extra value is extracted from the DOASA policy. The difference of $127,053,184 at the end of 2009 must be weighed up against the extra 700 GWh of water stored by the market at the end of 2009 as compared with the central plan. Assuming a future value of $50/MWh, this amounts to about $35 million, which gives the central plan a lower net savings in 2009 of $92 million or about 20%.

As observed in [19], Table 6 is flattering for CENTRAL, as the cost savings include the value that accrues from assuming perfect knowledge of inflows over all

336 trading periods in each week, a clairvoyance that is not enjoyed by the market solution. We can estimate a bound on this expected value of perfect information (EVPI) by the difference in total weekly savings and total daily savings as shown in Table 7.

	MARKET	EVPI	PERCENTAGE
2005	451,927,646	18,981,701	4.2%
2006	491,931,598	19,927,854	4.1%
2007	487,840,928	15,869,502	3.3%
2008	508,434,796	16,357,704	3.2%
2009	456,681,573	14,042,276	3.1%

Table 7: EVPI values estimated from the difference in savings between weekly CENTRAL and daily CENTRAL. The EVPI values should be added to the yearly CENTRAL fuel costs as the cost of day-to day uncertainty about inflows within a week.

Adjusting the annual figures for this EVPI and reducing the 2009 savings by $35 million gives the following adjusted figures.

	MARKET	CENTRAL	DIFFERENCE	PERCENTAGE	LOST LOAD
2005	451,927,646	383,849,283	68,078,363	15.1%	
2006	491,931,598	447,875,776	44,055,822	9.0%	12,999
2007	487,840,928	421,112,330	66,728,598	13.7%	
2008	508,434,796	454,837,610	53,597,186	10.5%	69,702,472
2009	456,681,573	378,670,665	78,010,908	17.1%	

Table 8: Final estimates of savings from CENTRAL solution computed using a DOASA policy updated each quarter and simulated over 5 years.

It is interesting to observe that the CENTRAL solution sheds about $70 million of load in 2008, which exceeds the fuel saving of $53.5 million earned in this year by this policy. One must bear in mind that the demand figures used in the models are from metered consumption, so they have already been reduced by any load shedding that actually occurred in 2008. The lost load reported here is in addition to this amount. The load is shed in CENTRAL to avoid violating the minzone constraints. The simulated storage trajectories in the simulation do not reach their minimum levels, and so one might regard this load shedding as being over cautious. The lost load means that the market outcome in 2008 is less costly than the central planned solution we have simulated, which has lower overall welfare.

In practice, energy shortages such as that observed in the simulation for 2008 are regarded as very undesirable by both planners and electricity market participants. Electricity regulators are charged by governments with ensuring that such shortages occur much more rarely than one might expect given the lost load costs we have been using. Ideally, market instruments that price reliability for customers might lead to the appropriate level of energy security. Until these mature, the level of reliability is the responsibility of the regulator. System reliability might be ensured by imposing constraints on operations (such as minzones) or by calling for public savings campaigns when lake levels get low. It appears from the results in Table 8 that forcing savings with minzone violations in a central plan is an imperfect mechanism for achieving this. On the other hand, if system shortages were modelled using a dynamic risk measure [15], then it is possible that a risk-averse central plan would perform better in dry years. We turn attention to this in the next section.

5 Improved benchmark policies

In this section we apply DOASA with different settings to the data from the previous section. The DOASA model we use in this section has several differences. In the first instance, we use a faster C++ implementation of the SDDP algorithm. DOASA in the previous section was implemented in AMPL Cplex, and each counterfactual run took days of computation. The C++ implementation is about 20 times faster. This enables us to compute more cuts per stage and also to simulate over a rolling horizon in which we solve DOASA every four weeks rather than every quarter (roughly three times as often). This means that we simulate the policy for only four weeks before re-solving to compute an updated policy. Although the data used in the models are the same, we work with seven reservoirs rather than the six of the previous sections.

In order to compare the new results wuth those from the previous section and to ensure that the market dispatches are consistent with water usage, we use the same water duty conversion factors that were computed in the INTER models of the previous section. The network model has three nodes and two transmission lines forming a simple radial representation of the New Zealand network. We simulate the policies obtained by DOASA in this smaller network rather than in the 18 node network of the previous section. This enables us to compute a benchmark with 300 cuts per stage in about 3 hours of CPU time per year of simulation.

As in the models of the previous section, we allow for the outages in both generation and transmission that were described in the previous section. To do this we assume full availability in the policy computation but then impose the observed and planned outages in simulation. This avoids any bias from plans that might anticipate these outages. The models were all computed assuming the historical demand as estimated from the EP models. The demand in 2008 and 2005 was reduced by savings campaigns and reductions in industrial load. To test the effect of this, we experimented with a model in which higher 2007 demand substituted for 2008. The results were similar in magnitude to the results reported below.[6]

None of the models we solve includes a minzone. However the C++ version of DOASA also allows us to compute risk-averse policies with varying levels of risk aversion. Risk is modelled using a convex combination of the expectation and average value at risk of future fuel and shortage cost (see e.g. [24]). In other words we use the risk measure

$$\rho(Z) = (1-\lambda)\mathbb{E}[Z] + \lambda \mathsf{AVaR}_{1-\alpha}[Z]$$

where $\lambda \in (0,1)$, Z represents the random future cost, and

$$\mathsf{AVaR}_{1-\alpha}[Z] = \inf_t \{t + \alpha^{-1}\mathbb{E}[(Z-t)_+]\}.$$

In order to enable the solution of a mutistage optimization problem, we must implement a dynamic version of this risk measure as described in [25] and [16]. The dynamic version uses a nested form of ρ, where the risk averse certainty equivalent of a random stream of costs, say Z_1, Z_2, Z_3, is computed using a nested formulation, which would be $\rho(Z_1 + \rho(Z_2 + \rho(Z_3)))$ in this example. A straightforward procedure for implementing this within SDDP algorithms is described in [15]. In our implementation the value of α is chosen to be $\frac{1}{M}$ where M is the number of scenarios used in each stage of DOASA. This means that the measure $\rho(Z)$ is equivalent to weighting all scenarios with equal probability $\frac{(1-\lambda)}{M}$ except for the most expensive scenario which receives weight $\frac{1}{M}(M\lambda - \lambda + 1)$.

5.1 Comparison with CENTRAL

5.1.1 Risk Neutral Planning

The first experiments were aimed at confirming the results from the previous experiments with CENTRAL. To do this we solved DOASA with 300 cuts over a 52-week period, and then implemented the policy for four weeks before re-solving with updated reservoir levels. We then compared the risk-neutral storage trajectory for the six South Island reservoirs against the historical storage of these in the market and against the trajectory for the six reservoirs that were modelled in the risk-neutral central plan. These are shown in Figure 5.

The SI CENTRAL trajectory (blue) was computed with penalties for violating a minzone. This leads to a slightly more conservative trajectory than the SI RISK NEUTRAL trajectory (green) which ignores the minzone. The green trajectory hits minimum storage levels in winter 2008. This means that \$41.52M of load is shed in the winter months of 2008. Savings from lower fuel costs in wet years

[6] The results of these experiments and other data from this study will be made publicly available on the EPOC website.

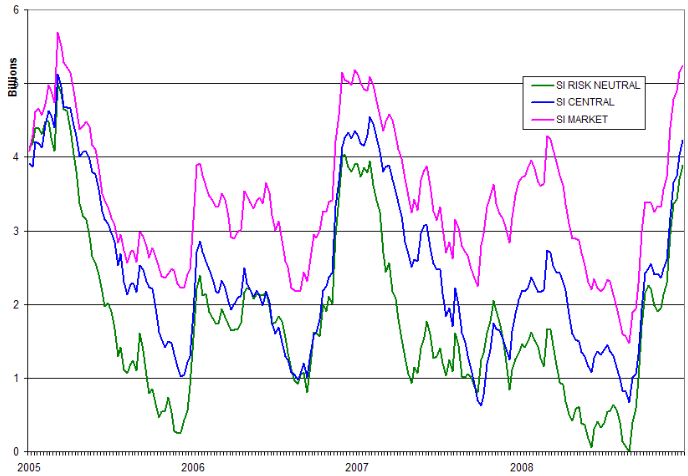

Figure 5: Historical total South Island (Manapouri, Hawea, Pukaki, Benmore, Tekapo, Ohau) storage trajectories, and those for the central plan computed in section 4 as well as the risk-neutral plan computed using DOASA with 4-weekly solves.

must be balanced by the \$41.52 M in shortage cost incurred in 2008. This is less than the shortage cost (\$70 M) incurred in the previous section (where a minzone was used). Observe that the risk neutral trajectory has less water remaining in storage at the end of 2008. This amounts to about 880GWh, or \$44.11 M at \$50/MWh.

When we account for the extra \$44.11 M incurred by the risk-neutral counterfactual we get annual fuel and shortage costs as shown in Table 10. The total short-term efficiency loss of the market over the four years is \$105.81 M.

	Annual costs (\$M)				
	MARKET	λ=0 Fuel	λ=0 Shortage	EVPI	Efficiency loss
2005	451.79	369.03	0.00	18.98	63.78
2006	490.99	442.73	0.00	10.93	37.33
2007	492.51	438.90	0.00	15.87	37.74
2008	508.49	439.74	41.52	16.36	-33.24

Table 10: Annual fuel cost for risk-neutral solution with 300 cuts. This incurs load shedding cost of \$41.52M in 2008. We add the EVPI values to the $\lambda = 0$ solution that has avoided these by anticipating intra-weekly inflows, and decrease the efficiency loss in 2008 by the difference in value of final storage.

The shortages in the risk neutral solution in 2008 lead to price spikes. We can derive the Benmore marginal water value from the simulation path. In this section prices are determined from the three-node network used in DOASA (rather than the 18 node network). The price at the South Island node is plotted in Figure 6.

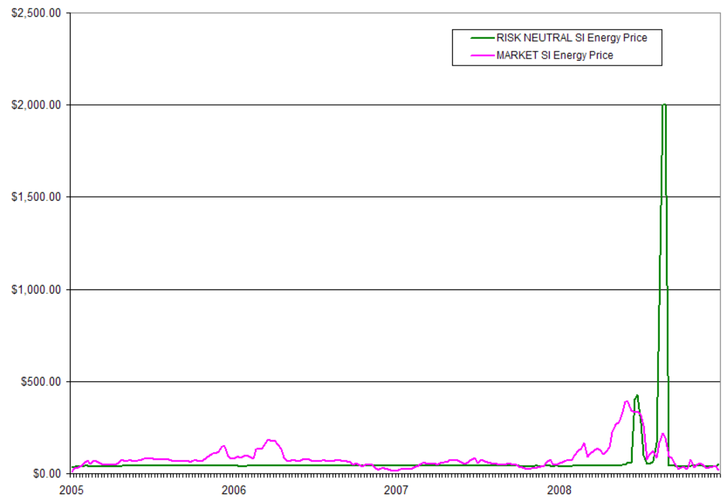

Figure 6: Weekly average South Island prices (green) under a simulated risk-neutral central plan compared with market historical averages at the Benmore node (pink).

5.1.2 Risk Aversion

We now consider increasing the level of risk aversion of the planner to high costs incurred by shortage. When risk aversion is added to DOASA the policies become more conservative. The plots of these are shown in Figure 7 for the market, the risk neutral plan ($\lambda = 0$), and a risk-averse plan ($\lambda = 0.5$). One can see that increasing the coefficient of risk aversion (λ) gives a risk-averse South Island trajectory (blue) that becomes closer to the storage trajectory of the market (pink). The difference in end values is 285.85GWh amounting to \$14.29 M at \$50/MWh. As one would expect on average, the annual thermal fuel cost increases as the level of risk aversion increases. Observe in this realization when a dry winter actually occurs, the fuel cost and water cost from the risk-averse plan is actually lower when summed over the four years of the simulation. These values and differences from the costs of the historical solution are shown in Table 11 and Table 12 below.

	Annual thermal fuel cost (\$M)		
	MARKET	λ=0	λ=0.5
2005	451.79	369.03	393.94
2006	490.99	442.73	451.07
2007	492.51	438.90	462.45
2008	508.49	439.74	437.32

Table 11: Annual fuel cost for different levels of risk aversion. The risk neutral solution ($\lambda = 0$) incurs load shedding cost of \$41.52M in 2008. The risk-averse

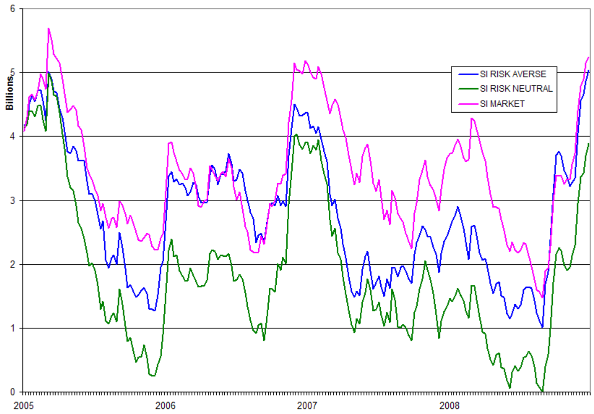

Figure 7: South Island storage trajectories for risk neutral ($\lambda = 0$) and risk-averse ($\lambda = 0.5$) policies

solution ($\lambda = 0.5$) incurs no load shedding.

	Annual costs ($M)				
	MARKET	λ=0.5 Fuel	λ=0.5 Shortage	EVPI	Efficiency loss
2005	451.79	393.94	0.00	18.98	38.87
2006	490.99	451.07	0.00	10.93	28.99
2007	492.51	462.45	0.00	15.87	14.18
2008	508.49	437.32	0.00	16.36	40.51

Table 12: Annual productivity gain for risk averse simulation. This can be compared with Table 10. We account for residual storage in the 2008 efficiency loss, so total efficiency losses of the market over four years are $122.55 M with respect to this plan ($105.61 M for the risk-neutral plan).

The South Island prices obtained by simulating the risk-averse solution are more conservative than those from the risk-neutral solution and remain quite stable. These are plotted in Figure 8.

This provides some evidence that a centrally planned solution might successfuly navigate the difficulties of a dry year without incurring shortages, or indeed leading to high electricity prices. The rentals earned by thermal plants under this solution are quite small compared with the market solution. This raises the problem of how to ensure that they cover their fixed costs. Clearly in such a system this "missing money" must be found from other sources.

5.2 Welfare effects

We now turn attention to studying electricity prices and welfare effects. The benefits that we calculate in this section are the difference between nominal rev-

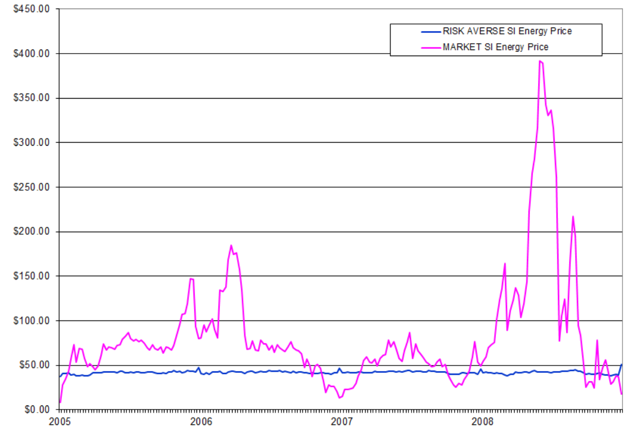

Figure 8: Weekly average South Island prices from risk averse model with $\lambda = 0.5$ (green) compared with historical Benmore prices (pink).

enue and fuel costs for the generators, the difference in value and payment for consumers, and the transmission rentals earned by the grid owner. These benefits should not be interpreted as annual profit or EBITDA for these entities, as such values must account for fixed costs that we are not including here. For this reason we will often refer them as *short-run benefits* or *rentals*. We have also not applied a 2008 price inflator to electricity prices in 2005-7 (as we did for costs). This means that welfare transfers from consumers to suppliers in the spot market will be understated in these years. (We include tables showing these transfers in 2008 prices in an appendix).

As shown in Figure 6, system marginal prices can become high when there is a threat of imminent shortage. Since prices are uniform, every generation unit gets paid the system marginal price in a given trading period. The model CP3 that we use to compute welfare effects is a version of $\mathrm{SP}_t(x, \omega)$ for each week t, where we represent a week by three load blocks, peak, shoulder, and offpeak giving $p = 1, 2, 3$, and we represent the opportunity cost of using water by cutting planes indexed by $k \in \mathcal{C}(t+1)$. The number of hours in each load block is denoted $\tau(p)$. We have three nodes UNI, LNI, and SI , denoted $i = 1, 2, 3$, and unsigned line flows y_{12} and y_{23}. We denote load shedding cost by $l_m(p)$, $m \in \mathcal{S}(i)$, $p = 1, 2, 3$, where each tranche $m \in \mathcal{S}(i)$ costs ψ_m per MWh. Here $S(i)$ is the set of nine tranches of load shedding available at each node as defined by Table 5.

We formulate CP3 as follows.

$$
\begin{array}{ll}
\text{CP3: min} & \sum_{p=1}^{3} \tau(p) \sum_{i\in\mathcal{N}} \sum_{m\in\mathcal{F}(i)} \phi_m f_m(p) + \sum_{i\in\mathcal{N}} \sum_{m\in\mathcal{S}(i)} \psi_m l_m(p) + \theta \\
\text{s.t.} & -y_{12}(p) + \sum_{m\in\mathcal{F}(1)} f_m(p) + \sum_{m\in\mathcal{H}(1)} \gamma_m h_m(p) + \sum_{m\in\mathcal{S}(1)} \psi_m l_m(p) = D_1(p), \\
& y_{12}(p) - y_{23}(p) + \sum_{m\in\mathcal{F}(2)} f_m(p) + \sum_{m\in\mathcal{H}(2)} \gamma_m h_m(p) + \sum_{m\in\mathcal{S}(2)} \psi_m l_m(p) = D_2(p), \\
& y_{23}(p) + \sum_{m\in\mathcal{F}(3)} f_m(p) + \sum_{m\in\mathcal{H}(3)} \gamma_m h_m(p) + \sum_{m\in\mathcal{S}(3)} \psi_m l_m(p) = D_3(p), \\
& \\
& x(4) = x(1) - \sum_p A(h(p) + s(p))\tau(p) + \omega, \;\; p = 1,2,3, \\
& x(1) = \bar{x}(1), \\
& \\
& \alpha_{t+1}^k + \beta_{t+1}^k x(4) \le \theta, \; k \in \mathcal{C}(t+1). \\
& \\
& -K_{12}(p) \le y_{12}(p) \le K_{12}(p), \quad -K_{23}(p) \le y_{23}(p) \le K_{23}(p), \\
& \\
& 0 \le f_m(p) \le a_m, \quad m \in \mathcal{F}(i), \, i = 1,2,3,, \\
& 0 \le l_m(p) \le d_m(p), \quad m \in \mathcal{S}(i), \;\; p = 1,2,3, \quad i = 1,2,3,, \\
& 0 \le h_m(p) \le b_m, \quad 0 \le s_m(t) \le c_m, \quad m \in \mathcal{H}(i), \, i = 1,2,3,, \\
& 0 \le x_m(4) \le r_m, \quad m \in \mathcal{H}(i), \, i = 1,2,3,.
\end{array}
$$

We denote the shadow prices on the first three constraints by $\pi_1(p)$, $\pi_2(p)$ and $\pi_3(p)$. We are now in a position to define welfare. We will do this by defining the hourly benefits to each agent in each load block p. The total benefits per week will then be these values multiplied by $\tau(p)$ and added over the three blocks.

The total of generation times system marginal price gives the total revenue per hour that is received by generators in load block p. We call this *generator revenue*, denoted $R(p)$. Thus

$$
R(p) = \sum_{i=1}^{3} \pi_i(p) \left(\sum_{m\in\mathcal{F}(i)} f_m(p) + \sum_{m\in\mathcal{H}(i)} \gamma_m h_m(p) \right).
$$

When electricity prices are different at different nodes, the system operator earns an hourly *transmission rental*, denoted $T(p)$. Since we do not have loops or account for transmission losses in our model, we have

$$
T(p) = |\pi_2(p) - \pi_1(p)| \, K_{12}(p) + |\pi_3(p) - \pi_2(p)| \, K_{23}(p),
$$

where the capacity $K_{ij}(p)$ in trading period p is adjusted for any outages. The total hourly *fuel cost* is denoted $F(p)$ and is

$$
F(p) = \sum_{i=1}^{3} \sum_{m\in\mathcal{F}(i)} \phi_m f_m(p),
$$

where the sum does not count load shedding cost. Thus hourly *total generator benefit* $G(p)$ is revenue minus fuel cost, whence

$$G(p) = R(p) - F(p).$$

The load shedding cost per hour is denoted $L(p)$ where

$$L(p) = \sum_{i=1}^{3} \sum_{m \in \mathcal{L}(i)} \psi_m l_m(p)$$

In our model, consumers at different locations pay the locational price times their load. The hourly consumer payment $P(p)$ is then

$$P(p) = \sum_{i=1}^{3} \pi_i(p)(D_i(p) - \sum_{m \in \mathcal{L}(i)} l_m(p))$$

The *total gross consumer benefit* $C(p)$ (i.e. their hourly welfare ignoring their payments) is the constant hourly gross welfare $J(p)$ that they get from uncurtailed demand minus their loss in hourly welfare from curtailed demand. Thus

$$C(p) = J(p) - L(p).$$

Thus the total hourly *net benefit of consumers* $N(p)$ is

$$N(p) = C(p) - P(p).$$

We can sum the hourly benefits in each period p, to get total system benefit $W(p)$ accruing to consumers, generators, and the grid owner respectively.

$$\begin{aligned} W(p) &= N(p) + G(p) + T(p) \\ &= J(p) - L(p) - P(p) + R(p) - F(p) + T(p). \end{aligned}$$

In each hour it also makes sense to track payments. The consumers pay $P(p)$. Since we do not model losses, the total amount of generation is $\sum_{i=1}^{3}(D_i(p) - \sum_{m \in \mathcal{L}(i)} l_m(p))$. The generators are paid

$$R(p) = \sum_{i=1}^{3} \pi_i(p) \left(\sum_{m \in \mathcal{F}(i)} f_m(p) + \sum_{m \in \mathcal{H}(i)} \gamma_m h_m(p) \right).$$

The difference is

$$
\begin{aligned}
P(p) - R(p) &= \sum_{i=1}^{3} \pi_i(p)(D_i(p) - \sum_{m \in \mathcal{L}(i)} l_m(p)) - \sum_{i=1}^{3} \pi_i(p) \left(\sum_{m \in \mathcal{F}(i)} f_m(p) + \sum_{m \in \mathcal{H}(i)} \gamma_m h_m(p) \right) \\
&= \sum_{i=1}^{3} \pi_i(p)((D_i(p) - \sum_{m \in \mathcal{L}(i)} l_m(p)) - \sum_{m \in \mathcal{F}(i)} f_m(p) - \sum_{m \in \mathcal{H}(i)} \gamma_m h_m(p)) \\
&= \pi_1(p)(-y_{12}(p)) + \pi_2(p)(y_{12}(p) - y_{23}(p)) + \pi_3(p)(y_{23}(p)) \\
&= (\pi_2(p) - \pi_1(p))y_{12}(p) + (\pi_3(p) - \pi_2(p))y_{23}(p) \\
&= T(p)
\end{aligned}
$$

as price differences only occur between nodes when $y_{ij}(p) = K_{ij}(p)$ or $y_{ij}(p) = -K_{ij}(p)$. Thus load payments contribute to generator revenue and transmission rentals.

We are now in a position to compare total hourly welfare in two different dispatch solutions, denoted by subscripts M and by C, for market and counterfactual. Thus since $J_C(p) = J_M(p)$ we get

$$
\begin{aligned}
W_C(p) - W_M(p) &= J_C(p) - L_C(p) - P_C(p) + R_C(p) - F_C(p) + T_C(p) \\
&\quad -(J_M(p) - L_M(p) - P_M(p) + R_M(p) - F_M(p) + T_M(p)) \\
&= -L_C(p) - P_C(p) + R_C(p) - F_C(p) + T_C(p) \\
&\quad -(-L_M(p) - P_M(p) + R_M(p) - F_M(p) + T_M(p)) \\
&= L_M(p) - L_C(p) + F_M(p) - F_C(p)
\end{aligned}
$$

when we substitute

$$
P_C(p) - R_C(p) = T_C(p) \quad \text{and} \quad P_M(p) - R_M(p) = T_M(p)
$$

The increase in total benefit then amounts to adding up the decreases in lost load cost and fuel cost.

We are also interested in transmission rentals but these pose some difficulties. In our counterfactual model with three nodes we may easily calculate transmission rentals. However the transmission rentals in the market are computed using a full SPD transmission network and so these are almost certain to be different from those in CP3, even with the same dispatch. We compare annual historical figures provided in documents on the Electricity Authority web site with our three-node estimates for a risk neutral solution shown in Table 14.

	Transmission rentals ($M)	
	MARKET	λ=0
2005	92.00	1.70
2006	61.00	0.25
2007	67.00	0.41
2008	210.00	908.62

Table 14: Total loss and constraint rentals for the market and for the CP3 solution. The annual historical figures are taken from [5].

These differences must be borne in mind when comparing the market and the counterfactuals.

5.2.1 Risk neutral results

We are now in a position to compare changes in short-run benefits under the counterfactual solution as compared with historical outcomes. We first study the risk neutral case. The benefit outcomes are summarized in Table 15 and Table 16.

	Annual consumer benefit ($M)		
	MARKET	λ=0	Benefit gain
2005	-2496.76	-1407.27	-1089.49
2006	-2650.23	-1445.19	-1205.04
2007	-1779.29	-1438.55	-340.74
2008	-4275.19	-3615.11	-660.08
	Annual generator benefit ($M)		
	MARKET	λ=0	Benefit gain
2005	1952.97	1017.56	935.41
2006	2098.24	991.28	1106.96
2007	1219.78	983.37	236.41
2008	3556.71	2164.76	1391.95
	Annual transmission benefit ($M)		
	MARKET	λ=0	Benefit gain
2005	92.00	1.70	90.30
2006	61.00	0.25	60.75
2007	67.00	0.41	66.59
2008	210.00	908.62	-698.62
	Total benefit ($M)		
	MARKET	λ=0	Benefit gain
2005	-451.79	-388.01	-63.78
2006	-490.99	-453.66	-37.33
2007	-492.51	-454.77	-37.74
2008	-508.49	-541.73	33.24

Table 15: Short run benefits in simulation of moving from counterfactual (competitive) solution to a market (historical) solution. Figures include avoiding shortage costs and value of extra stored water at the end of 2008.

	Short run benefit gains from operating market ($M)			
	Consumers	Generators	Grid owner	Total
2005	-1089.49	935.41	90.30	-63.78
2006	-1205.04	1106.96	60.75	-37.33
2007	-340.74	236.41	66.59	-37.74
2008	-660.08	1391.95	-698.62	33.24

Table 16: Summary of short-run benefit gains from moving to the market from the counterfactual solution

These results merit some discussion. As compared with the risk-neutral competitive benchmark, generators in the market over the four years of the study are about \$3.67 B better off. The consumers and the grid owner are worse off. Over these four years about \$105.61M is lost through inefficient dispatch. The total benefits (revenue minus fuel costs) earned by generators in the risk-neutral competitive benchmark is (from Table 15) about \$5.16B. Of this, \$2.16B is earned in the shortage periods of 2008 when the grid owner earns rentals of \$909M.

It is interesting to compare the figures obtained in Table 16 with those in the Wolak report [27] shown in Table 17.

Table 5.4: Annual decomposition of wholesale market revenues

Year	Wholesale revenues	Variable costs		Competitive rents		Market power rents	
		$ million	% of total	$ million	% of total	$ million	% of total
Counterfactual 1							
2001	2956.9	562.7	19.0%	906.9	30.7%	1487.4	50.3%
2002	1457.9	528.2	36.2%	900.9	61.8%	28.7	2.0%
2003	3052.9	581.8	19.1%	975.6	32.0%	1495.5	49.0%
2004	1371.1	516.8	37.7%	1117.7	81.5%	-263.4	-19.2%
2005	2904.2	697.9	24.0%	1255.7	43.2%	950.7	32.7%
2006	3118.7	780.4	25.0%	1570.7	50.4%	767.7	24.6%
1H 2007	1049.5	395.6	37.7%	784.4	74.7%	-130.5	-12.4%

Table 17: Decomposition of market revenues from the Wolak report (reproduced from [27, Table 5.4]). These figures ignore productive inefficiency. Counterfactual 1 is the benchmark that fixes hydro generation at historical levels.

This estimates "market power rents" in 2005 to be \$950.7 M under Counterfactual 1 which is close to the figure of additional benefit \$935.41 estimated from our analysis.

In an energy-only market, the benefits accrued in the risk-neutral competitive benchmark can be seen as payments to cover the fixed costs of generating plant. It is interesting to see in our experiments that these rentals are not equally shared. The total annual benefits (revenue minus fuel costs) averaged over 4 years that are earned by major generators in the risk-neutral counterfactual break down as follows.

	Annual benefit ($M)
Contact Energy	244.13
Genesis	90.98
Meridian	747.56
Mighty River Power	185.70
Trustpower	47.45

Table 18: Average annual benefits by generators in competitive benchmark solution

The rentals earned by Genesis in the counterfactual model accrue almost entirely from their hydro generation. It is tempting to suppose that their major thermal plant Huntly earns substantial rentals when prices in the South Island average $2000/MWh in some weeks of 2008. This would align with the theory of scarcity pricing covering fixed costs. In our model, when these prices occur, the transmission capacity from Huntly in the upper North Island to other regions is constrained (mainly by reductions from outages). The consequence is that Huntly's nodal price is its short-run marginal cost for many trading periods in these weeks. The rentals in these weeks are accrued by South Island generators and the grid owner (who makes $908.62 M over 2008).

It is somewhat perplexing that the very hydro-firming conditions for which Huntly was designed and built should provide it virtually no earnings to cover its fixed costs. The competitive market resolution to these seemingly perverse incentives would, at the margin,

1. decrease Huntly's (unused) capacity and
2. increase the transmission capacity

Both actions would increase the earnings of the asset owners at the margin until the thermal capacity was fully dispatched and earning a capacity rent, or the line was no longer constrained. Of course this argument is subject to the usual caveats of economies of scale and indivisibilties, but it does indicate that the competitive benchmark we have computed is not as perverse at it might first appear.

5.2.2 Risk averse results

We now present the results for the counterfactual risk-averse solution. Because South Island prices do not peak in this simulation, the transmission rentals are very modest as shown in Table 19.

	Transmission rentals ($M)	
	MARKET	λ=0.5
2005	92.00	2.11
2006	61.00	0.73
2007	67.00	0.56
2008	210.00	4.43

Table 19

The short-run benefit outcomes for the risk averse case are summarized in Table 20, and Table 21.

	Annual consumer benefit ($M)		
	MARKET	λ=0.5	Benefit gain
2005	-2496.76	-1416.87	-1079.89
2006	-2650.23	-1454.25	-1195.98
2007	-1779.29	-1449.38	-329.91
2008	-4275.19	-1377.65	-2897.54
	Annual generator benefit ($M)		
	MARKET	λ=0.5	Benefit gain
2005	1952.97	1001.84	951.13
2006	2098.24	991.52	1106.72
2007	1219.78	970.49	249.29
2008	3556.71	905.25	2651.46
	Annual transmission benefit ($M)		
	MARKET	λ=0.5	Benefit gain
2005	92.00	2.11	89.89
2006	61.00	0.73	60.27
2007	67.00	0.56	66.44
2008	210.00	4.43	205.57
	Total benefit ($M)		
	MARKET	λ=0.5	Benefit gain
2005	-451.79	-412.92	-38.87
2006	-490.99	-462.00	-28.99
2007	-492.51	-478.32	-14.18
2008	-508.49	-467.97	-40.51

Table 20: Short run benefits of moving from counterfactual risk-averse central plan to a market (historical) solution.

	Short run benefit gains from operating market ($M)			
	Consumers	Generators	Grid owner	Total
2005	-1079.89	951.13	89.89	-38.87
2006	-1195.98	1106.72	60.27	-28.99
2007	-329.91	249.29	66.44	-14.18
2008	-2897.54	2651.46	205.57	-40.51

Table 21: Summary of short run benefits of moving from counterfactual risk-averse central plan to a market (historical) solution.

It is easy to see that the risk-averse counterfactual plan is less efficient in 2005, 2006 and 2007, but makes more savings in 2008 when it avoids shortage costs. The substantial decrease in electricity prices in 2008 means that consumers over the four years make \$5.50 B less in short-run benefits in the wholesale market than if they paid prices computed from the counterfactual risk-averse dispatch. A total of \$4.96 B of this welfare transfer goes directly to generator short-run benefits. In the counterfactual, generators only make \$3.87 B in short-run benefits from 2005 to 2008, whereas our estimate of this figure in the market simulation is \$8.83 B.

6 Conclusions

In this paper we have described some experiments with stochastic optimization models of the NZEM that provide counterfactual outcomes for competitive markets. The policies obtained from these models have been backtested against historical outcomes over the years 2005-2008, and differences in generation and prices recorded. The benchmark policies are designed to maximize expected system welfare, and our calculations bear this out over the period 2005-2008 where the counterfactual solutions are more efficient on average. Of course this is based on one historical sequence of inflows out of many possible ones that could have occurred, so it is not possible to claim that the counterfactual model will always perform better. Indeed in 2008 we see it was less efficient.

It is interesting to speculate on the causes of the differences in prices and rentals observed in the market and the counterfactuals above. The Wolak report [27] makes a case for strategic exercise of market power as being the primary cause, but uncertainty and risk also play a role that must not be discounted. Electricity spot markets around the world are set up as one shot games, and work well when the time horizon is short enough so that all necessary information is known at the time of dispatch. Uncertainty (such as wind intermittency) causes problems for this model. Valuing the opportunity cost of water is also difficult in this setting. Our counterfactual model is an attempt to remove some of the bias associated with hindsight benchmarks. We have not eliminated this entirely as our model admits full clairvoyance of intra-week inflows, albeit reduced in efficiency calculations by an EVPI approximation. In addition, the benchmark model will still have some residual bias from relaxing other information constraints that will be present in a real setting.

As observed above, it is possible (at least in theory) for a Walrasian equilibrium to give a stochastic process of prices with respect to which every agent optimizes its own expected benefit with the outcome of maximizing total expected welfare. Such an equilibrium might give a sample path of prices as observed in Figure 6. As shown by [3], the stochastic process of prices that yields an equilibrium might be very complicated with none of the stagewise independence properties that make computing optimal policies easy for generators. We would contend that optimal electricity market design for systems with large amounts of hydro generation is not as well understood as the theory for purely thermal markets, and more remains to be done to improve the operation of these markets to get closer to welfare-optimizing outcomes. The different solutions adopted in various jurisdictions (e.g. New Zealand, Brazil, Colombia, France) are evidence of the different approaches to tackling this issue.

There are many enhancements that could be made to our models to make them more realistic and more accurate. The obvious improvement is to update the data set to encompass the last four years 2009-2012. This will be possible with the publication of updated hydrological series in the next published Centralized Data Set to appear in 2013. The fuel costs that we used could be increased to include other operation costs that make up short-run marginal cost, but these are relatively small. As mentioned above, we could also include stockpiles of coal and gas in the model which would alter the costs of thermal generation to include some possible loss in opportunity value.

The inflow processes used in DOASA are assumed to be stagewise independent. This means that a sequence of dry weeks will occur in the model with lower probability than we would expect. When reservoir levels are low, and low inflows persist, this assumption will tend to produce optimistic estimates of future costs. The marginal water values at low reservoir levels are therefore likely to be lower in our model than in a model with serial dependence. Whether a simulation of such a policy with DOASA would match historical outcomes is not known, and will depend on the particular model of dependence that is implemented. An aggregated reservoir model using constructive dual dynamic programming ([22],[23]) gives similar storage trajectories to historical outcomes over 2007-2008 (see [26]).

Another weakness of our recent experiments is the approximation to three load blocks. This attenuates the peak period demands when we expect to see very high prices. The simulations in section 4 used half hour trading periods and an 18 node network so the prices here might be a more accurate representation of what would see in a real risk-neutral central plan, bearing in mind that these prices are strongly affected by our choices of penalties for minzone violations. A more comprehensive simulation would include spinning reserve and other ancilliary services. Reserve in particular can have a significant effect on energy dispatch and prices. The recent development of vSPD, a full-scale public domain version of SPD (see [10]) raises the possibility of incorporating a full-scale dispatch simulation in CP48. This would be a useful set of experiments to form part of the market monitoring work stream of the Electricity Authority.

Our models could also be criticized for not including hedge contracts, but this is difficult to do as the contract information of electricity market participants is not in the public domain. When all market participants are perfectly competitive and risk-neutral, there is no incentive for any producer to alter behaviour. For such producers we would expect to see similar benefits as reported above even in the presence of contracts. On the other hand, the behaviour of competitive risk-averse agents will differ when contracts are included, and so our results in this setting will be misleading. If risk-neutral agents are able to exercise market power in the market then the inclusion of contracts may alter their payoffs in the short term. However for a rational expectations equilibrium with risk-neutral contract traders it it can be shown (see e.g. [13]) that the contract price for a baseload contract must be equal to the time-weighted spot price, so under these assumptions contracts can be expected to have no effect on welfare transfers in the long run.

References

[1] T. Alvey, D. Goodwin, M. Xingwang, D. Streiffert, and D. Sun. A security-constrained bid-clearing system for the New Zealand wholesale electricity market. *IEEE Transactions on Power Systems*, 13(2):340 – 346, 1998.

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[3] K. Barty, P. Carpentier, and P. Girardeau. Decomposition of large-scale stochastic optimal control problems. *RAIRO-Operations Research*, 44(03):167–183, 2010.

[4] T. Denne, S. Bond-Smith, and W. Hennessy. Coal prices in New Zealand markets, COVEC report for New Zealand Ministry of Economic Development, 2009. http://www.med.govt.nz/upload/68784/coal-price-report.pdf.

[5] EnergyLink. Losses and constraints excess projections. http://www.ea.govt.nz/dmsdocument/10014.

[6] L. Evans, S. Hogan, and P. Jackson. A critique of Wolak's evaluation of the NZ Electricity Market: Introduction and overview. *Available at SSRN 1968026*, 2011.

[7] D. Heath and H. Ku. Pareto equilibria with coherent measures of risk. *Mathematical Finance*, 14(2):163–172, 2004.

[8] O. Kauppi and M. Liski. *An empirical model of imperfect dynamic competition and application to hydroelectricity storage.* MIT Center for Energy and Environmental Policy Research, 2008.

[9] P. Lino, L.A.N. Barroso, M.V.F. Pereira, R. Kelman, and M.H.C. Fampa. Bid-based dispatch of hydrothermal systems in competitive markets. *Annals of Operations Research*, 120(1):81–97, 2003.

[10] EMI-Electricity Market Information System. New Zealand Electricity Authority. http://reports.ea.govt.nz/emiemi.htm.

[11] New Zealand Electricity Authority. Centralized Data Set, November, 2010. http://www.ea.govt.nz/industry/monitoring/cds/.

[12] New Zealand Ministry of Economic Development. New Zealand Ministry of Economic Development - Energy Data, June 2009. http://www.med.govt.nz/templates/{M}ultipage{D}ocument{TOC}21660.aspx.

[13] D.M. Newbery. Competition, contracts, and entry in the electricity spot market. *The RAND Journal of Economics*, 29(4):pp. 726–749, 1998.

[14] M. V. F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.

[15] A. Philpott, V. de Matos, and E. Finardi. On solving multistage stochastic programs with coherent risk measures. *Operations Research*, forthcoming, 2013.

[16] A.B. Philpott and V.L. de Matos. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. *European Journal of Operational Research*, 218(2):470 – 483, 2012.

[17] A.B. Philpott and Z. Guan. On the convergence of stochastic dual dynamic programming and other methods. *Operations Research Letters*, 36:450–455, 2008.

[18] A.B. Philpott and Z. Guan. Online companion for "Production inefficiency of electricity markets with hydro generation". Technical report, Electric Power Optimization Centre, 2010.

[19] A.B. Philpott, Z. Guan, J. Khazaei, and G. Zakeri. Production inefficiency of electricity markets with hydro generation. *Utilities Policy*, 18(4):174 – 185, 2010.

[20] A.B. Philpott and G. Pritchard. Financial transmission rights in convex pool markets. *Operations Research Letters*, 32(2):109 – 113, 2004.

[21] POCP. POCP database of planned outages. http://pocp.redspider.co.nz.

[22] E. G. Read. A dual approach to stochastic dynamic programming for reservoir release scheduling. *Dynamic Programming for Optimal Water Resources Systems Analysis (AO Esogbue, Ed.)*, pages 361–372, 1989.

[23] E. G. Read and M. Hindsberger. Constructive dual DP for reservoir optimization. In *Handbook of Power Systems I*, pages 3–32. Springer, 2010.

[24] R.T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.

[25] A. Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209:63–72, 2011.

[26] G. Telfar. Winter 2008: A Meridian perspective. `http://www.epoc.org.nz/ww2009.html`, 2009.

[27] F.A. Wolak. An assessment of the performance of the New Zealand wholesale electricity market. Technical report, New Zealand Commerce Commission, 2009.

7 Appendix giving differences of EMBER data from previous studies

This appendix documents some small changes in modelling assumptions and data estimation in the EMBER project as compared with the benchmark models described in [19],[18].

The CDS data are from the 2010 November version of the CDS provided by the New Zealand Electricity Authority (EA) [11] and span years 2005 to 2010, with hydrology data up to June 30 in 2010, and demand and generator data up to September 30 in 2010. Additional data were downloaded from the Ministry of Economic Development (MED, now MBIE) website [4],[12].

Demands in the 18 nodes are the aggregation of demands at demand points extracted from the CDS [11]. Some demands have negative values as the embedded generation exceeds the demands in aggregation. This is different from the previous project in which negative values we rounded to zero. Although demand can be negative, load shedding is the proportion of the positive demand only.

The prices are extracted from the CDS. When gnash commands return several prices for each node, the first price is arbitrarily chosen. There can be a small difference among the prices for each node. Given this, in estimating line derating in the EP model, we choose the maximum slope in calculating the ratio of marginal power sent and received instead of the actual slope at the particular power to obtain a conservative estimate of line derating.

Historical dispatches for offering generators and fixed generators as well as offer quantities for offering generators are extracted from the CDS [11]. The conversion factor for Manapouri is 1.518 in the CDS, which can be as high as 1.65 if we admit a 1.1 scaling factor. However the conversion factor calculated from the maximum historical dispatch of 885 MW at the maximum flow rate of 510 cumec is 1.7352. Thus we have set the nominal conversion factor to 1.65, and allow it to vary between 1.518 and 1.7352.

For the daily central and weekly central models, we use daily conversion factors estimated from the inter models. For the yearly central model, since the reservoir storages may be significantly different from those from the inter models in each week, the daily conversion factor are not exactly the same as those estimated. Thus we choose to use the weekly average of these daily conversion factors as the conversion factor throughout the week.

The thermal generator No. 3 in New Plymouth was dispatched and offered in some periods in 2008 after its decommissioning in 2007. Thus its decommissioning date is set to the end of 2008, with derating to restrict the capacity in each TP to be the maximum of historical dispatch and offer.

Fuel costs are real wholesale prices in quarters of a year up to the 3rd quarter in 2010, denominated in 2008 New Zealand dollars. The gas cost is available up to the 2nd quarter of 2009, and thus the costs for the remaining quarters are calculated such that their changes in percentages from the 2nd quarter in 2009 are the same as those of diesel costs. Note that since the diesel cost is 4 to 5 times of the gas cost, the diesel-fired generator Whirinaki has a running cost of at least \$250/MWh, substantially higher than \$84/MWh shown in the CDS. Our estimate is supported by the high historical prices at the TPs when Whirinaki were dispatched, and also online documents published by industrial companies.

The HVDC lines are represented in our model as a pair of lines running in opposite directions between HAY and a node A, followed by a pair between A and BEN, and also a pair running between HAY and a node B, followed by a pair between B and BEN. Each line has the same nominal capacity. However, the historical flows, which are power sent from HAY and BEN extracted from the CDS, were sometimes over the nominal capacity, up to 200MW. Thus the HVDC capacity is allowed to vary over time, and is computed as the maximum of the nominal capacity and the historical flow.

The HVDC forced outage data are extracted from the CDS to compute the HVDC derating. For a trading period when an outage occurred, the maximum of historical flows in the HVDC lines connecting the node A is computed, and the line derating for all lines connecting A are set to be the capacity minus the maximum. The same applies to the lines connecting the node B and their derating. The HVDC lines may be constrained even though there is no forced outage, in particular when reserve constarints are binding. These may be picked up by the price difference at the ends of HVDC, and such derating are estimated from prices and flows, as described in [19].

In the CDS, the daily average storage in cubic meters are available from 1990 to mid 2010. The Opus Consultants lake storage report in the CDS [11] describes the methodology in creating these data, and discusses the discrepancy in data between Opus and COMIT Hydro. (We noticed such discrepancies in using data from both sources in the project described in [19].) The report uses the lake levels from the power archive data and level-storage tables to create the storage. It also states that water over maximum control levels is possible and can be used for generation.

The nominal capacities of the six large lakes in our risk-neutral models are the storage at the normal maximum control levels specified in the report. Since historical storage occassionally exceeds these capacities, we use varying capacities in our models. The actual daily capacity for each reservoir is the maximum of the nominal capacity and historical storage, which doubles the capacities for Ohau and Manapouri in some time periods.

Updated values of reservoir inflows are provided by Opus in the CDS[11]. Some values are different from those in the preceding data set (October 2008 CDS). Inflows for Arapuni and Manareduced due to rating change in computation started on 23 Sep 1969 and 3 Mar 1976 in respective data sets, and those for Benmore due to the NIWA quarterly update were erroneously missed in the 2008 version.The inflow data from week 27 to 39 in 2010 are not available and thus those in the previous year are used when computing the release policy in 2009 (that looks 52 weeks ahead). Lake storages and scaling factors for conversion factors of hydro generators in these weeks are not available either, and thus those in the latest week are used.

There are negative inflow values for Clyde, Roxburgh, Waikaremoana and Manapouri. Negative values may be due to reservoir leaks or evaporation. For example, the minimum inflow of -90 for Manapouri with a 141km^2 surface area means 2.2 minimeters per hour in evaporation, which is reasonable. The negative values are retained in this project, which were set to zero in the data used for [19].

The penalty cost for flow bound violation is \$500/MWh, and the conversion factor in computation of energy from flow bound violation is set to be 0.2 MW/cumec.

The dispatching of thermal generator No. 3 in New Plymouth after decommissioning was not anticipated and thus is not taken into account in DOASA.

There was an outage in pole one of the HVDC line for a long period in 2008, which was not anticipated. Then policy generation would assume it was available until it was out, and assume it was available again after it was back to operation. However, the information hasn't been obtained, and thus this feature is not incorporated in the model.

The reservoir capacities in each week are the ones on the last day of the week in Central.

The inflow data from week 27 to 39 in 2010 are not available and thus those in the previous year are used. Lake storages and scaling factors for conversion factors of hydro generators in these weeks are not available either, and thus those in the latest week are used.

8 Appendix giving welfare transfers in 2008 dollars

In this appendix we give the results of welfare transfers when market electricity prices in 2005-2008 are expressed in 2008 real terms. Observe that is difficult to compare these with nominal values taken from years other than 2008 (such as the 2005 figure from the Wolak report).

8.1 Risk neutral values

We compute changes in short-run benefits under the counterfactual solution as compared with historical outcomes. Here 2005-2007 prices in the market are inflated by the following factors taken from [12].

2005	1.214
2006	1.113
2007	1.100
2008	1.000

Price inflation factors to express annual values in (2008) NZD.

We first study the risk neutral case. The benefit outcomes are summarized below.

	Annual consumer benefit ($M)		
	MARKET	λ=0	Benefit gain
2005	-3010.60	-1407.27	-1603.32
2006	-2942.05	-1445.19	-1496.86
2007	-1950.51	-1438.55	-511.97
2008	-4275.19	-3615.11	-660.08
	Annual generator benefit ($M)		
	MARKET	λ=0	Benefit gain
2005	2466.81	1017.56	1449.24
2006	2390.06	991.28	1398.78
2007	1391.01	983.37	407.63
2008	3556.71	2164.76	1391.95
	Annual transmission benefit ($M)		
	MARKET	λ=0	Benefit gain
2005	92.00	1.70	90.30
2006	61.00	0.25	60.75
2007	67.00	0.41	66.59
2008	210.00	908.62	-698.62
	Total benefit ($M)		
	MARKET	λ=0	Benefit gain
2005	-451.79	-388.01	-63.78
2006	-490.99	-453.66	-37.33
2007	-492.51	-454.77	-37.74
2008	-508.49	-541.73	33.24

Short run benefits (in 2008 NZD) in simulation of moving from counterfactual (competitive) solution to a market (historical) solution. Figures include avoiding shortage costs and value of extra stored water at the end of 2008.

	Short run benefit gains from operating market ($M)			
	Consumers	Generators	Grid owner	Total
2005	-1603.32	1449.24	90.30	-63.78
2006	-1496.86	1398.78	60.75	-37.33
2007	-511.97	407.63	66.59	-37.74
2008	-660.08	1391.95	-698.62	33.24

Summary of short-run benefit gains from moving to the market from the counterfactual solution

8.1.1 Risk averse results

We now present the results for the risk-averse solution expressed in 2008 dollar terms. Because South Island prices do not peak in this simulation, the transmission rentals are very modest as shown below.

	Transmission rentals ($M)	
	MARKET	λ=0.5
2005	92.00	2.11
2006	61.00	0.73
2007	67.00	0.56
2008	210.00	4.43

The short-run benefit outcomes for the risk averse case are summarized below.

	Annual consumer benefit ($M)		
	MARKET	λ=0.5	Benefit gain
2005	-3010.60	-1416.87	-1593.73
2006	-2942.05	-1454.25	-1487.80
2007	-1950.51	-1449.38	-501.14
2008	-4275.19	-1377.65	-2897.54
	Annual generator benefit ($M)		
	MARKET	λ=0.5	Benefit gain
2005	2466.81	1001.84	1464.96
2006	2390.06	991.52	1398.54
2007	1391.01	970.49	420.52
2008	3556.71	905.25	2651.46
	Annual transmission benefit ($M)		
	MARKET	λ=0.5	Benefit gain
2005	92.00	2.11	89.89
2006	61.00	0.73	60.27
2007	67.00	0.56	66.44
2008	210.00	4.43	205.57
	Total benefit ($M)		
	MARKET	λ=0.5	Benefit gain
2005	-451.79	-412.92	-38.87
2006	-490.99	-462.00	-28.99
2007	-492.51	-478.32	-14.18
2008	-508.49	-467.97	-40.51

Short run benefits of moving from counterfactual risk-averse central plan to a market (historical) solution (2008 NZD).

	Short run benefit gains from operating market ($M)			
	Consumers	Generators	Grid owner	Total
2005	-1593.73	1464.96	89.89	-38.87
2006	-1487.80	1398.54	60.27	-28.99
2007	-501.14	420.52	66.44	-14.18
2008	-2897.54	2651.46	205.57	-40.51

Summary of short run benefits of moving from counterfactual risk-averse central plan to a market (historical) solution (2008 NZD).