Submitted to *Operations Research*
manuscript (Please, provide the manuscript number!)

On cost recovery in energy-only electricity markets under uncertainty

Edward Anderson
Imperial College, London, e.anderson@imperial.ac.uk

Andy Philpott
Electric Power Optimization Centre, University of Auckland, Auckland, New Zealand, a.philpott@auckland.ac.nz

We examine cost recovery in competitive energy-only electricity markets when investment and operating decisions are made under uncertainty. Using a stochastic social-planning model with price-responsive demand, we show that when market prices are given by the shadow prices of the energy-balance constraints, then optimally chosen generation and storage assets will recover their investment costs through expected operating profits. The result extends the classical screening-curve argument to multistage investment and to a general class of technologies with intertemporal operating constraints. Applications include thermal generation, variable renewable generation, ramp-constrained plant, batteries, pumped storage, and hydroelectric generation. For technologies whose capacity can be expanded freely at the margin, expected operating profits exactly equal investment costs, supporting the system-optimal capacity under free entry. By contrast, hydroelectric resources may earn additional rents because uncontrolled inflows are scarce and cannot generally be expanded. The analysis clarifies the conditions under which energy-only prices support efficient investment across a broad range of electricity-system technologies.

Key words: electricity markets, capacity expansion, screening curve

1. Introduction

In this paper we reconsider a classical result for energy-only electricity markets: when generation capacities are chosen optimally and the market is perfectly competitive, prices based on short-run marginal costs allow owners of generation capacity to recover their investment costs through scarcity rents. We ask how far this result extends to stochastic systems containing variable renewable generation, storage, ramping constraints and hydroelectric resources, and whether it supports system-optimal investment through free entry.

The recovery of generation investment costs through energy-market revenues is well established in the theory of peak-load pricing. In an optimally chosen portfolio of generation technologies, marginal-cost pricing produces scarcity rents during periods in which capacity constraints bind, and these rents exactly cover the fixed costs of the capacity installed. Early analyses were given by Boiteux (1960) and Steiner (1957);

the same principle underlies the screening-curve analysis commonly used in electricity economics and is discussed, for example, by Green (2000) and Stoft (2002). These treatments typically use deterministic demand-duration curves or stylized static models in which technologies differ in their fixed and variable costs. A recent exception is Holmberg and Philpott (2026), which studies screening curves with variable renewable electricity modeled as a random variable.

The ability of energy-only markets to support efficient investment in practice has nevertheless been questioned. Price caps, market power mitigation, imperfect demand response and incomplete contracting may prevent scarcity prices from reaching the levels required for fixed-cost recovery - often described as a "missing-money" problem. This has motivated a substantial literature on resource adequacy and capacity mechanisms; see, for example, Joskow and Tirole (2007), Joskow (2008) and Newbery (2016). Our purpose is different. We will not be considering whether actual electricity markets satisfy the institutional conditions required for efficient investment. Instead we consider a competitive market in which investment can occur freely at the margin, and ask whether the physical and intertemporal characteristics of modern generation and storage technologies invalidate the classical cost-recovery argument.

Several papers have examined pricing and cost recovery in markets with uncertain or intermittent generation. The papers Pritchard et al. (2010), Cory-Wright et al. (2018) and Zakeri et al. (2019) study energy-only settlement mechanisms for market participants facing uncertain demand. In Philpott et al. (2016) and Ferris and Philpott (2022) the authors establish the connection between social planning and competitive equilibrium in multistage hydrothermal systems under uncertainty, including extensions to risk-averse agents with complete risk markets. Most closely related to the present paper, Korpås and Botterud (2020) derive optimality and cost-recovery conditions for thermal generation, variable renewable energy and storage using a generalized net-load-duration-curve model.

We extend this literature by formulating investment and dispatch on a general scenario tree and allowing price-responsive demand and stagewise capacity expansion. We develop a linear investment-operations model that encompasses thermal generation, variable renewable generation, ramp-constrained plant, batteries, pumped storage and hydroelectric reservoirs. For these models, we establish the recovery of fixed costs when capacities are set at their optimal levels from a social planning solution. We are able to show that cost recovery is very robust to different model formulations as long as they share some common features. For each technology, investment costs must be linear in capacity, and operating costs linear in controls. The technology must also be representable using linear constraints and constant conversion efficiencies, for example, we rule out increases in turbine efficiency with increased input. Our results broaden the class of technologies for which competitive energy-only prices can support efficient investment and cost recovery. For exposition purposes we will focus on a market at a single location. Introducing electricity transmission leads to locational prices and possible transmission rents, which brings up the problem of cost recovery of transmission assets. The same duality-based reasoning can be extended to transmission investment with locational prices, but we do not pursue that extension here.

Our discussion will be primarily framed in terms of a single generation plant whose capacity must be decided. Because capacity costs are assumed to be linear, the discussion carries over into a portfolio setting in which multiple plants with identical parameters could be constructed. When individual plants have a fixed or bounded size we can consider a social optimum in which the planner determines how many of these plants are to be built. When plant size is relatively small or there is some flexibility on individual plant sizes we can expect to recover the system-optimal capacity for this type of generation plant.

This setting makes it natural to consider decentralized entry and exit decisions by investors in individual plants. We ask whether this can be expected to result in a system-optimal amount of generation being built. When there are no barriers to entry, and given our assumption on linear investment costs at the plant level, it is clear that if a particular technology yields a positive profit then new plants of that type will be built (so that new investment occurs at the margin). Equally if a plant is losing money then we may suppose that its owner will reduce its capacity or withdraw it from the market, if there are no barriers to doing so.

Now we can see the impact of our results in this setting. We show in this paper that for many different types of generation there is no net profit when capacities are set at the system-optimal levels. If in addition we can show that reducing capacities below the system-optimal level leads to higher plant-level profit, and the reverse when capacities are increased, then this establishes that the system-optimal capacity is supported as a free-entry equilibrium when there are no barriers to entry or exit.

Our assumption will be that prices in the market are defined by the Lagrange multipliers of the energy-balance constraints in a system-optimal solution in which capacities are held constant. This is generally the case when the market is perfectly competitive, so generators supply their energy at its marginal cost. It is worth noting that arguments showing that free entry of new generation leads to a system-optimal choice of capacities cannot be carried over to the owner of a single plant choosing capacity when other generation capacity is held fixed: in this case the owner earns positive profit if its capacity is reduced below the system-optimal level.

For ease of exposition, all our results assume risk-neutral market participants. When entry at the margin is unconstrained, the expected rent from operations minus fixed costs will equal 0. This result extends to the case where agents optimize using coherent risk measures and risk markets are complete: then the risk-adjusted rent from operations minus fixed costs will equal 0 (see, e.g., Ehrenmann and Smeers (2011), Ferris and Philpott (2022), Kok et al. (2018)).

It is interesting to compare the optimal levels of wind power and hydroelectricity investment in a socially optimal solution. Wind generators do not pay for wind and the optimal investment makes zero profit in expectation. In contrast, if hydroelectric generators do not pay for inflows at their marginal value then they may make positive profits in expectation. Positive profits should incentivize entry in hydroelectric generation in the form of capacity expansion. However the positive profits come from inflows which cannot be expanded (unless a catchment is increased). Expanding turbine or reservoir capacity above their socially optimal levels will reduce welfare. The inability to expand inflows is a barrier to entry that may yield a strictly positive expected Ricardian rent for hydroelectric resources.

Although wind generators similarly do not pay for the energy provided by wind, the wind resource is different from inflows. The available wind energy scales proportionally with the capacity of the wind turbines that are built. Thus adding additional wind capacity is not constrained because, unlike inflows it does not "use up" the available resource.

We can summarize here the contributions of this paper:

1. We generalize the classic screening curve analysis that is typically used to show that in a competitive market with a uniform clearing price the cost of building capacity is exactly met by excess revenue for each generation type if there is a socially optimal portfolio. Our treatment covers stochastic behavior over time and allows capacity to be added in stages.

2. We show that the cost-recovery result applies to every generation plant whose capacity is chosen as part of the system optimization, regardless of which other plant capacities are also optimized and which are held fixed. When capacities are determined by a scalar quantity rather than a vector, and absent constraints arising from exogenous inputs a unilateral variation in one plant's capacity yields nonnegative profit below its system-optimal level and nonpositive profit above it. Thus the jointly system-optimal portfolio is supported by free entry and exit.

3. We show how this framework applies with renewable generation, both with and without storage. This allows us to show that for wind power, batteries and pumped storage the fundamental result holds.

4. We show that if generation depends on a scarce exogenous resource that does not expand with installed capacity, as with hydro inflows, then the generator's revenue may exceed that required for capacity construction.

The paper is laid out as follows. In the next section we consider the most basic model of investment under uncertainty to establish notation and a framework for analysis. Here the investment problem is linked to a corresponding operational model in which asset owners earn Ricardian rents from market clearing prices. In section 3.1 we study a single thermal generation plant and show that its cost is recovered by expected rents when there is a system optimal capacity choice. This leads to the free entry result of section 3.2. Section 3.3 extends this analysis to the case where capacity expansion occurs over several stages. In section 4 we formulate a general convex optimization model that we will use to establish all the following results. The remainder of section 4 applies this framework to variable renewable generation and run-of-river hydro, for which current decisions do not affect future states. In section 5 we turn to a range of environments in which an optimal decision requires consideration of future states. We deal successively with ramping constraints, batteries, pumped-storage and hydro reservoirs with inflow. Finally, section 6 concludes the paper by considering some extensions.

2. The basic model

We begin with a simple model in which uncertainty arises only from demand and generation costs. So, at this stage, we do not include intermittent renewable generation (such as wind).

Uncertainty. Uncertainty is represented in a general and flexible manner using a discrete-time stochastic process with a finite set of possible outcomes at each of T stages. This process is modeled via a *scenario tree* with nodes $n \in \mathcal{N}$ and leaf nodes $\mathcal{L} \subseteq \mathcal{N}$. The root node is indexed by $n = 0$, and for each non-root node $n \neq 0$, we denote its unique predecessor or parent node by $n-$. For each non-leaf node n, the set of immediate successors of n is denoted $n+$.

All parameters and decision variables are indexed by the node n at which they are realized. In particular, each node n is associated with a demand served of $d(n)$, and a probability of occurrence $p(n)$, with $p(0) = 1$. Probabilities divide in the natural way so that $p(n) = \sum_{m \in n+} p(m)$. We assume only nodes with $p(n) > 0$ are included. This formulation allows for arbitrary correlation structures between demand and costs.

Price-dependent demand. To capture price-responsive consumption, we assume that consumers are represented by a (collective) strictly increasing concave utility function $U_n(d)$ for electricity consumption d at node n. Variations in $U_n(\cdot)$ across nodes reflect differences in underlying demand conditions. The marginal utility $U_n'(d)$ represents consumers' willingness to pay for an incremental unit of electricity at node n. If U_n is not differentiable, there could be a range of supporting prices for the demand. We assume this does not occur, so that $U_n'(d)$ is uniquely determined and continuous. This makes $U_n'(d)$ a non-increasing continuous function of d (possibly with some horizontal sections). In practice, this utility function is reflected through demand-side bidding in the market; modeling demand via utility allows us to simplify the formulation. We restrict attention to non-negative consumption levels, so that $d(n) \geq 0$ for all $n \in \mathcal{N}$.

Generation technologies and capacity investment. Electricity generation is supplied by a portfolio of technologies indexed by a finite set $\mathcal{J}$. For each generation type $j \in \mathcal{J}$, the variable cost of production at node n is denoted by $c_j(n)$, reflecting, for example, uncertainty in fuel prices.

Investment and operation decisions are separated. Let b_j denote the capacity of generation type j installed ex ante, prior to the resolution of uncertainty. We assume that generation capacity has a linear investment cost, with κ_j denoting the cost per unit of capacity for type j. At each node n, generation output $y_j(n)$ must satisfy $0 \leq y_j(n) \leq b_j$, incurring a variable cost of $c_j(n)y_j(n)$ at that node.

Unmet demand. Electricity markets operate with an implied cost for demand that is not met. In our model this is handled through the utility function, since demand is chosen to maximize utility rather than being exogenously fixed. If there is a value VOLL for the value of lost load, then this provides an upper bound on the possible prices and hence the derivative of U_n for every node n. Since U_n is concave, we see that if the bound on the derivative is active over a region, then this implies some point $\tilde{q}(n)$ below which U_n becomes affine with derivative VOLL. Thus VOLL in our model is an endogenous cap on marginal utility, rather than being an exogenously imposed regulatory price cap.

System optimization. We are now ready to state the problem of minimizing expected cost (with utility treated as a negative cost). Some of the capacities are taken as fixed, and the others b_j for $j \in \mathcal{J}_v$ are to be determined as part of the optimization. The other variables are, for each node n, the demand served $d(n)$ and the generation quantities $y_j(n)$. The basic system optimization problem considered is as follows

$$
\begin{aligned}
\text{GENOPT} : \min \quad & \sum_{j \in \mathcal{J}_v} \kappa_j b_j + \sum_{n \in \mathcal{N}} p(n) \left(\sum_{j \in \mathcal{J}} c_j(n) y_j(n) - U_n(d(n)) \right) \\
\text{s.t.} \quad & \sum_{j \in \mathcal{J}} y_j(n) = d(n), \quad \text{for all } n \in \mathcal{N} \qquad [p(n)\pi(n)] \qquad (1) \\
& 0 \leq y_j(n) \leq b_j, \quad \text{for all } n \in \mathcal{N}, j \in \mathcal{J} \qquad [p(n)\rho_j(n)] \qquad (2)
\end{aligned}
$$

The non-negativity of $d(n)$ follows from non-negativity of $y_j(n)$ so we do not need to include that constraint explicitly. The only non-linear term here comes from U_n, and as we assume utilities are concave, this will be a convex problem. The square bracketed terms are the Lagrange multipliers on the constraints. It is convenient to let these be $p(n)\pi(n)$ and $p(n)\rho_j(n)$. With $p(n)$ included in the multiplier, $\pi(n)$ becomes the derivative of U_n and hence the appropriate price for energy.

The dispatch problem. Dispatch takes place after capacities are determined. This leads to a problem we call DISPATCH(b) where the vector b of capacities for each generation type is fixed. This social planning problem is identical to GENOPT, but with a reduced set of decision variables $y_j(n)$ and $d(n)$, (i.e., not including the capacities b_j) and without the term $\sum_{j \in \mathcal{J}} \kappa_j b_j$ in the objective.

This basic model is appropriate for thermal plants, but later we will discuss batteries and hydro storage, where the problem is more complicated. This will imply that DISPATCH(b) is formulated in a scenario tree over several periods subject to uncertain demand and energy supply, implying that the system operator needs to solve a multistage stochastic program, yielding optimal charge and discharge quantities as well as storage levels in each node of the scenario tree.

We work with the solution of the social planning problem DISPATCH(b). Under suitable completeness and price-taking assumptions, this allocation can be decentralized as a competitive equilibrium, with the market prices given by the Lagrange multipliers on the corresponding balance constraints. This is the standard duality argument underlying the welfare-theorem connection between competitive equilibria and social optima (see, e.g., Arrow and Debreu (1954)).

In our electricity market context we can be precise about the operation of the bid-based market that will result in the social planning allocation. We assume a perfectly competitive market by which we mean that agents make operating decisions on the basis of prices that are given (or anticipated) without making allowance for changes in prices that could occur as a result of their actions. Thus, throughout the paper, we assume no exercise of market power.

Since DISPATCH(b) is a convex optimization problem with linear constraints its optimal solution satisfies the following KKT conditions:

$$\pi(n) = U_n'(d(n)), \tag{3}$$

$$0 \le c_j(n) - \pi(n) + \rho_j(n) \perp y_j(n) \ge 0, \tag{4}$$

$$0 \le b_j - y_j(n) \perp \rho_j(n) \ge 0. \tag{5}$$

A price-taking generator j in node n seeing price $\pi(n)$ will solve the problem

$$\max_{0 \le y_j(n) \le b_j} \pi(n) y_j(n) - c_j(n) y_j(n),$$

giving the same conditions (4) and (5) as above.

Thus at node n if each generator bids its true costs $c_j(n)$ and the system operator uses a merit order dispatch on the basis of these bids, taking account of the utility function U_n to determine the optimal choice of demand $d(n)$, then the outcome will be the system-optimal solution to DISPATCH(b). Notice that this implies there is no need for a generator to make an estimate of the price before determining its bid. As we will see, this last observation will not hold for a storage agent, who can only determine their bid after making estimates of future prices.

Price formation. In this energy-only market a single price is set at each node and applies to all generation at that node. Typically this price is equal to the bid of the most expensive generator dispatched. However, when demand depends on the utility function U_n, the price may lie between two bids so that demand exactly equals supply. This can occur if there is a discontinuity in the price arising from exhausting the capacity of a cheaper generator and then requiring a more expensive generator for the next increment of demand.

Prices are unique. For both GENOPT and DISPATCH(b) the price at node n is given by $\pi(n)$, the dual variable associated with constraint (1). From the first-order condition for $d(n)$, we have $\pi(n) = U_n'(d(n))$ when $d(n) > 0$, and we assume this holds for every node n in an optimal solution. Our assumption on U_n being continuously differentiable implies that this price is unique when $d(n)$ is fixed.

Even though GENOPT and DISPATCH(b) may have multiple optimal solutions, in each case the set of prices, $\pi(n)$, are uniquely determined. To see this suppose we have two different solutions possibly with different capacity choices b_j (for GENOPT), with different generation quantities and different demands $d_1(n)$ and $d_2(n)$. As each problem is convex, every point on the line segment joining the two optimal solutions is also optimal with the same objective. Because every term in the objective is linear, apart from the terms involving U_n, we can deduce that the U_n terms are all affine between $d_1(n)$ and $d_2(n)$. Thus $U_n'(d_1(n)) = U_n'(d_2(n))$ and the prices for the two optimal solutions will match.

LEMMA 1. *If (b^*, d^*, y^*) is an optimal solution to GENOPT, and $(\bar{d}, \bar{y})$ is an optimal solution to DISPATCH(b^*), then $(b^*, \bar{d}, \bar{y})$ is an optimal solution for GENOPT with prices at each node that match the shadow prices for GENOPT at the (b^*, d^*, y^*) solution.*

Proof. As we have observed the DISPATCH(b) problem is equivalent to GENOPT if all capacities are determined (i.e., when $\mathcal{J}_v$ is empty), since then the extra term $\sum_{j\in\mathcal{J}} \kappa_j b_j$ in GENOPT is a constant. Thus any solution to DISPATCH(b) is system optimal for these fixed capacities, from which it follows that $(b^*, \bar{d}, \bar{y})$ is optimal for GENOPT when $b = b^*$. From our earlier remarks on the uniqueness of the prices, we deduce that the prices will match at each node. ∎

3. Classical static dispatch models

In this section we place the classical capacity expansion problem (where capacities are chosen for dispatchable thermal plants) into a framework in which operations occur in a scenario tree. The operations of each plant are assumed to occur over a given planning horizon (e.g., a year). The problem GENOPT chooses a capacity b_j for each plant $j \in \mathcal{J}_v$ and its output $y_j(n)$ and demand $d(n)$ in each node of the scenario tree to minimize the expected total social disbenefit over the time horizon. The total cost of capacity is linear in b_j and is levelized to a payment $\kappa_j b_j$ for the planning period.

3.1. Capacity decisions for thermal generation

We will consider the capacity choice for a single thermal generation plant. Similar arguments can be made with respect to a whole category of generation plants if they share the same cost characteristics. We label the plant in question as $j = 0$, which identifies it in the following version of GENOPT.

$$
\begin{aligned}
\text{GENOPT : min} \quad & \kappa_0 b_0 + \sum_{j\in\mathcal{J}_v\setminus\{0\}} \kappa_j b_j + \sum_{n\in\mathcal{N}} p(n)\left(\sum_{j\in\mathcal{J}} c_j(n) y_j(n) - U_n(d(n))\right) \\
\text{s.t.} \quad & y_0(n) + \sum_{j\in\mathcal{J}\setminus\{0\}} y_j(n) = d(n), \quad \text{for all } n\in\mathcal{N} && [p(n)\pi(n)] \\
& 0 \le y_0(n) \le b_0, \quad \text{for all } n\in\mathcal{N} && [p(n)\rho(n)] \\
& 0 \le y_j(n) \le b_j, \quad \text{for all } j\in\mathcal{J}, j\neq 0
\end{aligned}
$$

Now we can formulate the foundational result for a thermal generation plant. In later sections we will give equivalent results for other types of generation. We show that when capacities are chosen to be system optimal then the extra revenue earned by the plant as a result of prices being above the marginal cost will exactly match the investment cost for the capacity built. In the deterministic case with no demand elasticity (which is the standard result) this can be established using an explicit analysis of the parts of the overall demand curve that are met by different plant types - sometimes called a "screening curve" analysis.

Here we consider a group of thermal plants, $\mathcal{J}_v$ with jointly optimized capacities, and let the other generation plants have capacities that are fixed and not optimal. With all capacities fixed at these values, the revenue earned by plants in $\mathcal{J}_v$ will accrue from energy prices defined by the values of $\pi(n)$ at the optimal solution to DISPATCH(b). This gives the following result.

PROPOSITION 1. *If the capacities for generation plants $j \in \mathcal{J}_v$ are chosen at their optimal values b_j^* for the system optimization problem, GENOPT, then in the problem DISPATCH(b) the expected revenue for plant $0 \in \mathcal{J}_v$ earned at system marginal prices $\pi(n)$ is equal to the expected cost of generation plus the investment cost, $\kappa_0 b_0^*$.*

Proof. Consider the problem GENOPT with the Lagrangian to be minimized over non-negative b_j, $j \in \mathcal{J}_v$, $y_0(n)$, $d(n)$ and $y_j(n) \in [0, b_j]$, $j \neq 0$:

$$\begin{aligned} L = &\sum_{j \in \mathcal{J}_v} \kappa_j b_j + \sum_{n \in \mathcal{N}} p(n) \Big(\sum_{j \in \mathcal{J}} c_j(n) y_j(n) - U_n(d(n)) \Big) \\ &+ \sum_{n \in \mathcal{N}} p(n) \pi(n) \Big(d(n) - \sum_{j \in \mathcal{J}} y_j(n) \Big) + \sum_{n \in \mathcal{N}} p(n) \rho(n) (y_0(n) - b_0). \end{aligned}$$

Thus

$$L = b_0 \Big(\kappa_0 - \sum_{n \in \mathcal{N}} p(n) \rho(n) \Big) + \sum_{n \in \mathcal{N}} p(n) y_0(n) \left(c_0(n) + \rho(n) - \pi(n) \right) + L_{-0}$$

where L_{-0} collects terms not involving b_0 or y_0. The optimal solution minimizes L for appropriately chosen multipliers. The following complementarity conditions will hold:

$$0 \leq \kappa_0 - \sum_{n \in \mathcal{N}} p(n) \rho(n) \perp b_0 \geq 0 \tag{6}$$

$$0 \leq c_0(n) - \pi(n) + \rho(n) \perp y_0(n) \geq 0 \tag{7}$$

$$0 \leq b_0 - y_0(n) \perp \rho(n) \geq 0 \tag{8}$$

This is not a complete set of optimality conditions since we omit the optimization over $y_j(n)$ in the region $[0, b_j]$ for $j \in \mathcal{J}_v \setminus \{0\}$.

When $y_0(n) > 0$, the condition (7) gives $c_0(n) - \pi(n) + \rho(n) = 0$, so $\rho(n) = \pi(n) - c_0(n)$, which are the rents for plant 0 at node n. Notice that (8) implies there is positive rent only at nodes where $y_0(n)$ is at capacity b_0 (because $\rho(n) = 0$ if $y_0(n) < b_0$). Thus the expected rent for this thermal plant received over the time horizon is

$$\sum_{n \in \mathcal{N}} p(n) \left(\pi(n) - c_0(n) \right) y_0(n) = \sum_{n \in \mathcal{N}} p(n) \rho(n) b_0. \tag{9}$$

When we optimize over b_0 as well, so that $b_0 = b_0^*$ and (6) holds, then the sum of these terms must add up to κ_0 in expectation if $b_0^* > 0$. Thus at an optimal solution when $b_0 = b_0^*$ and $y_0(n) = y_0^*(n)$ we have

$$\sum_{n\in\mathcal{N}} p(n)\pi(n)y_0^*(n) - \sum_{n\in\mathcal{N}} p(n)c_0(n)y_0^*(n) = \kappa_0 b_0^* \tag{10}$$

Here we choose an optimal solution in which $y_0^*(n)$ matches the choice for the DISPATCH problem in the case that these are not unique. From Lemma 1 we know that when $b_j = b_j^*$, for $j \in \mathcal{J}_v$ the DISPATCH problem has a solution with $y_0(n) = y_0^*(n)$ and the price that plant 0 receives at node n is $\pi(n)$ from the GENOPT solution. Thus the expected revenue for plant 0 is $\sum_{n\in\mathcal{N}} p(n)\pi(n)y_0^*(n)$ and from (10) this is $\sum_{n\in\mathcal{N}} p(n)c_0(n)y_0^*(n) + \kappa_0 b_0^*$ as required. ∎

Observe that the multipliers $\pi(n)$ determine rents and are affected by the entire system, but the revenue calculation does not need to tie down the values of these multipliers: it is enough to use the complementary slackness involving y_0 in (7). Thus the result for generation plant 0 is independent of the constraints applying to generation outside the set $\mathcal{J}_v$.

Indeed, in a more general form of this problem we would allow for some generation resources to include storage so that generation amounts may be constrained by previous decisions as well as capacity built. We may suppose that the capacity constraints imply that the tree of possible values for $y_j(n)$ must be in the set $Y_j(b_j)$. Where there is storage the value chosen for $y_j(n)$ can affect the possible values for y_j at nodes lower in the tree, so that $Y_j(b_j)$ cannot be decomposed to a set of constraints operating at each node independently. The effect of this change is to replace the box constraints (2) with the more general constraints

$$\{y_j(n)\} \in Y_j(b_j), \quad \text{for all } j \in \mathcal{J}.$$

We can see from the proof of Proposition 1 that this makes no difference provided that the problem remains a convex optimization.

EXAMPLE 1. To make the result of Proposition 1 more concrete we illustrate this with an example. Consider a scenario tree with a root node $n = 0$ and two successor nodes $n = 1$ ("normal") and $n = 2$ ("peak"), with $p(0) = 1$, $p(1) = p(2) = 0.5$. Utility at each node is quadratic, $U_n(d) = a_n d - d^2/2$, so that $U_n'(d) = a_n - d$ and $\pi(n) = a_n - d(n)$, and we set $a_0 = 100$, $a_1 = 90$, and $a_2 = 150$.

Consider a single thermal generation plant with marginal cost $c_0 = 20$ and capacity cost $\kappa_0 = 20$, with no other generation present. The system-optimal capacity is $b_0^* = 90$. To see this, note that, absent a capacity constraint, demand at each node would be $a_n - c_0$, giving 80, 70, 130 at nodes $0, 1, 2$; since $130 > 90$ the plant is capacity-constrained only at node 2, giving dispatch

$$y_0(0) = 80, \qquad y_0(1) = 70, \qquad y_0(2) = 90,$$

and prices $\pi(0)=20,\ \pi(1)=20,\ \pi(2)=a_2-90=60$. Then optimality of b_0^* follows since $\rho(2)=\pi(2)-c_0=40$ and $\rho(0)=\rho(1)=0$, and condition (6) is satisfied:

$$\sum_n p(n)\rho(n)=0.5(40)=20=\kappa_0.$$

Then we can confirm the result of the proposition: expected revenue is $1(20)(80)+0.5(20)(70)+0.5(60)(90)=5000$, expected generation cost is $20(80+0.5\cdot 70+0.5\cdot 90)=3200$, and the difference, 1800, equals $\kappa_0 b_0^*=20\times 90$. ∎

In succeeding sections we consider a variety of generation and storage types. In each case we will solve a problem in which all other generation sources are absent, but the results carry over to any situation in which, besides the generation type considered, there is a portfolio of other generation resources leading to the appropriate additional terms in the objective and in the supply-demand balance constraint. These additional terms will not impact the complementarity relationships on which the results we establish are based.

Remark: Proposition 1 applies to a corresponding equilibrium for the formulation GENOPT. This assumes that there is no constraint on b_0 in GENOPT, i.e., there are no barriers to free entry of thermal plant into this market. If there is some binding upper bound on the capacity that can be built, e.g., a constraint of the form $b_0\le W_0$ then this will earn the generator an additional rent over and above that needed to cover its fixed cost. The constraint gives new optimality conditions

$$0\le W_0-b_0\perp\sigma\ge 0,$$

$$0\le\kappa_0+\sigma-\sum_{n\in\mathcal{N}}p(n)\rho(n)\perp b_0\ge 0,$$

which imply that the expected rent from plant operations exceeds its capital cost by σW_0.

Remark: We have established this result with a differentiable utility function, which is needed to avoid non-unique prices, since otherwise prices in the dispatch problem might not give exact cost recovery. But this result can also be established through a screening curve approach when demand is price-inelastic and there is no utility function involved (often this involves VOLL giving a price for very high demands). The proof will still apply provided that demand uncertainty is continuous rather than discrete, for in that case the possibility of the "wrong" prices is restricted to demand that is equal to the sum of a subset of capacities. This occurs with probability zero and hence does not affect the expected values.

3.2. Free entry/exit supports an optimal allocation

We now establish that the system-optimal amount of each generation type is supported by free entry and exit. In line with the rest of this section, we establish the key proposition for thermal generation. A more general result that applies to other technologies is given in Section 4.

PROPOSITION 2. *Fix the capacities of all plants other than plant* 0*. For each* $b_0 \geq 0$*, let* $\Pi_0(b_0)$ *denote the expected rent minus capital cost of plant* 0 *under an optimal solution to DISPATCH*(b_0)*. If* $b_0^* > 0$ *is a system-optimal capacity, then*

$$\Pi_0(b_0) \geq 0 \quad \textit{for } b_0 < b_0^*, \qquad \Pi_0(b_0) \leq 0 \quad \textit{for } b_0 > b_0^*.$$

Proof. Let $\Phi(b_0)$ denote the optimal value of DISPATCH(b_0). The function Φ is convex in b_0. For a given b_0, let $\rho_{b_0}(n)$ be the optimal Lagrange multiplier on the constraint $y_0(n) \leq b_0$, and define

$$V(b_0) := \sum_{n \in \mathcal{N}} p(n)\rho_{b_0}(n),$$

which gives $-V(b_0) \in \partial\Phi(b_0)$. The total system cost as a function of the capacity of plant 0 is

$$C(b_0) := \kappa_0 b_0 + \Phi(b_0).$$

It follows that $\kappa_0 - V(b_0) \in \partial C(b_0)$. Since $b_0^* > 0$ minimizes the convex function C, we have $0 \in \partial C(b_0^*)$. The subgradients of a convex function are monotone, and hence

$$\big(\kappa_0 - V(b_0)\big)(b_0 - b_0^*) \geq 0.$$

Consequently, $V(b_0) \geq \kappa_0$ when $b_0 < b_0^*$, and $V(b_0) \leq \kappa_0$ when $b_0 > b_0^*$. From (9) we have

$$\Pi_0(b_0) = b_0\big(V(b_0) - \kappa_0\big),$$

and this establishes the required signs. ∎

Now suppose that there are a large number of small plants that may enter the market. Each has the same cost characteristics as plant 0, but has a capacity bounded above by b_δ. If there are N such plants in the market then linearity of GENOPT allows a solution in which the generation of each individual plant is given by $y_0^*(n)/N$ where y_0^* is the optimal solution for the original problem with capacity $b_0 \leq N b_\delta$. The optimal capacity of this type of generation can be achieved with $N = \lceil b_0^*/b_\delta \rceil$ plants, each of which has capacity b_0^*/N. Then an application of Proposition 2 shows that a system optimal amount of generation from this generation type is supported by free entry and exit. At the system optimal solution each small plant covers its costs, and away from this system optimal solution there are weak incentives to adjust towards it.

3.3. Stagewise investment

In practice investment in generation capacity takes place in stages. We will show that the result of the previous section extends to a case where investment can be made at any stage. Let $z_j(n)$ be the investment in additional capacity added at node n, with the new capacity becoming available at node n, and let $\kappa_j(n)$

be the cost of this investment. We consider an optimal solution for the stagewise investment $z_j(n)$, for $n \in \mathcal{N}, j \in \mathcal{J}_v$, and take other $b_j(n)$ values as fixed. Then the system optimization problem becomes

$$\begin{aligned}
\min \quad & \sum_{n\in\mathcal{N}} p(n)\Big(\sum_{j\in\mathcal{J}_v}\kappa_j(n)z_j(n) + \sum_{j\in\mathcal{J}} c_j(n)y_j(n) - U_n(d(n))\Big) \\
\text{s.t.} \quad & \sum_{j\in\mathcal{J}} y_j(n) = d(n), \quad \text{for all } n\in\mathcal{N} && [p(n)\pi(n)] \quad (11)\\
& 0 \le y_0(n) \le b_0(n), \quad \text{for all } n\in\mathcal{N} && [p(n)\rho(n)] \quad (12)\\
& b_j(n-) + z_j(n) = b_j(n), \quad \text{for all } n\in\mathcal{N}, j\in\mathcal{J}_v && [p(n)\lambda_j(n)] \quad (13)\\
& 0 \le z_j(n), \quad \text{for all } n\in\mathcal{N}, j\in\mathcal{J}_v, && (14)\\
& 0 \le y_j(n) \le b_j(n), \quad \text{for all } n\in\mathcal{N}, j\in\mathcal{J}\setminus\{0\}.
\end{aligned}$$

Here we define $b_j(0-) = 0$, and we have written the $y_0(n)$ bound separately since we will be looking at the revenue for plant 0.

The dispatch problem is now formulated for the case where all $z_j(n)$ and $b_j(n)$ values are given and we omit the investment term in the objective. This means that we do not need to include the variables z_j, b_j or the constraints (13) and (14) and the DISPATCH problem remains as it was before and Lemma 1 still holds.

PROPOSITION 3. *If the sequence of capacity expansions for plants $j \in \mathcal{J}_v$ are chosen at their optimal values for the system optimization problem, then in the problem DISPATCH(b) the expected revenue for plant $0 \in \mathcal{J}_v$ earned at system marginal prices $\pi(n)$ is equal to the expected cost of generation plus the expected total investment cost, $\sum_{n\in\mathcal{N}} p(n)\kappa_0(n)z_0(n)$.*

Proof. For GENOPT, we have the additional constraint (13), and the Lagrangian to be minimized over non-negative $z_j(n), b_j(n)$, $j \in \mathcal{J}_v$, $y_0(n)$ and $y_j(n) \in [0, b_j]$, $j \neq 0$ is:

$$\begin{aligned}
L = & \sum_{n\in\mathcal{N}} p(n)\Big(\sum_{j\in\mathcal{J}_v}\kappa_j(n)z_j(n) + \sum_{j\in\mathcal{J}} c_j(n)y_j(n) - U_n(d(n))\Big) \\
& + \sum_{n\in\mathcal{N}} p(n)\pi(n)\Big(d(n) - \sum_{j\in\mathcal{J}} y_j(n)\Big) + \sum_{n\in\mathcal{N}} p(n)\rho(n)(y_0(n) - b_0(n)) \\
& + \sum_{n\in\mathcal{N}} p(n) \sum_{j\in\mathcal{J}_v} \lambda_j(n)(b_j(n) - b_j(n-) - z_j(n)).
\end{aligned}$$

We may pick out the term in $b_0(n)$ which is

$$\begin{aligned}
& b_0(n)\Big(p(n)\lambda_0(n) - p(n)\rho(n) - \sum_{m\in n+} p(m)\lambda_0(m)\Big), \text{ for } n \in \mathcal{N}\setminus\mathcal{L}, \\
& b_0(n)\Big(p(n)\lambda_0(n) - p(n)\rho(n)\Big), \text{ for } n \in \mathcal{L}.
\end{aligned}$$

The complementarity conditions (7) continue to hold and we have revised versions of (6) and (8) and further new complementarity conditions as follows

$$0 \le \kappa_0(n) - \lambda_0(n) \perp z_0(n) \ge 0, \tag{15}$$

$$0 \leq b_0(n) - y_0(n) \perp \rho(n) \geq 0, \tag{16}$$

$$0 \leq p(n)\lambda_0(n) - p(n)\rho(n) - \sum_{m \in n+} p(m)\lambda_0(m) \perp b_0(n) \geq 0, \text{ for } n \in \mathcal{N} \setminus \mathcal{L}, \tag{17}$$

$$0 \leq p(n)(\lambda_0(n) - \rho(n)) \perp b_0(n) \geq 0, \text{ for } n \in \mathcal{L}. \tag{18}$$

As in the proof of Proposition 1, if $y_0(n) > 0$, the condition (7) gives $c_0(n) - \pi(n) + \rho(n) = 0$, so $\rho(n) = \pi(n) - c_0(n)$. Thus the expected rent from plant 0 is

$$\sum_{n \in \mathcal{N}} p(n)(\pi(n) - c_0(n))y_0(n) = \sum_{n \in \mathcal{N}} p(n)\rho(n)y_0(n) = \sum_{n \in \mathcal{N}} p(n)\rho(n)b_0(n), \tag{19}$$

where we have used (16) to replace y_0 with b_0 in this sum. The result of Lemma 1 shows that the left hand side of (19) is exactly the expected revenue minus costs from the dispatch problem.

At optimal choices of $z_j(n) \geq 0$ and $b_j(n) \geq 0$, we have

$$\begin{aligned}
\sum_{n \in \mathcal{N}} p(n)\kappa_0(n)z_0(n) &= \sum_{n \in \mathcal{N}} p(n)\lambda_0(n)z_0(n) \\
&= \sum_{n \in \mathcal{N}} p(n)\lambda_0(n)(b_0(n) - b_0(n-)) \\
&= \sum_{n \in \mathcal{N} \setminus \mathcal{L}} \Big(p(n)\lambda_0(n) - \sum_{m \in n+} p(m)\lambda_0(m)\Big) b_0(n) + \sum_{n \in \mathcal{L}} p(n)\lambda_0(n)b_0(n) \\
&= \sum_{n \in \mathcal{N}} p(n)\rho(n)b_0(n),
\end{aligned}$$

where the first two equalities come from (15) and (13). The third equality is a rearrangement and the last is by application of the conditions (17) and (18). This establishes that the right hand side of (19) matches the expected cost of building the capacity for plant 0. ∎

COROLLARY 1. *For any node n the expected rents accrued from future dispatch will equal or exceed the expected future capital costs.*

Proof. Let $\mathcal{T}$ be the complete subtree rooted at node n. The expected capital cost given we are in node n is

$$\begin{aligned}
\frac{1}{p(n)} \sum_{m \in \mathcal{T}} p(m)\kappa_0(m)z_0(m) &= \frac{1}{p(n)} \sum_{m \in \mathcal{T}} p(m)\lambda_0(m)z_0(m) \\
&= \frac{1}{p(n)} \sum_{m \in \mathcal{T}} p(m)\lambda_0(m)(b_0(m) - b_0(m-)) \\
&= \frac{1}{p(n)} \sum_{m \in \mathcal{T} \setminus \mathcal{L}} \left(p(m)\lambda_0(m) - \sum_{k \in m+} p(k)\lambda_0(k)\right) b_0(m) \\
&\quad + \frac{1}{p(n)} \sum_{m \in \mathcal{L} \cap \mathcal{T}} p(m)\lambda_0(m)b_0(m) - \lambda_0(n)b_0(n-) \\
&\leq \frac{1}{p(n)} \sum_{m \in \mathcal{T}} p(m)\rho(m)b_0(m),
\end{aligned}$$

as required. The final inequality follows since $\lambda(m) \geq 0$ for $m \in \mathcal{L}$ from (18), and then induction using (17) shows that $\lambda(n) \geq 0$. ∎

The inequality in Corollary 1 arises because rents in $\mathcal{T}$ also contribute $\lambda_0(n)b_0(n-)$ to the investment costs accrued up to node $n-$.

In the following sections, we examine cost recovery for different types of generation technology. In Proposition 1 we saw how we can focus on a single generation plant, since other generation amounts appear in the supply balance constraint, but otherwise do not interact with the decisions on b_0 and $y_0(n)$. We will follow the same approach throughout this paper by assuming that all other sources of generation are fixed at some level $g(n)$ in node n. Here $g(n)$ may be chosen through the optimal solution to the social planning problem, or may be fixed as part of the problem set up.

4. A general model

We will now set up a model that is general enough to cover the particular cases that we consider in the rest of the paper.

There is a set $\mathcal{J}$ of plants (which may be generation, storage or hydro plants). Plant j has a capacity vector h_j, a control vector $u_j(n)$, a state vector $x_j(n)$ at the beginning of node n, and a state vector $\bar{x}_j(n)$ at the end of node n. We consider the system optimization problem, GENOPT, in which capacities for the set of plants $\mathcal{J}_v$ are chosen optimally, and all the operating variables u, x and d are also optimized, with given values for other fixed energy generation $g(n)$. This is

$$
\begin{array}{lll}
\text{GENOPT: } \min & \displaystyle\sum_{j\in\mathcal{J}_v} \theta_j^\top h_j + \sum_{n\in\mathcal{N}} p(n)\left(\sum_{j\in\mathcal{J}} c_j(n)^\top u_j(n) - U_n(d(n))\right) & \\
\text{s.t.} & \bar{x}_j(n) = K_j x_j(n) + A_j u_j(n) + w_j(n), \quad j\in\mathcal{J}, n\in\mathcal{N}, & [p(n)\lambda_j(n)] \\
& 0 \leq \bar{x}_j(n) \leq G_j h_j, \quad j\in\mathcal{J}, n\in\mathcal{N}, & [p(n)\mu_j(n)] \\
& x_j(n) = \bar{x}_j(n-), \quad j\in\mathcal{J}, n\in\mathcal{N}\setminus\{0\}, & [p(n)\tau_j(n)] \\
& x_j(0) = 0, \quad j\in\mathcal{J}, & \\
& g(n) + \displaystyle\sum_{j\in\mathcal{J}} H_j u_j(n) = d(n), \quad n\in\mathcal{N} & [p(n)\pi(n)] \\
& 0 \leq u_j(n) \leq B_j(n) h_j, \quad j\in\mathcal{J}, n\in\mathcal{N}, & [p(n)\rho_j(n)] \\
& u_j(n) \leq D_j x_j(n) + F_j h_j, \quad j\in\mathcal{J}, n\in\mathcal{N}, & [p(n)\eta_j(n)] \\
& h_j \geq 0, \quad j\in\mathcal{J}_v. &
\end{array}
$$

All vector inequalities are interpreted componentwise. The multipliers $\lambda_j(n)$, $\tau_j(n)$ and $\pi(n)$ are unrestricted in sign, while $\mu_j(n)$, $\rho_j(n)$, and $\eta_j(n)$ are non-negative. The constant matrices K_j and A_j determine the dynamics for plant j; H_j is a matrix with just one row multiplying the control vector u_j to give the output for plant j; and the matrices G_j, $B_j(n)$, D_j and F_j determine the upper bound constraints that apply to the states and controls for plant j.

The first constraint (λ_j) is the state dynamics with an extra exogenous vector $w_j(n)$ used for example to capture uncontrolled inflows in the hydro models. The second constraint (μ_j) is a constraint on the state that is linked to the capacity choices, used for example to model the choice of reservoir sizes. The constraint with multiplier ρ_j is a flexible form of the standard capacity constraint on the controls. The final constraint (η_j) is a linking constraint involving both control and states and is used for example to model ramping constraints.

We will assume that the optimization problem is feasible with finite optimal values, and that standard constraint qualifications are satisfied, so that the KKT conditions hold. Our results will be framed in terms of the prices occurring at the optimal solution to GENOPT. However we can observe that Lemma 1 also holds in this case, and there is a unique set of prices arising from GENOPT, with these being the same as those in the corresponding problem DISPATCH(h), where all capacities are fixed, with those chosen optimally in GENOPT being set to their optimal values.

We use the convention that $\lambda_j(n)$ and $\pi(n)$ are the marginal values from relaxing the right-hand side of the corresponding constraint. Equivalently, the Lagrangian contains the terms

$$p(n)\lambda_j(n)^\top \left(\bar{x}_j(n) - K_j x_j(n) - A_j u_j(n) - w_j(n)\right),$$

$$p(n)\pi(n)^\top \left(d(n) - g(n) - \sum_{j\in\mathcal{J}} H_j u_j(n)\right).$$

With this convention, the relevant optimality conditions are as follows. For the upper state-capacity constraint,

$$0 \le G_j h_j - \bar{x}_j(n) \perp \mu_j(n) \ge 0. \tag{20}$$

For the upper control-capacity constraint,

$$0 \le B_j(n) h_j - u_j(n) \perp \rho_j(n) \ge 0. \tag{21}$$

For the state-control linking constraint,

$$0 \le D_j x_j(n) + F_j h_j - u_j(n) \perp \eta_j(n) \ge 0. \tag{22}$$

The lower bound $u_j(n) \ge 0$, together with stationarity with respect to $u_j(n)$, gives

$$0 \le u_j(n) \perp c_j(n) - A_j^\top \lambda_j(n) - H_j^\top \pi(n) + \rho_j(n) + \eta_j(n) \ge 0. \tag{23}$$

For $n \ne 0$, stationarity with respect to $x_j(n)$ gives

$$-K_j^\top \lambda_j(n) - D_j^\top \eta_j(n) + \tau_j(n) = 0. \tag{24}$$

Stationarity with respect to $\bar{x}_j(n)$ gives

$$0 \le \bar{x}_j(n) \perp p(n)\lambda_j(n) + p(n)\mu_j(n) - \sum_{m \in n+} p(m)\tau_j(m) \ge 0, \tag{25}$$

where the summation is zero if n is a leaf node. Finally, the lower bound on h_j together with stationarity with respect to h_j, gives for $j \in \mathcal{J}_v$,

$$0 \le h_j \perp \theta_j - \sum_{n \in \mathcal{N}} p(n) \left(G_j^\top \mu_j(n) + B_j(n)^\top \rho_j(n) + F_j^\top \eta_j(n) \right) \ge 0. \tag{26}$$

The result below continues to hold when some of the constraint blocks are absent. For example, it applies when there are no state variables x and $\bar{x}$, when the state-control linking constraint is absent, or when only some components of the state or control vectors have upper bounds. In these cases the corresponding multipliers and matrix terms are simply omitted from the optimality conditions and from the proof.

THEOREM 1. *Suppose the capacity vectors h_j for $j \in \mathcal{J}_v$ are chosen at their optimal values h_j^* for the system optimization problem, GENOPT, and other capacities are fixed. Given an optimal solution to the corresponding problem DISPATCH(h), the expected operating surplus for every plant $j \in \mathcal{J}_v$ (expected revenue earned at system marginal prices $\pi(n)$ minus expected operating cost) is*

$$\theta_j^\top h_j^* + \sum_{n \in \mathcal{N}} p(n)\lambda_j(n)^\top w_j(n).$$

Thus the operating surplus equals investment cost, plus the value of the exogenous contribution to the state equation.

Proof. All variables in the proof are evaluated at an optimal solution of DISPATCH(h). From the generalized version of Lemma 1 we may choose all the variables satisfying the complementarity conditions for GENOPT so that they match the optimal solution to DISPATCH(h) and have the same prices. To simplify notation, we write h_j rather than h_j^*.

The expected operating profit of plant j is

$$R_j = \sum_{n \in \mathcal{N}} p(n) \left(H_j^\top \pi(n) - c_j(n) \right)^\top u_j(n).$$

By (23),

$$\begin{aligned} \left(H_j^\top \pi(n) - c_j(n) \right)^\top u_j(n) &= -\lambda_j(n)^\top A_j u_j(n) + \rho_j(n)^\top u_j(n) + \eta_j(n)^\top u_j(n) \\ &= -\lambda_j(n)^\top A_j u_j(n) + \rho_j(n)^\top B_j(n) h_j + \eta_j(n)^\top \left(D_j x_j(n) + F_j h_j \right), \end{aligned}$$

using (21) and (22). We can substitute for $A_j u_j(n)$ from the state equation, and therefore

$$\begin{aligned} \left(H_j^\top \pi(n) - c_j(n) \right)^\top u_j(n) = &- \lambda_j(n)^\top \bar{x}_j(n) + \lambda_j(n)^\top K_j x_j(n) + \lambda_j(n)^\top w_j(n) \\ &+ \rho_j(n)^\top B_j(n) h_j + \eta_j(n)^\top D_j x_j(n) + \eta_j(n)^\top F_j h_j. \end{aligned}$$

Summing over nodes gives

$$R_j = \sum_{n\in\mathcal{N}} p(n)\Big[-\lambda_j(n)^\top \bar{x}_j(n) + \lambda_j(n)^\top K_j x_j(n) + \eta_j(n)^\top D_j x_j(n) \\ + \rho_j(n)^\top B_j(n) h_j + \eta_j(n)^\top F_j h_j + \lambda_j(n)^\top w_j(n)\Big].$$

By (24), for $n \neq 0$,

$$\lambda_j(n)^\top K_j x_j(n) + \eta_j(n)^\top D_j x_j(n) = \tau_j(n)^\top x_j(n). \tag{27}$$

The root contribution is zero since $x_j(0) = 0$. Hence

$$R_j = \sum_{n\in\mathcal{N}} p(n)\Big[-\lambda_j(n)^\top \bar{x}_j(n) + \rho_j(n)^\top B_j(n) h_j + \eta_j(n)^\top F_j h_j + \lambda_j(n)^\top w_j(n)\Big] \\ + \sum_{n\in\mathcal{N}\setminus\{0\}} p(n)\tau_j(n)^\top x_j(n).$$

Using $x_j(n) = \bar{x}_j(n-)$, the final term can be rewritten as

$$\sum_{n\in\mathcal{N}\setminus\{0\}} p(n)\tau_j(n)^\top x_j(n) = \sum_{n\in\mathcal{N}\setminus\mathcal{L}} \sum_{m\in n+} p(m)\tau_j(m)^\top \bar{x}_j(n). \tag{28}$$

By (25),

$$\bar{x}_j(n)^\top \left[p(n)\lambda_j(n) + p(n)\mu_j(n) - \sum_{m\in n+} p(m)\tau_j(m)\right] = 0.$$

Hence

$$\sum_{m\in n+} p(m)\tau_j(m)^\top \bar{x}_j(n) = p(n)\big(\lambda_j(n) + \mu_j(n)\big)^\top \bar{x}_j(n),$$

where the summation is zero when n is a leaf node. Therefore

$$-p(n)\lambda_j(n)^\top \bar{x}_j(n) + \sum_{m\in n+} p(m)\tau_j(m)^\top \bar{x}_j(n) = p(n)\mu_j(n)^\top \bar{x}_j(n). \tag{29}$$

It follows that

$$R_j = \sum_{n\in\mathcal{N}} p(n)\left[\mu_j(n)^\top \bar{x}_j(n) + \rho_j(n)^\top B_j(n) h_j + \eta_j(n)^\top F_j h_j\right] + \sum_{n\in\mathcal{N}} p(n)\lambda_j(n)^\top w_j(n).$$

Using (20),

$$\mu_j(n)^\top \bar{x}_j(n) = \mu_j(n)^\top G_j h_j. \tag{30}$$

Thus

$$R_j = \sum_{n\in\mathcal{N}} p(n)\left[\mu_j(n)^\top G_j h_j + \rho_j(n)^\top B_j(n) h_j + \eta_j(n)^\top F_j h_j\right] + \sum_{n\in\mathcal{N}} p(n)\lambda_j(n)^\top w_j(n). \tag{31}$$

Finally, using (26) gives

$$\theta_j^\top h_j = \sum_{n\in\mathcal{N}} p(n)\left[\mu_j(n)^\top G_j h_j + \rho_j(n)^\top B_j(n) h_j + \eta_j(n)^\top F_j h_j\right].$$

Substituting this expression into the expression for R_j yields

$$R_j = \theta_j^\top h_j + \sum_{n\in\mathcal{N}} p(n)\lambda_j(n)^\top w_j(n).$$

Restoring the notation $h_j = h_j^*$ gives the result we want. ∎

The linearity of investment costs is an important assumption here. If $\theta_j^\top h_j$ is replaced by a differentiable convex investment cost $C_j(h_j)$ then the condition (26) becomes

$$\nabla C_j(h_j^*) = \sum_{n\in\mathcal{N}} p(n)\left(G_j^\top \mu_j(n) + B_j(n)^\top \rho_j(n) + F_j^\top \eta_j(n)\right)$$

on positive components. The proof then shows that

$$R_j = \nabla C_j(h_j^*)^\top h_j^* + \sum_{n\in\mathcal{N}} p(n)\lambda_j(n)^\top w_j(n).$$

Thus when $w_j = 0$, revenue exceeds investment costs by $\nabla C_j(h_j^*)^\top h_j^* - C_j(h_j^*)$ which is positive when C_j is strictly convex with $C_j(0) = 0$.

Theorem 1 also allows us to establish a bound on what can be achieved with any feasible dispatch policy. For a plant j we say that a set of dispatch decisions $\{\tilde{u}_j(n), \tilde{x}_j(n), \tilde{\bar{x}}_j(n)\}$ for $n \in \mathcal{N}$ is h_j-feasible if it satisfies its own technology constraints, that is,

$$\tilde{\bar{x}}_j(n) = K_j\tilde{x}_j(n) + A_j\tilde{u}_j(n) + w_j(n), \qquad 0 \le \tilde{\bar{x}}_j(n) \le G_j h_j, \qquad \tilde{x}_j(n) = \tilde{\bar{x}}_j(n^-),$$

$$0 \le \tilde{u}_j(n) \le B_j(n)h_j, \qquad \tilde{u}_j(n) \le D_j\tilde{x}_j(n) + F_j h_j, \qquad \tilde{x}_j(0) = 0,$$

for all $n \in \mathcal{N}$, and $\tilde{u}_j(n)$ depends only on information available at node n, i.e., it is non-anticipative.

COROLLARY 2. *Fix the capacities h_j^* for $j \in J_v$ at their system-optimal values, and let $\pi(n)$, $n \in \mathcal{N}$, be the corresponding prices from an optimal solution to DISPATCH(h^*). Suppose plant j operates as a price taker, and its dispatch decisions do not affect $\pi(n)$. Then for any h_j^*-feasible dispatch policy $\{\tilde{u}_j(n), \tilde{x}_j(n), \tilde{\bar{x}}_j(n)\}$ the expected operating surplus under $\tilde{u}_j$ satisfies*

$$\sum_{n\in\mathcal{N}} p(n)\big(H_j^\top \pi(n) - c_j(n)\big)^\top \tilde{u}_j(n) \;\le\; \theta_j^\top h_j^* + \sum_{n\in\mathcal{N}} p(n)\,\lambda_j(n)^\top w_j(n),$$

with equality for a system-optimal dispatch policy.

Proof. For prices $\pi(n)$ fixed at their values from GENOPT, consider the problem faced by plant j alone:

$$\max_{u_j, x_j, \bar{x}_j} \sum_{n\in\mathcal{N}} p(n)\big(H_j^\top \pi(n) - c_j(n)\big)^\top u_j(n)$$

subject to $\{u_j(n), x_j(n), \bar{x}_j(n)\}$, $n \in \mathcal{N}$ being h_j^*-feasible.

This is a linear program: the objective is linear in $(u_j, x_j, \bar{x}_j)$ and the feasible region is a polyhedron defined over the scenario tree, since none of plant j's constraints involve the decisions of any other plant. The optimal solution $u_j^*(n)$, $x_j^*(n)$, $\bar{x}_j^*(n)$ identified in the proof of Theorem 1, together with the multipliers $\lambda_j(n)$, $\mu_j(n)$, $\rho_j(n)$, $\eta_j(n)$, $\tau_j(n)$ from the optimal solution of GENOPT, satisfies the KKT conditions (20)–(25) for this LP. Because the problem is a linear program, KKT conditions are sufficient as well as necessary for global optimality, so u_j^* attains the maximum of the above problem over all h_j^*-feasible policies, including $\tilde{u}_j$. Hence

$$\begin{aligned}\sum_n p(n)\big(H_j^\top \pi(n) - c_j(n)\big)^\top \tilde{u}_j(n) &\leq \sum_n p(n)\big(H_j^\top \pi(n) - c_j(n)\big)^\top u_j^*(n) \\ &= \theta_j^\top h_j^* + \sum_n p(n)\lambda_j(n)^\top w_j(n),\end{aligned}$$

where the final equality is Theorem 1. ∎

This corollary shows that Theorem 1 identifies the frontier of what a price-taking generation or storage asset can achieve for operating surplus, at the system-optimal prices. Any feasible dispatch that is not optimal at these prices earns strictly less; such loss of optimality may arise, for example, from incorrect models of future prices. Hence, when $w_j(n) = 0$, any non-optimal dispatch will fail to recover the investment cost. The absence of an explicit single-period market mechanism achieving DISPATCH(h^*), discussed in Section 5, does not undermine Theorem 1 as an upper bound on achievable profit.

4.1. Generalizing the free entry/exit result

We begin by establishing the equivalent to Proposition 2 for the general model.

PROPOSITION 4. *Fix the capacities of all plants other than plant j. For any $h_j \geq 0$, consider an optimal solution to GENOPT with h_j fixed, yielding dual variables $p(n)\mu_j(n)$, $p(n)\rho_j(n)$, $p(n)\eta_j(n)$. Let*

$$V(h_j) = \sum_{n\in\mathcal{N}} p(n)\left[G_j^\top \mu_j(n) + B_j(n)^\top \rho_j(n) + F_j^\top \eta_j(n)\right].$$

If h_j^* is a system-optimal capacity, then

$$(V(h_j) - \theta_j)^\top (h_j^* - h_j) \geq 0.$$

Proof. Examining the right-hand side of the constraints of GENOPT containing h_j we see that $-V(h_j)$ is a subgradient of $\Phi(h_j)$, the optimal value of GENOPT with capacity h_j, excluding investment cost. Define $C(h_j) := \theta_j^\top h_j + \Phi(h_j)$, so that $\theta_j - V(h_j)$ is a subgradient of $C(h_j)$. Hence

$$C(h_j^*) \geq C(h_j) + (\theta_j - V(h_j))^\top (h_j^* - h_j).$$

Then optimality of h_j^* implies that

$$(\theta_j - V(h_j))^\top (h_j^* - h_j) \leq 0,$$

from which the result follows. ∎

Suppose $w_j(n) = 0$ for every $n \in \mathcal{N}$. Observe from (31) that $V(h_j)$ is then a vector giving the rent per unit capacity for each type of capacity defined by h_j. By Proposition 4, at any capacity vector h_j, there is a weak first-order incentive to adjust towards the system-optimal capacity vector h_j^*.

If $\sum_n p(n)\lambda_j(n)^\top w_j(n) > 0$, then the plant makes a positive profit at the system optimal choice h_j^*. Observe that entry in this case appears profitable even if $h_j = h_j^*$, since additional rent would accrue from the marginal value of inflows. If the plant had to forgo this value (e.g., by paying a levy for inflows) then entry towards h_j^* would be (weakly) profitable as in the previous case.

Now we consider what can be established when there is free entry of new plants, and free exit of loss making plants for the general model which has more than one capacity choice for the same plant, for example, a storage capacity chosen at the same time as a turbine capacity. As in subsection 3.2, we study entry by considering a model in which we replace plant j with a large number of small plants having the same aggregate capacity and each having the same cost characteristics as plant j. We write g_j for the vector of capacities for these small plants, and assume that the first component of g_j is subject to an upper bound of b_δ. Thus a capacity choice for each small plant takes the form $g_j = (g, \bar{g}_j)$, where $g \leq b_\delta$ and $\bar{g}_j$ denotes the vector of choices of other components. Now observe that, when $w_j(n) = 0$, we can match an optimal solution $h_j^* = (h^*, \bar{h}_j^*)$, $x_j^*(n)$ and $u_j^*(n)$ for the single combined plant, with a solution for the problem with N small plants by simply dividing each element in the original solution by N, so that $g_j = h_j^*/N$. Assuming that the optimal solution has $h^* > 0$, taking $N = \lceil h^*/b_\delta \rceil$ will ensure that the individual plants satisfy $g \leq b_\delta$.

Moreover the total energy supplied is the same and so prices $\pi(n)$ are matched. This implies that the optimality conditions are satisfied for the individual plants. Thus, with a price taking assumption, there is no incentive to change the capacity components $\bar{h}_j^*/N$ from these values. Each small plant covers its investment costs, and there is no incentive for other plants to enter, or for these plants to exit.

If there is just a single capacity to be chosen (e.g., for a grid scale battery) then this argument establishes that the optimal system solution is supported in an equilibrium with free entry and exit for many small plants. When each plant has multiple capacity choices then the result will hold provided that the individual plants have a single capacity choice (perhaps reservoir size) constrained, while other capacity elements (perhaps turbine size) can be freely chosen.

4.2. Variable renewable energy

Consider a wind farm or solar array with capital cost θ per unit of capacity. We seek the optimal capacity a of this generation plant given that the energy available per unit capacity at node n (from wind or sun) is $\varphi(n)$. There is no cost of generation for these resources. Let $z(n)$ denote the energy supplied in node n. The social optimization problem given other fixed energy generation $g(n)$ is

$$\begin{array}{llrll}
\text{GENOPT:} & \min & \theta a - \sum_{n\in\mathcal{N}} p(n)\,(U_n(d(n))) & & \\
& \text{s.t.} & z(n)+g(n) = d(n), & n\in\mathcal{N}, & [p(n)\pi(n)] \\
& & z(n) \leq a\varphi(n), & n\in\mathcal{N}, & [p(n)\rho(n)]
\end{array}$$

where all decision variables are nonnegative.

PROPOSITION 5. *If the capacity a of a variable renewable generation plant is chosen at its optimal value a^* for the system optimization problem, GENOPT, then in the problem DISPATCH(a) the expected revenue for the plant earned at system marginal prices $\pi(n)$ is equal to the expected cost of generation plus the investment cost θa^*.*

Proof. We will apply Theorem 1 with just one plant in $\mathcal{J}_v$. We set

$$h_j = a, \quad u_j(n) = z(n), \quad c_j(n) = 0, \quad H = 1, \quad B(n) = \varphi(n)$$

and we do not have state variables or constraints in this case. Since there are no exogenous terms $w_j(n)$ here, the result follows. ∎

4.3. Run-of-river hydro

A run-of-river hydro plant is very similar to a wind or solar plant. There is no cost of generation for the hydro plant. Here we denote the flow arriving at the station by $\omega(n) \geq 0$. The generation of the station is bounded by its capacity a and by $\omega(n)$. Thus any flow above a is spilled. The formulation is:

$$\begin{array}{llrll}
\text{GENOPT:} & \min & \theta a - \sum_{n\in\mathcal{N}} p(n)\,(U_n(d(n))) & & \\
& \text{s.t.} & z(n)+g(n) = d(n), & n\in\mathcal{N}, & [p(n)\pi(n)] \\
& & z(n) \leq \omega(n), & n\in\mathcal{N}, & [p(n)\lambda(n)] \\
& & z(n) \leq a, & n\in\mathcal{N} & [p(n)\rho(n)]
\end{array}$$

with decision variables a, d and z all nonnegative.

PROPOSITION 6. *Suppose the capacity a of a run-of-river station is chosen at its optimal value a^* for the system optimization problem, GENOPT. Then in the problem DISPATCH(a) the expected revenue for the station earned at system marginal prices $\pi(n)$ is greater than or equal to the investment cost θa^*.*

Proof. We will use Theorem 1 with just one plant having its capacity optimized. In this case we set

$$h_j = a, \quad u_j(n) = [z(n) \;\; s(n)]^\top, \quad c_j(n) = [0 \;\; 0]^\top, \quad H = [1 \;\; 0],$$

$$K = 0, \quad A = [-1 \;\; -1], \quad w_j(n) = \omega(n), \quad G = 0,$$

where $s(n)$, the slack variable in the $\omega(n)$ constraint, is the spill. The matrix B is of reduced dimension and has the single element 1 corresponding to an upper bound on $z(n)$, and the upper bound on $s(n)$ is absent. The state-control linking constraint is absent here. Since $G = 0$ all the state variables $x_j(n)$ and $\bar{x}_j(n)$ are zero and the state dynamics constraint is used to ensure that $z(n) + s(n) = \omega(n)$ where $s(n)$ is a slack variable. Then complementarity shows $\lambda(n) \geq 0$ and the additional term in the profits is $\sum_{n\in\mathcal{N}} p(n)\lambda(n)^\top\omega(n)$ which is non-negative establishing the result. ∎

In this case we find that there is an additional amount paid to the run-of-river operator over and above the cost of capacity that is related to the inflows through the shadow prices $\lambda(n)$. Thus the difference will be zero if, at the optimal solution, the generation quantity $z(n)$ is strictly less than the inflow $\omega(n)$ at every node n where this is positive.

It is interesting to consider a chain of N smaller generating stations on the river each of capacity a/N costing θa in total and delivering total energy $z(n)$ in node n. The inflow is passed down the chain, so each station receives $\omega(n)$. Provided that $\omega(n)$ is bounded away from zero, we see that for sufficiently large N the generation $z(n)/N$ of each station would be strictly less than $\omega(n)$ when this is positive and so we would have $\lambda(n) = 0$ and no additional rent. In effect the additional rent from inflows arises from the assumption in GENOPT that we build only one generating station on the river, which could be interpreted as a barrier to entry.

5. Dynamic dispatch problems

In this section we turn to problems in which the scenario tree plays a more prominent role, and optimal decisions on generation or storage require consideration of future states. In this environment we need to be careful about a market interpretation of DISPATCH(b). There is no simple way for a market that operates in single periods to yield a system optimal result since future prices will influence the optimal decision.

We may choose to assume that the system operator has complete information allowing an accurate determination of future prices and chooses the optimal solution to DISPATCH(b) to account for these. An alternative is to suppose that each agent can accurately estimate all the future prices $\pi(n)$, $n \in \mathcal{N}$ and their probabilities, and then offers an optimal bid (both price and quantity). When dealing with the classical static problem, the optimal offer for a generation plant in node n is its entire capacity bid at its marginal cost (assumed to be constant in this paper). This offer will yield the same dispatch as the solution to DISPATCH(b).

In contrast, for a dynamic problem the quantity dispatched in any state affects the future state (e.g., energy storage), so the optimal bid needs to include a quantity component chosen as part of the optimization (see, e.g., Philpott et al. (2025)). When the price process is stagewise independent the optimal offer at state $x = x(n)$ can be constructed from a marginal value of x that yields a nondecreasing supply curve equal to the expected marginal cost of generation. For a more general price process, this marginal cost might depend on both the generation quantity and the current price realization, which could make the optimal dispatch decreasing with current price. Then the optimal solution to DISPATCH(b) will not be achievable by a nondecreasing offer curve.

5.1. Ramping

Consider a scenario tree with nodes $n \in \mathcal{N}$ and a set of J plants with ramping constraints. Each plant generates $u_j(n)$, $j \in \mathcal{J}$. The system optimization problem GENOPT seeks plant capacities a_j for $j \in \mathcal{J}_v$, consumption $d(n)$, generation $u_j(n)$, $j \in \mathcal{J}$ to minimize expected social cost. The generation in the next period cannot ramp up by more than $\gamma_j a_j$. Let $x_j(n)$ denote the generation at the start of node n. Then we have

$$
\begin{aligned}
\text{GENOPT:} \quad \min \quad & \sum_{j\in\mathcal{J}_v} \theta_j a_j + \sum_{n\in\mathcal{N}} p(n) \left(\sum_{j\in\mathcal{J}} c_j(n) u_j(n) - U_n(d(n)) \right) \\
\text{s.t.} \quad & u_j(n) \le x_j(n) + \gamma_j a_j, \quad j \in \mathcal{J}, n \in \mathcal{N}, && [p(n)\eta_j(n)] \\
& x_j(0) = 0, \quad j \in \mathcal{J}, \\
& x_j(n) = u_j(n-), \quad j \in \mathcal{J}, n \in \mathcal{N} \setminus \{0\}, && [p(n)\tau_j(n)] \\
& g(n) + \sum_{j\in\mathcal{J}} u_j(n) = d(n), \quad n \in \mathcal{N}, && [p(n)\pi(n)] \\
& 0 \le u_j(n) \le a_j, \quad j \in \mathcal{J}, n \in \mathcal{N}. && [p(n)\rho_j(n)]
\end{aligned}
$$

PROPOSITION 7. *If the capacities a_j for $j \in \mathcal{J}_v$ for thermal plants with ramping constraints are chosen at their optimal values a_j^* for the system optimization problem, GENOPT, then in the corresponding problem DISPATCH(a) the expected operating surplus for plant $j \in \mathcal{J}_v$ earned at system marginal prices $\pi(n)$ is equal to the investment cost $\theta_j a^*$.*

Proof. To apply Theorem 1 in this case we need to set

$$h_j = a_j, \quad K = 0, \quad A = 1, \quad H = 1, \quad G = 1, \quad w_j(n) = 0, \quad D = 1, \quad F = \gamma_j.$$

We omit the control capacity constraint involving B. Then the state dynamics ensures that $\bar{x}_j(n) = u_j(n)$. We can apply the theorem to show that the expected operating profit is equal to the investment cost as required. ∎

5.2. Battery

We now turn to problems in which storage plays a part. As is the case with ramping, we cannot achieve an optimal solution for DISPATCH(b) through the battery operator making an offer at a single price when the system operator uses merit order dispatch to determine whether the battery is called upon for generation (with a different price offered as part of a demand bid to determine whether the battery is charged).

Consider choosing the size h of a battery, given other fixed energy generation $g(n)$. Let ψ denote the capital cost per unit of storage capacity. The battery of capacity h can charge at a maximum rate $\alpha_C h$ and discharge at a maximum rate $\alpha_D h$. Let $u_C(n)$ denote the charge rate in node n and $u_D(n)$ the discharge rate. We model round-trip inefficiency by assuming charging by $u_C(n)$ requires $\beta u_C(n)$ units of energy where $\beta > 1$. The state of charge in the battery at the beginning of n is denoted $x(n)$ and the state of charge in the battery at the end of n is denoted $\bar{x}(n)$. We assume zero charge at the beginning of the planning horizon. This gives

$$\begin{array}{llrlr}
\text{GENOPT:} & \min & \psi h - \sum_{n\in\mathcal{N}} p(n)U_n(d(n)) & & \\
& \text{s.t.} & x(n) - u_D(n) + u_C(n) = \bar{x}(n), & n \in \mathcal{N}, & [p(n)\lambda(n)] \\
& & x(0) = 0, & & \\
& & x(n) = \bar{x}(n-), & n \in \mathcal{N} \setminus \{0\}, & [p(n)\tau(n)] \\
& & g(n) + u_D(n) - \beta u_C(n) = d(n), & n \in \mathcal{N}, & [p(n)\pi(n)] \\
& & u_D(n) \le \alpha_D h, & n \in \mathcal{N}, & [p(n)\rho(n)] \\
& & u_C(n) \le \alpha_C h, & n \in \mathcal{N}, & [p(n)\sigma(n)] \\
& & \bar{x}(n) \le h, & n \in \mathcal{N}. & [p(n)\mu(n)]
\end{array}$$

All variables are assumed nonnegative.

PROPOSITION 8. *If the capacity h of a battery is chosen at its optimal value h^* for the system optimization problem, GENOPT, then in the corresponding problem DISPATCH(h) the expected revenue for the battery earned at system marginal prices $\pi(n)$ is equal to the investment cost ψh^*.*

Proof. As before we apply Theorem 1 with just one capacity being optimized, and here we have

$$u_j(n) = [u_D(n)\ \ u_C(n)]^\top, \quad h_j = h, \quad K = 1, \quad A = [-1\ \ 1], \quad w_j(n) = 0$$

$$H = [1\ \ -\beta], \quad c_j(n) = [0\ \ 0]^\top, \quad B(n) = [\alpha_D\ \ \alpha_C]^\top, \quad G = 1, \quad \theta_j = \psi,$$

with the state-control linking constraint absent. This establishes the result. ∎

EXAMPLE 1 (Continued) We can extend Example 1 to this case. We add a lossless battery ($\beta = 1$, $\alpha_C = \alpha_D = 1$) of capacity h and unit cost $\psi = 4$, and take the dispatch quantities from Example 1 as a fixed background generation $g(n) = (80, 70, 90)$ for $n = 0, 1, 2$. We will see that an optimal policy has the battery charging fully at node 0 and discharging fully at both node 1 and node 2. So,

$$d(0) = 80 - h, \qquad d(1) = 70 + h, \qquad d(2) = 90 + h,$$

and hence $\pi(0) = 20 + h$, $\pi(1) = 20 - h$, and $\pi(2) = 60 - h$.

The optimal choice of capacity in GENOPT is given by $h^* = 8$, which we can check by noting that the derivative of the welfare objective is given by $-\pi(0) + 0.5\,\pi(1) + 0.5\,\pi(2) - \psi$. Thus the prices are $\pi(0) = 28$, $\pi(1) = 12$ and $\pi(2) = 52$. Since $\pi(1)$ and $\pi(2)$ are positive this confirms that there is discharge at both nodes.

Now we can calculate the expected rent. At $h^* = 8$, the expected rent is given by

$$-\pi(0)h^* + p(1)\pi(1)h^* + p(2)\pi(2)h^* = -28(8) + 0.5(12)(8) + 0.5(52)(8) = 32,$$

which equals the investment cost $\psi h^* = 32$, confirming the result of the proposition. ∎

5.3. Pumped-storage

The pumped-storage model is essentially the same as the battery model. The difference is that we choose the optimal size a of the generation/pumping plant and assume that the volume that can be pumped (u_C) is constrained to be at most δa, with $\delta < 1$. This gives

$$\begin{array}{llrll}
\text{GENOPT:} & \min & \multicolumn{3}{l}{\theta a + \psi h - \displaystyle\sum_{n\in\mathcal{N}} p(n) U_n(d(n))} \\
& \text{s.t.} & x(n) - u_D(n) + u_C(n) = \bar{x}(n), & n \in \mathcal{N}, & [p(n)\lambda(n)] \\
& & x(0) = 0, & & \\
& & x(n) = \bar{x}(n-), & n \in \mathcal{N}\setminus\{0\}, & [p(n)\tau(n)] \\
& & g(n) + u_D(n) - \beta u_C(n) = d(n), & n \in \mathcal{N}, & [p(n)\pi(n)] \\
& & u_D(n) \le a, & n \in \mathcal{N}, & [p(n)\rho(n)] \\
& & u_C(n) \le \delta a, & n \in \mathcal{N}, & [p(n)\sigma(n)] \\
& & \bar{x}(n) \le h, & n \in \mathcal{N}, & [p(n)\mu(n)]
\end{array}$$

where all the decision variables are nonnegative.

PROPOSITION 9. *Suppose the capacity a of a pumped-storage station is chosen at its optimal value a^*, and its reservoir capacity h is chosen at its optimal value h^* for the system optimization problem, GENOPT. Then in the corresponding problem DISPATCH(a, h) the expected revenue for the pumped-storage operator earned at system marginal prices $\pi(n)$ is equal to the investment cost $\psi h^* + \theta a^*$.*

Proof. We apply Theorem 1 with just a single optimized capacity vector

$$h_j = [h\ a]^\top, \quad u_j(n) = [u_D(n)\ u_C(n)]^\top, \quad \theta_j = [\psi\ \theta]^\top, \quad c_j(n) = [0\ 0]^\top,$$

$$w_j(n) = 0, \quad K = 1, \quad H = [1\ -\beta], \quad B(n) = \begin{pmatrix} 0 & 1 \\ 0 & \delta \end{pmatrix}, \quad G = [1\ 0], \quad A = [-1\ 1].$$

This gives the result we require. ∎

5.4. Hydro reservoir with inflow

We now consider a hydro reservoir with random inflow $\omega(n) \geq 0$, feeding a generation station with spill $s(n)$ and water release $z(n)$ that produces energy $\nu z(n)$. Spill must be nonnegative but has no upper bound. The reservoir and station together are regarded as a single hydro plant. The system optimization problem GENOPT seeks a reservoir capacity h, station capacity a, consumption $d(n)$, release $z(n)$, and spill $s(n)$ to minimize expected social cost. Storage in the reservoir at the beginning of n is $x(n)$, and the storage $\bar{x}(n)$ at the end of n must remain within its capacity. Given other fixed generation $g(n)$ the system optimization problem is

$$\begin{array}{llll}
\text{GENOPT:} & \min & \theta a + \psi h - \sum_{n \in \mathcal{N}} p(n) U_n(d(n)) & \\
& \text{s.t.} & \bar{x}(n) = x(n) + \omega(n) - z(n) - s(n), \quad n \in \mathcal{N}, & [p(n)\lambda(n)] \\
& & x(0) = 0, & \\
& & x(n) = \bar{x}(n-), \quad n \in \mathcal{N} \setminus \{0\}, & [p(n)\tau(n)] \\
& & g(n) + \nu z(n) = d(n), \quad n \in \mathcal{N}, & [p(n)\pi(n)] \\
& & 0 \leq z(n) \leq a, \quad n \in \mathcal{N}, & [p(n)\rho(n)] \\
& & 0 \leq \bar{x}(n) \leq h, \quad n \in \mathcal{N}, & [p(n)\mu(n)] \\
& & 0 \leq s(n), \quad n \in \mathcal{N}. &
\end{array}$$

PROPOSITION 10. *If the capacity a of the hydro generation station is chosen at its optimal value a^* and its reservoir capacity h is chosen at its optimal value h^* for the system optimization problem GENOPT, then in the corresponding problem DISPATCH(a, h) the expected revenue for the generation station earned at system marginal prices $\pi(n)$ is equal to the investment cost $\psi h^* + \theta a^*$ plus the expected value of uncontrolled inflows into its reservoir.*

Proof. Apply Theorem 1 with a single capacity vector in $\mathcal{J}_v$ and

$$\theta_j = [\theta\ \ \psi]^\top,\ \ h_j = [a\ \ h]^\top,\ \ u_j(n) = [z(n)\ \ s(n)]^\top,\ \ c_j(n) = [0\ \ 0]^\top,\ \ w_j(n) = \omega(n),$$

$$K = 1,\ \ A = [-1\ \ -1],\ \ H = [\nu\ \ 0],\ \ G = [0\ \ 1].$$

As in the run-of-river case, the control-capacity constraint is imposed only on the $z(n)$ component, giving $z(n) \leq a$; there is no upper bound constraint on $s(n)$. The state-control linking constraint is absent here. This shows that expected operating profits are

$$\theta a^* + \psi h^* + \sum_{n\in\mathcal{N}} p(n)\lambda(n)\omega(n).$$

Then complementarity with $s(n)$ shows $\lambda(n) \geq 0$ and the additional term is non-negative. ∎

EXAMPLE 1 (Continued) Returning to the example, we replace the battery with a hydro station on the same scenario tree and with the same fixed background generation $g(n) = (80, 70, 90)$. The station has turbine (release) capacity a, reservoir capacity h, no operating cost, and receives an uncontrolled inflow $\omega(0) = 8$ at the root node only, with $\omega(1) = \omega(2) = 0$. Investment costs are $\theta = 10$ per unit of turbine capacity and $\psi = 8$ per unit of reservoir capacity. The efficiency $\nu = 1$.

The optimal solution here is to set $h^* = a^* = 5$, and then $z(0) = 3$, and $z(1) = z(2) = 5$. Thus $\bar{x}(0) = 5$, and $x(1) = x(2) = 5$ with $\bar{x}(1) = \bar{x}(2) = 0$. This gives $d(0) = 83$, $d(1) = 75$ and $d(2) = 95$. Optimality follows by checking that the following multipliers satisfy the KKT conditions.

$$(\pi(0),\pi(1),\pi(2)) = (17,15,55), \quad (\lambda(0),\lambda(1),\lambda(2)) = (17,15,35), \quad (\rho(0),\rho(1),\rho(2)) = (0,0,20),$$
$$(\tau(1),\tau(2)) = (15,35), \quad (\mu(0),\mu(1),\mu(2)) = (8,0,0).$$

The turbine and reservoir capacity stationarity conditions are $\sum p(n)\rho(n) = \theta$ and $\sum p(n)\mu(n) = \psi$. As $\bar{x}(0) > 0$ we need $\lambda(0) + \mu(0) - (1/2)\tau(1) - (1/2)\tau(2) = 0$. Also, since $z(n) > 0$ we need $\lambda(n) + \rho(n) = \pi(n)$, $n = 0, 1, 2$. The remaining conditions follow immediately from complementarity and non-negativity.

The expected revenue at the fixed prices is $\pi(0)z(0) + p(1)\pi(1)z(1) + p(2)\pi(2)z(2) = 226$, which matches

$$\theta a^* + \psi h^* + \lambda(0)\omega(0) + (1/2)\lambda(1)\omega(1) + (1/2)\lambda(2)\omega(2) = 10(5) + 8(5) + 17(8) = 226,$$

confirming Proposition 10. ∎

We can also consider a cascade of J hydro reservoirs, indexed by j where $j = 1$ is the top reservoir and $j = J$ is the bottom. Each reservoir has its own random inflow $\omega_j(n)$ and in addition a station $j > 1$ receives $s_{j-1}(n) + z_{j-1}(n)$, being the spill and water release from the upstream reservoir $j - 1$. We can solve the overall system optimization problem in which all variables (including duals) are indexed by j. A natural model is to suppose that water transfers on the cascade are priced, so that the operator of plant $j > 1$ pays the upstream operator for these transfers at the appropriate price $\lambda_j(n)$ and is paid by the downstream operator for $s_j(n)$ and $z_j(n)$ at the price $\lambda_{j+1}(n)$. We can show that in this case we recover the result of Proposition 10 for each individual generation station.

6. Discussion and conclusions

Using stochastic programming, we have shown how classical screening-curve analysis extends naturally to technologies subject to uncertainty. The shadow prices on the supply-demand constraints in an economic dispatch problem produce Ricardian rents that pay for system-optimal generation capacities. The stochastic programming model extends screening curve analysis to wind, solar, pumped storage and batteries in a natural way. The results that we establish depend on our formulation of GENOPT the system capacity expansion problem. In the case of hydro generation the model determines the optimal size of a single generation facility on a river or reservoir catchment. This facility captures inflows and restricts additional investment in a way that wind and solar capacity does not, thereby creating a barrier to hydro entry and generating more rent than is needed to pay for the optimal capacity choice.

The additional rents accrued by hydro generators from inflows may result in positive profits at the socially optimal investment. If these additional rents were appropriated then there would be no effect on the optimal investment or operation of the hydro system, at least to first order. Appropriation would not be acceptable if inflows to hydroelectricity stations were viewed as property rights to water that are sold to the owners of the generation assets when the new capacity is built. One might imagine the price of this sale occurring in an auction in which the rights owner sells them to the highest bidder. In a competitive equilibrium the maximum sale price of inflow rights would then be $\sum_{n\in\mathcal{N}} p(n)\lambda(n)\omega(n)$. Any higher price would make an optimal hydro generation investment unprofitable.

The correspondence between a socially optimal solution and an energy-only market equilibrium relies on market completeness: all commodities should be priced. To recover social optimality, generation stations at different points on a river chain must either be coordinated by a single operator, or their operators must trade water at market prices if the stations are operated independently. Our analysis also assumes that all asset owners are risk neutral. If agents are endowed with nested coherent risk measures and risk markets are complete then we can apply the dynamic risked equilibrium analysis of Ferris and Philpott (2022). In this setting agents will trade Arrow-Debreu securities to yield revenues and costs that cover their investment costs in risk-adjusted expectation when evaluated using the appropriate social risk measure.

Linearity of costs is a key assumption in our analysis. Each plant is assumed to have a constant marginal cost of generation. If this cost were strictly increasing then marginal pricing yields a convexity rent even when the plant is not at capacity. Similar rents accrue when marginal investment costs are increasing. In this case plants recover their optimal capacity times their marginal investment cost at this capacity, as we show in Section 4. An alternative but more realistic investment cost model satisfies economies of scale, where marginal investment costs are decreasing. In this case we lose convexity of GENOPT, which invalidates nearly all the arguments we have used above.

There are cases, for example when reservoir inflows are drawn from a continuous distribution, where it is more natural to use a continuous state space rather than a finite scenario tree. This can be viewed informally as the limiting case in which the size of the scenario tree at each time stage becomes large. To formalize this, we would need to specify a filtration describing the information revealed over time, as well as conditional distributions for the evolution of the state. The resulting problems become infinite dimensional, and the decision variables become functions of the state rather than finite-dimensional vectors indexed by the nodes of the scenario tree. In our setting, it is natural for marginal utility to be bounded above by a value of lost load. We also typically have bounded generation, storage and investment decisions. These bounds provide the kind of compactness needed for convergence of finite approximations, although a full convergence result would also require regularity assumptions on the transition laws. We do not pursue these details here, but simply note that our restriction to finite scenario trees can be interpreted as an approximation to more general stochastic models.

References

Arrow K, Debreu G (1954) Existence of an equilibrium for a competitive economy. *Econometrica* 22(3):265–290.

Boiteux M (1960) Peak-load pricing. *The Journal of Business* 33(2):157–179.

Cory-Wright R, Philpott A, Zakeri G (2018) Payment mechanisms for electricity markets with uncertain supply. *Operations Research Letters* 46(1):116–121.

Ehrenmann A, Smeers Y (2011) Generation capacity expansion in a risky environment: a stochastic equilibrium analysis. *Operations Research* 59(6):1332–1346.

Ferris M, Philpott A (2022) Dynamic risked equilibrium. *Operations Research* 70(3):1933–1952.

Green R (2000) Competition in generation: The economic foundations. *Proceedings of the IEEE* 88(2):128–139.

Holmberg P, Philpott A (2026) Long-run equilibrium effects of variable renewable electricity. Technical report, www.epoc.org.nz.

Joskow P (2008) Capacity payments in imperfect electricity markets: Need and design. *Utilities Policy* 16(3):159–170.

Joskow P, Tirole J (2007) Reliability and competitive electricity markets. *The RAND Journal of Economics* 38(1):60–84.

Kok C, Philpott A, Zakeri G (2018) Value of transmission capacity in electricity markets with risk averse agents. Technical report, www.epoc.org.nz.

Korpås M, Botterud A (2020) Optimality conditions and cost recovery in electricity markets with variable renewable energy and energy storage. *MIT Center for Energy and Environmental Policy Research* 2020–005.

Newbery D (2016) Missing money and missing markets: Reliability, capacity auctions and interconnectors. *Energy Policy* 94:401–410.

Philpott A, Ferris M, Mays J (2025) Electricity dispatch and pricing using agent decision rules. Technical report, www.epoc.org.nz.

Philpott A, Ferris M, Wets R (2016) Equilibrium, uncertainty and risk in hydro-thermal electricity systems. *Mathematical Programming* 157(2):483–513.

Pritchard G, Zakeri G, Philpott A (2010) A single-settlement, energy-only electric power market for unpredictable and intermittent participants. *Operations Research* 58(4-part-2):1210–1219.

Steiner PO (1957) Peak loads and efficient pricing. *The Quarterly Journal of Economics* 71(4):585–610.

Stoft S (2002) *Power System Economics: Designing Markets for Electricity* (John Wiley).

Zakeri G, Pritchard G, Bjorndal M, Bjorndal E (2019) Pricing wind: a revenue adequate, cost recovering uniform price auction for electricity markets with intermittent generation. *INFORMS Journal on Optimization* 1(1):35–48.