Portfolio Short Run Marginal Cost of Electricity Supply in Half Hour Trading Intervals

Technical Paper

Author: Adam McHugh

11 January 2008

Economic Regulation Authority
WESTERN AUSTRALIA

Important Notice

A full copy of this document is available from the Economic Regulation Authority web site at www.era.wa.gov.au.

For further information, contact:

Economic Regulation Authority
Perth, Western Australia
Phone: (08) 9213 1900

Contents

1 Introduction

This paper has been prepared by Adam McHugh, Manager of Projects, in the Economic Regulation Authority's Competition Markets and Electricity Division, in the interests of advancing the debate in economic regulation. The paper seeks to assist market participants in the understanding of short run marginal cost (**SRMC**). In doing so, it sets out to identify what costs may be included in a firm's short run portfolio supply curve calculation and how SRMC may be estimated. The views presented herein are those of the author and should not be taken as reflecting the views of the Authority, individual members of the Authority, the Authority's Secretariat, or members of other organisations.[1]

Rules pertaining to the Wholesale Electricity Market (**WEM**) in Western Australia (**WA**) are, in part, designed to replicate the outcomes of a competitive market. Specifically, generators are required to offer electricity at the SRMC of production. Clause 6.6.3 of the *Wholesale Electricity Market Amending Rules (December 2006)* (referred to in this document as the **market rules**) states:

> 6.6.3. A Market Generator must not, for any Trading Interval, offer prices within its Portfolio Supply Curve that do not reflect the Market Generator's reasonable expectation of the short run marginal cost of generating the relevant electricity when such behaviour relates to market power.

The roles of the Economic Regulation Authority (**Authority**) and the Independent Market Operator (**IMO**) with respect to clause 6.6.3. are outlined in clause 2.16.9. and its various subclauses as follows:

> 2.16.9. The Economic Regulation Authority is responsible for monitoring the effectiveness of the market in meeting the Wholesale Market Objectives and must investigate any market behaviour if it considers that the behaviour has resulted in the market not functioning effectively. The Economic Regulation Authority, with the assistance of the IMO, must monitor:
>
> …
>
> (b) inappropriate and anomalous market behaviour, including behaviour related to market power and the exploitation of shortcomings in the Market Rules or Market Procedures by Rule Participants including, but not limited to:
>
> i. prices offered by a Market Generator in its Portfolio Supply Curve that do not reflect the Market Generator's reasonable expectation of the short run marginal cost of generating the relevant electricity;
>
> …
>
> 2.16.9A. The IMO must assist the monitoring activities identified in clause 2.16.9(b)(i) by examining prices in STEM Submissions, including Standing STEM Submissions, used in forming STEM Bids and STEM Offers against information collected from Rule Participants in accordance with clauses 2.16.6 and 2.16.7.
>
> 2.16.9B. Where the IMO concludes that prices offered by a Market Generator in its Portfolio Supply Curve may not reflect the Market Generator's reasonable expectation of the short run marginal cost of generating the relevant electricity and the IMO considers that the behaviour relates to market power the IMO must:
>
> (a) as soon as practicable, request an explanation from the Market Participant which has made the relevant STEM Submission; and

[1] Comments and/or feedback in the development of this paper came from Peter Kolf and Robert Pullella of the Economic Regulation Authority, Dora Guzeleva of the Independent Market Operator, and Duncan Farrow of Murdoch University, and are gratefully acknowledged. The author takes sole responsibility for any errors that may remain.

(b) advise the Economic Regulation Authority of its conclusions. The IMO advice must outline the reasons for the IMO's conclusions.

…

2.16.9G. Where the Economic Regulation Authority determines that prices in the Portfolio Supply Curve, subject to the investigation, did not reflect the Market Generator's reasonable expectation of the short run marginal cost of generating the relevant electricity, the Economic Regulation Authority must request that the IMO applies to the Energy Review Board for an order for contravention of clause 6.6.3.

2.16.9H. Where the IMO receives a request under clause 2.16.9G the IMO must refer the relevant matter to the Energy Review Board requesting that a civil penalty be imposed on the relevant Market Participant.

In summary, the Authority, with the assistance of the IMO, is responsible for monitoring the effectiveness of the market in meeting the Wholesale Market Objectives and must investigate any market behaviour if it considers that the behaviour has resulted in the market not functioning effectively. If a firm with market power submits a portfolio supply curve that does not reflect that firm's reasonable expectation of SRMC for any given trading interval and the Authority determines that to be the case, the matter must be referred by the IMO to the Energy Review Board requesting that a civil penalty be imposed on the relevant market participant.[2]

A non-technical version of this paper entitled 'Short Run Marginal Cost' (Economic Regulation Authority, 2008) invites public submissions on the matters raised. A copy of both versions are available on the Authority's web site.

1.1 Economic efficiency

Economic theory labels a market as *efficient* if the net benefits (i.e. the total benefits minus the total costs) created in that market reach a notional maximum. An *inefficient* market wastes resources, either because:

a) more goods could have been created with the same amount of resources, and/or;

b) goods are not allocated in accordance with the tastes and preferences of consumers.

That is, in the short run, economic efficiency requires:

a) technical efficiency;

b) allocative efficiency, and;

c) in the long run, dynamic efficiency.

Under theoretical assumptions that describe a *competitive* market, efficiency will occur naturally. These assumptions are:

1) No market participant has the ability to independently influence the clearing price by virtue of market share or control (i.e. market power);

2) There is no strategic behaviour in the market;

3) All participants are able to make fully informed decisions;

[2] A firm's market power generally corresponds to its ability to influence the clearing price. Future work will focus on this concept more thoroughly.

4) There are no costs or benefits that accrue to third parties (i.e. externalities).

If these conditions are met, economic theory suggests that the price of electricity, or any other good or service, will fall to SRMC. That is to say, price will be set by the cost of the most expensive MWh of electricity produced during the relevant period of generation.

However, electricity markets are rarely, if ever, fully competitive. In the absence of sufficient competition, an alternative is to replicate the competitive (and therefore efficient) outcome through economic regulation and good market management.

Condition 3 above, for example, can be met by timely and centralised release of market information by a market operator (the IMO in the case of WA). Condition 4 can be met by taxes, subsidies or other forms of regulation that adjust the private costs of the firm so as to 'internalise' any external effects. This paper however, is concerned with conditions 1 and 2, which may, in theory, be met at relatively low regulatory cost by putting in place legislation that requires firms to offer their output at SRMC. Provided firms act in accordance with this rule (and provided conditions 3 and 4 are also appropriately managed), the resulting price will be one that equals SRMC at the efficient equilibrium between supply and demand.

1.2 The definition of short run marginal cost (SRMC)

The marginal concept in economics refers to the rate at which one quantity changes with respect to extremely small increases in another quantity. It follows therefore that SRMC is simply defined as *the change in short run total cost for an extremely small change in output*.[3,4]

[3] Equally, SRMC can be defined as the change in short run *variable* cost for a small change in output.

[4] The smaller the change in output used as a reference, the more accurate will be the measure of SRMC.

Box 1.1 Mathematical definition of SRMC

Consider the following generic function for an individual plant or a portfolio:

$$C = C(Q), \qquad Q \geq 0,$$

where short run total cost (denoted C) for a plant or a portfolio of plants is a function of its output Q.

The most precise mathematical definition of SRMC is given by the first derivative of the short run total cost function:

$$C'(Q) = \frac{dC}{dQ}, \qquad Q \geq 0,$$

where SRMC (denoted C') is a function of the plant's output. That is, C' is the instantaneous rate of change of C with respect to Q as a continuous function.

This function can be very closely approximated by means of the following difference equation:

$$\dot{C} = \frac{\Delta C}{\Delta Q}, \qquad Q \geq 0,$$

where SRMC (denoted $\dot{C}$) is given as the change in cost over the change in quantity. Use of this discrete method, however, is only appropriate for very small changes in Q such that $\dot{C}$ comes very close to C' - i.e. the smaller the change in Q used, the closer the result will be to that of the first derivative method.

2 Short run cost concepts

Since SRMC relates to the rate of change of short run total cost, an understanding of what constitutes short run total cost is a pre-requisite for determining SRMC. Once such an understanding is developed, SRMC can be derived mathematically. Statistical methods can then be employed enabling data to be entered into an optimisation model (see Section 5 and Appendix 1). Unfortunately, a disparity between accounting concepts of cost and economic concepts of cost can often lead to confusion at the conceptual level. Economic regulators expect participants to use the economic concept of cost in their interpretation of the market rules. Part of the scope of this paper therefore is to state clearly the economic concept of cost in a way that corresponds to these expectations.

The economic understanding of cost incorporates a number of fundamental concepts that are not immediately intuitive. These are:

- the distinction between the short run and the long run;
- the distinction between sunk and avoidable costs;
- the distinction between fixed, variable, shutdown, and total costs, and;
- the concept of opportunity cost.

This section has two purposes:

1) to describe the above concepts; and
2) to provide clarification as to what should be incorporated in the economic determination of SRMC.

2.1 The distinction between the short run and the long run

The distinction between the short run and the long run in economics is only superficially about any notion of duration or time. Rather, the distinction has more to do with whether or not there is full scope to substitute a more cost effective input for a less cost effective input.

An electricity generator that is able to choose any technique within budget for combining all the various inputs such as plant and machinery, fuel, labour, land, etc. has much more flexibility to minimise cost than a firm that is committed to a certain quantity of one or more inputs. If the price of one type of input changes relative to another, a manager with the full flexibility to utilise any affordable technique can simply choose one that employs less of the more expensive input and more of the cheaper one. If fuel becomes cheap, for example, it may become economic to sell off some expensive assets to provide the necessary cash to purchase more fuel: this is fuel input substituting for capital input.

Alternatively, with rising fuel costs, the installation of fuel saving equipment may become economic: this is capital input substituting for fuel input. When a firm is able to optimise its expenditure in any way it pleases, it is able to choose that *particular* mix of inputs that produce output at the minimum technologically feasible cost. *This can only occur over the long run, when the firm is free of any irrevocable commitments to input*. In other words, the long run is how long it takes for the technologically feasible mix of inputs to be able to be (re)arranged in any combination a manager pleases.

A firm that is irrevocably committed to quantities of one or more input type does not have the same flexibility. In this case the firm is said to face a *short run* decision. *In the short run, the firm must optimise as best it can given that it cannot vary certain quantities of input*. In other words, the short run means that a manager is stuck with some amount of input that cannot be swapped for a different type of input. A firm's portfolio of generating facilities, for example, cannot be varied in the short run. Similarly, sunk labour contracts constitute a short run constraint.

2.2 Defining economic cost: The opportunity cost concept

An economic decision is one in which the decision maker acknowledges all the *alternative opportunities* that are available to create benefits. The question that is always asked in a rational economic decision is therefore "what do I have to give up by taking this course of action?" which is equivalent to asking "is this course of action going to create more net benefits (profit) for myself (my firm) than any other possible course of action?". If the answer to this second question is "yes" then the choice is an *economic* one. If the answer is "no", then the choice is *uneconomic* because an opportunity to create greater net benefits has been forfeited. All economic costs incorporate this notion of opportunity cost.

Since an economic decision takes into account the benefits that have to be given up to achieve *greater* benefits, a further, economically important, question must be asked: "what net benefits (profit) would I have created for myself (my firm) if I had taken the next best course of action?". This question defines the concept of opportunity cost. An opportunity cost is a *potential* benefit that is forgone to create greater benefits. The opportunity cost of an input decision is therefore equal to the benefits that could have been derived from employing resources in their next best use.

To clarify the economic concept of cost, consider the following example. A generating plant, if started, will bring in revenue of $1.2M over a period. If the operating costs over that period were to include past fuel expenditures of $0.6M plus other expenditures of $0.4M, it could be said that the plant would make a $0.2M accounting profit.

However, before committing to operating the plant, it is appropriate to ask: "what alternatives are there?". If the value of the fuel required to operate the plant rose on the current spot market to $0.9M, assuming the firm has the opportunity to sell the fuel, the choice to produce electricity would be uneconomic. In this case, the firm could make an accounting profit of $0.3M by selling the fuel as opposed to a $0.2M accounting profit by selling electricity. In other words, the firm would profit by $0.1M more by selling the fuel and not operating the plant. In *economic* terms, if the plant were to operate it could make an economic loss of $0.1M.

As the above example suggests, the past purchase price or past contract price of fuel does not represent a short run economic cost. Rather, the current price that can be obtained by on-selling a quantity of fuel represents what is given up by using it to produce electricity. It follows that if the fuel cannot be immediately on-sold then the opportunity cost of that fuel for a firm in a competitive market may be zero unless there exists a real prospect to either: (a) sell the fuel at some point in the future; or (b) use the fuel in the future production of electricity at higher than current market values.[5] In either case, the

[5] It is noted that if a plant is withdrawn as a result of fuel being withheld then penalties, including withdrawal of capacity credits under the Reserve Capacity Mechanisms, may apply and would need to be taken into consideration. Note, however, that from the regulator's point of view it would be inappropriate to include regulatory penalties in the calculation of SRMC under the legislated SRMC approach as this would result in increased prices, thus reducing the incentives the penalties are intended to instill in producers.

current spot price, or alternatively the current market price for fuel contracts, still provides the best basis for the opportunity cost of fuel if the view is taken that the current price reflects the rational expectations (based on the latest information available to the market) of all future prices for the resource.[6] However, this forgone benefit would need to be adjusted downwards by storage costs, inflation, and the interest that could have been earned while waiting for a buyer if the fuel had been converted to cash immediately.

2.3 Sunk costs

As stated in Section 2.1, firms in the short run face either technological or contractual commitments to input. In economics, the costs associated with these commitments are considered *sunk*. They cannot be avoided or reversed in the short run.[7]

Importantly, *the economic cost of a sunk commitment is zero*. As already stated, opportunity cost describes the potential benefits that are forgone by employing a resource in its next best alternative use. By definition, expenditures that are sunk have no alternative use since they cannot be avoided in the short run. It follows that sunk costs have no influence on economic decision making in the short run. Rather, the short run decision is entirely about current expenditure on marginal operating costs compared to market price.

Take, for instance, the decision to build a gas fired power station as opposed to the decision to sell electricity from a gas fired power station that has already been built. The former is a long run decision: the firm must decide whether it believes the investment will be worthwhile based on the present value of the expected flow of future revenues and costs, with the certainty of revenue and cost flow typically guaranteed by long run contracts with take-or-pay provisions (so as to ensure the recovery of sunk costs).

Once the plant is built however, the decision as to how much electricity should be produced becomes a short run decision and the plant's construction costs become relegated to the past. In a competitive market, the firm would face a market price that it cannot control: a “take it or leave it” proposition. Therefore, the optimal economic decision for determining how much electricity should be produced, over and above that for which take-or-pay contracts may exist, is determined simply by whether supply of an additional MWh increases the firm's bank balance or decreases it. This depends solely on whether or not the price that is available for incremental amounts of electricity exceeds the immediate costs of production.

Figure 2.1 provides an example. A firm operating in a competitive market is offered a price of $30 per MWh to supply electricity. If the various current expenditures that the firm must outlay to produce incremental MWh's of electricity are lower than $30, the sale of that electricity will see the firm's profits going up. If, however, an incremental MWh of supplied electricity entails more than $30 of current expenditure, the firm's profits will get progressively smaller for each additional MWh produced. Thus, while the value of the firm before the short run decision is made will have been affected by past decisions on capital

---

6 According to rational expectations theory, the market price in the t'th period, designated P_t , is given by: $P_t = P_t^e + u_t$, where P_t^e is the expected price in the t'th period, and where u_t is a random error term that is independent of the current price and has an expected value of zero (see: Muth, J. (1961), ‘Rational Expectations and the Theory of Price Movements’, *Econometrica*, Vol. 29, No. 3, pp. 315-335).

7 In the long run there are no sunk costs because, by definition, there are no irrevocable commitments to input. There are, however, redundant investments, the scrap value of which may be equal to or close to zero.

expenditure, the amount previously invested will have absolutely no bearing on the best decision that the firm can make *today* to see the value of the firm head in the right direction. If, for example, the firm seeks to recover all of its capital costs by asking for a higher price, the buyers in a competitive market will simply get their electricity elsewhere.

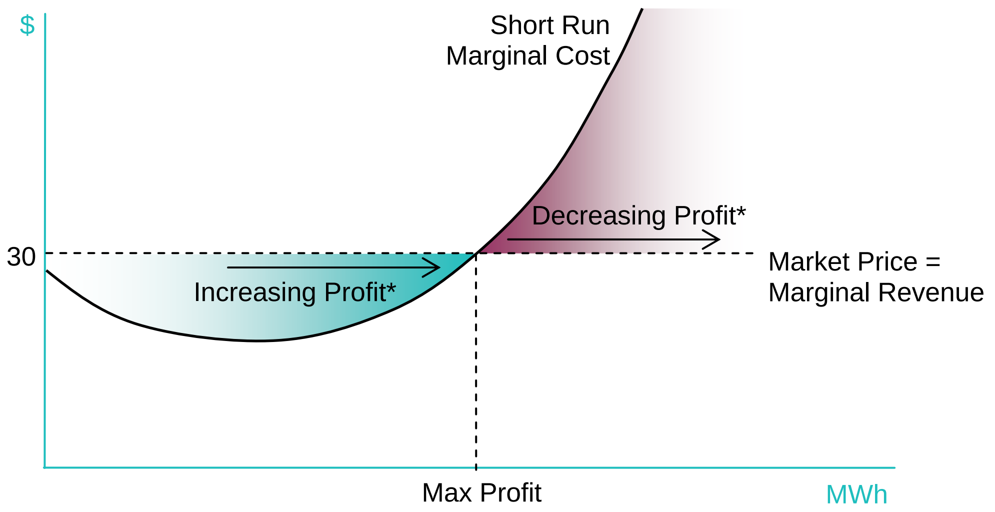


Figure 2.1 – Profit maximising short run supply decision in a perfectly competitive market

* Note: the short run marginal cost curve does not account for avoidable fixed costs such as those associated with the startup of a plant during a trading interval. These must be subtracted from any profit or added to loss for the trading interval. Avoidable fixed costs are independent of output and incurred as a lump sum. They differ from sunk costs in that the latter are incurred prior to the current trading interval and so are unavoidable in the current trading interval.

To define in summary:

- **Sunk costs**. A sunk cost is a fixed cost that cannot be avoided in any given (short run) time period.

2.4 Avoidable costs

An avoidable cost is an expenditure to which the firm is *not* irrevocably committed or that can be reversed over the time period in question. Therefore, unlike a sunk cost, the economic (opportunity) cost of an avoidable cost will be of positive amount because the money or the resources saved by avoiding the cost could potentially be used to generate benefits elsewhere. Avoidable costs can be broken down as follows:

- **Avoidable fixed costs**. An avoidable fixed cost is an expenditure that must be borne by the firm if it chooses to produce any amount of output in a given time period. That is, an avoidable fixed cost is an expenditure that remains constant for any level of output above zero but that can be avoided by producing zero output. An avoidable fixed cost should not be confused with a sunk cost, which cannot be avoided in the short run even if the firm were to choose not to produce.[8] The most relevant avoidable fixed cost in the context of electricity production is the cost of starting a plant.

- **Shutdown costs.** A shutdown cost is an expenditure that must be borne by the firm if it has invested in plant, is operating that plant and chooses to produce zero

[8] While all sunk costs can be considered fixed costs, not all fixed costs are sunk (see Wang, H. & Yang, B (2001), “Fixed and Sunk Costs Revisited”, *Journal of Economic Education*, Vol. 31, No. 2, Spring, pp. 178-185).

output in a given time period. A shutdown cost is therefore the opposite of an avoidable fixed cost in that it can only be avoided by producing non-zero output. In the context of electricity production, a plant already operating in a half hour trading interval will suffer a short run shutdown cost if demand for its output does not equal or exceed its minimum technical output (**mingen**).[9] The shutdown cost corresponds to the maximum amount a firm would be willing to pay to avoid the shutdown. This will include any loss in net revenue that could be avoided if the plant remained operational.

- **Variable costs.** A variable cost is a cost that varies with the level of output (beyond mingen) that the firm produces over the relevant time period. Prominent variable costs in the context of electricity production are fuel costs, operating costs, costs associated with wear and tear on plant and equipment. These costs may either vary directly in proportion with output or otherwise change in steps as output increases (e.g. labour costs). Most variable costs increase with output but there are some that decrease with output (i.e. some variable costs are avoidable in the short run by producing higher levels of output). If, for example, an electricity generator has commitments to provide steam to nearby industries, any per unit amount of compensation that they must pay to that firm if they do not provide the steam constitutes a decreasing variable economic cost.

- **Total cost.** Total economic cost is the variable cost for a given level of output plus any avoidable fixed costs or shutdown costs. Sunk expenditures are also a component of total cost, but because they have a zero opportunity value they do not affect its calculation.

Figure 2.2 describes a taxonomy of cost categories consistent with the discussion above. Figure 2.3 provides sequential questions that enable a clear categorisation of the short run total cost components outlined in this section.

	Variable	***Not Variable***
Avoidable	**Variable Costs**	**Avoidable Fixed Costs** **Shutdown Costs**
Not Avoidable		**Sunk Costs**

Figure 2.2 – Categorisation of short run economic costs associated with electricity generation with given examples.

[9] Mingen is the minimum amount of output (and hence input) required to keep a plant running.

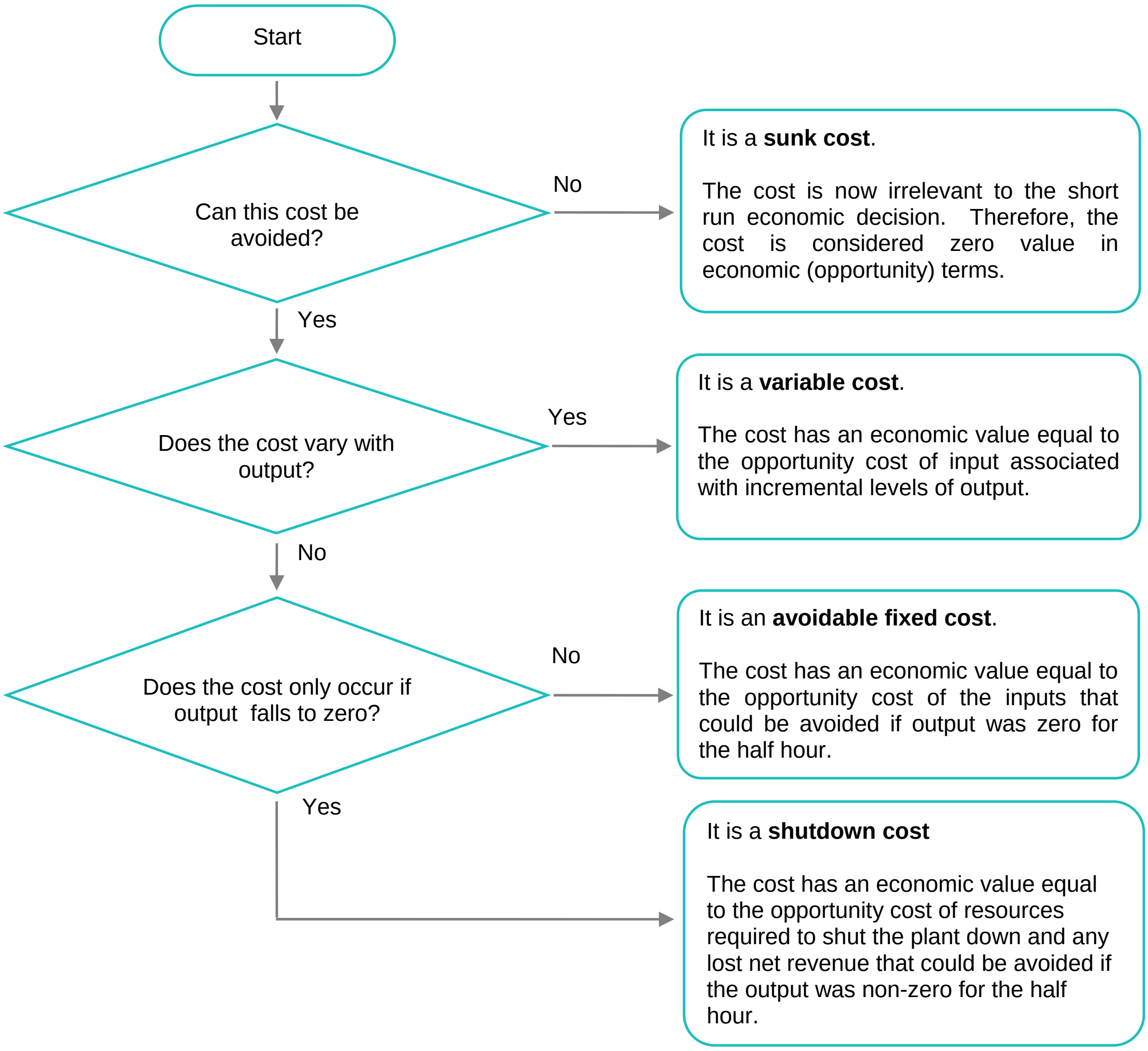


Figure 2.3 – Sequential question method of categorising the components of short run total economic cost for a half hour trading interval.

Below are examples of the costs included in each of the categories identified in Figure 2.2 above.

Variable costs

- Value of saleable fuel (at the current market rate) used to produce electricity output.
- Wear and tear on plant and equipment directly attributable to the production of output.
- The expected costs of plant failure - i.e. the probability of plant failure multiplied by the cost of plant failure, where the probability of plant failure increases with the level of output.[10]
- Value of water and other inputs used to produce electricity output.

[10] Such costs are associated with unplanned maintenance and include not only repair costs, but also lost net revenues as a result of plant unavailability, as well as financial penalties imposed for unscheduled outage. It should be noted that if the probability of unplanned outage is not correlated with high levels of output, the risk of equipment failure becomes an avoidable fixed cost as opposed to a variable cost.

- Costs as a result of per unit financial penalties imposed for not meeting contracted output commitments.[11]
- Cost of pollution control permits, per unit taxes and other output linked regulatory instruments.

Avoidable Fixed Costs

- Fuel, water and other operating costs of starting a plant within the current trading interval.

Shutdown costs

- Fuel, water, and operating and maintenance costs of shutting a plant down within the current trading interval.
- Lost net revenue in future trading intervals as a result of the plant lying idle while awaiting future startup (as opposed to incurring costs resulting from the plant operating at mingen).
- Includes costs associated with a future startup minus mingen costs over the intervening period (adjusted for any avoided cost of base-load).

Sunk costs

- Capital cost of plant and equipment.
- Cost of starting a plant prior to the current trading interval.
- Cost of fuel purchased on long term contracts that cannot be on-sold to third parties.
- Labour costs (apart from avoidable labour costs, such as overtime).[12]

[11] Note: where such costs relate to regulatory penalties intended to influence a firm's private decision (e.g. penalties for not meeting obligations under the reserve capacity mechanism or for renewable energy certificates) they are not a cost a regulator would allow to be included in SRMC calculations.

[12] Because the hiring of labour occurs over periods long than a half hour or even a day, such costs are sunk in the short run.

3 Unit commitment

The time it takes to start plants up and shut them down again constrains the availability of various plants for any given trading interval and therefore affects the day ahead estimation of portfolio SRMC. It is reasonable to assume that this unit commitment decision is motivated by the dual incentives of reliability and profit. Therefore, the unit commitment plan, under the assumption of profit maximising behaviour, becomes the foundation upon which a reasonable expectation of portfolio SRMC is built. What follows in this section is a brief discussion of unit commitment considerations and how these impact upon the reasonable, day ahead, expectations of half hourly economic cost.

3.1 Short run time frames: The 24 hour trading day and the half hour trading interval

By the very nature of the product, the scheduling of electricity supply must be considered with reference to very small time intervals. This is because electricity dispatch requires a near perfect instantaneous balancing between supply and demand otherwise the system will move outside of a narrow technical envelope and collapse at great cost.[13] Consequently, market equilibrium is an engineered solution in the first instance and the economically efficient price the retrospective determination of a market operator.

Typically, an electricity market is cleared on a half hourly basis (or more frequently), involving both planned and real time events. In Western Australia, market participants submit portfolio supply and demand schedules on a day ahead basis to the IMO. Electricity generators, therefore, make economic short run decisions that correspond to strictly predefined temporal bounds. The predefined bounds in the Western Australian WEM are a 24 hour trading day (8am to 8am) divided up into 48 half hourly trading intervals. To the extent that real time events differ from the submitted day ahead portfolio supply and demand schedules, a balancing mechanism settles any differences.

3.2 Linkages between unit commitment and short run economic cost

In the State's wholesale electricity market there are essentially two relevant short run questions facing electricity generators:

1) How should resources be allocated over the trading day?
2) What is the profit maximising output for each half hour trading interval?

These two questions are inextricably linked. If the primary goal of the firm is to maximise profit over the trading day, this can only be achieved under competitive conditions if the SRMC of electricity supply is set at marginal revenue in each half hourly period. However, the optimal allocation of resources through the trading day will affect the SRMC of supply in each trading interval. The profit motive will cause a firm to attend to this allocation of resources, known as unit commitment, as efficiently as possible. Hence, the assumption of profit maximising behaviour by a firm simplifies analysis.

[13] Electricity supply tolerances are defined and regulated in technical rules, which are very narrow to maintain the integrity of the system. Electricity market equilibrium is therefore quite artificial and heavily dependent on engineering and information technology for its existence.

The profit motive, however, can be a two edged sword. While it provides firms with the incentive to minimise their costs of production, a profit maximising firm with the ability to raise price by virtue of market power can extract excessive profits at the expense of consumers, resulting in less net benefits for society as a whole. This is to say that in the presence of market power the profit maximising level of output will not be economically efficient.

A firm operating in a competitive (efficient) market, however, would find the profit maximising level of output occurs where price equals SRMC in each half hour trading interval. This would produce a socially optimal outcome.[14] The approach taken in the WEM is to attempt to replicate this outcome by requiring generators to submit their electricity supply offers so as to reflect *SRMC in each trading interval*. To the extent this can be implemented, enforced, and provided the economic definition of cost is used, the level of profit will reflect an efficient market outcome thereby arresting the problem of market power.

3.3 The mix of generation technologies

Figure 3.1a shows a typical diurnal pattern of electricity output. Given the periodic pattern of demand a mix of technologies is required to continuously balance system output with load over a trading day. Generation technologies can be categorised as either intermittent, base-load, mid-merit or peaking plants. The following characteristics generally apply:

a) Currently, plants that use intermittent resources such as wind and solar provide a relatively uncertain (stochastic) level of output in a trading interval, but have zero fuel costs and so are typically the lowest SRMC plants. This is particularly likely to be the case where environmental and technological regulations such as renewable energy certificates, tradable pollution permits, and/or carbon taxes are significant.

b) Base-load plants have low variable costs but high startup and shutdown costs, and have a relatively limited capability to ramp output up or down to follow load. Base-load generation in WA often uses coal to produce steam to be converted into electricity.

c) Mid-merit or load following units are normally medium cost plants both in terms of variable cost, startup cost and shutdown cost, and are quite effective at adjusting (ramping) output up or down to cover reasonably rapid changes in demand. In WA mid-merit plants are usually natural gas fuelled turbines or co-generation plants.

d) Peaking plants in WA are typically based upon gas turbine technology. These units are very effective at following load and can be started up and shutdown quickly and cheaply. However, they are usually the highest variable cost plants in a portfolio, particularly when distillate is used as the fuel input.

[14] Assuming any externalities are internalised through supplementary government policy.

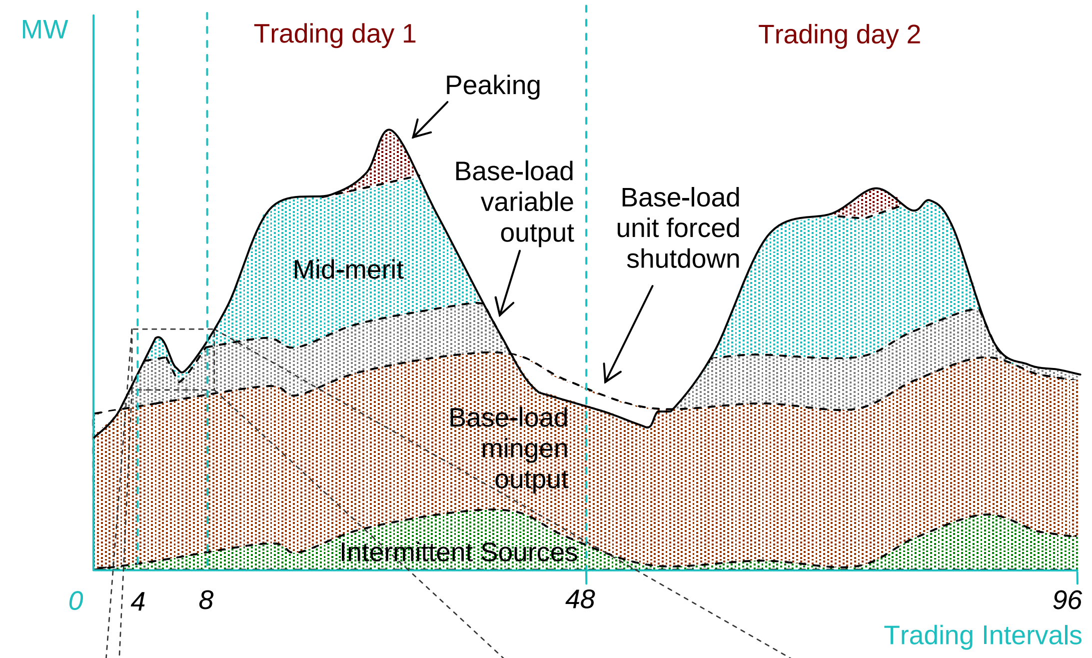


Figure 3.1a – Electricity output over two trading days

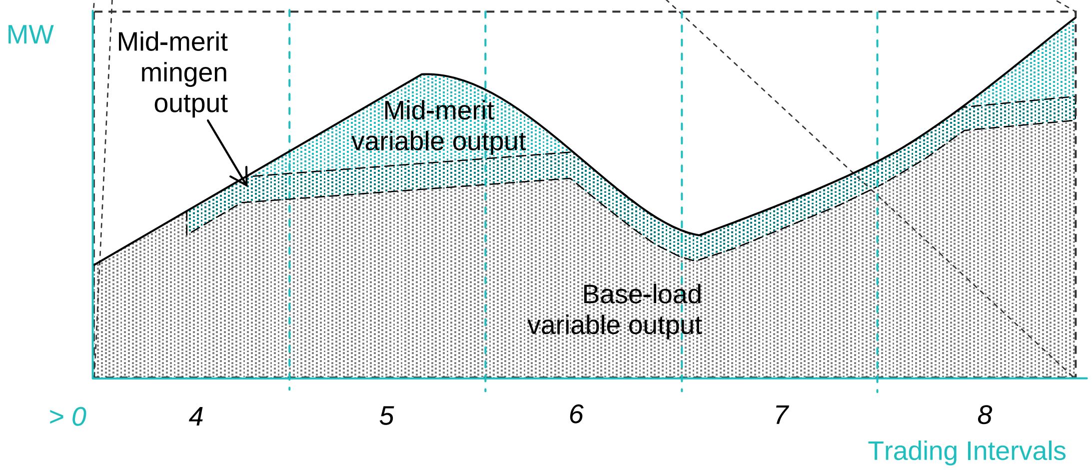


Figure 3.1b – Electricity output over five trading intervals

3.4 Shutdown costs vs. mingen costs

In maximising profits over a trading day, a critical consideration is the comparison between the cost of a future startup for a plant (which can be large) and the cost of keeping the plant running at mingen over the intervening period. The primary example is base-load generation. Whenever possible, coal fired base-load facilities will continue to operate even when electricity demand is low. This is because the costs associated with shutting a coal fired base-load plant down are large.

The opportunity costs of forcing such a plant below mingen will include not only the immediate costs associated with taking the plant offline but also the cost of starting the plant up again when it is required.[15] The time that it takes to have such a plant come back into operation can be considerable. If this causes the plant to be unavailable when it is needed there will be an additional opportunity cost associated with lost revenue in future trading intervals while the plant is lying idle. In other words, while within the trading interval it may be cheaper to shut a plant down than to run the plant, it may not be the best decision over the trading day. Therefore, the impact on cost in future trading intervals must be considered in the current decision. For this reason, and for reasons of security and reliability, coal fired plants are, ideally, only shut down for scheduled maintenance.[16]

The previous startup of a plant (i.e. in a prior half hour trading interval) constitutes a sunk expenditure over the short run time period of half hourly electricity supply. That is, once a plant is started, the associated expenditure is committed and can no longer be avoided. Conversely, the costs of operating a plant at mingen or above are not sunk because they can be avoided by shutting the plant down for the relevant half hour trading interval. Note, however, that the shutdown cost can be either greater than or less than the cost of operating at mingen, depending on the particular technology involved. Where the cost of operating a plant at mingen is substantially lower than the cost of shutting the plant down, a shutdown decision over a temporary period of low demand may prove uneconomic. In any case, the standard opportunity cost criterion applies: the value of resources in their alternative use should be fully accounted for. For example, the decision as to whether the mid-merit plant in Figure 3.1b should be shutdown prior to trading interval 7, or alternatively, left operating at mingen, would have to account for the avoided cost of variable base-load (that is, the cost of base-load output the mid-merit plant will offset if it is not shutdown).

3.5 Time averaging of cost estimations

Technological (metering) constraints typically require electricity output to be measured in units of *power*. Conversely, market measures of electricity supply require a comparison to be made between price and *energy*. As such, a suitable method must be adopted to bridge between inter-interval representations such as Figure 3.1a and Figure 3.1b (which give electricity output in units of power on their vertical axes against time on the horizontal axis) to intra-interval representations such as Figure 2.1 (which gives electricity output in units of energy on the horizontal axis against monetary units on the vertical axis).

Because electricity supply systems require output to be continuously balanced against load, it is extremely unlikely, given the volatility of electricity demand, that the system's output will be constant with time, even over small trading intervals. As a result some degree of time-averaging will be required to convert dynamic electricity supply data (i.e. power with respect to time over a trading day), to static, economic, representations of electricity supply (i.e. energy with respect to price over a trading interval). This is likely to produce a level of inaccuracy in cost estimations whenever mid-merit plants and, in particular, peaking plants, which have the ability to ramp output up or down relatively

[15] To avoid costly damage to steam turbines associated with expansion and contraction, venting steam at low demand is not an option for most base-load plants. If this were technologically feasible, it would be cheaper to operate a coal fired base-load plant at mingen and vent steam during periods of low demand for electricity rather than shut it down.

[16] The costs alluded to here can be avoided if a firm commits itself to a low (possibly negative) bidding strategy that ensures dispatch. This can be thought of as a payment to the market that ensures against another producer dispatching energy into the system. This discourages alternative sources of supply enough to enable the base-load plant to be kept above mingen. A profit maximising firm would be willing to pay the market up to the opportunity cost of shutting the plant down not to supply electricity.

quickly to follow load, are required to be operated to a high degree of intra-interval variability in output.

A related constraint arises because of the relative inability of some power stations to ramp output up or down quickly enough to respond to rapid fluctuations in electricity demand. As already mentioned in Sections 3.3 and 3.4 above, base-load plants are designed to operate continuously, typically at their rated capacity, and are only shut down for planned or unplanned maintenance. Moreover, while some base-load power stations are able to be run at reduced output, others cannot ramp output up or down at all, and those that can vary output will do so at a much slower rate than mid-merit and peaking plants. Therefore, there is likely to be a limit to a base-load plant's feasible variability in output from one trading interval to the next. In this way, optimisation of plant and equipment over the trading day may impose limits to the range of output that some plants can offer for particular trading intervals.

Finally, it should be noted that demand cannot be perfectly predicted a day ahead, and the optimal allocation of generation resources is required to be reactive to market conditions in real time. Therefore, plant availability, and the range of output deemed to be available from particular plants, may need to be adjusted to reflect response time considerations.

4 Short run economic costs for individual generating facilities over a half hour trading interval

This section will show how the concepts of short run total cost and its SRMC derivative apply to individual generating facilities within an economic price-quantity framework. Two broad categories of plant are considered: those operating with sunk costs and those for which operations comprise an avoidable fixed cost component. Geometric interpretations are supported by mathematical models (see Appendix 1). Cost information at the plant level becomes the input for an optimisation process (outlined in Section 7 below) enabling the construction of a portfolio SRMC curve, which in turn becomes the basis for the portfolio supply curve as required by clause 6.6.3 of the market rules.

At least three general categories of plant-specific short run total cost curve have been postulated, as an *a priori* set of hypotheses, for statistical based economic analysis : those based on cubic functions, those based on quadratic functions, and those based on linearly increasing functions.[17] SRMC is the rate of change (i.e. the first derivative) of the short run total cost function. This suggests, as a starting point, three corresponding categories of plant level SRMC curve to consider: quadratic functions, linearly increasing functions, and linear constants. In other words, in a cost-output relationship, SRMC curves can be considered either U shaped, increasing with output, or flat. The immediate discussion (below) compares each of these postulated functions side by side. Appendix 1 applies the optimisation process with reference to a hypothetical portfolio of plants, each with U shaped SRMC functions.

It should be understood that the exact nature of SRMC functions can only be determined by observation. Evidence collected by the author pertaining to Western Australian and Californian plants suggests typically U shaped plant-specific SRMC functions.[18] Therefore, pending statistical analysis of data pertaining to Western Australian generating facilities, each of the three categories of function will be considered in the following discussion.

4.1 Short run total cost for a plant with sunk startup costs and small shutdown costs occurring below mingen

By definition, a plant can only face shutdown costs if it is operating during the trading interval with the cost of prior startup considered sunk. Figure 4.1 depicts three possible versions of the short run total cost curve over a half hour trading interval for a single

[17] Johnston, J. (1952), 'Statistical Cost Functions in Electricity Supply', *Oxford Economic Papers*, New Series, Vol. 4, No.1, pp. 68-105. This study considers the cost-quantity relationship of steam turbine plants in the UK in both short run and long run analyses. It finds statistical evidence for a reasonably constant short run marginal cost function. Note, however, that Johnston defines the short run purely in terms of a fixed capital stock, as opposed to the predefined bounds of a half hour trading interval, and relies solely on aggregated output and total cost data as opposed to technologically detailed modelling of electricity supply.

[18] A U shaped SRMC curve is consistent with the law of diminishing returns and is therefore one that is commonly proposed in economic theory. Moreover, U shaped SRMC functions are consistent with the engineering studies of incremental heat rate for fossil fuel plants, thus incorporating a major physical/technological characteristic relevant to electricity generation cost in the short run. See, for example, Schweppe, F., Caramanis, M., Tabors, R. & Bohn, R. (1988), *Spot Pricing of Electricity*, Kluwer Academic Publishers, Boston, pp. 283-285.

generation plant with sunk startup costs and small shutdown costs relative to the total cost of operating the plant at mingen. The important features of Figure 4.1 are:

1) The curves are non-declining with output.
2) The curves get very steep at high levels of output as the thermodynamic limits of the capital in place are approached at K.[19] This will occur to the right of some manufacturer recommended output rating for the plant (denoted R).
3) The tangents of the short run total cost curves represent their rate of change. The slope of each tangent represents the SRMC of the plant at that level of output.
4) The steeper the slope of a tangent line, the higher the rate of change and the higher the SRMC.
5) Below mingen output inevitably falls to zero thereby incurring a shutdown cost. In Figure 4.1 this effect occurs between zero output and output M. The slope of the tangent over this section will be zero.[20]

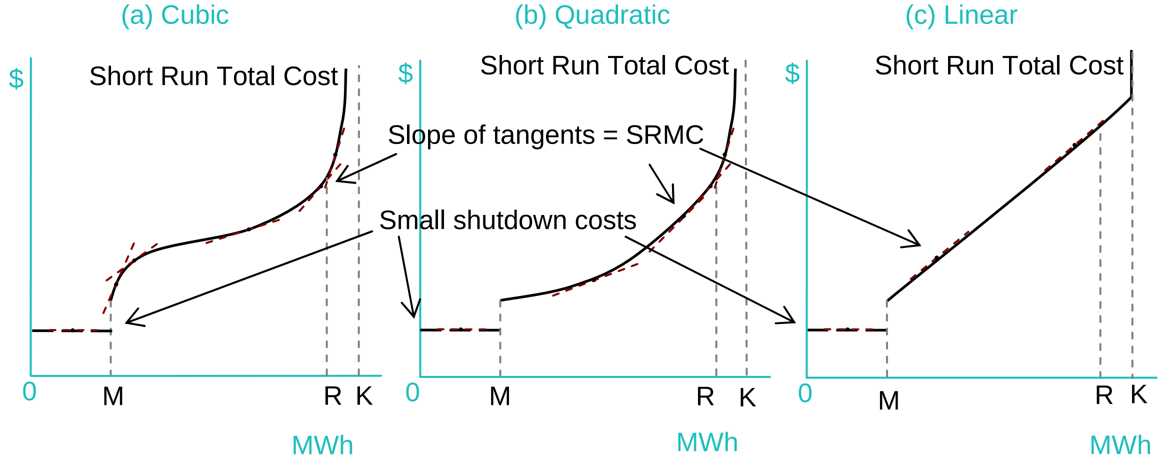


Figure 4.1 – Alternative short run total cost curves for a plant with sunk startup costs and small shutdown costs occurring below mingen

4.2 SRMC for an individual plant with sunk startup costs and small shutdown costs occurring below mingen

Once the short run total cost curve for a generating facility is determined for a half hourly trading interval, it is then a matter of finding the rate of change of that curve to determine the SRMC of the plant for the trading interval. This was shown in Figure 4.1 as the slope of the tangent of the short run total cost curve. Figure 4.2 converts this detail into SRMC curves, with diagrams (a), (b) and (c) derived from the correspondingly labelled diagrams in Figure 4.1.

19 The economic cost of operating at output levels close to a plant's thermodynamic limit can be thought of in a probabilistic sense: as the plant is pushed outside its normal range of operation (e.g. beyond R in Figure 4.1) the likelihood of very costly failure of equipment becomes increasingly significant, thus resulting in the exponential increase in the expected costs of generating the relevant electricity (see Section 5). There will, therefore, be some upper bound to short run total cost that represents the worst possible failure of equipment. This cost will include any lost revenues that result from plant outages.

20 The plant cannot operate in the range of output below mingen, as indicated by the dashed curve.

1) The SRMC curves do not exist up to the mingen quantity of output (M in Figure 4.2).

2) There is a sharp spike in marginal cost when output is reduced below mingen (output M in Figure 4.2).[21]

3) The SRMC curves can be increasing or decreasing over a range of output, but eventually get very steep with increasing output as thermodynamic limits are approached at K (and beyond the manufacturer's recommended output rating at R).

4) Because the short run total costs in Figure 4.1 are non-declining, the slope of the total cost curve must also be non-declining. Therefore, SRMC in Figure 4.2 must be of non-negative value.

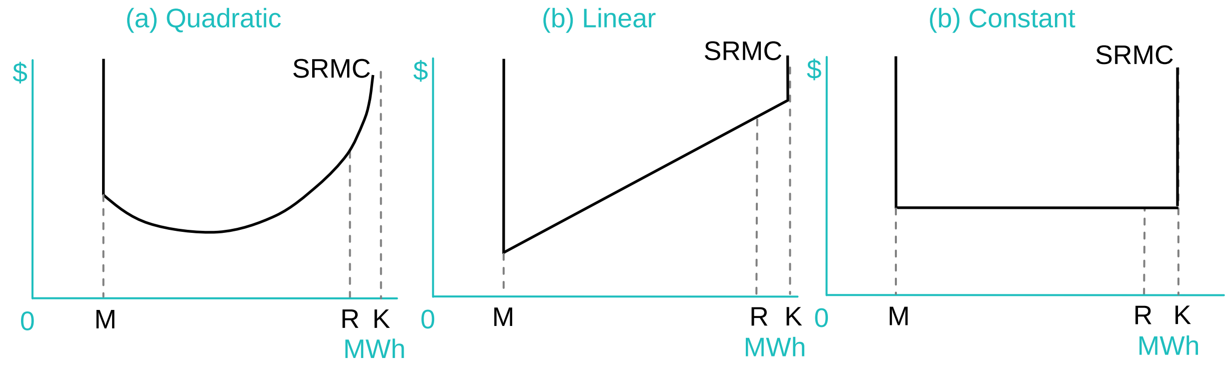


Figure 4.2 – Alternative SRMC curves for a plant with sunk startup costs and small shutdown costs occurring below mingen

4.3 Short run total cost for a plant with sunk startup costs and large shutdown costs occurring below mingen

Three theoretical versions of the short run total cost curve for a plant with substantial shutdown costs relative to the total cost of operating the plant at mingen is depicted in Figure 4.3. If the facility is forced below mingen during the relevant half hour trading interval it will be forced to shutdown incurring large costs. Beyond mingen, the plant faces normal variable costs. The important features of Figure 4.3 are:

1) Because production is not possible below mingen quantity M shutdown costs will occur if production falls below this level of output.

2) Greater than mingen quantity M the curves are non-declining, eventually becoming very steep at high levels of output as the thermodynamic limits of the capital in place are approached at K (and beyond the manufacturer's rating for the plant at R).

[21] The first derivative of a vertical increase or decrease in total cost is mathematically undefined. However, using a difference equation will give the spike in total cost a SRMC value. For example, with ΔC = 1MWh, where ΔC denotes the change in total cost, SRMC (at just below mingen in Figure 4.1) will equal the difference between the cost of shutting the plant down and the total cost of operating the plant at mingen. If the average SRMC up to mingen is then taken, the value over this range will be equal to $\Delta C / \Delta Q$, where ΔQ denotes the change in quantity between zero output and mingen output.

3) The tangents of the short run total cost curves represent their rate of change and therefore their SRMC.

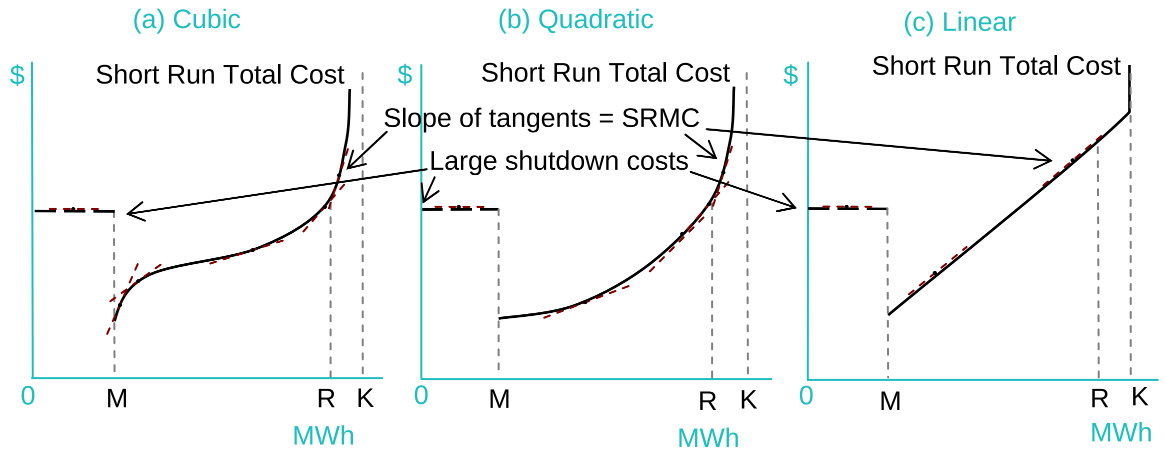


Figure 4.3 – Alternative short run total cost curves for a plant with large shutdown costs occurring below mingen

4.4 SRMC for a plant with sunk startup costs and large shutdown costs occurring below mingen

As always, the SRMC of a plant is determined by finding the slope of the tangents to the short run total cost curve for the trading interval. Diagrams (a), (b) and (c) in Figure 4.4 are derived by this process from the correspondingly labelled diagrams in Figure 4.3. However, because the short run total cost curves are vertical at mingen, SRMC as a first derivative will be undefined at this quantity: i.e. a vertical line is infinitely steep.[21] This can be thought of as the narrow downward spike in SRMC as illustrated in each of the diagrams in Figure 4.4. Because this occurs at negative prices, it will only intercept a downward sloping demand curve at output below M. The important features of Figure 4.4 are:

1) Production is not possible up to mingen output M.

2) Just below mingen output M, SRMC is of negative value due to the large shutdown cost and is highly inelastic.

3) Beyond mingen output M, a SRMC curve can be increasing or decreasing with output but eventually gets very steep as thermodynamic limits are approached at K (in particular, beyond the manufacturer's rating for the plant at R).

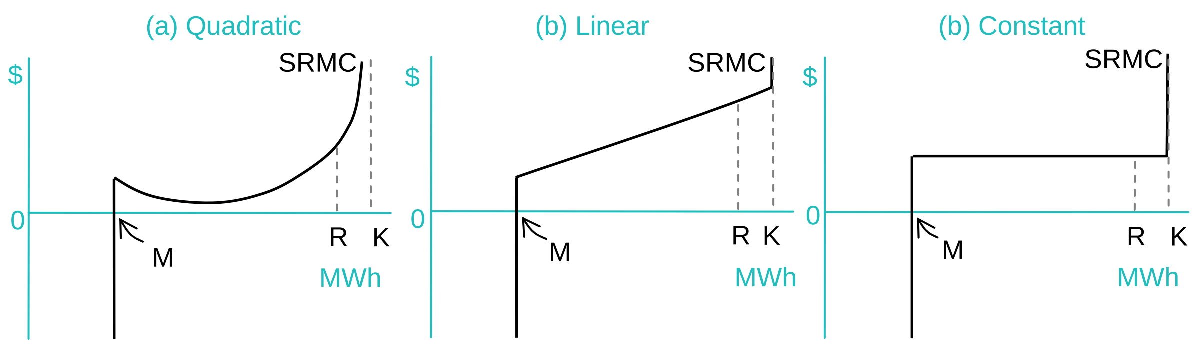


Figure 4.4 – Alternative SRMC curves for a plant with sunk startup costs and large shutdown costs occurring below mingen

4.5 Short run total cost for a plant with avoidable fixed costs

The theoretical short run total cost curve for a plant with some level of avoidable fixed cost is depicted in Figure 4.5. As stated above, in the case of electricity generation, the avoidable fixed cost component for a plant not already operating prior to the trading interval is essentially a startup cost. Here, particular attention should be paid to avoiding any double counting of variable costs. For example, only that amount of fuel expended in startup which is additional to variable fuel use should be counted in the startup costs. The important features of Figure 4.5 are:

1) If the plant is to produce output in the half hour, it must produce beyond the mingen level M.
2) The curves are non-declining with output.
3) Beyond the manufacturer's rating for the plant at output R, the curve gets very steep as the thermodynamic limits of the capital in place are approached at output K.
4) The short run total cost curves comprise variable costs plus the avoidable fixed costs of startup.[22]
5) The tangents of the short run total cost curves represent their rate of change. The slope of each tangent represents the SRMC of the plant at that level of output.
6) The higher the slope of a tangent line, the higher the rate of change and the higher the SRMC.

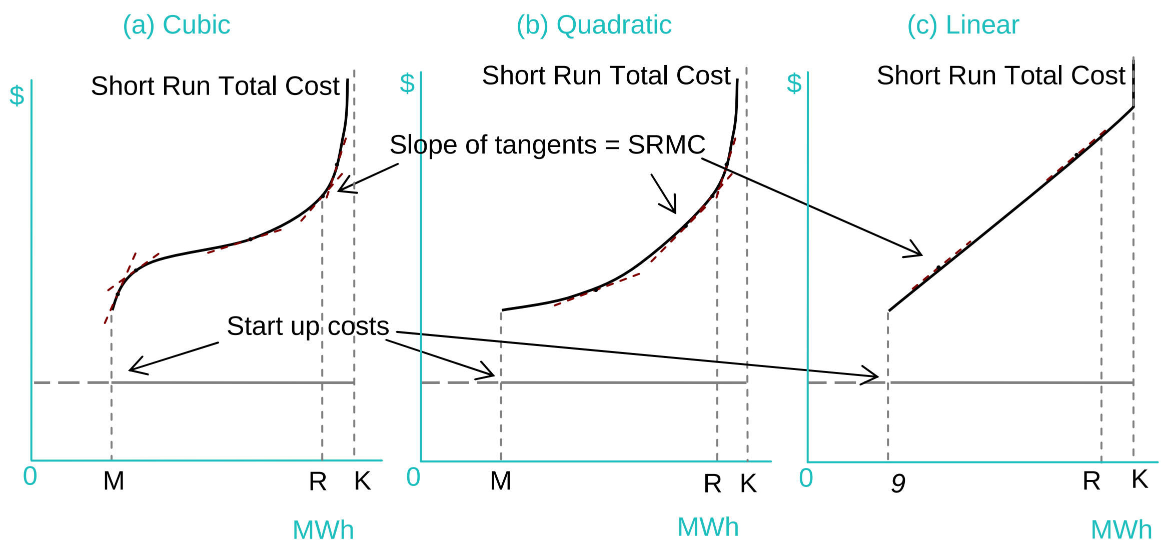


Figure 4.5 – Alternative short run total cost curves for a plant with avoidable fixed costs

4.6 SRMC for a plant with avoidable fixed costs

The derivation of the SRMC curves for a plant requiring startup in a half hour trading interval is depicted in Figure 4.6. The important features of Figure 4.6 are:

[22] Note: shutdown costs should also be added to short run total costs in the event where a peaking plant is required to both startup and shutdown within the same trading interval.

1) The SRMC curves can be increasing or decreasing over a range of output, but they eventually get very steep with increasing output beyond the manufacturer's rating R as thermodynamic limits are approached at K.

2) Because short run total and short run variable costs are non-declining, the slope of the short run total cost curve must also be non-declining. Therefore, SRMC must be of non-negative value.

3) The very first unit of output, produced at mingen (denoted M in Figure 4.6), incurs large fixed costs as well as variable costs thus producing a sharp spike in marginal cost.

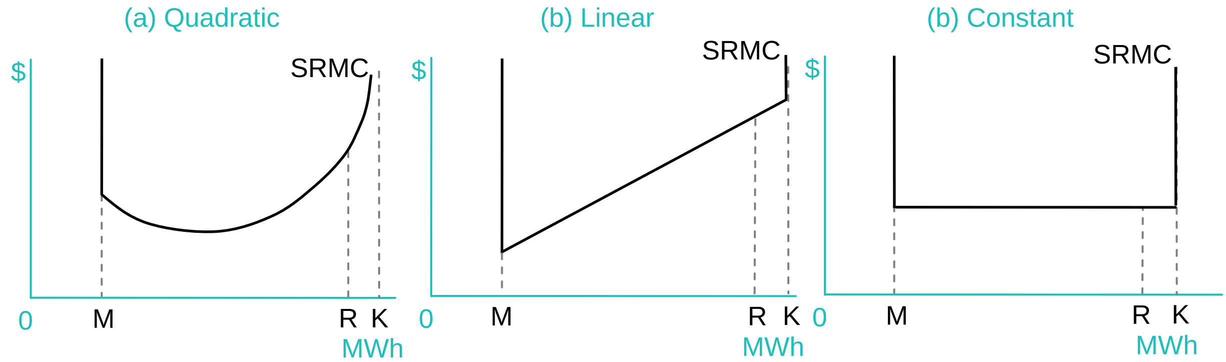


Figure 4.6 – Alternative SRMC curves for a plant with avoidable fixed costs

4.7 Short run total cost for a plant with contracted commitments to output

In some cases a firm may be pre-committed to a contracted level of output over a half hour interval from a certain plant. The case of an electricity generator that has contractual commitments to provide steam to nearby industries provides an example. The breach of such a commitment will incur costs, one form of which may be a financial penalty (stipulated in the contractual arrangements) that is levied per unit of output dishonoured. This would result in a cost situation which can be represented by a generating facility specific, downward sloping, penalty curve.

A theoretical short run total cost curve for a plant with a contracted commitment to output is depicted in Figure 4.7a as the vertical summation of the plant's normal short run total operating cost curve with the penalty curve that applies for not meeting contracted output. As stated above, this can result in a downward sloping cost-quantity relationship. Figure 4.7b is similar to Figure 4.7a, but also comprises an avoidable fixed cost component for plant startup. In both figures, output S represents the contractual commitment to output for the half hour interval. Note that, while the diagrams apply to the case of a plant with a U shaped SRMC curve, the vertical summation method can be similarly applied to other functional forms. The important features of Figure 4.7a and Figure 4.7b are:

1) The short run total cost curve is equal to the total short run operating costs plus any short run cost of penalties for not meeting output S.

2) Up to quantity S the short run total cost curve decreases with output, beyond quantity S the short run total cost curve increases with output.

3) The curve gets very steep at high levels of output as the thermodynamic limits of the capital in place are approached at K.

4) The tangents of the short run total cost curve represent its rate of change and therefore its SRMC.

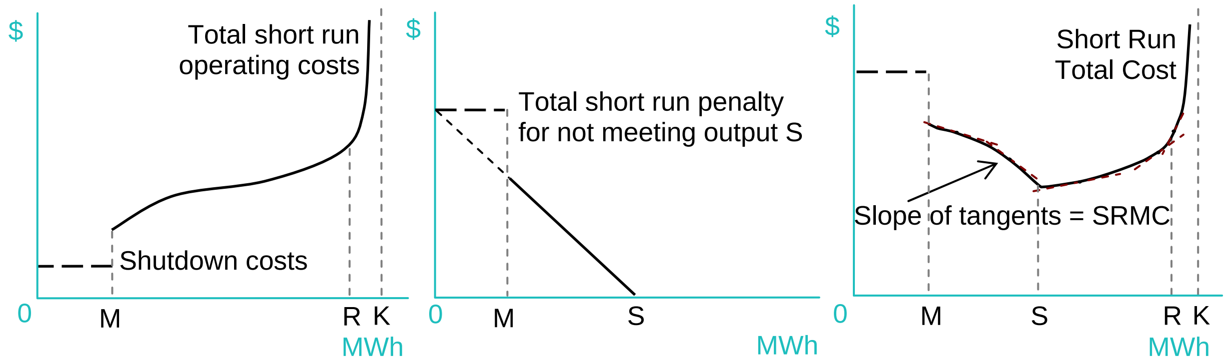


Figure 4.7a – Short run total cost curve based on a cubic function for a plant with sunk startup costs that is subject to penalties for not meeting contracted output

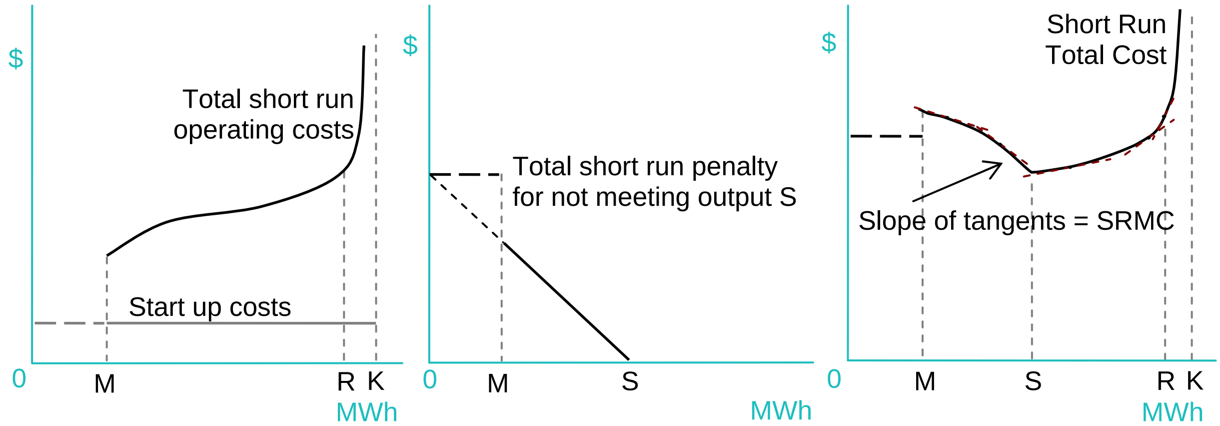


Figure 4.7b – Short run total cost curve (based on a cubic function) for a plant with avoidable fixed costs that is subject to penalties for not meeting contracted output

4.8 SRMC for a plant with contracted commitments to output

Again, the SRMC of a plant is determined by finding the rate of change of the short run total cost curve for the trading interval. Because the penalty curve can be downward sloping, the short run total cost curve can also be downward sloping over the portion of output where the penalties apply (i.e. up to quantity S in Figure 4.7a, Figure 4.7b, and in Figure 4.8a and Figure 4.8b). This results in negative SRMC as illustrated in Figure 4.8a and Figure 4.8b, the important features of which are:

1) SRMC can be negative as a result of the financial penalties incurred through not meeting contractual commitments at output S.

2) The SRMC curve can be increasing or decreasing over a range of output, but eventually gets very steep with increasing output as thermodynamic limits are approached at K beyond the manufacturer's rating for the plant at R.

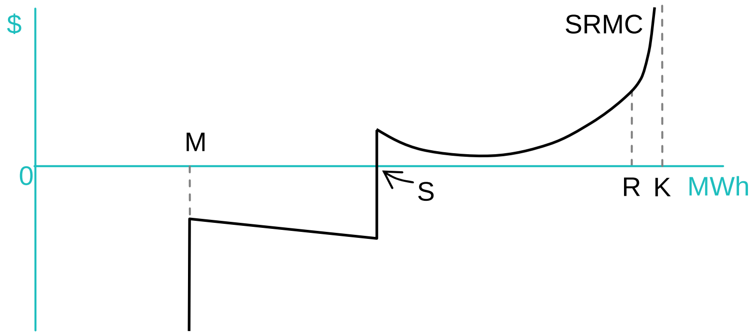


Figure 4.8a –SRMC curve (based on a U shaped function) for a plant with sunk startup costs that is subject to penalties for not meeting contracted output

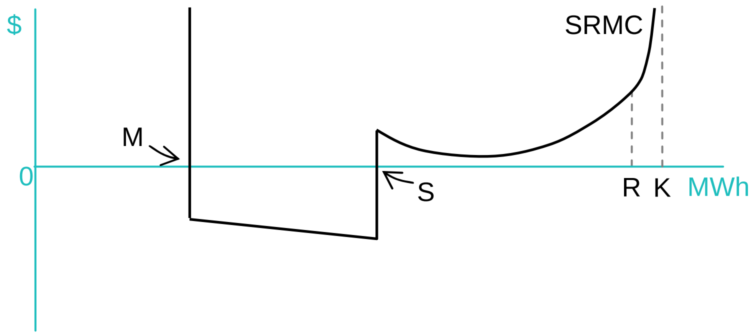


Figure 4.9b –SRMC curve (based on a U shaped function) for a plant with avoidable fixed costs that is subject to penalties for not meeting contracted output

5 Cost measurement and estimation methods

This section turns to the issue of measurement and draws upon engineering based descriptions of electricity supply in economic modelling. Given the short run components of economic cost identified in Section 2 of this paper, the short run total cost curve for an individual plant will comprise of:

a) a variable cost component based upon empirical observation of fuel and operational and maintenance inputs; and

b) if the plant requires startup, an avoidable fixed cost component determined by empirical measurement of plant startup costs; or

c) a shutdown cost component that applies for a plant already started if that plant's output falls below mingen.

Means by which empirical data can be used to estimate these cost components are described below. The goal is to construct a short run total cost curve for each plant in a firm's portfolio. Plant-specific SRMC can be derived as the rate of change of the plant's short run total cost curve.

Below is a list of the minimum types of cost data requiring collection and estimation for each generating facility in the WEM:

1) input-output, average heat rate, or efficiency data;
2) planned and unplanned outage costs;
3) startup costs; and
4) shutdown costs.

In addition to the above, the following data would also improve the estimation of SRMC:

1) Reliability data by plant or plant type with respect to generation load.
2) Half hourly water use by plant with respect to plant output.

Provided an effective metering technology exists, the relationship between an input variable and electricity production can be estimated via regression analysis. Econometric techniques allow a mathematical function to be estimated that not only describes the relationship, but also indicates the degree of confidence that can be attached to the estimation.

5.1 Determination of variable cost components

5.1.1 Fuel input

In electricity supply, fuel requirements dominate all other components of short run cost. Therefore, fuel input and fuel efficiency become the primary focus of empirical analyses. Engineers typically collect input-output data for each plant by measuring the fuel requirement (in GJ/hr) for given levels of electricity output (in MW) that the plant sends out.[23] Alternatively, plant efficiency data can be collected from which the input-output

[23] The Input-output relationship is "…the data that is actually measured in the field", Klein, J. (1998), *The Use of Heat Rate Curves in Production Cost Modeling and Market Modeling*, Electricity Analysis Office,

relationship can be mathematically derived. Both types of data are usually measured at a number of discrete levels of output for each plant. This is sometimes referred to as point data, segment data, or valve best point data.

Table 5.1 provides an example of input-output data for the Potrero 3 gas turbine plant in California, with point 0 corresponding to the origin of the input-output curve, point 1 corresponding to mingen output, points 2 through 4 corresponding to intermediate levels of output in the operational range of the plant, and point 5 corresponding to the plant's rated capacity.[24] Basic algebra enables easy conversion of the input-output relationship to average heat rate and/or efficiency measures. Average heat rate is energy input divided by energy output, whereas efficiency is energy output divided by energy input (i.e. the inverse of average heat rate) given as a percentage after converting to a common unit of measurement.

Table 5.1 – Input-output and heat rate data for the Potrero 3 gas turbine plant

	Output		Input-Output	Av. Heat Rate	Efficiency
	% Rated Cap.	***MW***	***MWh/GJ***	***GJ/MWh***	***%***
Point 0	0	0	*Undefined*	*Undefined*	*Undefined*
Point 1	23	47	543.31	11.56	31.14
Point 2	25	52	590.63	11.36	31.70
Point 3	50	104	1080.81	10.39	34.64
Point 4	80	166	1687.64	10.17	35.41
Point 5	100	207	2135.05	10.31	34.90

Source data: Klein, J. (1998)

An input-output curve can be estimated by fitting a cubic function to the data. This can be done routinely on statistical or spreadsheet packages. Figure 5.1 depicts the fitted input-output curve and the relevant equation for the Potrero 3 gas turbine plant using Excel.

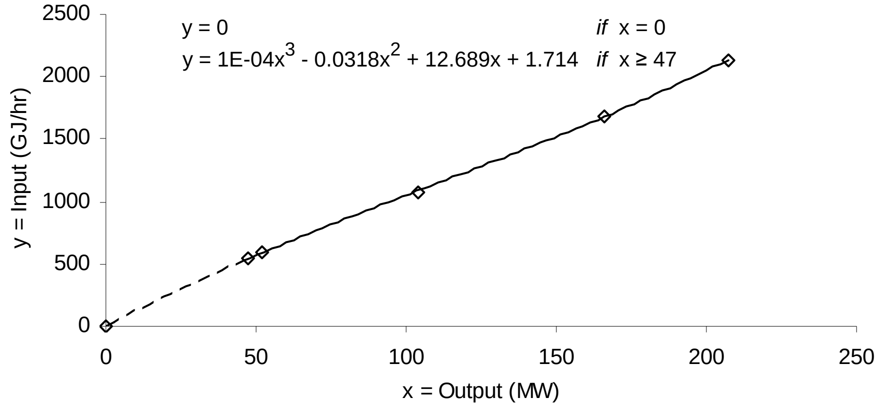


Figure 5.1 – Input-output curve for the Potrero 3 gas turbine

California Energy Commission. See also Schweppe, F., Caramanis, M., Tabors, R. & Bohn, R. (1988), *Spot Pricing of Electricity*, Kluwer Academic Publishers, Boston, p. 283-285.

[24] Data sourced from: Klein, J. (1998), *The Use of Heat Rate Curves in Production Cost Modeling and Market Modeling*, Electricity Analysis Office, California Energy Commission.

If a constant load is assumed over a half hour period, a fitted input-output curve can be converted to a total fuel consumption curve by multiplying both sides of the fitted equation by hours and dividing by two. This is then multiplied by the opportunity cost of fuel to produce a total fuel cost curve for a trading interval. For example, assuming an opportunity cost of fuel of $9.00/GJ results in the half hourly fuel cost curve for the Potrero 3 gas turbine depicted in Figure 5.2.

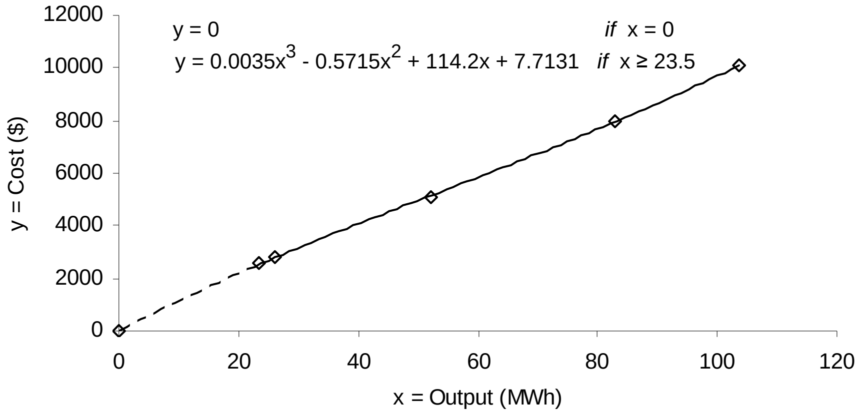
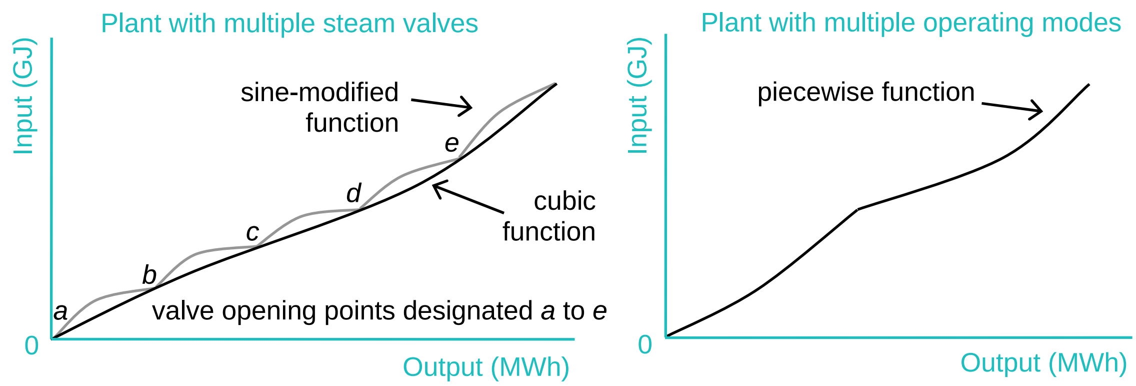


Figure 5.2 - Half hourly fuel cost curve for the Potrero 3 gas turbine

For some plants, the actual relationship between fuel energy input and electrical energy output may not be smooth. For example, in the case of a coal fired plant with multiple steam valves that are opened sequentially, a more exact representation of the input-output representation may be to refine a quadratic or cubic function by a sine function.[25] Similarly, the input-output curve for a plant with multiple operating modes such as a combined cycle plant may be more in the form of a piecewise function rather than a fitted cubic function. Both cases are depicted in Figure 5.3.



Sources: Kim et al, (2002) and Wood, A. & Wollenberg, B. (1996)

Figure 5.3 - Possible input–output curve of plant with multiple steam valves (e.g. large thermal generators) and plant with multiple operating modes (e.g. combined cycle generators).

However, given that the measurement of input-output data for a plant is taken at a limited number of points, it is unlikely that the requisite level of detail would be available to improve the input-output representation. Therefore, in the context of portfolio SRMC

[25] Kim, J. Shin, D. Park. J, & Singh, C. 'Atavistic genetic algorithm for economic dispatch with valve point effect', *Electric Power Systems Research*, 62, 2002, pp 201-207.

estimation, an approximation of these complex curves might be made by use of a simpler function, such as a quadratic function, a cubic function, or quadratic functions over two ranges of output. Alternatively, more detailed input-output data could be sought from the generating firm and the problem solved numerically on a spreadsheet.

5.1.2 Operational and maintenance costs: Planned outages

Where output causes a costly deterioration of equipment, wear and tear can be thought of as a productive input, and thus can be described by an input-output curve similar to that of fuel. Equipment reliability is often the major focus of engineering based studies. However, such data is usually not of a form that can be readily placed into an economic short run price-quantity framework, except by averaging the data. There is, for example, no obvious way by which to conduct a regression analysis of wear and tear against equipment load. Therefore, in the case of normal wear and tear of plant and equipment covered by scheduled maintenance, a practical approach is to divide the cost of a scheduled maintenance service by the total amount of MWh produced by the plant between such services. This averaging can then be adjusted by the time value of money: the further away the scheduled maintenance is in time, the more discounted will be the per MWh cost of wear and tear in the current trading interval. Note that this method assumes a linear relationship between load and wear and tear on plant and equipment.

For example, during the period between scheduled maintenances a plant produces 1,500MWh of electricity. If the total cost of a scheduled maintenance, in net present value terms, is $7,080, the average operational and maintenance cost for the current period would be:

$$\frac{7,080}{1,500} = 4.73,$$

or $4.73 per MWh. Figure 5.4 transposes these operational and maintenance cost assumptions onto the Poterero 3 gas turbine, with the points marked corresponding to half hourly operation at the mingen and rated capacity outputs of the plant.

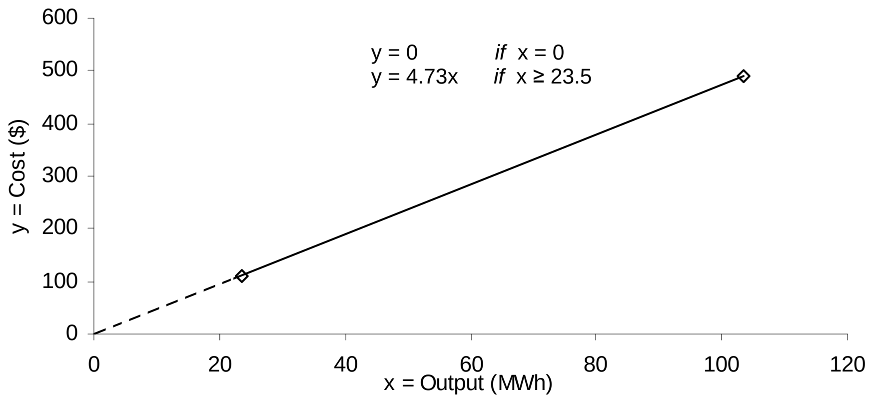


Figure 5.4 – Assumed half hourly operational and maintenance costs for the Potrero 3 gas turbine

5.1.3 Operational and maintenance costs: Unplanned outages

The extent to which a statistical relationship exists between the probability of an unplanned outage event and a plant's load is relevant to the modelling of short run variable cost. Equipment failure leading to an unplanned outage can result in substantial financial loss, both to a firm and to society as a whole. From the point of view of the firm, such costs include not only direct expenditure on unscheduled maintenance, but also any loss of net revenue that would have been generated by the plant if it were still in service, along with any regulatory penalties imposed. Therefore, the average value of an unplanned outage is likely to be very large.

Of particular interest is whether operating near or beyond a plant's rated capacity leads to an increased rate of equipment failure. If evidence can be found of such a relationship, the average cost of an outage event can be multiplied by the probability of it occurring to develop an expected cost function. In other words, correlation between the probability of equipment failure and plant output will see unplanned outage expenses treated as variable costs. Conversely, if a statistical relationship between output and the probability of failure cannot be established, unplanned outage expenses should be added to the short run total cost function as an avoidable fixed cost component.

Figure 5.5 suggests one feasible probability relationship. It seems apparent that, given plants can and do occasionally operate beyond their rated capacity, there must be some reliability or other variable cost related reason why firms tend to avoid operating a plant at this level. Thus, one might postulate a rapidly rising probability of plant failure near and beyond rated capacity.

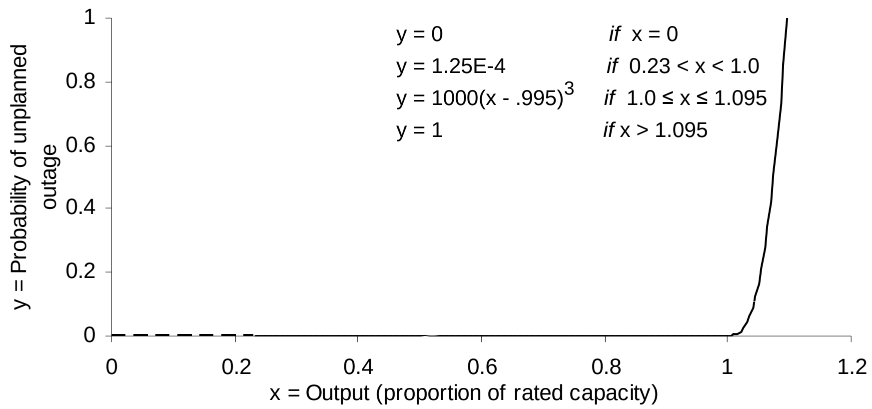


Figure 5.5 – Hypothetical probability of outage for a gas turbine plant

An expected value of failure could be applied to any plant where such a relationship is found. For example, using an assumed cost of $35,000 for an unplanned outage, simple algebra converts the cumulative distribution function of Figure 5.5 to the variable factor cost function of Figure 5.6. Note that in Figure 5.5 the bottom axis refers to the proportion of rated capacity whereas the bottom axis of Figure 5.6 refers to actual output in MWh, after applying the assumptions of Figure 5.5 to the Poterero 3 gas turbine.

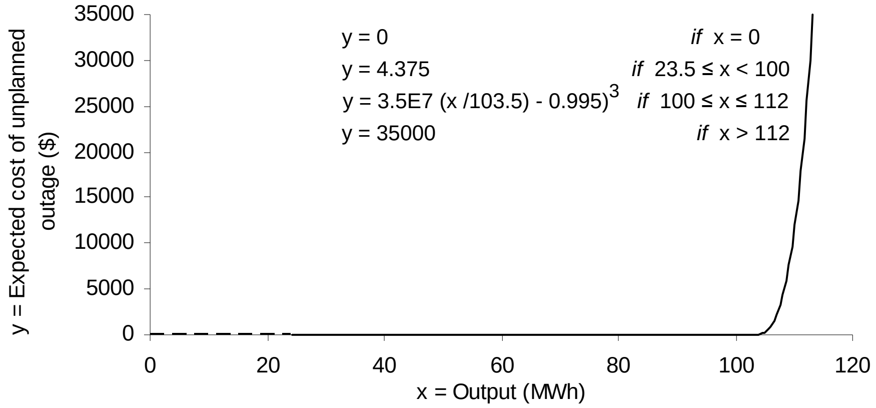


Figure 5.6 – Hypothetical expected cost of unplanned outage applied to the Poterero 3 gas turbine plant

It must be recognised that, in practice, some instances of plant failure, for example those attributable to human error, will be associated more with startup than with high levels of output. Therefore, before the correct relationship can be established, time series and cross sectional data will need to be sought indicating a plant's load immediately prior to an unplanned outage event. Depending on the number of unplanned outages per year, it may take a considerable period of time to collect enough observations for a statistical model to be developed. Any relationship is likely to be first detected in aggregated, market-wide data of plant load as a percentage of rated capacity. This would essentially develop an average value that could be applied to all the plants in a firm's portfolio.

As such, until a correlation with output can be demonstrated, it is proposed that costs relating to unplanned outages be treated as correlated with startup. In other words, until an output based cumulative distribution function is developed, expected costs associated with the risk of plant failure should be treated as a fixed cost component as opposed to a variable cost component. That is, the expected cost of an unplanned outage should be included as part of startup costs. This would entail the averaging of plant failure cost across total startups per plant over a time period. In this case, however, it would not be appropriate to include regulator imposed financial penalties in the determination of expected cost, because these would then translate to higher STEM prices, a result which is contrary to the intention of the penalties: i.e. the enhancement of risk avoidance incentives.

5.1.4 Water input

Water input into fossil fuel based electricity generation is, generally, not considered a significant cost component and therefore can be ignored without much impact on SRMC calculations.[26] However, where required, a similar framework to fuel input or operational and maintenance input can be used to determine the relationship between volumetric water consumption and electricity output (by plant and half hour trading interval). Clearly, water metering is a well established technology and therefore, as with fuel input data, one would expect water flow data to be collected by a firm. This, however, may not be accurate to the half hour, in which case averaging will need to be conducted with reference to longer timeframes.

[26] See, for example, figures given in: National Energy Technology Laboratory, *Power Plant Water Usage and Loss Study*, August 2005.

Once the required amount of water input to produce a given level of electricity output is known, that amount can then be multiplied by the current price of water (i.e. its opportunity cost) to determine the relationship between the factor cost and electricity output. Studies on water use by electricity plants often report the results as an average value in gallons or litres per MWh. This constrains estimation to a linear function and thus may conceal an economically important non-linear input-output relationship. Therefore, it would be better to collect actual field data from generating firms, rather than to rely on aggregated data.

5.1.5 Other variable costs

As regulation in the area of pollution control permits, renewable energy certificates and carbon taxes continues to evolve, the costs associated with such instruments are likely to increase in significance. Where appropriate, these should be included in economic cost calculations. However, from the regulator's point of view, it should be noted that any penalties, actual or expected, associated with not meeting regulatory obligations should not be allowed to be passed on to consumers. That is, regulatory penalties should not be included in legislated SRMC calculations.

5.2 Determination of non-variable cost components

5.2.1 Startup costs

Startup costs include the opportunity costs of fuel, water, internal power, additional labour and lost asset value directly attributable to the startup, including the expected costs associated with the risk of plant failure (if it is not deemed to be correlated with output). Startup is defined as the period from zero output to mingen. Therefore, to avoid double counting, the variable cost component between zero output and mingen output for fuel, water, and wear and tear on plant and equipment is not included in the startup cost calculation. In other words, only those costs additional to those associated with variable output should be included in the startup cost calculation. This is because the short run total factor cost of fuel, water and operation and maintenance already include such costs, as is evident in the relevant diagrams in Section 4 of this paper.

Startup costs are placed in three categories depending upon the period elapsed since the plant's previous shutdown: hot starts, warm starts, and cold starts. The decision as to which of these should be used in short run total cost calculations will reflect the firm's unit commitment plan, thus relating to the 'reasonable expectations' allowance in clause 6.6.3 of the Market Rules.

Startup costs are described by a constant function, as in the hypothetical example of Figure 5.7 (assumed for the Poterero 3 gas turbine), where the marked points correspond to the plant's mingen and rated capacity output levels.

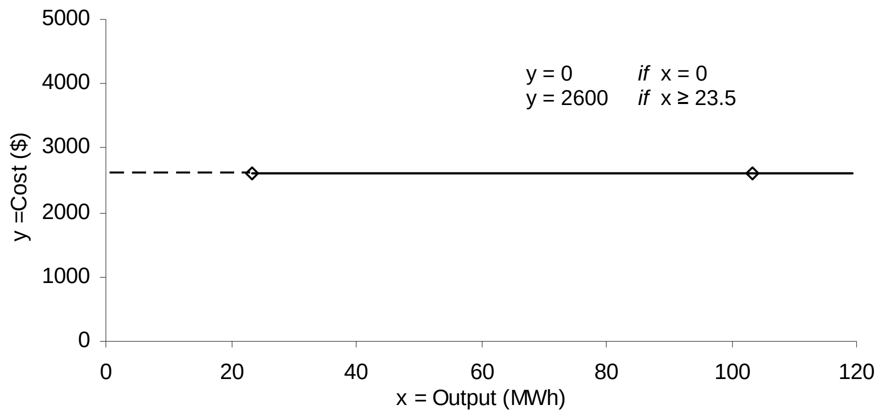


Figure 5.7 – Assumed startup costs for the Poterero 3 gas turbine

5.2.2 Shutdown costs

For most plants, the direct costs of shutdown are comprised of fuel and internal power components and are not significant. However, the decision to shutdown and inevitably start the plant up again must be weighed against the alternative cost of operating a plant at mingen over the intervening period (minus any avoided cost of variable base-load generation). Shutdown costs are likely to be significant in the case of coal fired power stations. They are avoided only if the plant's output remains at or above mingen. As with startup costs, only those costs additional to those associated with variable output should be included in the shutdown cost calculation. To prevent double counting, the variable costs associated with the output a plant produces as its generation is gradually allowed to slow down to zero are not added to shutdown costs.[27] Figure 5.8 provides a hypothetical example of shutdown costs assumed for the Poterero 3 gas turbine, where the marked point corresponds to zero output.

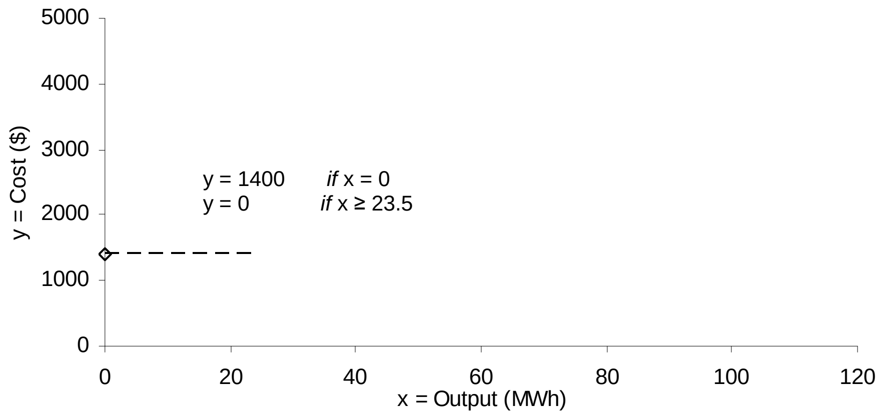


Figure 5.8 – Assumed shutdown costs for the Poterero 3 gas turbine

[27] Where a facility such as a base-load plant (which typically operates as a facility with sunk startup costs in every trading interval) needs to be shutdown periodically for maintenance, the cost of its shutdown and startup should not be considered a non-variable cost. These are costs that, while paid in a lumpy fashion, accrue due to wear and tear on plant and equipment and are therefore variable costs. In this case it is correct to average the cost over the intervening period and then adjust this average for the time value of money (see Section 5.3.2 above).

5.3 Derivation of a plant's short run total cost curve

5.3.1 Plants with sunk startup costs

For a plant already started prior to the half hour trading interval, the short run total cost curve is calculated by vertical summation of all the short run total factor costs (i.e. variable fuel costs, variable water costs, and variable operational and maintenance costs) in excess of mingen, plus the avoidable non-variable cost of shutdown. This is accomplished via the following two step procedure:

1. The half hourly fuel cost curve, the half hourly operational and maintenance cost curve, and any other relevant half hourly factor cost curves (e.g. water) are added together to create an interim variable cost curve for the plant.
2. Shutdown costs are then added to the curve.

5.3.1.1 Example based on Poterero 3 gas turbine heat rates

Using the assumptions given in Sections 5.1 and 5.2 for the Poterero 3 gas turbine, the steps for generating the plant's short run total cost curve (with sunk startup costs) are preformed as follows:

Step 1

An interim curve is determined by the vertical summation of Figure 5.2 and Figure 5.4. The result is depicted in Figure 5.9.

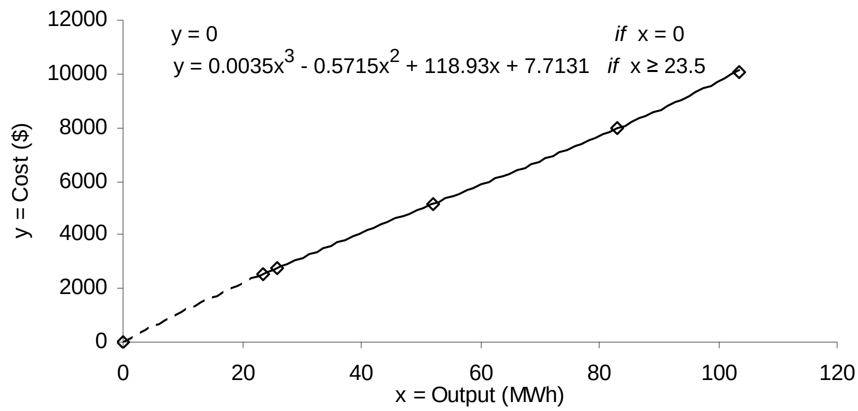


Figure 5.9 – Interim variable cost curve (based on Poterero 3 gas turbine heat rates)

Step 2

The shutdown costs appropriate for that half hour interval (given reasonable expectations of unit commitment) are then added to the curve. Here it is assumed the shutdown cost is $1,400, as given in Figure 5.8. The result, depicted in Figure 5.10, is the plant's short run total cost curve for the half hour.

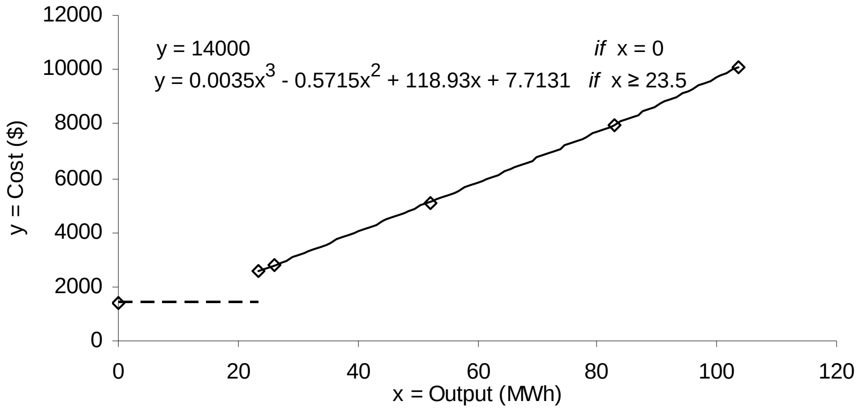


Figure 5.10 – Short run total cost curve (based on Poterero 3 gas turbine heat rates) for a plant with sunk startup costs

5.3.2 Plants requiring startup to produce in a half hour trading interval

For a plant not started prior to the half hour trading interval, the short run total cost curve is calculated by vertical summation of all the short run total factor costs (i.e. variable fuel costs, variable water costs, and variable operational and maintenance costs) in excess of mingen, plus the avoidable fixed cost of startup. This is accomplished via a two step procedure similar to that given in Section 5.3.1 above:

1. The half hourly fuel cost curve, the half hourly operational and maintenance cost curve, and any other relevant half hourly factor cost curves (e.g. water) are added together to create an interim variable cost curve for the plant.
2. Startup costs are then added to the curve.

Example based on Poterero 3 gas turbine heat rates

Using the assumptions given in Sections 5.1 and 5.2 for the Poterero 3 gas turbine, the steps for generating the plant's short run total cost curve (with sunk startup costs) are preformed as follows:

Step 1

This step is identical to that outlined in Section 5.3.1 above.

Step 2

The startup costs appropriate for that half hour interval (given reasonable expectations of unit commitment in relation to a cold, warm or hot start) are then added to the curve. Here it is assumed the startup cost is $2,600, as given in Figure 5.7. The result, depicted in Figure 5.11, is the short run total cost curve.

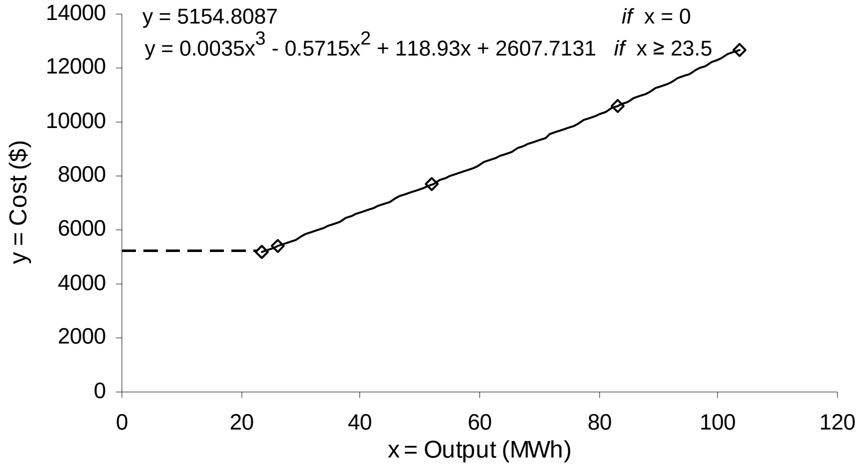


Figure 5.11 – Short run total cost curve (based on Poterero 3 gas turbine heat rates) for a plant requiring startup

5.4 Derivation of a plant's SRMC curve

The optimisation process given in Section 7 and Appendix 1 below, does not require the derivation of plant-specific SRMC. Rather, it optimises plant-specific short run total cost to generate an optimal portfolio short run total cost curve, and then derives the optimal portfolio SRMC from this curve. However, if desired, a plant-specific SRMC can be produced by taking the first derivative of the plant level short run total cost curve. In the case of the hypothetical plant given above in Sections 5.2 and 5.3 (i.e. based on Poterero 3 gas turbine heat rates at a \$9.00/GJ gas price), the derivation of plant-specific SRMC results in the curve depicted in Figure 5.12 below.

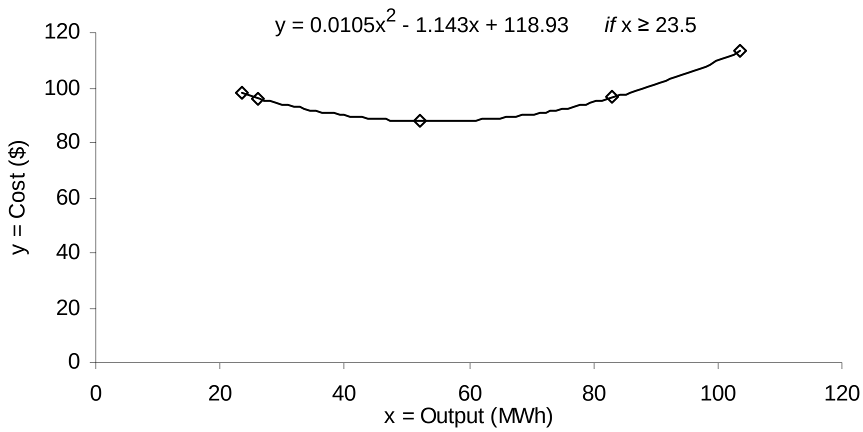


Figure 5.12 - SRMC curve (based on Poterero 3 gas turbine heat rates)

6 Line losses

The spatial location of a plant relative to load and the physical nature and topography of transmission infrastructure will affect the plant's effectiveness in supplying electricity to the market. This is because the transmission of electricity over distance is subject to loss. In practice, a relatively efficient plant that is distant from consumers may be more expensive at meeting demand when compared with a relatively inefficient plant that is located close to consumers.[28] Therefore, the quantity of electricity sent down a transmission line must be adjusted by a loss calculation before the optimisation process described in Section 7 and Appendix 1 below is applied.

6.1 Line losses and the SRMC optimisation problem

In a general sense, there are three main determinants of transmission loss in an electricity supply system. These are outlined below.

1) *Losses are proportional to line resistance*. The line resistance has a constant value in the short run as determined by ambient conditions and the physical nature of a transmission line (a major aspect of this being the distance between generation and load). Therefore, a generating firm must take line resistance as given for a particular half hour trading interval.

2) *Losses are inversely proportional to the square of line voltage*. Line voltage is a parameter determined by system management requirements. While it may be influenced in the short run by actions of generating firms, for all practical purposes each firm must take line voltage as given for a half hour trading interval.

3) *Losses are proportional to the square of the power flow along a line*. Power flow is a variable parameter in the short run and is determined by the magnitude of load during a trading interval as well as the relative location of that load *vis-à-vis* generation. Hence, if the network distribution of load for each trading interval were apparent and predictable, the choice of power flow would be within the complete control of the entity responsible for plant dispatch in a day ahead market.

It can be seen from the above that the only determinant relevant to the short run economic decision is power flow. This is because, from a generating firm's point of view in a short run market, net power flow is variable at the portfolio level to the extent that the share of portfolio output can be allocated to different plants within the half hour. Therefore, power flow is a lever upon which a generating firm can optimise its resource allocation in a half hourly price-quantity framework.

Given the above, plant optimisation is complicated by the fact that, because increased power flow results in a greater-than-proportional increase in line losses, the total loss from a combination of plants will be greater than the sum of its parts. Figure 6.1 provides a demonstration. Assume that generators A, B, C and D are identical plants with similarly identical short run total cost curves, but that while generators A and B are distant from the 380MW load at the reference node (i.e. bus 2) and therefore subject to loss, generators C

[28] Good overviews of line loss economics are provided by: Stoft, S. (2002), *Power System Economics*, IEEE Press, New Jersey, pp. 415-423, and by; Wood, A. & Wollenberg, B. (1996), *Power Generation, Operation and Control*, 2nd. ed., Wiley-Interscience, New York.

and D are not. Losses along the line are equal to $0.0002P_{\{A,B\}}^2$, where $P_{\{A,B\}}$ is the combined quantity of power sent out by generators A and B.

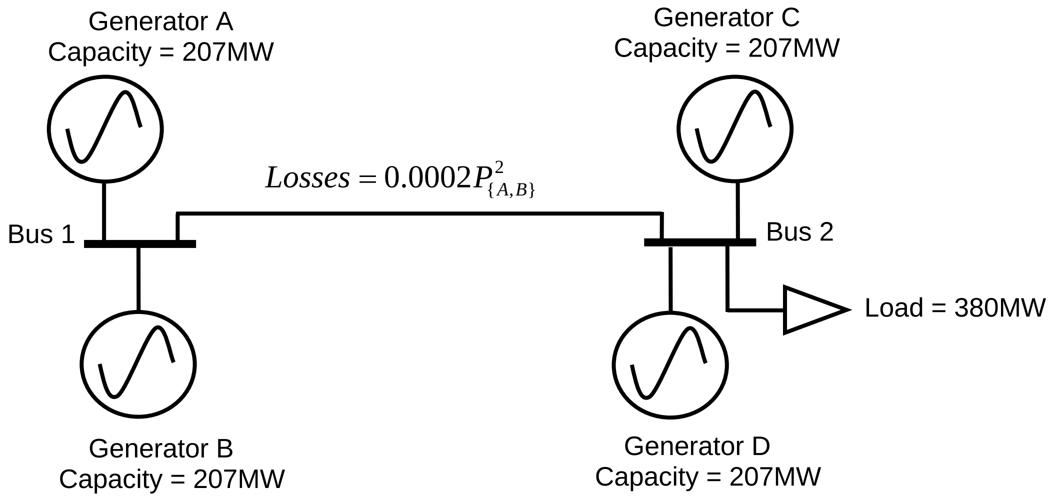


Figure 6.1 – A simple generation and transmission system

Table 6.1 – Two plant optimisation with line losses

	Plant combination to meet 190MWh demand at bus 2 in Figure 6.2		
	Plants C and D	**Plants A and C**	**Plants A and B**
Optimal portfolio SRMC	**$105.53**	**$111.57**	**$136.13**
Optimal plant share of energy arriving at bus 2	*Plant C*: 95MWh *Plant D*: 95MWh	*Plant A*: 88.5MWh *Plant C*: 101.5MWh	*Plant A:* 95MWh *Plant B*: 95MWh
Optimal plant share of energy sent out	*Plant C*: 95MWh *Plant D*: 95MWh	*Plant A*: 92MWh *Plant C*: 101.5MWh	*Plant A:* 103.5MWh *Plant B*: 103.5MWh
Total energy generated	190MWh	193.5MWh	207MWh
Total energy generated at bus 1	0MWh	93MWh	207MWh
Total losses	0MWh	3.5MWh	17MWh

Using the Poterero 3 gas turbine given in Section 5 above as the basis for the four identical plants in the system described by Figure 6.1, gives the results summarised in Table 6.1.[29] Each of the optimal two plant combinations are described after taking line losses into account. By assuming demand is constant over the trading interval, the optimisation problem is able to be converted to the usual price-energy relationship, i.e. where 380MW of power is treated as 190MWh of energy demand for the half hour.

[29] In each case, the plant is modelled as already started prior to the half hour interval – i.e. shutdown costs are included in short run total cost, but startup costs are not.

The starting point is to consider the situation where generators C and D are the only plants available for the trading interval. In this case the 190MWh of demand is met without any line loss. Each plant produces 95MWh of energy to meet the load and the optimisation method can be conducted without modification to the plants' short run total cost curves. If, however, only generators A and C, (or alternatively, any two plant combination comprising a plant located at bus 1 and a plant located at bus 2) are available for the half hour trading interval, line losses will need to be taken into account for the energy sent out at bus 1. Figure 6.2 describes how, in these cases, the correct optimisation method first requires the short run total cost curve for generator A or B to be shifted back by an amount equal to any line losses. This enables the energy arriving at bus 2 to be evaluated in an 'apples with apples' comparison of the plants' total cost curves. Not taking line losses into account will produce erroneous results.

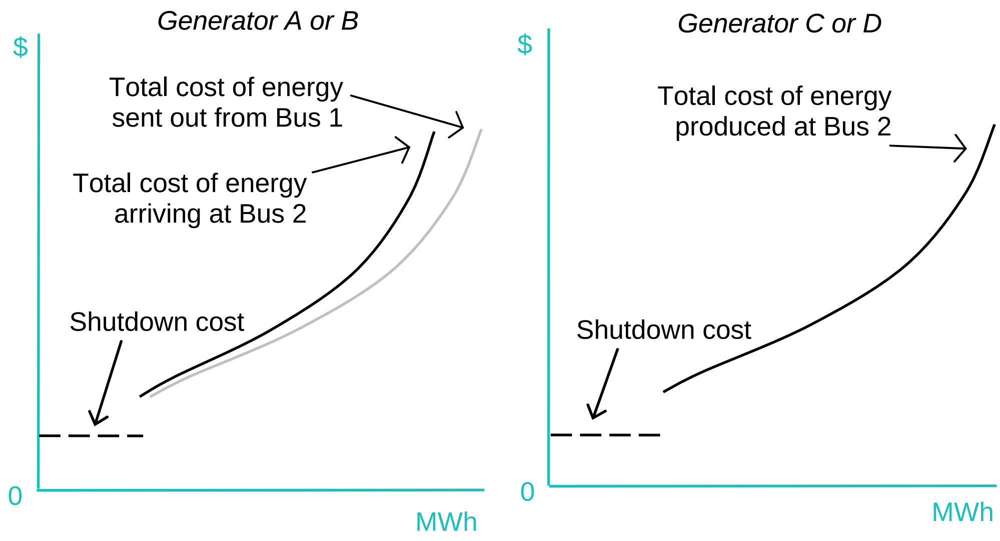


Figure 6.2 – The backward shift in plant level short run total cost to account for line loss.

To elaborate, as Table 1 indicates, taking line losses into account when generators A and C are combined results in an optimal plant share of energy sent out compromising 92MWh from plant A and 101.5MWh from plant C, at an optimal portfolio SRMC of $111.57. However, not taking line losses into account will lead to the same portfolio SRMC estimation as that of combining generators C and D, i.e. $105.53, an error of $6.04 or about 5.7 per cent. Moreover, line losses of 3.5MWh of energy will be unaccounted for in the calculation. Such errors are greatly magnified as more power is sent down the line and losses increase at a quadratic rate. If, for example, only generators A and B (as described in Figure 6.1) are available to meet system demand, Table 6.1 shows that 17MWh of energy will be unaccounted for and the SRMC estimation error will be $30.60 or about 29 per cent.

6.2 Line loss factors in the WEM

It should be clear from the above discussion that an accurate consideration of line losses is critical in determining the optimal short run economic dispatch of plant.[30] However, under the previous market structure of a vertically integrated monopoly there existed no need to allocate the financial responsibility for line losses among competing economic agents. Therefore, methods for determining the financial allocation of line losses are, worldwide, a relatively recent development, having arisen under industry deregulation. All

[30] Similarly, the optimal location of transmission and generation investment in the long run requires an accurate assessment of line losses.

such accounting methods have entailed a difficult trade-off between accuracy and tractability, given the dynamic spatial and temporal irregularity between load and generation, the inability to physically separate one generator's transmission loss from another's when sharing a line, and the nonlinearity of the loss function. A universal approach to tackling the allocation problem is yet to emerge. Given the large amounts of money involved, the particular method chosen for a given market has often invoked debate.

The approach taken in the WEM is to collect yearly load and generation data at each node weighted by half hour and to compress this data into a single annual metric, known as a loss factor, for each node. The loss factor is an annualised static number that attempts to compare the tendency for loss at any given node in the network relative to the system's reference node (defined in the market rules as the Muja 330 bus-bar). A loss factor of 1.00 suggests that, on average, the node in question is equally effective at meeting system demand compared to the reference node. Conversely, a loss factor that differs from 1.00 indicates that the average network distribution of supply and demand causes the node to be, on average, either more or less effective at meeting system demand compared to the reference node. In practice, the meter reading for each generator is multiplied by the loss factor to give an adjusted output, and it is this adjusted output against which payment is made for electricity generation. For example, a loss factor of 0.95 would suggest the average effective output at the node is 5 per cent less than that at the reference node, whereas a loss factor of 1.05 would suggest the average effective output at the node is 5 per cent more than that at the reference node. Loss factors for each node are posted on the IMO's web site, as are the rules pertaining to their determination.[31]

It is clear that the loss factor methodology was not designed for the purposes of half hourly optimisation between generation facilities. Hence, given that the multiplication of a loss factor against nodal output is a linear operation, it must be recognised that its use in a short run optimisation framework will produce a nonlinear error that increases the further away nodal output diverges from its average value. By way of example, assume the line loss factor determined for Bus 1 in Figure 6.1 is deemed to be 0.96 based on an average half hourly output over the year of 100MWh generated at that node. For no error to occur under these circumstances, generation at bus 1 for the half hour would have to be exactly 100MWh. If, however, 190MWh were produced at bus 1 applying the loss factor of 0.96 would suggest losses of 7.6MWh. Yet, as Table 6.1 shows, this figure is incorrect. More power sent down the line results in more than proportional losses: a nonlinear relationship. Actual losses would be 17MWh, 223.7 per cent greater than the amount estimated by use of the linearised loss factor.

Therefore, errors should be expected if the loss factor method is used for SRMC optimisation. However, given the immense complexity of the supply and demand relationship in instantaneous network distribution terms, it is highly unlikely that, in the context of day ahead predictability, line loss estimation error will ever be fully reduced. The dynamics of power flow throughout a power system can change moment by moment with locational shifts in demand that are very difficult to anticipate. As such, given that the requirements of clause 6.6.3 of the market rules pertain to the "reasonable expectations" of SRMC in a day ahead market, the use of line loss factors might be tolerable at present. However, in time, a truly marginal, nonlinear, line loss metric could be developed by cost minimising firms and/or government agencies. It should be noted that errors caused by the use of line loss factors for short run optimisation will tend to be quadratic with output.

[31] Wholesale Electricity Market Procedure (2006), *Market Procedure for Determining Loss Factors,* available IMO website: www.imowa.com.au/Attachments/Loss%20Factor%20Procedure.pdf

Development of a metric that causes such errors to be small and linear with output would be a clear improvement.[32]

[32] The author may explore this possibility at a future date.

7 Portfolio short run cost over a half hour time frame

In the short run, a typical electricity generation firm will have a *portfolio* of plants to utilise in the production of electricity. Each plant in such a portfolio is likely to differ significantly in its cost characteristics, utilising different fuels at different prices and at varying levels of thermal efficiency which itself depends on factors such as plant design, vintage and level of maintenance. Therefore, a firm faces a cost minimising decision directly related to how much of its portfolio output should be shared between each facility. An optimal mix of output from the various plants will minimise the portfolio short run total cost in a half hour trading interval. In other words, to maximise profit over a half hour interval, electricity should be sourced in such a way that each plant is producing a level of output that minimises the portfolio short run total cost of meeting demand. This results in an *optimal portfolio total cost curve*. The rate of change of this curve determines the *optimal portfolio SRMC* curve.

7.1 Short run total cost for a portfolio of plants with sunk startup costs

In considering the construction of the optimal portfolio short run total cost curve, the simplest case is where all available plants have already been started prior to a half-hour trading interval. Under sunk startup cost assumptions, the following sections will describe how the optimal proportion of each plant's contribution to portfolio output will vary as portfolio output varies.

7.1.1 Example 1: Linearly increasing total cost technology, portfolio short run total cost determined by a SRMC order of merit

Figure 7.1 depicts a portfolio of plants that have already been started prior to the relevant half hourly trading interval, each of which will be required to generate electricity in subsequent trading intervals. The portfolio consists of three generating facilities with linearly increasing short run total cost functions: one low cost, one mid cost and one high cost plant.

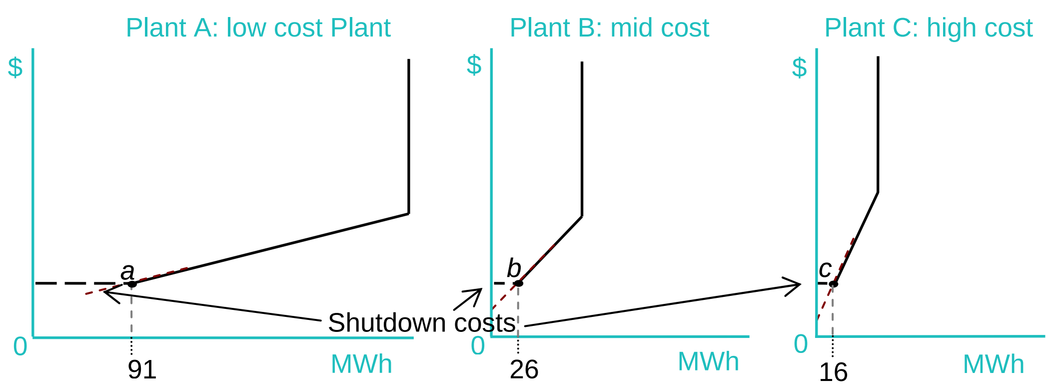


Figure 7.1 – A portfolio of three linear short run total cost plants with sunk startup costs and equalised shutdown and mingen costs between plants

So as to enable a relatively less complicated introduction to the calculation of optimal portfolio short run total cost, shutdown costs in Figure 7.1 are treated as being: (a) equal for each plant, and; (b) equal to the total cost at mingen for each plant.[33] The treatment of differing shutdown cost assumptions and differing total costs at mingen between plants will be considered in Section 7.1.6.

The linear cost assumption at the plant level in Figure 7.1 combined with an assumption of equalised shutdown costs reduces the determination of the optimal portfolio short run total cost curve to a relatively trivial problem. This is because optimisation can be achieved by building an aggregated short-run total cost curve from Figure 7.1 according to a strict, plant-specific SRMC order of merit. In doing so, ensuing quantities of output (moving to the right for each plant) are sourced so as to cause the total cost of the combined electricity generation (i.e. the portfolio short run total cost curve) to rise at the slowest possible rate. In other words, electricity is sourced from the plant with the lowest sloped short run total cost curve right up to its maximum level of output, before moving on to the next lowest SRMC plant. In the example of Figure 7.1 it is assumed that, for any given level of output, the SRMC of plant A is less than the SRMC of plant B, which in turn is less than the SRMC of plant C. Therefore, a cost minimising firm would set a rule that, whenever possible, variable output should be sourced from plant A first, plant B second, and plant C last.

Table 7.1 describes the constraints affecting the optimal sourcing of portfolio output from the three available plants depicted in Figure 7.1 for various ranges of portfolio output, given the aforementioned assumptions. For any level of portfolio output left of the pooled mingen of all three plants (i.e. outputs $91 + 26 + 16 = 133$ from Figure 7.1), the decision must incorporate the shutdown cost of at least one plant. Conversely, if portfolio output is greater than the pooled mingen of all three plants, no plant is required to shutdown, although the option remains available.

The optimal portfolio short run total cost curve derived from this optimisation process is depicted in Figure 7.2. The diagram shows a saw tooth pattern for the curve for a portfolio level of output left of the pooled mingen of all three plants, reflecting differing optimal combinations of variable output and shutdown from each plant as mingen constraints come into play. Right of the pooled level of mingen, energy is sourced incrementally in a SRMC order of merit: first from plant A up to its maximum capacity, then from plant B up to maximum capacity, and finally from plant C up to its maximum capacity.

[33] The total costs for the three plants, A, B and C, were modelled as follows:

$$C_A = \begin{cases} 2{,}000 & if\ Q_A = 0 \\ 20Q_A & if\ 91 \le Q_A \le 340 \end{cases},$$

$$C_B = \begin{cases} 2{,}000 & if\ Q_B < 26 \\ 75Q_B & if\ 26 \le Q_B \le 90 \end{cases}, \text{ and}$$

$$C_C = \begin{cases} 2{,}000 & if\ Q_C = 0 \\ 150Q_C & if\ 16 \le Q_C \le 50 \end{cases},$$

where C_i denotes the total cost of plant *i*, and Q_i denotes the output of plant *i*, $\{i \mid A, B, C\}$.

Table 7.1 – The optimal choice of plant operation and shutdown for the portfolio of plants given in Figure 7.1

Discrete portfolio output range	***Determining factor for bottom of output range***	***Run constraints***	***Optimal decision Shutdown vs. Run***	
MWh			***Plant***	***Decision***
0 - 15	Zero output.	All three plants must shutdown.	**A**	*Shutdown*
			B	*Shutdown*
			C	*Shutdown*
16 - 25	Mingen plant C.	Plants A and B must shutdown.	**A**	*Shutdown*
			B	*Shutdown*
			C	*Run*
26 - 41	Mingen plant B.	Plant A must shutdown and if plant B runs, plant C must shutdown.	**A**	*Shutdown*
			B	*Run*
			C	*Shutdown*
42 - 90	Mingens plants A + B.	Plant A must shutdown.	**A**	*Shutdown*
			B	*Run*
			C	*Run*
91 - 106	Mingen plant A.	If plant A runs, plants B and C must shutdown.	**A**	*Run*
			B	*Shutdown*
			C	*Shutdown*
107 - 116	Mingens plants A + C.	If plant A runs, plant B must shutdown.	**A**	*Run*
			B	*Shutdown*
			C	*Run*
117 - 132	Mingens plants A + B.	If plant A runs, plant C must shutdown.	**A**	*Run*
			B	*Run*
			C	*Shutdown*
133 +	Mingens plants A + B + C (pooled mingen of the three plants).	No run constraints.	**A**	*Run*
			B	*Run*
			C	*Run*

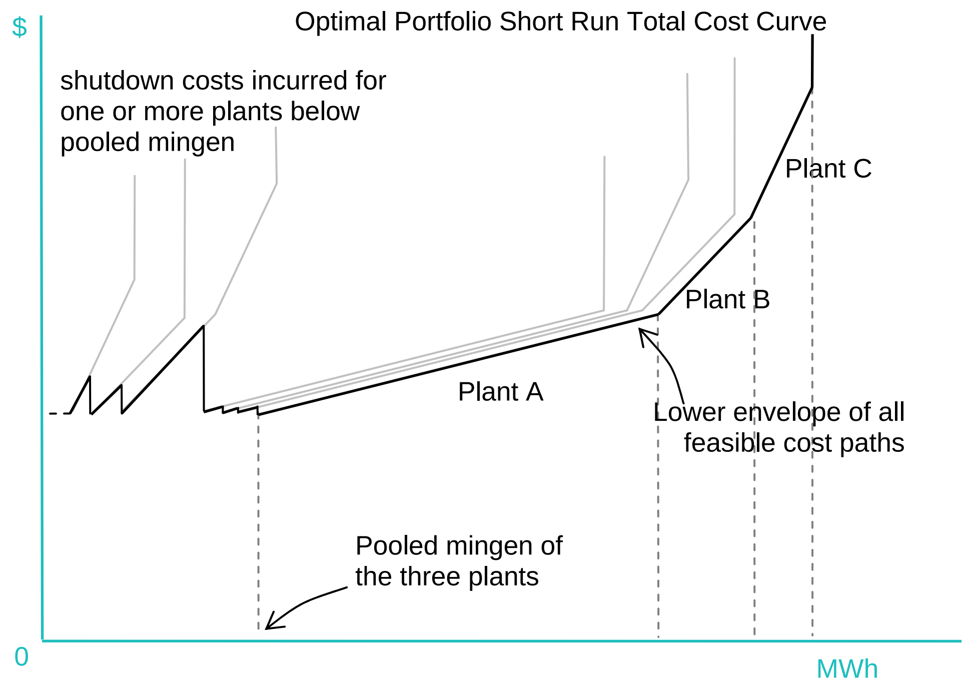


Figure 7.2 – Optimal portfolio short run total cost curve derived by combining linearly increasing total cost technology plants with equalised shutdown and mingen costs

7.1.2 Optimal portfolio SRMC

Once the short run total cost curve for a portfolio of plants is determined for a half hour trading interval, it is then a matter of finding the rate of change of that curve to determine the SRMC of the portfolio for the trading interval. This is defined in the same way for a portfolio of plants as it is for a single plant, i.e. by the slope of the tangents to the short run total cost curve.

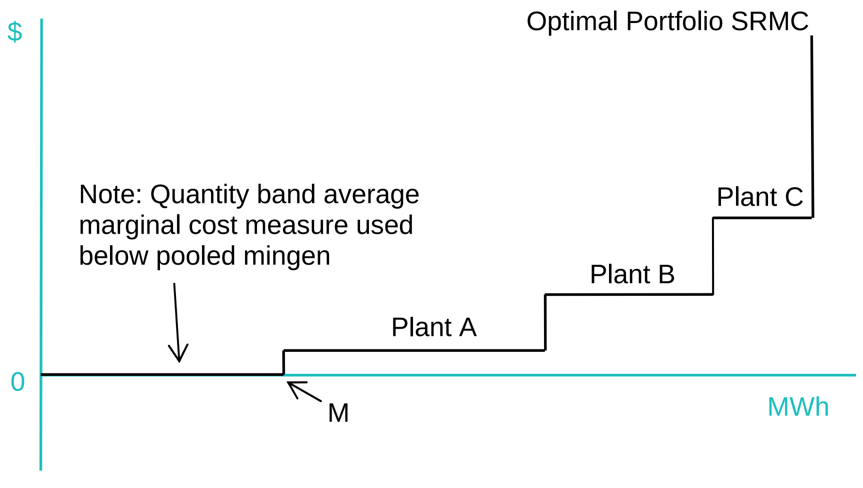


Figure 7.3 – Optimal portfolio SRMC comprised of plants with a linearly increasing total cost technology and equalised shutdown and mingen costs between plants

Figure 7.3 depicts the SRMC curve for the portfolio of plants described in Figure 7.1. Portfolio output M corresponds to the sum of all three plants' mingen levels. The results are intuitive. With each plant exhibiting a flat SRMC curve, the portfolio SRMC curve becomes a step function, with each plant stacked in its SRMC order of merit. Note that, below the pooled mingen level, where shutdown costs cause sudden jumps in the short run total cost curve, SRMC is treated as the change in price over the change in quantity from output zero to output M.[34]

7.1.3 Example 2: Cubic based short run total cost technology, the (sub-optimal) SRMC order of merit approach

As with Section 7.1.1 above, Figure 7.4 depicts a portfolio of plants that have already been started prior to the relevant half hourly trading interval. However, this time the portfolio consists of generating facilities with short run total cost curves based on a cubic functional form. In this case, the method applied in Section 7.1.1 right of the pooled mingen level of output of all three plants (i.e. building an aggregated short-run total cost curve from Figure 7.4 according to a strict SRMC order of merit) produces a suboptimal curve.

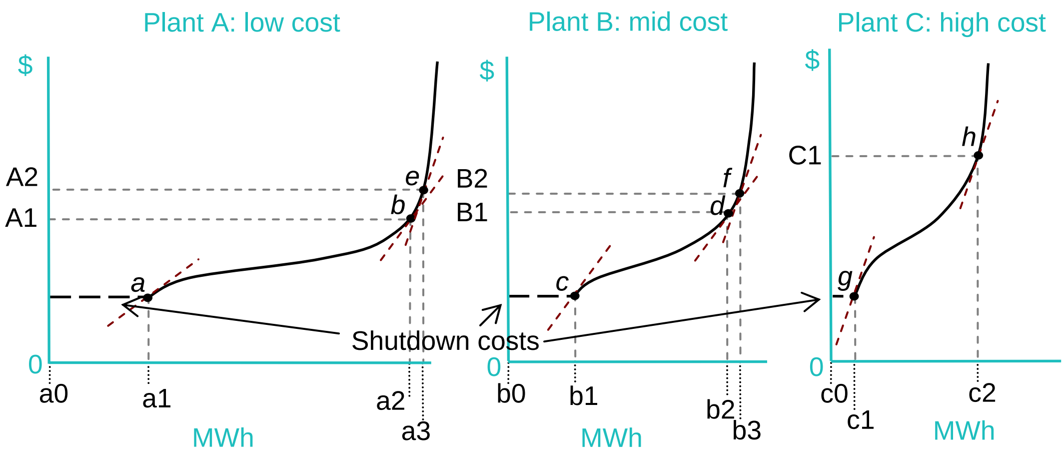


Figure 7.4 – A portfolio of three plants exhibiting a cubic short run total cost technology with sunk startup costs and equalised shutdown costs between plants

To demonstrate: ensuing quantities of output (moving to the right for each plant) are again sourced so as to cause the total cost of the combined electricity generation (i.e. the portfolio short run total cost curve) to rise at the slowest possible rate.[35] For example, consider portfolio output right of the pooled level of mingen for the portfolio of plants in Figure 7.4 (i.e. a1+b1+c1). Because the slope of the tangent at point *a* is less than the slope of the tangent at point *c* or *g*, plant A is first in SRMC order of merit. This is so up until output a2, where the slope of the tangent at point *b* is equal to the slope of the tangent at point *c*. Here the incremental cost of electricity from plant A becomes higher than that from plant B. This sends a signal that fuel should now be burned in plant B in lieu of plant A to meet the increasing load. This remains the case up until output b2 where it is better to increase plant B and plant A fuel input in staggered unison (corresponding to SRMC merit) as the portfolio's output increases. Once outputs b3 and a3 are reached

---

[34] Using a difference equation of $\dot{C} = \dfrac{\Delta C}{\Delta Q}$ with even increments of $\Delta Q = 1$ results in several large and negative sharp spikes in portfolio SRMC left of M in Figure 7.3 and Figure 7.8, and left of *m* in Figure 7.11. However, as a general rule, $\sum_{0}^{M} \dot{C} \Big/ M = \dfrac{\Delta C}{\Delta Q}$ if $\Delta Q = M$. While this corresponds to the discrete estimation of SRMC (see Box 1.1), for large ΔQ, it should actually be regarded as a quantity band average marginal cost measure.

[35] The same simplifying assumptions of Section 7.1.2 remain: startup costs are considered sunk, each plant's mingen levels are the same, and the shutdown costs for each plant are equal to the total cost at mingen.

however, the slopes of total cost for both plant A and plant B become greater than the slope of total cost for plant C at point *g*. It is then cheaper to supply additional quantities of electricity from plant C only. Eventually, when point *h* is reached, the slowest possible rate of total cost increase with respect to quantity is obtained by ramping up output from all three facilities in staggered SRMC merit. In this final stage each plant is nearing its theoretical output capacity and so total costs for the portfolio as a whole are rising very rapidly.

The portfolio short run total cost curve derived from this SRMC order of merit process is depicted in Figure 7.5. The points labelled above the curve indicate where a plant's output begins to be ramped up and added to the portfolio. Conversely, points labelled below the curve indicate where a plant's output is held constant over a subsequent portion of the portfolio's output. The labels in Figure 7.5 correspond to the points with the same labels in Figure 7.4. For output lower than the pooled level of mingen the same method was applied as that given in Table 7.1 resulting in a similar saw tooth pattern as that given in Figure 7.2.

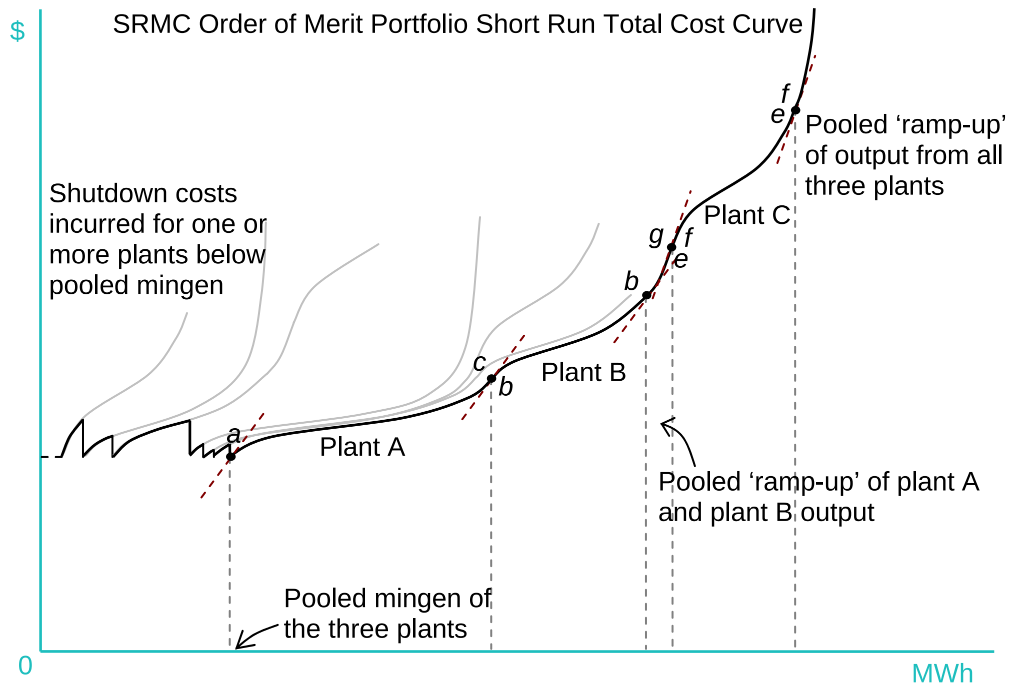


Figure 7.5 – Sub-optimal portfolio short run total cost curve derived by combining cubic function based plants in a SRMC order of merit

However, the stringent use of a SRMC order of merit (i.e. one based on the idea that output is ramped-up as more output is required) neglects the fact that, for non-linear short run total cost functions such as those in Figure 7.4, plant level SRMC is non-constant with output.[36] Consider the cost path based on a strict SRMC order of merit overlayed in Figure 7.6 with an *optimal* cost path specific to portfolio output Q. For portfolio output M through to X, both curves coincide. However, for a portfolio output greater than X they

[36] One common error is to think of economic representations such as those described by Figure 7.4 and Figure 7.5 as dynamic, as opposed to static, models. In reality, while Figure 7.4 and Figure 7.5 describe a cost-quantity relationship within a half hour slice of time they describe no aspect of temporal change. Hence linear thinking in terms of output along a SRMC path is not actually appropriate: i.e. it is not correct to think of optimisation in terms of one source of output being brought on 'earlier' or 'later' than another source of output. Rather, over the half hour period, the three plants operate simultaneously, each at its own average output throughout the interval. The optimal mix of outputs from the available plants will produce the lowest cost curve for a particular level of portfolio output.

begin to diverge with the dashed curve representing the SRMC merit order case and the solid line describing an optimal cost path to portfolio output Q. The reason for this divergence is that, given portfolio output Q, optimality requires a higher share of plant B generation than that provided by the SRMC merit order short run total cost curve.

More specifically, in the sub-optimal SRMC order of merit short run total cost curve case, the mix of generation that provides portfolio output Q consists of:

a) the mingen levels of output M, plus

b) an additional amount of plant A output equal to S minus M, plus

c) an additional amount of plant B output equal to Q minus S;

whereas the *optimal* mix of generation that provides portfolio output Q consists of:

a) the mingen levels of output M, plus

b) an additional amount of base-load output equal to X minus M, plus

c) an additional amount of plant B output equal to Q minus X.

The optimal mix of portfolio output at Q results in a lower cost compared to the sub-optimal mix by an amount equal to p2 minus p1.

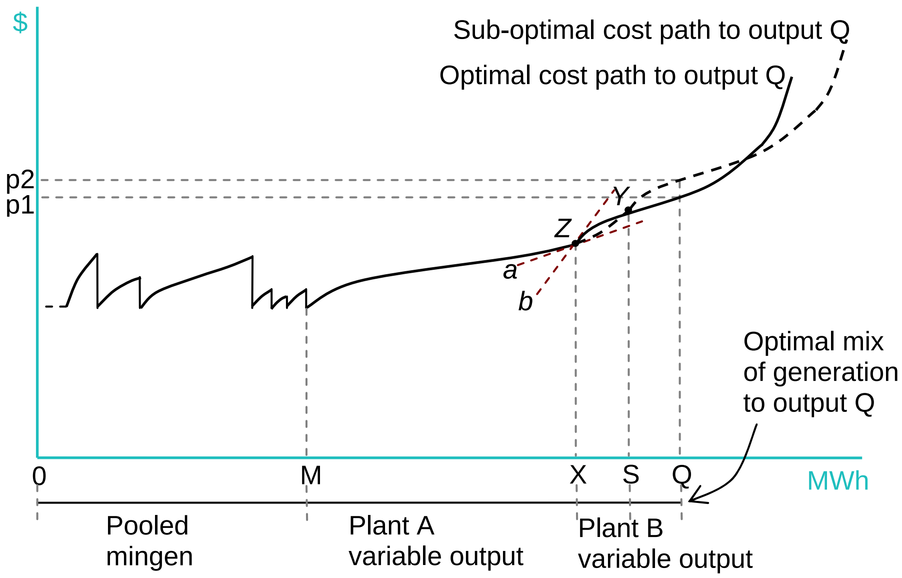


Figure 7.6 – An optimal portfolio short run total cost path to portfolio output Q compared to a SRMC order of merit portfolio short run cost path

7.1.4 Optimal short run total cost curve: The lower envelope of all feasible cost paths

The particular optimal cost path to output Q shown in Figure 7.6 was obtained by connecting the plant B and plant A short run total cost curves at the point marked *Z* (as opposed to the point marked *Y* in the sub-optimal case).[37] Here the breakaway from the

[37] Note that more than one optimal cost path might be able to describe an optimal generation mix for a given level of output. Hence the term ‘an’ optimal price path to a specific level of portfolio output is preferable to the term ‘the’ optimal price path to a specific level of portfolio output. In Figure 7.6, for example, an alternative optimal cost path to that shown might have been drawn as a pooled ramp up between plant A and plant B diverging from the sub-optimal curve at a portfolio output less than X.

strict SRMC order of merit rule can be seen, because at point *Z* tangent *a* (along the sub-optimal curve using plant A output) is not as steep as tangent *b* (along the optimal curve using plant B output). Note also that for the portfolio output between X and S the SRMC merit order short run total cost curve is lower than the optimal path shown. This indicates that each level of portfolio output has its own optimal generation mix associated with it. It follows that the lower envelope of all feasible price paths describes the *optimal portfolio short run total cost curve*. An example is depicted in Figure 7.7 which corresponds to the set of plants depicted in Figure 7.4.

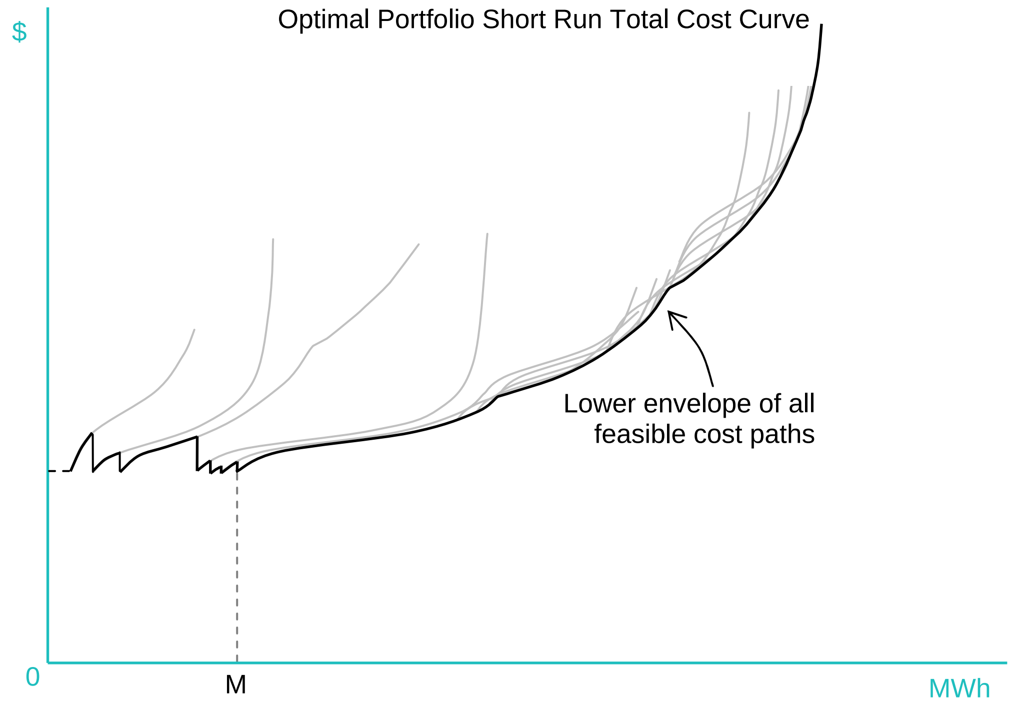


Figure 7.7 – The optimal portfolio short run total cost curve as the lower envelope of all feasible cost paths

7.1.5 Optimal portfolio SRMC

Figure 7.8 depicts the SRMC curve for the portfolio of plants described in Figure 7.9. Portfolio output M corresponds to the sum of all three plants' mingen levels. The saw tooth shape that is evident in the portfolio SRMC curve comes about mainly due to the heat rate dominated cost assumptions that result in a "U" shaped SRMC curve for fossil fuel based plants. These occur at the particular tipping points in portfolio output when it first becomes optimal to include a higher merit plant in the generating mix. The exact location of these points along the portfolio quantity axis depends upon the unique confluence of the SRMC curves associated with the available plants. As with Figure 7.3, portfolio SRMC below pooled mingen is treated as the change in price over the change in quantity from output zero to output M.[34, 38]

[38] See Appendix 1 (Section A1.02) below for a mathematical interpretation of a similar set of plants to those given in this example.

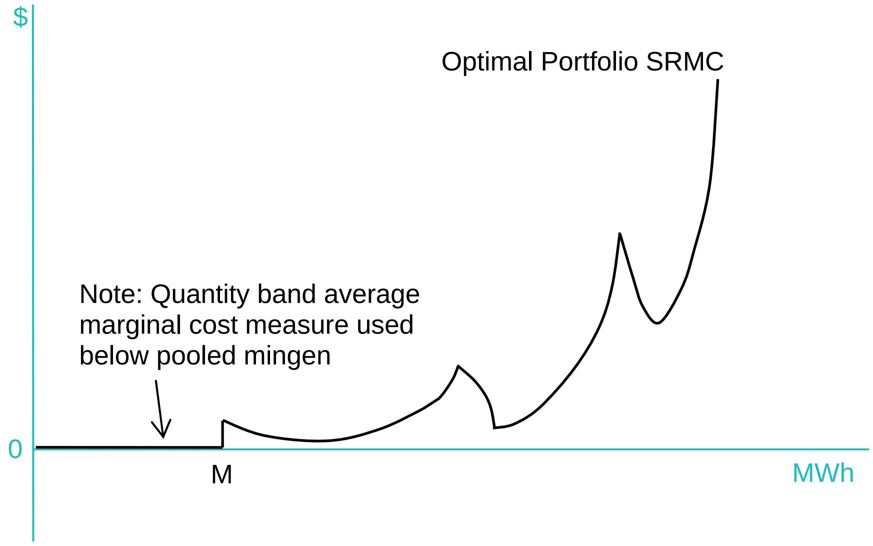


Figure 7.8 – Optimal portfolio SRMC comprising plants with technology based on a cubic short run total cost function

7.1.6 Example 3: The effect of differing shutdown and mingen costs between plants

Plants of varying design, vintage, and level of maintenance are likely to face differing shutdown costs and mingen costs. Base-load facilities, for example, are likely to have large shutdown costs relative to mid-merit or peaking plants. Figure 7.9 depicts a portfolio of linear cost plants that have already been started prior to the relevant half hourly trading interval. However, this time the portfolio consists of generating facilities with differing shutdown cost assumptions.[39]

[39] The total costs for the three plants, A, B and C, were modelled as follows:

$$C_A = \begin{cases} 18{,}000 & if\ Q_A = 0 \\ 20Q_A & if\ 91 \le Q_A \le 340 \end{cases},$$

$$C_B = \begin{cases} 1{,}000 & if\ Q_B = 0 \\ 75Q_B & if\ 26 \le Q_B \le 90 \end{cases}, \text{ and}$$

$$C_C = \begin{cases} 2{,}100 & if\ Q_C = 0 \\ 150Q_C & if\ 16 \le Q \le 50 \end{cases},$$

where C_i denotes the total cost of plant i, and $Q_i \ge 0$ denotes the output of plant i, $\{i \mid A, B, C\}$

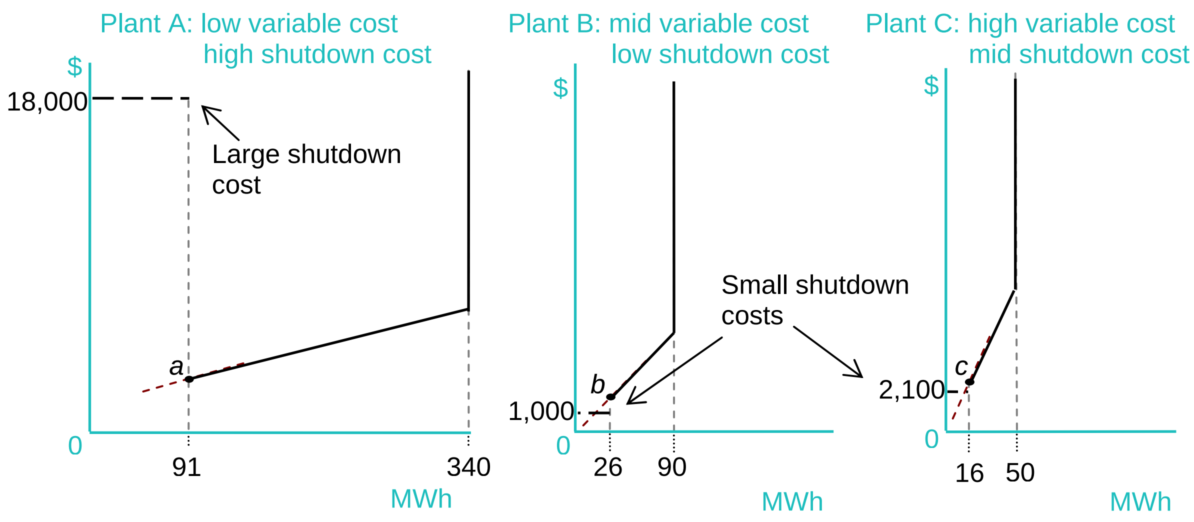


Figure 7.9 – A portfolio of three linear short run total cost plants with sunk startup costs and differing shutdown costs between plants

As with the example given in Figure 7.1, for any given level of output, the SRMC of plant A in Figure 7.9 is less than the SRMC of plant B, which in turn is less than the SRMC of plant C. However, the differing shutdown costs between plants in Figure 7.9 means that, unlike the example in Figure 7.9, where shutdown costs were equalised with mingen costs for each plant, a comparison must now be made between the cost of running a plant at mingen (or above) and shutting it down. Indeed, despite the fact that the variable costs of the plants in Figure 7.9 are similar to those in Figure 7.1, the change in shutdown costs changes the optimal shutdown decision significantly.

Table 7.2 describes the constraints affecting the optimal sourcing of portfolio output from the three available plants, depicted in Figure 7.9, for various ranges of portfolio output. Compare Table 7.1 with Table 7.2. The two tables are identical up until a portfolio level of demand of 116MWh. In Table 7.2, however, the differing shutdown costs between each plant means it is best to shut plant B down between 117MWh and 359MWh of portfolio output, and to run all three plants for portfolio output of 360MWh or above. In contrast, Table 7.1 indicates that it is best to shut plant C down between 117MWh and 132MWh of portfolio output, and beyond 133MWh of portfolio output it is best to keep all three plants running.

The optimal portfolio short run total cost curve derived from the optimal shutdown decision for the plants in Figure 7.9 is depicted in Figure 7.10 (contrast with Figure 7.2). The diagram shows that shutdown costs impact upon optimal short run total cost even for levels of portfolio output beyond the pooled mingen of all three plants (i.e. 133MWh in Figure 7.10). In fact, in Figure 7.10 the minimum total cost in the range of portfolio output between 133MWh and 359MWh is achieved by shutting down plant B and running plants A and C. In other words, the particular combination of variable and shutdown costs given in this example means it is only optimal to keep all three plants running if demand exceeds 359MWh.

Table 7.2 – The optimal choice of plant operation and shutdown for the portfolio of plants given in Figure 7.9

Discrete portfolio output range	***Determining factor for bottom of output range***	***Run constraints***	***Optimal Shutdown vs. Run decision***	
MWh			***Plant***	***Decision***
0 - 15	Zero output.	All three plants must shutdown.	**A**	*Shutdown*
			B	*Shutdown*
			C	*Shutdown*
16 - 25	Mingen plant C.	Plants A and B must shutdown.	**A**	*Shutdown*
			B	*Shutdown*
			C	*Run*
26 - 41	Mingen plant B.	Plant A must shutdown and if plant B runs, plant C must shutdown.	**A**	*Shutdown*
			B	*Run*
			C	*Shutdown*
42 - 90	Mingens plants A + B.	Plant A must shutdown.	**A**	*Shutdown*
			B	*Run*
			C	*Run*
91 - 106	Mingen plant A.	If plant A runs, plants B and C must shutdown.	**A**	*Run*
			B	*Shutdown*
			C	*Shutdown*
107 - 116	Mingens plants A + C.	If plant A runs, plant B must shutdown.	**A**	*Run*
			B	*Shutdown*
			C	*Run*
117 - 132	Mingens plants A + B.	If plant A runs, plant C must shutdown.	**A**	*Run*
			B	*Shutdown*
			C	*Run*
133 - 359	Mingens plants A + B + C (pooled mingen of the three plants).	No run constraints.	**A**	*Run*
			B	*Shutdown*
			C	*Run*
360 +	Total cost with plants A and C running and B shutdown exceeds total cost of all plants running.	No run constraints.	**A**	*Run*
			B	*Run*
			C	*Run*

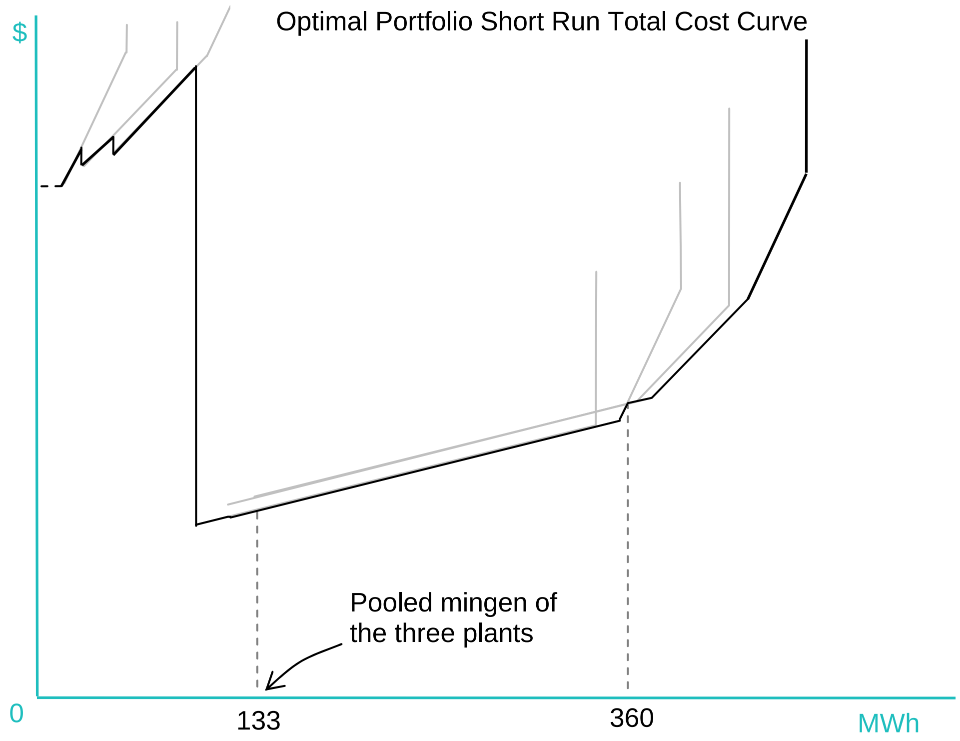


Figure 7.10 – Optimal portfolio short run total cost curve derived by combining linearly increasing total cost technology plants in a SRMC order of merit

7.1.7 Optimal portfolio SRMC

Figure 7.11 depicts the SRMC curve for the portfolio of plants described in Figure 7.4. Portfolio output M corresponds to the sum of all three plants' mingen levels, while portfolio output *m* corresponds to the mingen level of plant A (i.e. the low variable cost, high shutdown cost, plant). The sharp spikes that are evident in the portfolio SRMC curve come about at the particular tipping points in portfolio output where a change occurs in the optimal plant shutdown decision (see Table 7.2 above). Portfolio SRMC below plant A mingen is treated as the change in price over the change in quantity from portfolio output zero to portfolio output *m*.[34]

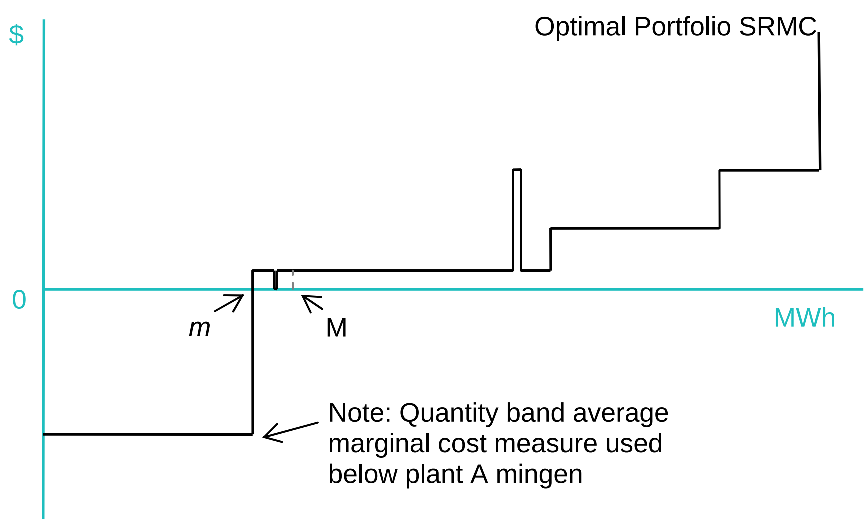


Figure 7.11 – Optimal portfolio SRMC comprised of plants with a linearly increasing total cost technology and with sunk startup costs

7.2 The equimarginal principle

The method applied in Section 7.1.4 is consistent with a fundamental principle of economics: the equimarginal principle. In the current context this principle refers to the fact that, for a given level of portfolio output, portfolio short run total cost will be minimised when each of the plants that have been started operate at the same SRMC. In other words, with multiple sources of production, optimisation requires a distribution of output between each source so that SRMC is equalised across them. This principle remains in place regardless of the combination of available plants and the particular shape of their total cost curves.

Consider the portfolio of two plants given in Figure 7.12. Assume demand for electricity for the half hour is at 100MWh. One (sub-optimal) way of producing this amount of portfolio output would be to supply 50MWh from each plant. However, given that plant A has a SRMC of $10 at 50MWh compared to $18 for plant B at 50MWh, reducing production from plant B by 1MWh would save $18 at a cost of $10 if it were replaced by a MWh of production from plant A. The overall marginal saving in this case would be $8. For a cost minimising firm, this process of marginal reallocation of production from the high marginal cost plant to the low marginal cost plant would continue until the plants' marginal costs were equalised. In Figure 7.12, this occurs when output from plant A is at 60MWh and output from plant B is at 40MWh.

Alternatively, Figure 7.13 shows how the equimarginal principle applies in the case of two constant SRMC technology plants. Again, if the demand for electricity is assumed to be 100MWh for the half hour trading interval, 50MWh of supply from each plant would be sub-optimal. This is because, increasing output from plant A to its capacity of 60MWh costs $100, while decreasing output from plant B by the same amount saves $180: a net saving of $80. Because plant A's SRMC curve becomes infinite at its capacity of 60MWh, the marginal cost of both plants is equalised at a SRMC of $18.

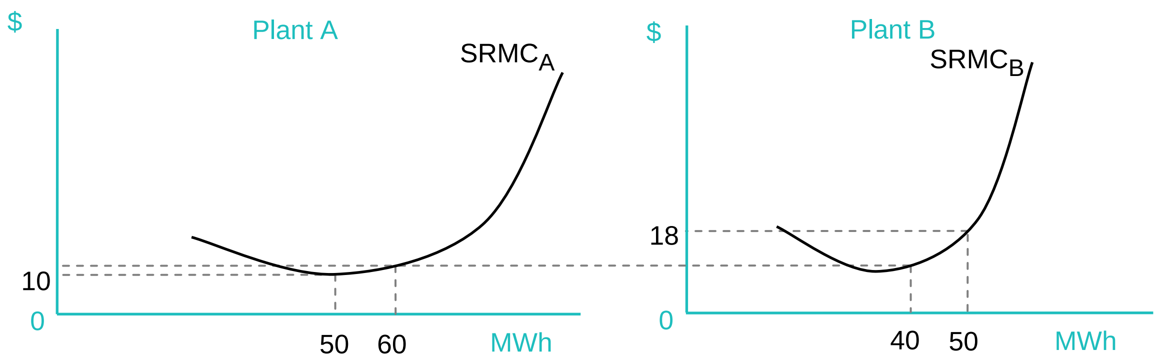


Figure 7.12 – The equimarginal principle for U shaped SRMC plants

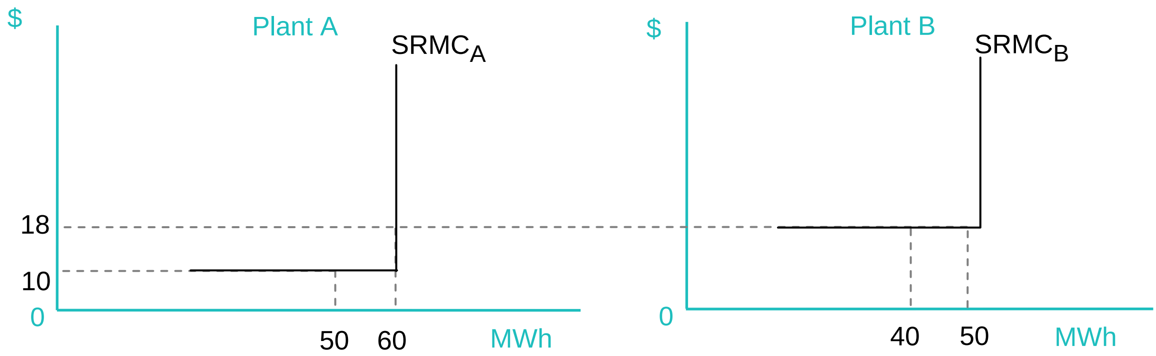


Figure 7.13– The equimarginal principle for constant SRMC plants

7.3 Short run total cost for a portfolio comprising one or more plants with avoidable fixed costs

When some plants in a portfolio are not already operating prior to a half hour trading interval, a firm faces the decision as to whether or not the costs associated with starting those plants should be incurred. While startup costs are a fixed cost at the level of the plant, they can be thought of as variable costs at the level of the portfolio. The following sections will provide examples of optimisation in the case where one or more plants in a portfolio are available, but not started, prior to the half hour trading interval.

7.3.1 Example 4: Linearly increasing total cost technology with differing shutdown costs and startup costs between plants

In Figure 7.14 a firm with four linearly increasing total cost technology plants in its portfolio has two already started prior to the half hour interval and two that could be started up, if required, to meet expected demand during the interval. The costs of startup for the first two plants are considered sunk, while for the two unstarted plants all costs are avoidable (by not engaging in startup) with startup costs comprising an avoidable fixed cost component. The firm faces the decision as to whether or not the costs associated with starting additional plants to meet that demand should be incurred. Note that plant C awaiting startup has a lower SRMC than the already operating plant B, as indicated by the lesser slope of its short run total cost curve.[40]

[40] The total costs for the four plants A, B, C and D were modelled as follows:

$$C_A = \begin{cases} 18{,}000 & if\ Q_A = 0 \\ 20Q_A & if\ 91 \le Q_A \le 330 \end{cases},$$

$$C_B = \begin{cases} 1{,}400 & if\ Q_B < 26 \\ 75Q_B & if\ 26 \le Q_B \le 95 \end{cases},$$

$$C_C = \begin{cases} 2{,}100 & if\ Q_C = 0 \\ 15Q_C & if\ 26 \le Q_C \le 95 \end{cases}, \text{ and}$$

$$C_D = \begin{cases} 18{,}000 & if\ Q_A = 0 \\ 150Q_A & if\ 12 \le Q_A \le 340 \end{cases},$$

where C_i denotes the total cost of plant *i*, and Q_i denotes the output of plant *i*, $\{i \mid A, B, C, D\}$.

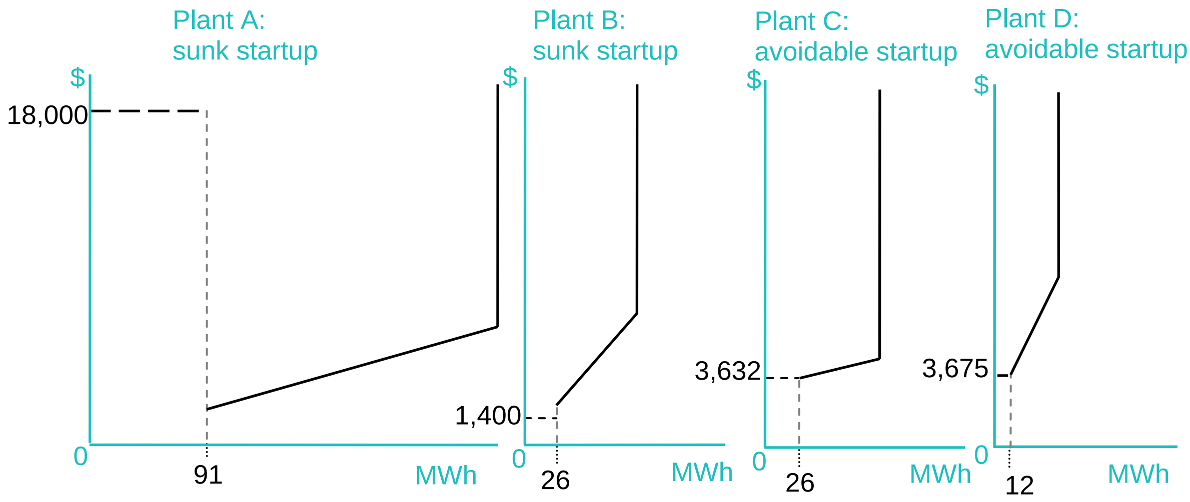


Figure 7.14 - A portfolio comprising plants with technology based on a cubic short run total cost function: two with sunk startup costs and two with avoidable fixed startup costs

The lower envelope of all feasible cost paths for the portfolio of plants in Figure 7.14 is described by Figure 7.15. Despite the apparent difference between the sunk cost and avoidable fixed cost cases, the decision as to whether or not to incur startup costs is the same optimisation problem as presented in Section 7.1.4. That is, a profit maximising firm will seek to minimise the total cost of generating a given portfolio quantity of electricity over the half hour interval. Hence, the derivation of the optimal portfolio short run total cost curve as the lower envelope of all feasible cost paths applies equally well to the case where avoidable fixed costs are included in the short run total costs for one or more generating facilities in the portfolio.

The various optimal cost paths in Figure 7.15 were obtained:

a) by considering the optimal shutdown decision for plants A and B.

b) by connecting the short run total cost curve for plant C (with avoidable fixed startup and mingen costs) to the curve that describes the optimal mix of production between plant A and plant B; and

c) by connecting the short run total cost curve for plant D to the curve that describes the optimal mix of production between plant A, plant B and plant C.

For each given level of demand an optimal mix of output from the available plants produces the required electricity at the lowest feasible cost. Note the ‘sudden jumps’ in optimal short run total cost as a result of the startup costs, in particular for plant D. These result from the assumption of linearity up to capacity in the plant level short run total cost functions.

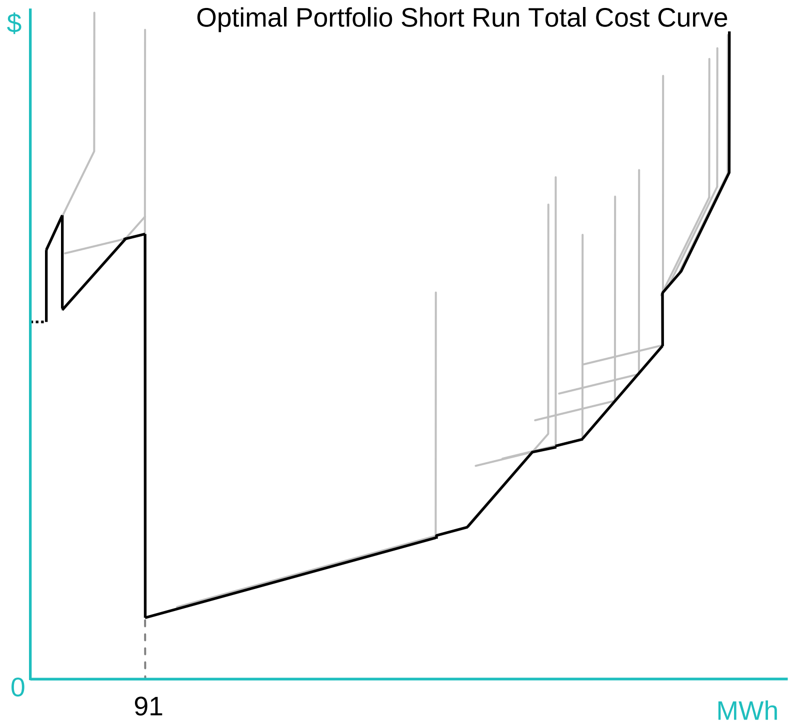


Figure 7.15 - The lower envelope of all feasible short run total cost paths in a portfolio of four plants with linear short run total costs and with two plants having avoidable fixed costs

Table 7.3 describes the constraints affecting the optimal sourcing of portfolio output from the four plants depicted in Figure 7.14 for various ranges of portfolio output. For any level of portfolio output left of the pooled mingen of the two plants already started (i.e. mingen outputs $A + B = 91 + 26 = 117$ from Figure 7.14), the decision must incorporate the shutdown cost of at least one of those two plants. Similarly, constraints apply to the decision to start a plant up. It is only beyond the pooled mingen of all four plants (i.e. outputs $91 + 26 + 26 + 12 = 155$ from Figure 7.14) where there are no constraints as to what plants can be utilised to meet demand. Note that such constraints are only one aspect of the cost minimisation decision for a given level of portfolio output. That is, the optimal decision as to what plants should operate or shutdown may differ at varying levels of output, even within a range of output under which identical run constraints apply.

Table 7.3 – The optimal choice of plant operation and shutdown for the portfolio of plants given in Figure 7.14

Discrete portfolio output range	***Determining factor for bottom of output range***	***Run constraints***	***Optimal decision***			
			Shutdown vs. Run		***Startup & Run vs. Don't Startup***	
MWh			***Plant***	***Decision***	***Plant***	***Decision***
0 - 11	Zero output.	No plant can run.	**A**	*Shutdown*	**C**	*Don't Startup*
			B	*Shutdown*	**D**	*Don't Startup*
12 - 25	Mingen plant D.	Plants A, B and C cannot run.	**A**	*Shutdown*	**C**	*Don't Startup*
			B	*Shutdown*	**D**	*Startup & Run*
26 - 37	Mingen plant B.	Plant A must shutdown;	**A**	*Shutdown*	**C**	*Don't Startup*
		if plants B, C or D run, no other plant can run.	**B**	*Run*	**D**	*Don't Startup*
38 - 51	Mingens plants (B = C) + D.	Plant A must shutdown;	**A**	*Shutdown*	**C**	*Don't Startup*
		if plants B or C run, the other plant cannot run.	**B**	*Run*	**D**	*Don't Startup*
52 - 63	Mingens plants B + C.	Plant A must shutdown;	**A**	*Shutdown*	**C**	*Don't Startup*
		if plants B and C run, plant D cannot run.	**B**	*Run*	**D**	*Don't Startup*
64 - 77	Mingens plants B + C +D.	Plant A must shutdown.	**A**	*Shutdown*	**C**	*Don't Startup*
			B	*Run*	**D**	*Don't Startup*
78 - 90	Plants A + B shutdown costs exceed plant C total cost.	Plant A must shutdown.	**A**	*Shutdown*	**C**	*Startup & Run*
			B	*Shutdown*	**D**	*Don't Startup*
91 - 101	Mingen plant A .	If plant A runs, no other plant can run.	**A**	*Run*	**C**	*Don't Startup*
			B	*Shutdown*	**D**	*Don't Startup*
102 - 116	Mingens plants A + D.	If plants A and D run, no other plant can run.	**A**	*Run*	**C**	*Don't Startup*
			B	*Shutdown*	**D**	*Don't Startup*
117 - 128	Mingens plants A + (B = C).	If plants A and B or C run, no other plant can run.	**A**	*Run*	**C**	*Don't Startup*
			B	*Shutdown*	**D**	*Don't Startup*
129 - 142	Mingens plants A + (B = C) + D.	If plants A and D run, and B or C run, the other plant cannot run.	**A**	*Run*	**C**	*Don't Startup*
			B	*Shutdown*	**D**	*Don't Startup*
143 - 154	Mingens plants A + B + C.	If plants A, B and C run, plant D cannot run.	**A**	*Run*	**C**	*Don't Startup*
			B	*Shutdown*	**D**	*Don't Startup*
155 - 330	Mingens plants A + B + C + D.	No run constraints.	**A**	*Run*	**C**	*Don't Startup*
			B	*Shutdown*	**D**	*Don't Startup*
331 - 404	Maximum output of plant A exceeded.	No run constraints.	**A**	*Run*	**C**	*Don't Startup*
			B	*Run*	**D**	*Don't Startup*
405 - 519	Plants A + B total costs exceeds plant B shutdown cost + plants A and C total costs.	No run constraints.	**A**	*Run*	**C**	*Startup & Run*
			B	*Shutdown*	**D**	*Don't Startup*
520 +	Maximum output plants A + B + C	No run constraints.	**A**	*Run*	**C**	*Startup & Run*
			B	*Run*	**D**	*Startup & Run*

7.3.2 Optimal portfolio SRMC

Figure 7.16 depicts the discrete portfolio SRMC curve for the portfolio of plants described in Figure 7.14.[34] Note that the step function is not monotonically increasing. For example, there is a decrease in cost as a result of the startup of plant C which has a lower SRMC than both plants already operating. Moreover, an off-the-scale spike in portfolio SRMC is evident at the point where it is optimal to start plant D up to meet demand.

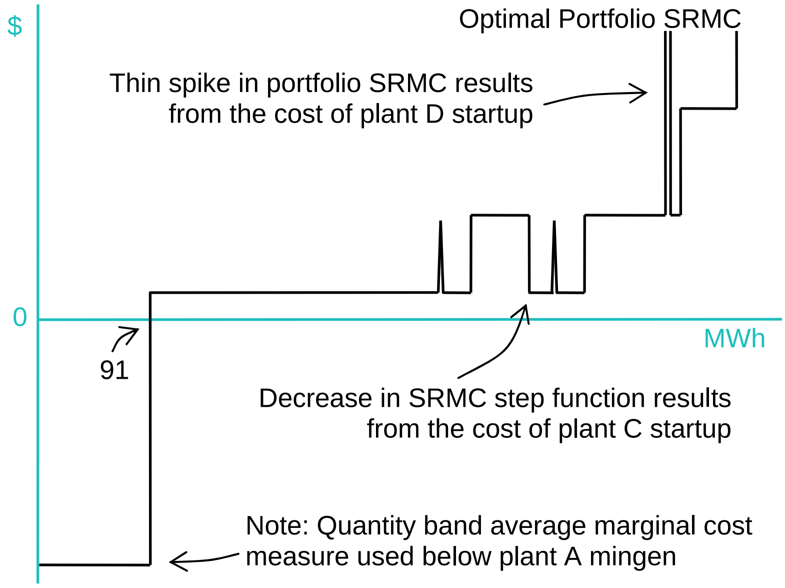


Figure 7.16 - Optimal portfolio SRMC for a cubic-based total cost function technology with one or more plants having avoidable fixed costs

7.3.3 Example 5: Portfolio of plants with cubic based total cost functions and with differing shutdown costs and startup costs between plants

Figure 7.17 describes a firm with four plants in its portfolio: two with sunk startup costs, and two that can be started when optimal. Each plant has a short run total cost function that is based on a cubic functional form.

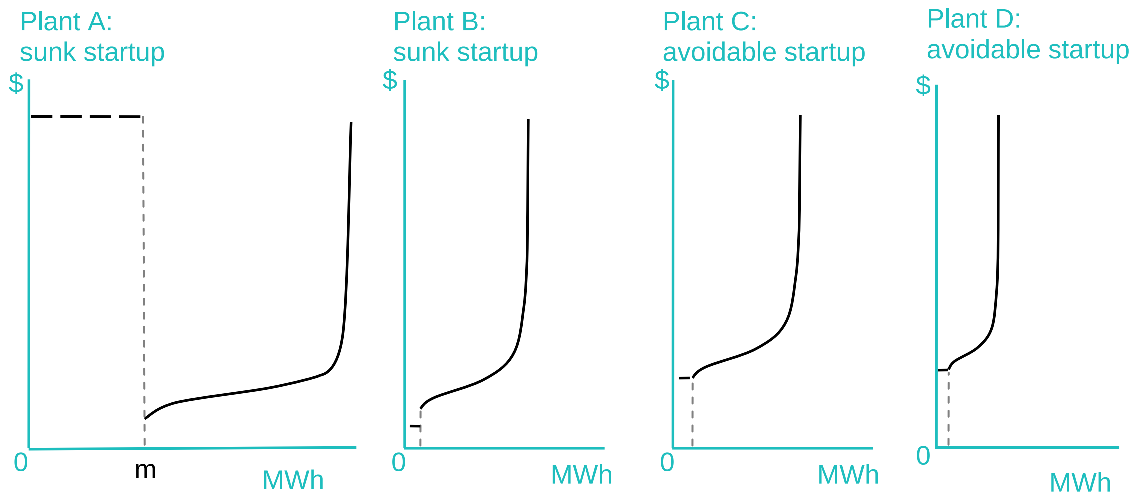


Figure 7.17 – A portfolio comprising plants with technology based on a cubic short run total cost function: two with sunk startup costs and two with avoidable fixed startup costs

Figure 7.18 reveals the optimal short run total cost curve corresponding to the portfolio of plants described by Figure 7.17. Note that, as a result of the optimisation process, and despite the large avoidable fixed costs associated with the two plants requiring startup, there is no sudden jump in the optimal portfolio short run total cost curve. The lower envelope of all feasible cost paths produces a relatively smooth curve.

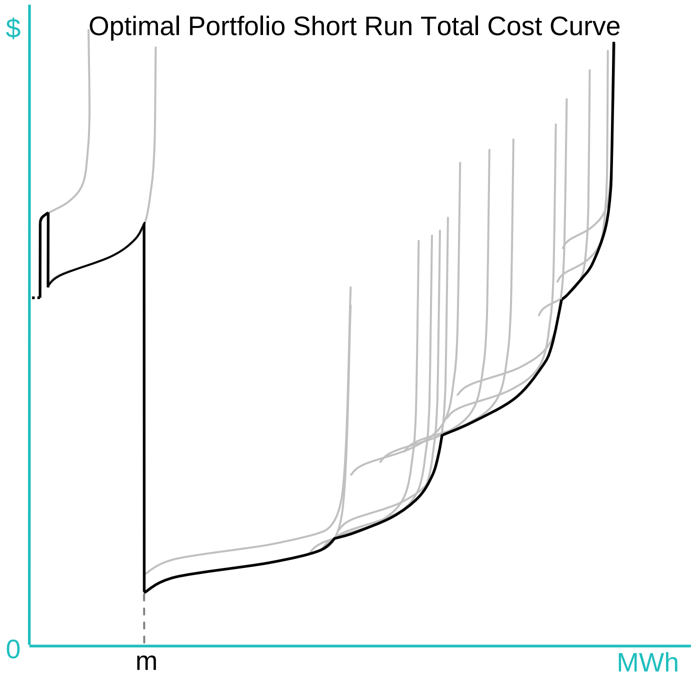


Figure 7.18 – The lower envelope of all feasible short run total cost paths for a portfolio of four plants with cubic based short run total costs and with two plants awaiting startup

7.3.4 Optimal portfolio SRMC

The saw tooth pattern that is evident in the SRMC curve for the portfolio of plants given in Figure 7.17 is remarkably similar to that described in Section 7.1.5, despite the introduction of startup costs for half of the plants in the portfolio.[41] Again, the saw teeth

[41] Note that the portfolio depicted in this section has one more plant to that in Section 7.1.5.

occur at the particular tipping points in portfolio output when it first becomes optimal to include output from a higher cost plant in the generating mix.[42] In the case of a plant already started, this applies to non-mingen variable output only (mingen output is already included in the mix). For plants not yet started however, the tipping points can be quite dramatic because of the sudden addition of a mingen quantity of output to the mix at zero marginal cost. This, combined with the U shape of plants' SRMC curves, causes the rate of increase in optimal portfolio short run total cost to decline abruptly before increasing again.[43]

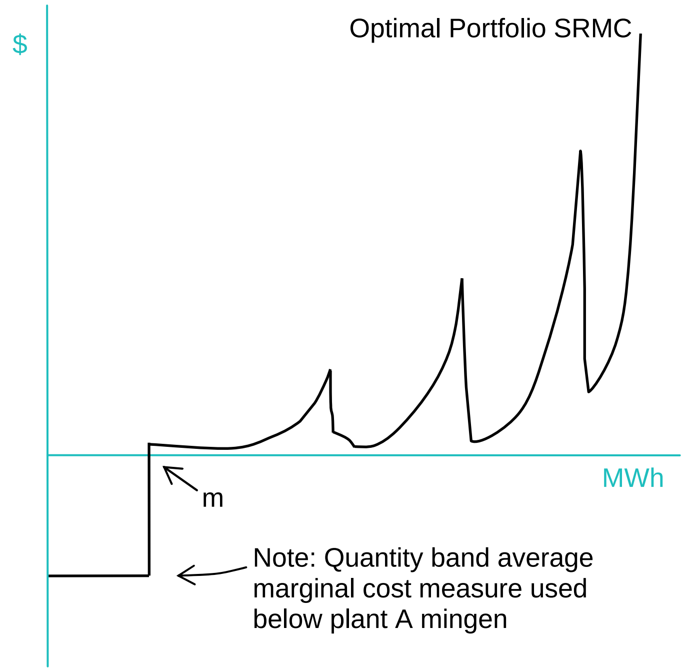


Figure 7.19 – Optimal portfolio SRMC for a cubic-based total cost function technology with one or more plants having avoidable fixed costs

[42] A (perhaps) surprising finding that results from the optimisation process is that, where plants have U shaped SRMC functions, startup costs do not trigger a sudden jump in portfolio short run total cost and therefore do not represent an off the scale spike in SRMC. The saw teeth spikes that do occur are similar in scale for a portfolio comprising plants with sunk startup costs as those for a portfolio comprising one or more plants that require startup within the relevant half hour trading interval.

[43] See Appendix 1 (Section A1.03) below for a mathematical interpretation of a similar set of plants to those given in this example.

8 Discussion

The purpose of this section is to discuss the various findings of the preceding sections and the modelling conducted in the Appendices of this paper. Much consideration is given to market efficiency and the effectiveness of the market rules given the likely shape of the portfolio SRMC curve. Measurement difficulties are also outlined, with specific regard given to the issue of data quality. Finally, a projection is given in relation to the competitive evolution of the market and the likely duration of the need for the SRMC rule requirement.

8.1 Estimation of plant level functional form and its effect on portfolio SRMC

The use of mathematical modelling, together with spreadsheet software, enables graphical depiction of the general features of portfolio SRMC in electricity generation. In Appendix 1 of this paper, the specific functional form chosen to describe fossil fuel plant short run total cost was a piecewise function that comprised of: (a) some component of mingen at either zero economic cost (in the case of a plant with sunk startup costs) or at a positive avoidable fixed cost (in the case of a plant requiring startup); and (b) a rational function with a cubic function as the numerator and a linear function as the denominator giving a "∼" shape to the curve. Provided three key assumptions are maintained, choosing the optimal mix of generation out of a portfolio of plants leads to the 'saw tooth' shaped portfolio SRMC pattern described in this paper. First, a plant's SRMC curve is assumed to initially decrease with improving thermal efficiency beyond some mingen amount. Second, an individual generating facility faces a capitalised thermodynamic limit in terms of output capacity that can be approached, but not exceeded, by the SRMC function. Third, a plant's SRMC will eventually increase as this limit is approached.

These assumptions are consistent both with the economic law of diminishing marginal returns and engineering based observations of thermal efficiency in fossil fuel plants. However, as is evident from the analysis in Section 5, a different choice of functional form will change the general features of the portfolio SRMC curve generated from the optimisation process described herein.[44] For example, the author has conducted a similar optimisation exercise to that conducted in Appendix 1 on a portfolio of plants assumed to have sunk startup costs, low shutdown costs, and monotonically increasing SRMC functions. The result was a portfolio SRMC function that was also monotonically increasing. In contrast, as described in Section 7 above, an assumption of linearity in plant level short run total cost results in a step function that may or may not increase monotonically depending upon the requirement to start a plant up to meet demand.

The important concept that connects the result of portfolio optimisation to the short run total cost curves of individual generating facilities is that the lower envelope of all feasible costs paths for a portfolio of plants provides the optimal portfolio short run total cost curve, with portfolio SRMC being the slope of this curve. At every point along the optimal portfolio short run total cost curve, all plants operating have equalised SRMC, as is consistent with the equimarginal principle. Any combination of plant that diverges from this principle will be inefficient. Therefore, in SRMC terms, for an efficient firm, there is no such thing as 'a most expensive plant' in operation during a trading interval. The obvious exception occurs under the assumption of plant level short run total cost linearity across

[44] This fact points towards the importance of data collection and empirical analysis.

the whole portfolio. In this case, the portfolio SRMC step function that results may be considered to have an identifiably marginal or most expensive plant.

8.2 Basic efficiency criteria with respect to a ‘saw tooth’ SRMC function

The fact that portfolio SRMC may not be monotonically increasing does impose some additional criteria to the setting of efficient output. This, however, does not prove inconsistent with the concept of market efficiency occurring where price equals SRMC. In Figure 8.1 one can imagine a competitive industry where market forces determine price P for a trading interval. In the competitive environment all firms are price takers so no output can be supplied above this price and profit maximising behaviour will correspond to the efficient outcome. However, because price P passes through the saw tooth shaped increasing portfolio SRMC curve at five different levels of output, profit maximisation has a more complex decision criteria than that usually described in microeconomic textbooks. In Figure 8.1, assuming average revenue is greater than average cost at all levels of output, the firm will maximise its producer surplus by choosing output Q3 if areas (c + e) > (b + d) and [(c + e) – (b + d)] > (c – b). Otherwise the firm will choose output Q2 if area c > b, or Q1 if it is not. This is the best a firm can do in an efficient market because in the competitive environment market price cannot be forced up by any given firm. There is, for example, no scope for additional profit by adding a margin to the SRMC curve for ‘commercial risk’ or for ‘fuel cost recovery’ on long run contracts. In a competitive market such thinking will result in reduced profits. However, in a market that is influenced by market power, a similar approach will push the price up, resulting in monopoly rents for the firm at the expense of consumers, thus reducing the net benefits of the market as a whole. Such behaviour is unacceptable under the market rules.

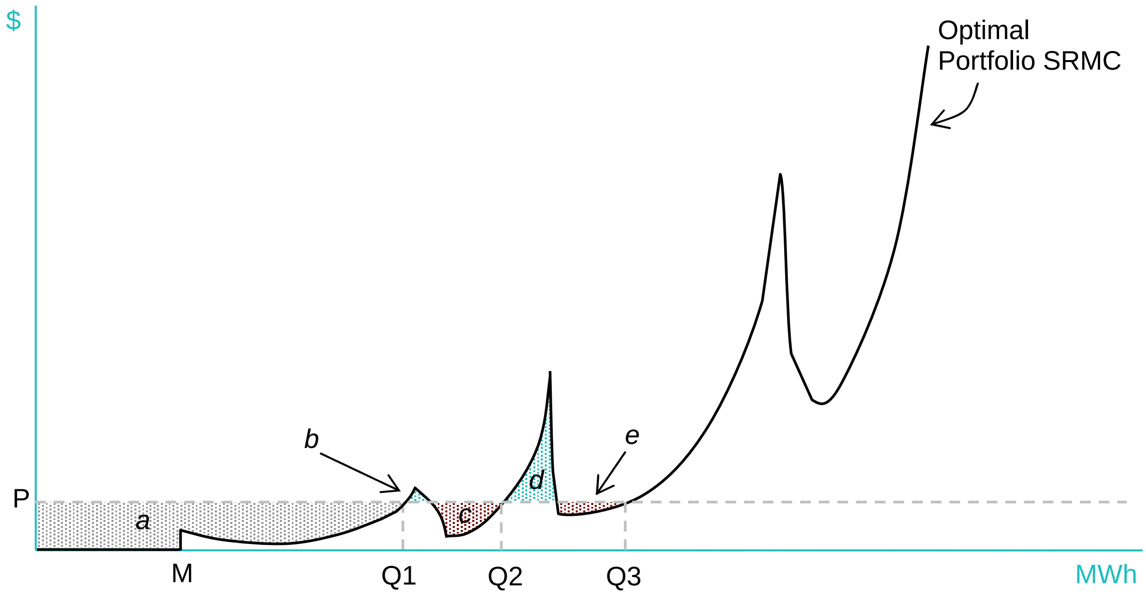


Figure 8.1 – The efficient output choice for a firm with a portfolio SRMC curve that does not increase monotonically

A saw tooth portfolio SRMC curve does have implications with regards to the effectiveness of the market rules. To the extent the word “reflect” in clause 6.6.3 of the market rules (see Section 1 above) is deemed to be equivalent to the word “reproduce”, there exists a conflict between various requirements of the market rules - specifically: the requirement to offer electricity at SRMC and the requirement to do this in monotonically increasing steps in price intervals bounded by fuel specific caps. Obviously, monotonically increasing steps make the algorithms that generate market prices easier to construct. However, any disparity between a firm's offer curve and its actual SRMC will fail to achieve the economic efficiency aspects of the market objectives.

To use a common analogy, a strict interpretation of the word "reflect" would require generators to fit a square peg into a round hole. If a generator constructs its portfolio offer curve in the manner described in the partial equilibrium framework of Figure 8.2, for example, the result is a market price of Pm as opposed to the efficient price of Pe and thus a deadweight loss to society of area b + c. Furthermore, because the optimal portfolio SRMC curve lies below the portfolio offer curve at the level of marginal benefit (as represented by the bid curve), the result is a large transfer of wealth from consumers to producers. This amounts to area a + b, thus failing market objectives aimed at encouraging economic efficiency.

If the choice as to where a step should lie is left open to the interpretation of a generator, it would be economically rational for that firm to choose a configuration that maximises its economic rent to the cost of consumers. The potential for gaming such an outcome will be directly correlated with the firm's ability to avoid regulated civil penalties as much as the scope of those penalties. In this context, the fact that a series of monotonically increasing steps cannot be fitted to the true nature of a generator's SRMC takes on obvious importance. The problem is further exacerbated by the requirement to fit offer curves within stipulated price floors and price caps. If, operating under these contradictions, a portfolio offer curve such as that depicted in Figure 8.2 were deemed to be noncompliant with the market rules, generators may, without clear guidance, claim they are being lured into a straw man argument. Conversely, if the portfolio offer curve depicted in Figure 8.2 were deemed to be consistent with the market rules, the outcome may approach that of full monopoly pricing, leaving the regulated SRMC rule wanting as a market efficiency instrument.

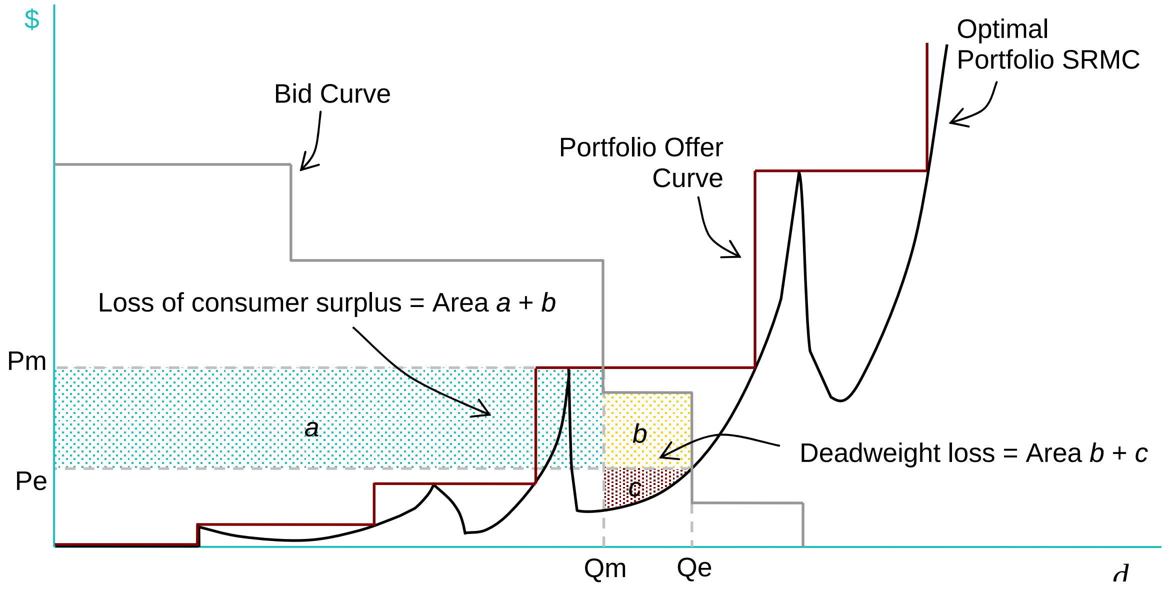


Figure 8.2 – Example of total and consumer surplus losses that can occur where a monotonically increasing portfolio offer curve cannot be fitted to an optimal portfolio SRMC curve which is not monotonically increasing over the full range of output

The extent to which this problem can be mitigated depends on the number of plants available to a monitored firm in a trading interval as well as the interpretation of the market rules. Firstly, a firm with many plants (and thus one likely to have a degree of market power) may face a saw tooth pattern that can be divided into a series of relatively horizontal bands. As such, a large number of available plants may naturally limit the extent to which monopoly profits can be extracted via an offer curve strategy such as that depicted in Figure 8.2. Secondly, if the average marginal cost over a quantity band were deemed to "reflect" SRMC, a second best outcome would be achieved in economic efficiency terms, on average, over repeated iterations. In fact, a quantity band average marginal cost interpretation could be applied to any series of offer steps that a firm places through a saw tooth SRMC curve, regardless of the steepness of its general upwards trend. Such an interpretation would also smooth out any thin spikes in SRMC that might result from startup (shutdown) costs associated with bringing a plant online (offline) during a trading interval.

Figure 8.3 provides an example. A firm divides its output into five steps. In calculating the price of each step, the firm takes the average of the portfolio SRMC curve over the range of output that corresponds to each step. For example, the step between outputs *a* and *b* has an average portfolio SRMC of price *p*. In other words, price *p* represents the average value of the portfolio SRMC curve bounded by outputs *a* and *b*. The same procedure is followed in the construction of other steps. The exception is where the average portfolio SRMC is higher than a fuel specific price cap over a range of output. In this case the price of the step equals the price cap. Box 8.1 defines this approach in mathematical terms.

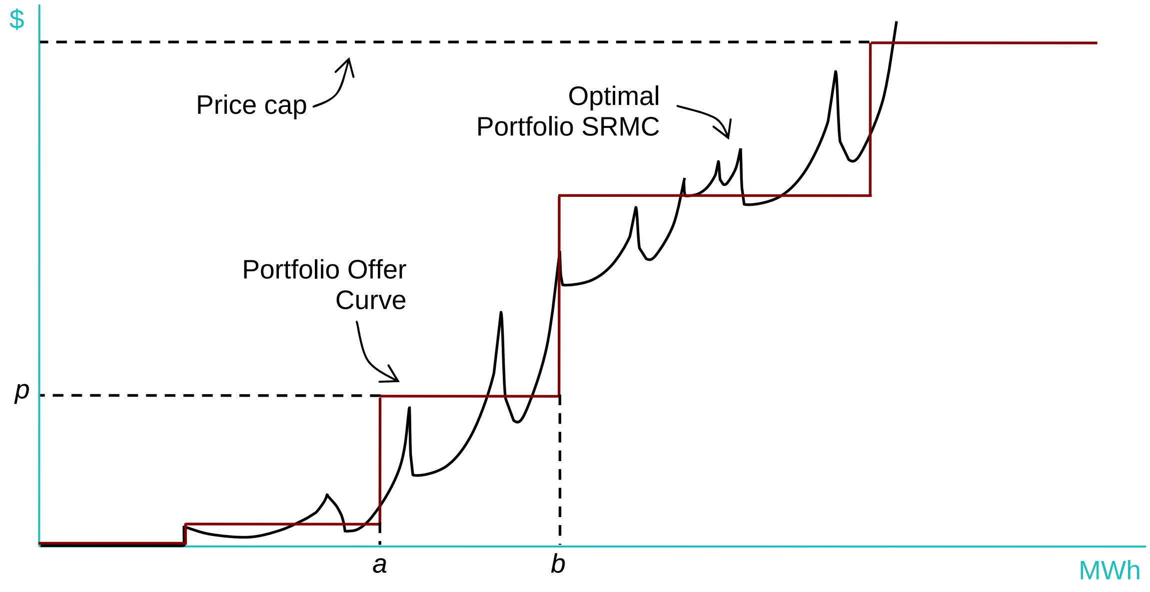


Figure 8.3 – A quantity band average marginal cost approach to reflecting portfolio SRMC with an offer curve comprising non-monotonically decreasing steps.

Box 8.1 Mathematical definition of quantity band average portfolio SRMC steps

Consider the following generic function for a portfolio of plants:

$$C'(Q) = \frac{dC}{dQ}, \qquad Q \geq 0,$$

where portfolio SRMC (denoted C') is a function of portfolio output Q and is the first derivative of the short run total cost function. In this case, the average portfolio SRMC on the interval [*a*,*b*] is given by:

$$\frac{1}{b-a}\int_a^b C'(Q)\,dQ,$$

provided the specified definite integral exists.

In most cases, the portfolio SRMC will not be continuous over a range of output. It is therefore appropriate to approximate the above by means of the following discrete method:

$$\frac{1}{b-a}\sum_{i=1}^{n} C'(Q_i)\Delta Q,$$

where portfolio SRMC (denoted C') over the interval [*a*,*b*] is divided into *n* subintervals between *i* and *n*, each of width ΔQ. The above is the average portfolio SRMC function for the *n* subintervals and the given values of Q_i. The greater the value of *n*, the closer the result will be to that of the definite integral method.

8.3 Empirical issues

Turning to the question of empirics, measurement difficulties arise due to the unavailability and quality of suitable data from market participants. For example, the real time metering of power stations for both fuel input and generation output is "…inherently inaccurate and unsuitable for…" SRMC estimation techniques (Waters, 2007; per com).[45] The main issue in this regard is the lag between fuel use and electricity output, which prohibits an accurate half hourly estimation of the relevant input-output relationship. However, a good substitute exists in the unit heat rate curves which are available to market participants. Using the cubic estimation technique described in Section 5 of this paper, these curves can be readily converted into a total fuel cost curve by half hour, which can in turn be converted into a marginal factor cost curve for the trading interval. This method allows practical evaluation of the most significant of the short run variable costs associated with electricity supply.

The quality of startup cost, shutdown cost, operational and maintenance cost, and line loss data, however, remains an issue. Thus far, the Authority is in possession of such information in its summary form, rather than in a form that reveals the method by which it

[45] Waters, J. Verve Energy, Per com., email, 9th August 2007.

was derived. This does not assist auditing on the part of the Authority to determine whether the data is robust by engineering and economic standards.

A further problem arises due to the gap between engineering and economic based representations of electricity supply. The software packages typically used by generators to optimise plant and equipment operate as a black box, i.e. without the user's detailed understanding of how its output is generated. Another constraint is that SRMC is an economic concept that is not well understood within the electricity industry. This second factor was evident when Short & Swan (2002) attempted to estimate the SRMC of generators in the Australian national electricity market. As they put it, they were forced "…to use the information revealed by the generators themselves through their publicly available supply offers as a means of estimating their generator marginal costs" because of a lack of "…accurate and detailed information on marginal costs for Australian generation stations…".[46] In reference to Short & Swan's paper, and to marginal cost based measures of market power more generally, Brennan (2002) pointed towards the common, economically erroneous, use of "…the average variable or operating cost of the 'last' generator that would be dispatched to meet energy demand" [47] as a proxy for marginal cost in models of electricity generation.[48] ROAM Consulting, which was commissioned to model SRMC in the WEM, sees the "textbook definition" of SRMC "…as those costs of a firm that vary directly with the short run level of production".[49] This approach is consistent with Brennan's concerns. A similar difficulty arises from the use of the PowrSym3 tool by Verve Energy in determining its "…portfolio marginal cost for each half hour… as the operating cost of the most expensive generating unit dispatched in that half hour".[50]

8.4 Real competition vs. replicated competition

The SRMC rule creates the paradox of taking a centralised, legislated approach to regulation while at the same time seeking an efficient, decentralised, market-based allocation of resources. Conversely, under real competition, market forces tend towards an efficient outcome without the need for participants to be aware of their SRMC. Ideally, the condition of real competition is preferable to the condition of replicated competition. Hence, the legislated SRMC approach should be thought of as a temporary, second best, measure that is expected to be removed at such time when the first best condition (i.e. the dissolution of market power) can be achieved.

While sound economic science is vital in good market design the main feature of a well designed market is that it should not rely on the participants' sound application of economic science. However, clause 6.6.3 of the market rules obliges participants to implement advanced economic techniques to determine their portfolio SRMC curve.

[46] The assumption of perfect competition precludes this method in any evaluation of the abuse of market power by participants (see, Short, C. & Swan, A. (2002), *Competition in the Australian National Electricity Market*, ABARE, Current Issues 02.01, January, p. 4).

[47] Brennan, T. (2002), *Preventing Monopoly or Discouraging Competition? The Perils of Price-Cost Tests for Market Power in Electricity*, Resources for the Future Discussion Paper 02-05, p. 2. See also Brennan, T. (2003), 'Mismeasuring Electricity Market Power', *Regulation*, Vol. 26, No. 1, pp. 60-65.

[48] Brennan's criticisms are relevant if SRMC is either U shaped (the standard theoretical description) or linearly increasing. If, however, SRMC can be approximated by a horizontal straight line, SRMC will equal short run average variable cost. Therefore, in the latter case, short run average cost would be a reasonable proxy for SRMC.

[49] ROAM Consulting (2007), *SRMC STEM Offer Modelling*, Report IMO 00010/MA to Independent Market Operator, 22 May, p. 1.

[50] Verve Energy (2007), *STEM Pricing Using PowrSym*, Report No VE 02/07, January, p. 3. Document not publicly available: permission given for this quotation. Note: the statement quoted may reflect the view of the paper's author rather than Verve Energy as an organisation.

Problems associated with this requirement were anticipated by Ruff (2002) when a similar proposal to clause 6.6.3 of the market rules was tabled in California before its Senate Judiciary Committee:

> The fundamental economic and policy flaw in… [the legislated SRMC approach] …is its focus on market outcomes rather than market structure, just the opposite of the approach generally taken in competition policy and law. As a general matter, competition policy does not try to determine what the outcome of a truly competitive market would be and then impose that outcome on a basically uncompetitive industry. Instead, competition policy tries to create an industry structure that is inherently competitive because it has many, competing players, or at least to induce the players in the industry to act as though they were competitive, and then lets the competitive process determine prices and other features of the industry. A policy focused on market structure and (secondarily) behaviour is far more feasible to implement than a policy focused on market outcomes, and is economically preferable because it allows the operations and evolution of the industry to be determined by competition rather than by calculations or judgments of competition authorities and judges.[51]

It should be noted that the legislated SRMC proposal was rejected in California. Indeed, to the author's best knowledge, no wholesale electricity market beyond WA polices SRMC in this manner. Economists tend to view the encouragement of competition via industry disaggregation as a longer term objective.[52] That is, the actual or threatened entry of competitors into a market, over time, not only forces prices down towards the efficient level, but also favours those firms that are able to lower their costs. Given that a firm's economic management skills are as much a part of this natural selection process as any other aspect of business operations, it should not be the role of a regulator to perform the function on the firm's behalf. This would be contrary to the concept of allowing the profit motive to drive technical change and gains in efficiency.

The introduction of competition into a market should be thought of as an evolutionary process both at the level of the market and at the level of regulation, with the fostering of real competition being the best long term regulatory strategy. That is, regulatory resources are best directed towards the promotion of a market that can evolve to *actual* competitiveness, thus enabling a move away from the need to replicate efficiency under conditions of acute market power. Hence the SRMC rule is viewed as a temporary measure that seeks to emulate an efficiency concept which would otherwise occur naturally in a market under real competition.

In WA, the competitive process is currently constrained by the State's geographical isolation and the dominance of the incumbent public generating firm. The upside to this is that there is much scope for structural improvement, given that the dominant player currently has in excess of 40 generating facilities at its disposal. In WA, the evolution of energy market policy has followed a deregulation process starting with the break up of a combined gas/electricity utility into separate gas and electricity utilities, through to the current situation where the electricity industry is vertically disaggregated, with the publicly owned generator constrained by legislation preventing it from adding to its portfolio beyond an aggregate generating capacity of 3,000MW (not including renewable energy sources).[53] In August 2007, the total installed capacity in the SWIS stood at 4,106MW.

---

[51] Ruff, L. (2002), *Statement of Larry E. Ruff, PhD on California State Senate Bill No. 2000 Unlawful Electric Power and Natural Gas Practices*, Before the Senate Judiciary Committee, April 23, p. 7.

[52] See, for example, Winston, C. (1998), 'U.S. Industry Adjustment to Economic Deregulation', *Journal of Economic Perspectives*, Vol. 12, No. 3, pp. 89-110; and Borenstein, S. & Bushnell, J. (2000), 'Electricity Restructuring: Deregulation or Reregulation?', *Regulation*, Vol. 23, No. 2, pp. 46-52.

[53] *Ministerial Direction under the Electricity Corporations Act 2005*, Effective 1 April 2006, Signed 21st March 2006.

An installed capacity of 4,786MW was anticipated by mid 2010 (see Figure 8.4).[54] In August 2007, Verve Energy's installed capacity in the SWIS stood at 3,240MW.[55]

Therefore, in a static policy environment (with respect to further market disaggregation), the evolution towards a truly competitive market would be expected to occur gradually over the next one to two decades. A linear projection beyond the current schedule for plant commissioning and decommissioning, while maintaining Verve Energy's capacity constant at 3,000MW, indicates Verve Energy's market share may reach 50 per cent by early 2016.[56] It is anticipated that the SRMC rule may still be required well beyond this date until such time that Verve Energy's share of capacity is matched by one or more generating firms.

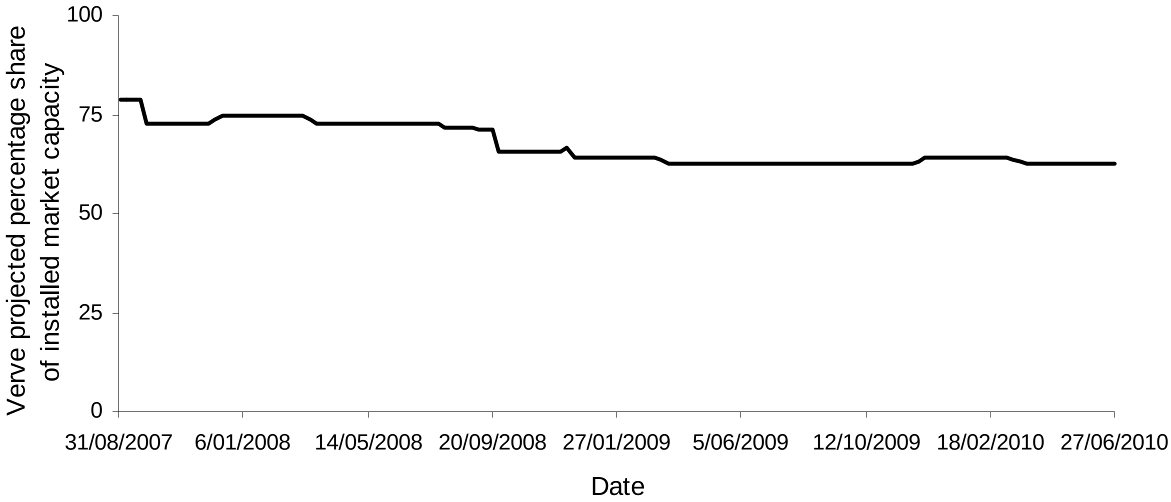


Figure 8.4 – Expected time evolution of Verve Energy's market share based on System Management projections

[54] System Management (2007), *Medium Term Projected Assessment of System Adequacy Report*, August, Available IMO website: http://www.imowa.com.au/10_5_1_n_mt_pasa.htm

[55] *Verve Energy Home Page*, Accessed 24 August 2007, Available: http://www.verveenergy.com.au/subContent/companyProfile/Company_Profile.html

[56] The accuracy of this estimate will depend on factors such as economic growth, technological change and changes to government policy.

9 Conclusions

In the WEM, and in wholesale electricity markets more generally, an efficient rationing of scarce resources will only occur if: (a) generators are paid their marginal costs of production, while; (b) consuming firms pay (equivalent to) a competitive price for their business inputs. In other words, efficiency will only occur if the wholesale price of electricity equals its SRMC of supply. A wholesale price above SRMC will lower the wealth generation potential of the market as a whole. Critically, given that no market operates in isolation, an inefficient wholesale electricity price will distort the broader economy, leading to a widely distributed loss of wealth.

The presence of market power in the WA electricity industry has resulted in regulation intended to replicate the outcomes of a competitive market. Clause 6.6.3 of the market rules requires generators to offer their electricity at a price that reflects their reasonable day ahead expectations of SRMC. These offers must be submitted as a portfolio supply curve in an increasing sequence of steps and under the constraint of price floors and fuel specific price caps. To the extent that the shape and range of the portfolio supply curve under the above constraints does not match the portfolio SRMC curve, an inconsistency arises in the requirements of the market rules. In effect, there is a trade-off between the desirability of a simple user-friendly information technology system through which bid and offer information is submitted to the IMO, the actual nature of cost, and the requirements of economic efficiency.

However, use of the word “reflect” in clause 6.6.3 of the market rules provides scope for flexibility. If the portfolio supply curve were determined by means of averaging marginal cost over a submitted quantity band, a second best outcome would be averaged out over repeated iterations. The conditions for this improve as the number of generating facilities in the portfolio increases, thus tending to be better suited to firms with market power. Therefore, provided price caps are based on worst case SRMC scenarios, a portfolio supply curve could be fitted to SRMC in a manner that is consistent with the full requirements of the market rules.

SRMC modelling has great potential in delivering commercial gains through improvement in business operations that have not yet been exposed to the rigour of a competitive market. Developing an understanding of a generator’s cost structure should, therefore, be seen not only as a social benefit but also as a private benefit to market participants. Projections suggest that the current condition of acute market power is likely to persist in the WEM in the medium term at least, with acute market power defined as the absolute majority share of installed capacity in the SWIS by a single generating firm. Hence the SRMC monitoring approach appears likely to remain a key part of the regulatory regime for some time.

The information provided within this paper is intended to improve the understanding of the SRMC concept within the Western Australian electricity industry. Clearly, information flow of this nature is requisite for the effectiveness of clause 6.6.3 of the market rules. Likewise, a flow of information in the opposite direction (i.e. from market participants to the Authority) will be of critical importance. Legislative powers granted to the Authority under the *Electricity Industry (Wholesale Electricity Market) Regulations 2004* provide the Authority with an ability to request such information. The quality of the data remains an area of priority. However, provided the data input is sound, the modelling techniques developed by the IMO and the Authority in conjunction with this paper place the Authority in a good position to determine a firm’s reasonable expectations of portfolio SRMC.

References

Borenstein, S. & Bushnell, J. (2000), 'Electricity Restructuring: Deregulation or Reregulation?', *Regulation*, Vol. 23, No. 2, pp. 46-52.

Brennan, T. (2002), *Preventing Monopoly or Discouraging Competition? The Perils of Price-Cost Tests for Market Power in Electricity*, Resources for the Future Discussion Paper 02-05.

Brennan, T. (2003), 'Mismeasuring Electricity Market Power', *Regulation*, Vol. 26, No. 1, pp. 60-65.

Economic Regulation Authority (2008), *Short Run Marginal Cost*, State of Western Australia.

Electricity Industry (Wholesale Electricity Market) Regulations 2004, Western Australia, Available, Available World Wide Web, URL: http://www.eriu.energy.wa.gov.au/

Johnston, J. (1952), 'Statistical Cost Functions in Electricity Supply', *Oxford Economic Papers*, New Series, Vol. 4, No.1, pp. 68-105.

Kim, J. Shin, D. Park. J, & Singh, C. 'Atavistic genetic algorithm for economic dispatch with valve point effect', *Electric Power Systems Research*, 62, 2002, pp 201-207.

Klein, J. (1998), *The Use of Heat Rate Curves in Production Cost Modeling and Market Modeling*, Electricity Analysis Office, California Energy Commission.

Market Procedure for Determining Loss Factors, Wholesale Electricity Market Procedure (2006), Available IMO website: www.imowa.com.au/Attachments/Loss%20Factor%20 Procedure.pdf.

Ministerial Direction under the Electricity Corporations Act 2005, Government of Western Australia, Effective 1 April 2006, Signed 21st March 2006.

Muth, J. (1961), 'Rational Expectations and the Theory of Price Movements', *Econometrica*, Vol. 29, No. 3, pp. 315-335.

National Energy Technology Laboratory, *Power Plant Water Usage and Loss Study*, August 2005

ROAM Consulting (2007), *SRMC STEM Offer Modelling*, Report IMO 00010/MA to Independent Market Operator, 22 May.

Ruff, L. (2002), *Statement of Larry E. Ruff, PhD on California State Senate Bill No. 2000 Unlawful Electric Power and Natural Gas Practices*, Before the Senate Judiciary Committee, April 23.

Schweppe, F., Caramanis, M., Tabors, R. & Bohn, R. (1988), *Spot Pricing of Electricity*, Kluwer Academic Publishers, Boston.

Short, C. & Swan, A. (2002), *Competition in the Australian National Electricity Market*, Australian Bureau of Agricultural and Resource Economics, Current Issues 02.01, January.

Stoft, S. (2002), *Power System Economics*, IEEE Press, New Jersey.

System Management (2007) , *Medium Term Projected Assessment of System Adequacy Report*, August, Available IMO website: http://www.imowa.com.au/10_5_1_n_mt_pasa.htm.

Verve Energy (2007), *STEM Pricing Using PowrSym*, Report No VE 02/07, January.

Verve Energy Home Page, Accessed 24 August 2007, Available: http://www.verveenergy.com.au/subContent/companyProfile/Company_Profile.html.

Wang, H. & Yang, B (2001), "Fixed and Sunk Costs Revisited", *Journal of Economic Education*, Vol. 31, No. 2, Spring, p. 178-185.

Waters, J. Verve Energy, Per com., email, 9th August 2007.

Wholesale Electricity Market Amending Rules (December 2006), Western Australia, Available IMO website: http://www.imowa.com.au/market_rules.htm.

Wholesale Electricity Market Procedure (2006), *Market Procedure for Determining Loss Factors,* Available IMO website: www.imowa.com.au/Attachments/Loss%20Factor%20Procedure.pdf.

Winston, C. (1998), 'U.S. Industry Adjustment to Economic Deregulation', *Journal of Economic Perspectives*, Vol. 12, No. 3, pp. 89-110.

Wood, A. & Wollenberg, B. (1996), *Power Generation, Operation and Control*, 2nd. ed., Wiley-Interscience, New York.

APPENDICES

Appendix 1. Optimisation examples

A1.01 Individual plant models

Consider the following generic function for an individual plant:

$$C_i = C_i(F_i, S_i, Q_i, K_i, M_i; \alpha_{j,i}), \qquad F_i \geq 0,\ M_i \geq 0,\ 0 \leq Q_i < K_i,\ i = 1,\ldots,n,\ j = 1,\ldots,m.$$

where short run total cost (denoted C) for plant i of n plants is a function of its output Q_i, any mingen output M_i, avoidable fixed costs F_i, shutdown costs S_i, the maximum thermodynamic output capacity of the facility K_i (under current ambient conditions), and where $\alpha_{j,i}$ constitutes a set of m parameters for plant i.

The mathematical formulations that follow are intended to provide theoretical descriptions of short run total cost and SRMC for individual plants and are for illustrative purposes only – i.e. the functional forms are not based on actual data. Indeed, the functions were deliberately made strongly convex to better demonstrate the method with a small number of plants (actual plant level SRMC functions may appear less convex). Nonetheless, the specific assumptions that are made – which are transparent within the functional forms themselves – reflect, in a general sense, thermodynamic principles and the law of diminishing returns. That is, the SRMC functions initially decrease with improving thermal efficiency before increasing asymptotically to a maximum output capacity.

The following functional form for plant-specific short run total cost is used forthwith:

$$C_i = \begin{cases} S_i & \text{if } Q_i = 0 \\ \dfrac{\alpha_{1,i}(Q_i - M_i)^3 + \alpha_{2,i}(Q_i - M_i)^2 + \alpha_{3,i}(Q_i - M_i)}{K_i - Q_i} + \alpha_{4,i} + F_i & \text{if } M_i \leq Q_i < K_i \end{cases}$$

Taking the first derivative results in the SRMC function:

$$C_i' = \frac{dC_i}{dQ_i} = \frac{(K_i - Q_i)\left[3\alpha_{1,i}(Q_i - M_i)^2 + 2\alpha_{2,i}(Q_i - M_i) + \alpha_{3,i}\right] + \alpha_{1,i}(Q_i - M_i)^3 + \alpha_{2,i}(Q_i - M_i)^2 + \alpha_{3,i}(Q_i - M_i)}{(K_i - Q_i)^2} \qquad \text{if } M_i \leq Q_i < K$$

for,

$$0 \leq M_i \leq K_i,\ M_i \leq Q_i < K_i.$$

Consider five individual plants where $\{i \mid 1,2,3,4,5\}$. First, a low cost plant with sunk startup costs:

$$C_1 = \begin{cases} 18{,}000 & \text{if } Q_1 = 0 \\ \dfrac{\frac{1}{16}(Q_1-92)^3 - 18(Q_1-92)^2 + 1{,}750(Q_1-92)}{291-Q_1} + 5{,}000 & \text{if } 92 \le Q_1 \end{cases}$$

$$\Rightarrow C_1' = \frac{(291-Q_1)\left[\frac{3}{16}(Q_1-92)^2 - 36(Q_1-92) + 1{,}750\right] + \left[\frac{1}{16}(Q_1-92)^3 - 18(Q_1-92)^2 + 1{,}750(Q_1-92)\right]}{(291-Q_1)^2} \qquad \text{if } Q_1 \ge 92 .$$

Second, a mid cost plant with sunk startup costs:

$$C_2 = \begin{cases} 1{,}400 & \text{if } Q_2 = 0 \\ \dfrac{\frac{1}{4}(Q_2-12)^3 - 56(Q_2-12)^2 + 3{,}200(Q_2-12)}{111-Q_2} + 2{,}500 & \text{if } 12 \le Q_2 \end{cases}$$

$$\Rightarrow C_2' = \frac{(111-Q_2)\left[\frac{3}{4}(Q_2-12)^2 - 112(Q_2-12) + 3{,}200\right] + \left[\frac{1}{4}(Q_2-12)^3 - 56(Q_2-12)^2 + 3{,}200(Q_2-12)\right]}{(111-Q_2)^2} \qquad \text{if } Q_2 \ge 12 .$$

Third, a mid cost plant that has not yet been started, technologically identical to plant 2:

$$C_3 = \begin{cases} 0 & \text{if } Q_3 = 0 \\ \dfrac{\frac{1}{4}(Q_3-12)^3 - 56(Q_3-12)^2 + 3{,}200(Q_3-12)}{111-Q_3} + 2{,}500 + 1{,}100 & \text{if } 12 \le Q_3 < 111 \end{cases}$$

$$\Rightarrow C_3' = \frac{(111-Q_3)\left[\frac{3}{4}(Q_3-12)^2 - 112(Q_3-12) + 3{,}200\right] + \left[\frac{1}{4}(Q_3-12)^3 - 56(Q_3-12)^2 + 3{,}200(Q_3-12)\right]}{(111-Q_3)^2} \qquad \text{if } Q_3 \ge 12 .$$

Fourth, a high cost plant with sunk startup costs:

$$C_4 = \begin{cases} 1{,}400 & if \;\; Q_4 = 0 \\ \dfrac{\frac{1}{2}(Q_4 - 7)^3 - 100(Q_4 - 7)^2 + 3{,}400(Q_4 - 7)}{50 - Q_4} + 2{,}600 & if \; 7 \le Q_4 < 50 \end{cases}$$

$$\Rightarrow C_4{}' = \frac{(111 - Q_4)\left[\frac{3}{2}(Q_4 - 7)^2 - 200(Q_4 - 7) + 3{,}400\right] + \left[\frac{1}{2}(Q_4 - 7)^3 - 100(Q_4 - 7)^2 + 3{,}400(Q_4 - 7)\right]}{(50 - Q_4)^2} \qquad if \;\; Q_4 \ge 7\,.$$

Fifth, a high cost plant that has not yet been started, technologically identical to plant 4:

$$C_5 = \frac{\frac{1}{2}(Q_5 - 7)^3 - 100(Q_5 - 7)^2 + 3{,}400(Q_5 - 7)}{50 - Q_5} + 2{,}600 + 1{,}400 \qquad if \; 7 \le Q_5 < 50$$

$$\Rightarrow C_5{}' = \frac{(111 - Q_5)\left[\frac{3}{2}(Q_5 - 7)^2 - 200(Q_5 - 7) + 3{,}400\right] + \left[\frac{1}{2}(Q_5 - 7)^3 - 100(Q_5 - 7)^2 + 3{,}400(Q_5 - 7)\right]}{(50 - Q_5)^2} \qquad if \;\; Q_5 \ge 7\,.$$

Geometric representations of the functions are provided on the following page (Figure A1.1 through to Figure A1.10).

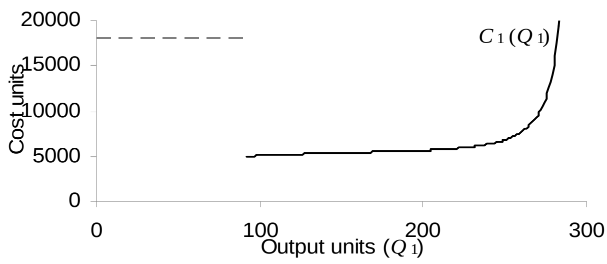


Figure A1.1

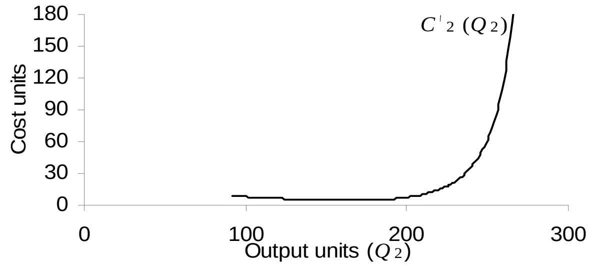


Figure A1.2

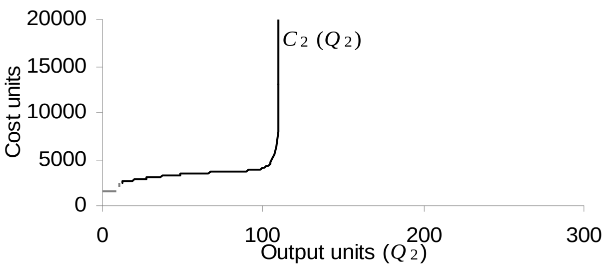


Figure A1.3

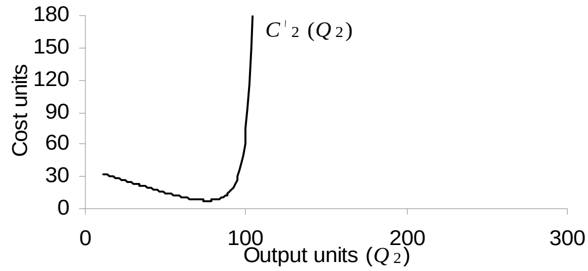


Figure A1.4

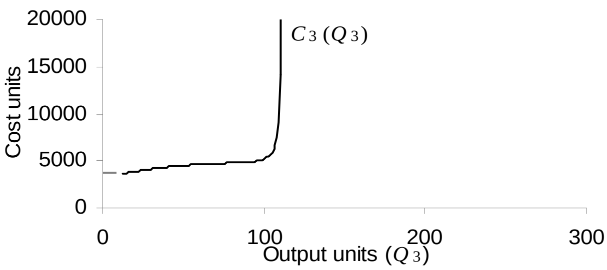


Figure A1.5

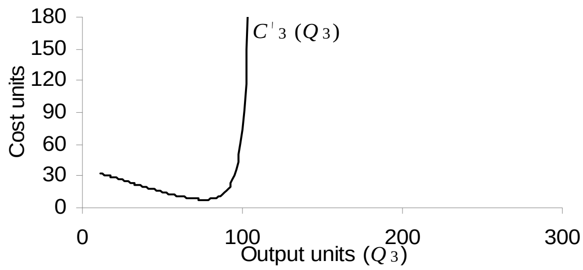


Figure A1.6

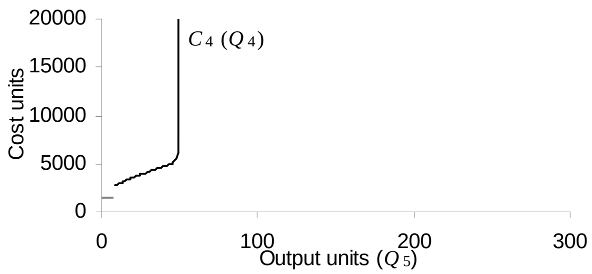


Figure A1.7

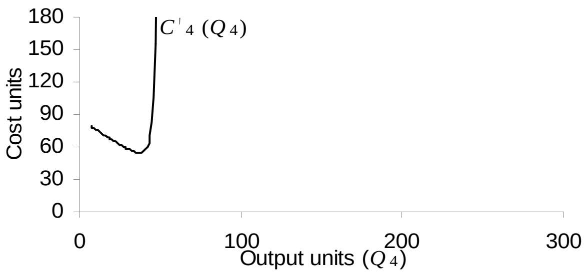


Figure A1.8

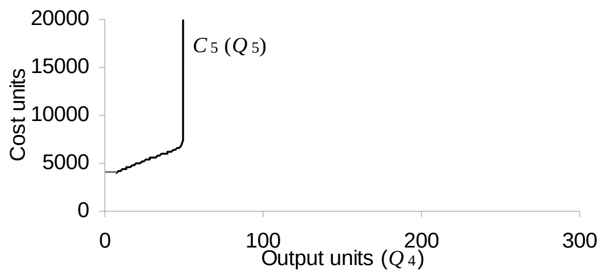


Figure A1.9

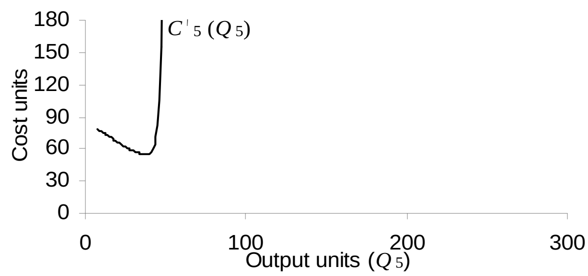


Figure A1.10

A1.02 Model of a portfolio comprising plants with sunk costs

Consider a simple portfolio of three plants corresponding to those with sunk startup costs given in Section A1.01: i.e. plant 1, plant 2 and plant 4. Portfolio short run total cost $C_{\{1,2,4\}} = C(\sigma)$, is a composite function that combines the short run total cost functions of each plant $C_1(Q_1)$, $C_2(Q_2)$ and $C_4(Q_4)$ at a portfolio quantity of total output $\sigma = Q_1 + Q_2 + Q_4$.

Developing the optimal curve $C_{\{1,2,4\}}(\sigma)$ is a two step process. First, the function $C_{\{1,2\}}(\sigma)$ provides an interim curve that combines the low cost and the mid cost functions $C_1(\sigma)$ and $C_2(\sigma)$ as a composite function with $Q_4 = 0$, i.e. $\sigma = Q_1 + Q_2$. This is performed as follows:

$$C_{\{1,2\}}(\sigma, p) = C_1(p\sigma) + C_2[(1-p)\sigma],$$

where p is the proportion of low cost plant production in the portfolio for the half hour period, i.e. $Q_1 = p\sigma$ and $Q_2 = (1-p)\sigma$. The optimal interim curve is then developed by minimising incremental $C_{\{1,2\}}(\sigma)$ as follows:

$$\begin{aligned} \frac{d}{dp}C_{\{1,2\}} &= C_1^{'}(p\sigma) - C_2^{'}\left[(1-p)\sigma\right] \\ &= 0. \end{aligned}$$

Therefore:

$$C_1^{'}(p\sigma) = C_2^{'}\left[(1-p)\sigma\right],$$

which is the equimarginal principle of economics.

Let p^0 be the optimal proportion of output from plant 1 for any given σ and $C^0_{\{1,2\}}(\sigma)$ be the optimised total cost function of plants 1 and 2, then:

$$C^0_{\{1,2\}}(\sigma) = C_1(p^0\sigma) + C_2\left((1-p^0)\sigma\right).$$

A graphical depiction of a spreadsheet model of the optimal interim curve is provided by Figure A1.11.

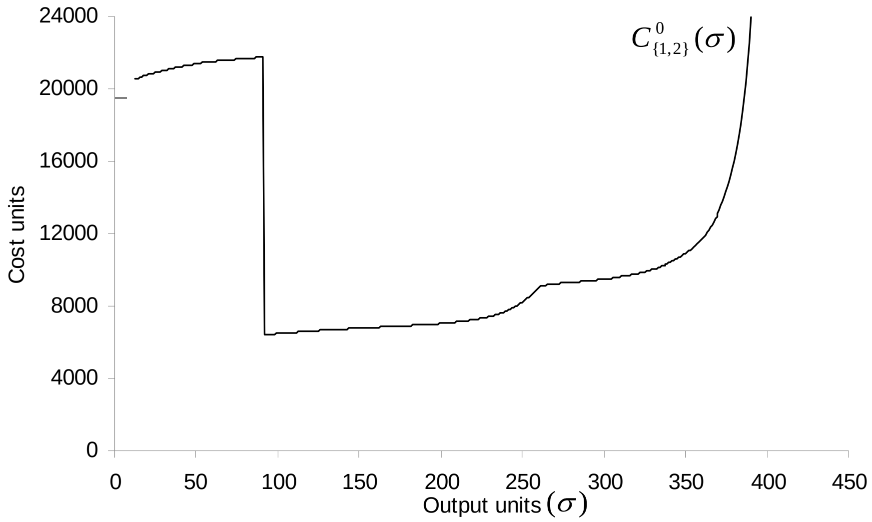


Figure A1.11

The second step is similar to the first, where the optimal interim curve $C^0_{\{1,2\}}(\sigma)$ is combined with $C_4(\sigma)$ to form a composite function that describes portfolio short run total cost $C_{\{1,2,4\}}(\sigma)$. This is performed as follows:

$$C_{\{1,2,4\}}(\sigma, p) = C_4(p\sigma) + C^0_{\{1,2\}}\big((1-p)\sigma\big)$$

where p is now the proportion of high cost plant production in the portfolio for the half hour period. The optimal portfolio short run total cost curve is then developed by minimising incremental $C_{\{1,2,4\}}(\sigma)$ as follows:

$$\frac{d}{dp}C_{\{1,2,4\}} = C_4{}'(p\sigma) - C^0_{\{1,2\}}{}'\big[(1-p)\sigma\big]$$

$$= 0$$

$$\therefore \qquad C_4{}'(p\sigma) = C^0_{\{1,2\}}{}'\big[(1-p)\sigma\big]$$

Again, this corresponds to the equimarginal principle of economics. It follows:

$$C^0_{\{1,2,4\}}(\sigma) = C_4\big(p^0\sigma\big) + C^0_{\{1,2\}}\big((1-p^0)\sigma\big),$$

where p^0 is the optimal proportion of output from plant 4 for any given σ and $C^0_{\{1,2,4\}}(\sigma)$ is the optimised total cost function of plants 1, 2 and 4.

A spreadsheet model of the optimal portfolio short run total cost curve is provided by Figure A1.12.

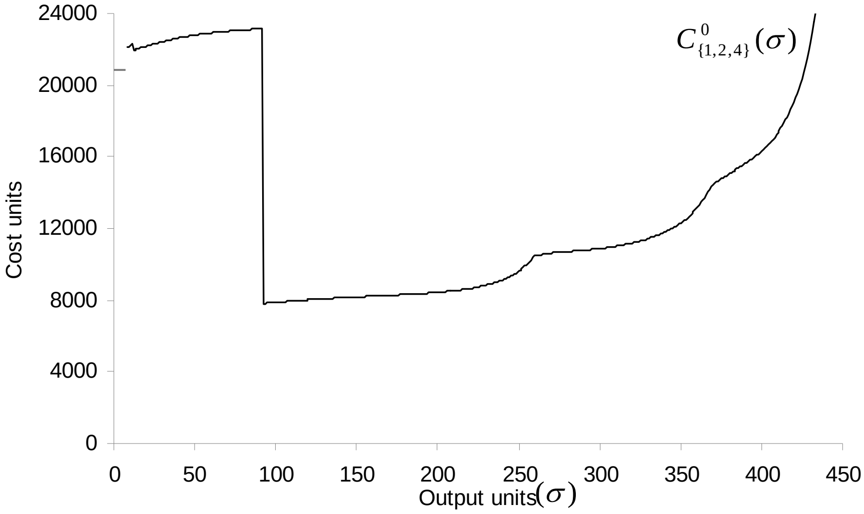


Figure A1.12

In Figure A1.13, a spreadsheet approximation of $\frac{d}{d\sigma}C^{0}_{\{1,2,4\}}$ is given by $\frac{\Delta C^{0}_{\{1,2,4\}}}{\Delta\sigma}$ using $\Delta\sigma = 1$ when $\sigma > 92$, and $\Delta\sigma = 92$ with $\Delta C^{0}_{\{1,2,4\}} = C^{0}_{\{1,2,4\}}(0) - C^{0}_{\{1,2,4\}}(92)$ when $0 \le \sigma \le 92$.

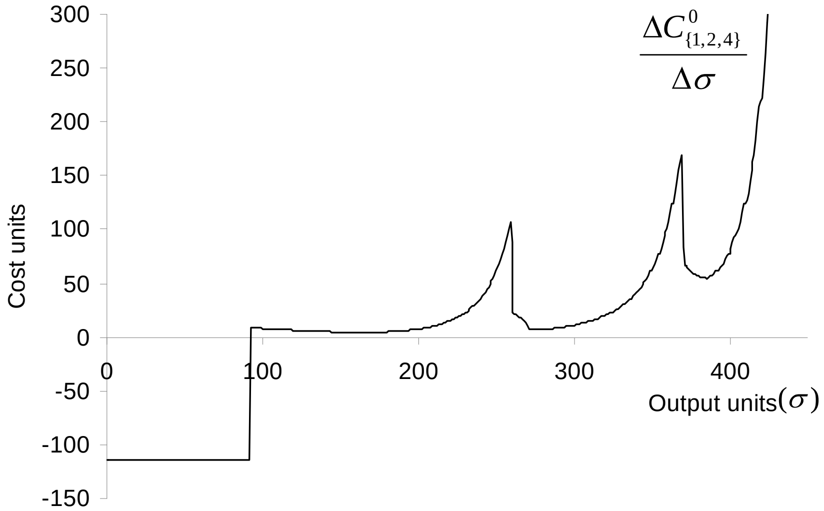


Figure A1.13

A1.03 Portfolio model with a mix of sunk cost and avoidable fixed cost plants

Consider a portfolio of four plants plant 1, plant 2 and plant 3 and plant 5 described in Section A1.01. Portfolio short run total cost $C_{\{1,2,3,5\}} = C(\sigma)$, is a composite function that combines the short run total cost functions of each plant $C_1(Q_1)$, $C_2(Q_2)$, $C_3(Q_3)$ and $C_5(Q_5)$ at a portfolio quantity of total output $\sigma = Q_1 + Q_2 + Q_3 + Q_5$.

The optimal interim curve from Section A1.02, $C^0_{\{1,2\}}(\sigma)$, which describes the optimal combination of output from plant 1 and plant 2, is used as a baseline. Developing the portfolio short run total cost function $C_{\{1,2,3,5\}}(\sigma)$ is a two step process. First, the function $C_{\{1,2,3\}}(\sigma)$ combines the functions $C^0_{\{1,2\}}(\sigma)$ and $C_3(\sigma)$, creating a composite function with $Q_5 = 0$, i.e. $\sigma = Q_1 + Q_2 + Q_3$. This is performed as follows:

$$C_{\{1,2,3\}}(\sigma, p) = C_3(p\sigma) + C^0_{\{1,2\}}\left[(1-p)\sigma\right],$$

where p is the proportion of production from plant 3 in the portfolio for the half hour period, i.e. $Q_3 = p\sigma$ and $Q_1 + Q_2 = (1-p)\sigma$. The optimal curve is then developed by minimising incremental $C_{\{1,2,3\}}(\sigma)$ as follows:

$$\frac{d}{dp}C_{\{1,2,3\}} = C_3{}'(p\sigma) - C^0_{\{1,2\}}{}'\left[(1-p)\sigma\right]$$

$$= 0$$

which corresponds to the equimarginal principle of economics:

$$C_3{}'(p\sigma) = C^0_{\{1,2\}}{}'\left[(1-p)\sigma\right].$$

The optimal proportion of output from plant 3, denoted p^0, determines the optimised short run total cost function, $C^0_{\{1,2,3\}}(\sigma)$, of plants 1, 2 and 3, as follows:

$$C^0_{\{1,2,3\}}(\sigma) = C_3\left(p^0\sigma\right) + C_{\{1,2\}}\left((1-p^0)\sigma\right).$$

A spreadsheet model of the optimal short run total cost curve $C^0_{\{1,2,3\}}(\sigma)$ is provided by Figure A1.14.

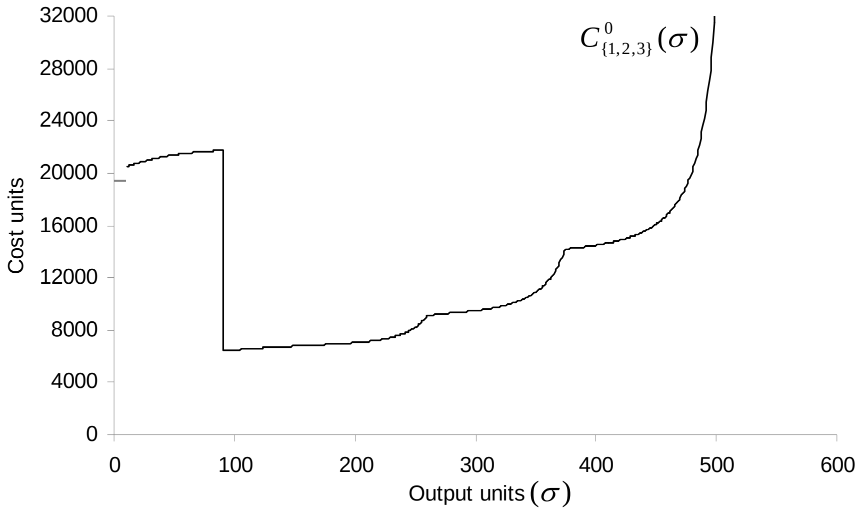


Figure A1.14

The second step is similar to the first step, where $C^0_{\{1,2,3\}}(\sigma)$ is combined with $C_5(\sigma)$ to form the composite portfolio short run total cost function $C_{\{1,2,3,5\}}(\sigma)$:

$$C_{\{1,2,3,5\}}(\sigma, p) = C_5(p\sigma) + C^0_{\{1,2,3\}}\left((1-p)\sigma\right)$$

where p is now the proportion of high cost plant (i.e. plant 5) production in the portfolio for the half hour period, i.e. $Q_5 = p\sigma$ and $Q_1 + Q_2 + Q_3 = (1-p)\sigma$. The optimal portfolio short run total cost curve is then developed by minimising incremental $C_{\{1,2,3,5\}}(\sigma)$ as follows:

$$\frac{d}{dp}C_{\{1,2,3,5\}} = C_5{'}(p\sigma) - C^0_{\{1,2,3\}}{'}\left[(1-p)\sigma\right]$$

$$= 0$$

$$\therefore \qquad C_5{'}(p\sigma) = C^0_{\{1,2,3\}}{'}\left[(1-p)\sigma\right].$$

It follows:

$$C^0_{\{1,2,3,5\}}(\sigma) = C_5\left(p^0\sigma\right) + C^0_{\{1,2,3\}}\left((1-p^0)\sigma\right),$$

where p^0 is the optimal proportion of output from plant 5 for any given σ and $C^0_{\{1,2,3,5\}}(\sigma)$ is the optimised total cost function of plants 1, 2, 3 and 5.

A spreadsheet model of the optimal portfolio short run total cost curve is provided by Figure A1.15.

Figure A1.15

In Figure A1.16, a spreadsheet approximation of $\frac{d}{d\sigma}C^0_{\{1,2,3,5\}}$ is given by $\frac{\Delta C^0_{\{1,2,3,5\}}}{\Delta\sigma}$ using $\Delta\sigma = 1$ when $\sigma > 92$, and $\Delta\sigma = 92$ with $\Delta C^0_{\{1,2,4,5\}} = C^0_{\{1,2,4,5\}}(0) - C^0_{\{1,2,4,5\}}(92)$ when $0 \le \sigma \le 92$.

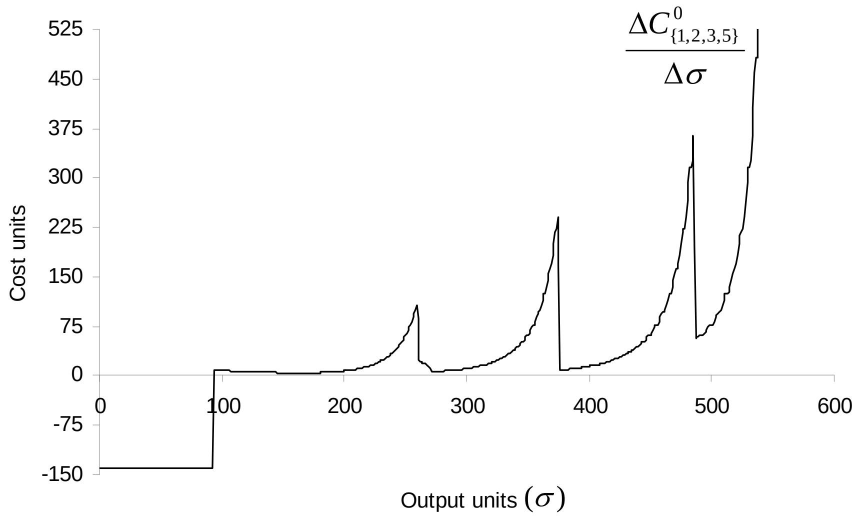


Figure A1.16