REVIEW OF MAE RULES FOR THE BRAZILIAN
WHOLESALE ELECTRICITY MARKET

Prepared by
Shmuel S. Oren, Ph.D.
1293 Alvarado, Rd.
Berkeley, California, 94705 USA
oren@ieor.berkeley.edu

This report was prepared for ANEEL,
The Brazilian Electricity Regulatory Agency under project
BRA/98/019
United Nations Development Programme
Contract No. 2000/003150

June 9, 2000
SUMMARY

This report provides a review of the proposed MAE rules for the Brazilian wholesale electricity market as described in the document distributed by ASMAE. The focus of this report is on two topics:

1. The process for setting the wholesale market price for electricity referred to as the "MAE purchase price" described in Chapter 3 of the MAE rules.

2. The rationale and calculation of capacity payments to generators, which are augmented on a prorata basis to the wholesale purchase price to produce the MAE selling price of electricity. These calculations are described in Chapter 11 and in Appendix G of the MAE rules.

The following recommendations are made:

1. If MAE prices are based on shadow prices from DESSEM, guarantee that every dispatched generator will cover its operating cost (incremental cost + no load + startup costs) in a specified scheduling interval (e.g. 24 hours). The calculation of incurred startup and no-load costs and of revenue shortfalls should be based on startup and no-load cost parameters declared by generators for a long time period (e.g. a week, a month or longer). Make up payments can be either set to the revenue shortfall in the scheduling interval or set to reimburse each generator for its incurred startup and no-load cost on top of its energy revenues that are based on the shadow prices. The cost of lump sum compensation to generators should be recovered through a service charge (uplift).

2. If MAE prices are not based on DESSEM shadow prices then initial schedule prices in each period should be based only on incremental cost of the most expensive unit operating in the submarket or importing energy to the submarket after excluding units that operate due to intertemporal constraints such as ramp rate limit or inflexibility. Start-up and no-load cost parameters (fixed over weekly or monthly periods) should be used to calculate specific startup and no-load costs
of each operating unit. Generators should receive lump sum payments covered by uplift charges to either offset revenue shortfalls over a 24 hour period or to reimburse each generator for its incurred start-up and no-load costs. Excess cost of unit operating out of merit in M type periods due to ramp constraints should be loaded to the peak periods and reflected as a peak price addition to the initial schedule price (as proposed).

3. MAE should switch to a multi-settlement system in which the ex-ante prices are applied to scheduled generation based on ONS day ahead schedule whereas ex-post prices based on actual dispatch and redeclared generator parameters will be applied to deviations of the actual dispatch from the day ahead schedule. In view of the high percentage of hydro in the Brazilian system a three-settlement system will probably not have much benefit beyond what a two settlement system can provide. Such an approach provides market incentives for participants to respond efficiently to uncertain demand and supply. Moreover, it mitigates incentives for gaming and reduces uncertainty for generators and buyers. The proposed approach also has the effect of replacing administrative penalties for deviation with market-based penalties.

4. Exempt eligible demand side bids (whether dispatched or not) from the capacity fee component of the MAE price. This will provide the proper incentive for demand side participation in shortage mitigation. Since demand side bids forgo the protection provided by the capacity fee they should be exempt from that fee. Since it is assumed in MAE pricing that contracts already include capacity payments the above rationale suggests that contracted demand that is bid for curtailment at a specified incremental cost should be paid the capacity fee portion of the MAE price when it is not dispatched and should be paid the full MAE selling price when dispatched (unless the interruptibility provision is part of the contract in which case one should assume that such payments have been accounted for in the contract).
5. Eliminate capacity payment component from MAE price and impose a requirement that load serving entities supplement their forward contract with energy call options or one sided contract for difference up to \((1+X)\) of their monthly peak load, where \(X\) is the planning reserves requirement. The prices for the call option are negotiated directly between the load and the generators. Load serving entities can meet their contracting obligation outside their submarket but the reference MAE price for a call option and the exercise point correspond to submarket of the load. In other words the load can contracts with a generator in another submarket to deliver power in its submarket when the MAE price reaches the strike price or it can settle financially. This approach solves the problem of distribution of capacity payments across submarkets. Generators that cannot meet their contractual obligations during interruptions are liable for VLL while generators that generate above their contractual obligations during shortfalls should receive VLL for their surplus generation.
1. Introduction

This report will provide a review of the proposed MAE rules for the Brazilian wholesale electricity market as described in the document distributed by ASMAE. The focus of this report is on two topics:

1. The process for setting the whole sale market price for electricity referred to as the "MAE purchase price" described in Chapter 3 of the MAE rules.

2. The rationale and calculation of capacity payments to generators, which are augmented on a prorata basis to the wholesale purchase price to produce the MAE selling price of electricity. These calculation are described in Chapter 11 and in Appendix G of the MAE rules.

In my discussion below I will attempt to summarize and examine the conceptual framework of the price setting and capacity payment calculation without going into the detailed mathematical formulae. My analysis will focus on the economic rationale for the different price components, the validity of the economic signals from an efficiency and equity perspective and on incentives and gaming opportunities created by the proposed pricing rules. Based on my analysis I will provide comments and recommendation for improvement.

The ASMAE is responsible for accounting, pricing and other contractual matters for electricity that flows through the system managed by ONS, the Brazilian system operator. ONS dispatches all plants participating in the electricity market using a hydro-thermal unit commitment optimization program (DESSEM) which is designed to optimize the dispatch schedule of all hydroelectric and thermal generation plans over a rolling horizon of 168 half hourly intervals (to my knowledge this model is not yet operational). DESSEM employs hydrological information and water values produced by a medium term planning model (DECOMP) and a long term planning model (NEW WAVE). While there is no formal auction in the Brazilian energy market, DESSEM dispatches are based on inputs from generators who declare cost parameters, dispatch constraints and availability of their units on a day ahead basis for each settlement period.
(half hour) and can redeclare availability over the course of the day. Hence, in principle generators day ahead declarations can be viewed as bids in a multipart auction which are used for optimal dispatch by ONS and for financial settlements by MAE. Specifically, for thermal generators, these "bids" that are used in DESSEM specify: Incremental energy cost, Start-up cost, No load cost, Must run levels, Ramp-up and Ramp-down constraints and Availability. Generators also disclose their initial supply contracts but that information should not affect the dispatch. An interesting feature of DESSEM and the longer term planning models is the use of rationing as if it was a supply technology. This is implemented by creating a supply function for "rationing" based on economic analysis of rationing cost. Introducing such a synthetic "rationing technology" will have the desired affect of reducing the use of water during water shortage periods which is reflected by high water value. However, such a scheme will have the desired economic efficiency results only if that rationing technology is indeed treated as any other generator and allowed to set market prices. A similar approach can be used to incorporate demand side bids into the scheduling algorithm or assign an incremental cost of VLL to involuntary curtailment.

Since DESSEM shadow prices form the basis for MAE prices it is important to understand what is accounted for (or will be accounted for when the program is fully operational) in that model. Typically, a hydro-thermal unit commitment optimization minimizes over the 168 scheduling periods an objective function that includes incremental thermal generation cost, startup costs, no load costs, the cost of water used for hydro generation (by reservoir), the cost of rationing and of demand side bids summed over all dispatchable resources. Startup costs of thermal units will typically depend on the length of time a unit has been turned off hence the initial conditions of each plant must be accounted for in the objective function for each time the optimization time window is reset. The Brazilian system is segmented into four submarkets that are connected through congestion prone transmission lines. Hence, the dispatch optimization is subject to the following types of constraints.
a) Hydro water balance constraints, Reservoir and river flow constraints, Hydro production functions and other specific hydro constraints by submarkets, reservoir and individual generator

b) Supply and demand balance including imports and exports in each submarket

c) Ramp, availability, min-max generation and inflexibility constraints on each units.

d) Transmission constraints on flows between submarkets (transmission constraints within submarkets are ignored in this optimization but are accounted for in the ONS real time dispatch).

Typical unit commitment optimization models also include a spinning reserve constraint requiring that maximum generation level of committed plants exceed demand by some specified percentage (typically 7%). According to information I obtained from the developers of DESSEM, however, such a constraint is not important in Brazil due to its heavy reliance on Hydro that can be ramped up quickly.

The above optimization problem is solved in DESSEM by a Lagrangian relaxation method that will produce an optimal schedule for each dispatchable resource accounting for all the cost components and constraints. As a byproduct the optimization produces shadow prices on the demand-supply balance constraints that in each submarket for each half hour (settlement period). It is important to realize that these shadow prices will account for any necessary adjustment to incremental costs due to intertemporal constraints such as ramp rate constraints or various intertemporal hydro constraints. The shadow prices will also account for transmission constraints, availability and inflexibility. However, the shadow prices will not reflect fixed costs such as startup and no load costs, unless a unit needs to be turned on at the margin.
2. Determining the MAE Price.

The Brazilian wholesale electricity market is divided into four sub-markets and transmission constraints within sub-markets are ignored for the purpose of price setting. Payments arising from constraining generators because of within sub-market congestion are settled ex-post with generators and are recovered through the system service charge.

Agents can enter into bilateral contracts and also participate in an insurance mechanism called the energy reallocation mechanism (ERM) which offers quantity certainty to generators through assured energy sales levels. Generators dispatched above their assured energy sales are paid only cost for the excess generation while those producing below this level receive an operating margin for the difference. Agents are net sellers in the MAE if their metered generation is more than their initial contracted generation. They are buyers otherwise. In case of rationing all contracted quantities are reduced proportionally so that contracted and spot transactions are equally exposed to rationing risk.

The MAE rules specify two alternative approaches to setting the MAE wholesale electricity purchase price. That price is then augmented by capacity fees to obtain the MAE selling price paid to uncontracted generation. Demand is charged the MAE purchase price and a system service charge. The capacity fees which will be discussed later provide for capacity payments to generators. The first approach to setting the MAE purchase price sets the half-hourly purchase price for electricity in each submarket to the corresponding shadow price produced by DESSEM. The alternative procedure employs the scheduling and dispatch information provided by ONS and calculates the prices through a series of steps and adjustments that will be discussed below. As we shall see the two approaches are inconsistent in the way they account for startup and no load costs. In either case prices are calculated twice. Ex-ante prices based on day ahead forecasted loads, forecasted inflows and declared availabilities are only used as indicative prices. All price calculations are repeated ex-post with actual load information, actual inflows and redeclared availabilities to obtain the ex-post prices that are used for settlements. The issue of ex-ante vs. ex-post pricing, its implications and recommendations will be addressed later.
2.1 Shadow Price Option:

As indicated above DESSEM shadow prices employed by the first pricing option will reflect adjustment due to intertemporal constraints. So for instance, the high incremental cost of a unit that is ramping up to meet peak demand is automatically shifted to the peak period. Likewise, the incremental cost of an out of merit inflexible unit will be spread over other periods so that the inflexibility cost is shifted to time periods that benefit from operating the inflexible unit. The shadow prices, however, do not reflect fixed costs incurred by generators such as startup and no load. Consequently, some dispatched units may not be able to cover their operating costs if they are compensated for the energy produced based on the shadow prices. One possible approach is to perform a revenue adequacy test for each generator and make lump sum payments to generators so as to cover their revenue shortfall. This approach may have some incentive compatibility problems, however, those can be overcome by limiting how often generators can revise their no load and start up costs. A scheme that will accomplish the revenue adequacy objective but is more generous to generators is to pay to each generators a supplement (above energy revenues that are based on the shadow prices) which will reimburse them for the start-up cost and no-load cost they incur. An important design issue in implementing such a "makeup" procedure is the time interval used for calculation of the revenue adequacy. In PJM, for instance dispatched generators are guaranteed to break even in every 24 hour period. The money needed for lump sum payments to generators that do not cover their startup and no load costs is raised as uplift on energy selling prices.

Recommendation:

If Mae prices are based on shadow prices from DESSEM, guarantee that every dispatched generator will cover its operating cost (incremental cost + no load + startup costs) in a specified scheduling interval (e.g. 24 hours). The calculation of incurred start-up and no load costs and of revenue shortfalls should be based on startup and no-load cost parameters declared by generators for a long time period (e.g. a week, a month or longer). Make up payments can be either set to the revenue shortfall in the scheduling
interval or set to reimburse each generator for its incurred start-up and no-load cost on top of its energy revenues that are based on the shadow prices. The cost of lump sum compensation to generators should be recovered through a service charge (uplift).

2.2 MAE Pricing

The alternative approach to using DESSEM shadow prices is to compute the MAE prices directly from the raw information submitted by generators and the ONS scheduling. In this procedure the indicative or ex-ante purchase price is set to the maximum of *schedule prices* of generating stations in a sub-market and the sub-market import prices. We will examine how each is calculated in turn.

Schedule prices are designed to include start-up, no-load, variable or incremental costs incurred by generators and any adjustments made for peak periods. Schedule prices are set in two stages. Initial schedule prices are set to reflect only costs incurred by generators. Adjustments are then made for peak periods and periods in which a generator was changing load at its maximum rate. This usually happens when a generator is being ramped up to or down from a peak period in which it is required that this generator is running. During ramp-up and ramp-down, the generator may be operating out-of-merit and is not allowed to set the MAE price.

To determine how start-up and no-load costs will be allocated, generator flexibility over start-up periods is considered. Start-up periods are a continuous set of periods preceded and followed by periods with no generation. Generators submit a "must run" (MR) level of generation in each settlement period, a half hour interval, which is used to determine flexibility. Initial MR flags are set to 'inflexible' if MR levels are strictly positive in any settlement period and stay inflexible through the end of the start-up period. They are set to 'flexible' otherwise. Final MR flags are set to next period's flag if both period have strictly positive generation. They are set to initial MR flags otherwise. These rules imply that a unit which declares inflexibility in any one settlement period in a start-up period will be considered inflexible for the entire start-up period. Thus, a unit could have a MR level of zero in a settlement period and still be considered inflexible because it declared inflexibility in some other period.
Start-up costs are only considered if a unit is flexible. They are assigned to the first settlement period in a start-up period, though the schedule price in any period reflects the average start-up cost over the start-up period. No-load cost is accounted for in those periods in which generators declare a MR level of zero. Incremental costs are accounted for when a unit is flexible or is producing above its MR level. They are not compensated when a unit is running at its MR level.

In periods where the generator is flexible, *initial schedule price* is set as the sum of loss factor adjusted incremental cost, no-load cost per MWh in a settlement period and start-up cost per MWh over the start-up period to which the settlement period belongs. Start-up costs of inflexible generators are not included in the initial schedule price. No-load costs are accounted for those periods with a MR level of zero and incremental costs are accounted for those periods in which the generator is scheduled to produce strictly above its MR level.

To determine if a generator will be considered in setting the MAE price in any settlement period, periods are classified as peak (P), maximum load change (M) or a trough (T) for each generator. Period types are also used in accounting for peak price additions. Essentially, P periods are those, which are reached with maximum rate load changes and lowered from with maximum rate load changes. Exceptions are when generators are being ramped-up down from P periods in adjacent settlement days. Peak price additions are not allowed across settlement days and periods in which a generator is ramped-up to meet a peak in the next settlement day or where it is ramped-down from a peak in the previous settlement day are considered as P periods. M periods are those, which take place at maximum rate load change to meet or lower from a peak period in the same settlement day. A T period is one for which the first period before it where load changes at maximum rate is when load is being reduced and the first period after it where load changes at maximum rate is when load is increasing.

The ex-ante MAE purchase price is set in two stages. First, initial schedule prices of generators with period type flags set to P or T are considered in calculating an unadjusted MAE price. The unadjusted MAE price is set at the maximum of these initial
schedule prices in a sub-market and sub-market import prices, which are loss adjusted purchase prices of sub-markets from where this sub-market is importing energy.

In the second stage excess costs over the unadjusted MAE price of generators in M periods are loaded on to their P periods using what is called a peak price addition. MAE prices are then set using initial schedule prices for generators in T periods and peak price addition adjusted initial schedule prices for generators in P periods. Generators in M periods are not considered in setting the MAE price. To ensure that generators are not paid these excess costs twice an ex-post deduction of the difference in the derived MAE price and the unadjusted MAE price is made from generation in M periods.

The indicative MAE selling price is set to the indicative MAE purchase price, however that price has no significance to the ultimate financial settlements. The ex-post MAE selling price which is used in the financial settlements is calculated by adding the ex-post capacity fee to the ex-post MAE purchase price. The selling price is set to a proportion of the value of lost load if a flag for interruption of service is active in a sub-market.

Some of adjustments described above dealing with inflexibility declarations and active ramp constraints parallel the calculations that are performed automatically in computing the DESSEM shadow prices. These adjustments can be justified on the grounds that cost consequences of intertemporal constraints should be shifted to the periods that benefit from the intertemporal optimization. For example the cost of dispatching inflexible units or ramping units in periods were in which they are out of merit should be charged against periods that benefit from the operation of those units. It thus follows that such units should not set the prices during periods in which they wouldn't be dispatched if it had not been for the intertemporal constraints on their operation. By contrast, the inclusion of no-load and start-up cost in the calculation of the schedule prices is at odds with the shadow price approach and has no economic justification. Even after all the exclusions that calculation has the implication of setting market clearing prices based on a time averaged rather than the instantaneous marginal cost of the most expensive unit. A spot price that is based on a time averaged cost is in
some sense a contradiction and a distortion of efficient economic signal for production and consumption of electricity.

While accounting for startup and no load costs is important for efficient multi-period dispatch these should be treated as sunk cost in calculating the marginal cost of meeting incremental demand which sets the efficient price signal in each period. It is legitimate to account for no-load and start-up cost to verify the revenue adequacy of each dispatched resource and provide lump sum payments to generators that do not cover their startup and no load costs. However, it is economically inefficient to implement fixed cost recovery by raising the spot price. Such an approach distorts marginal cost signals and rewards those generators that do not incur the fixed cost and already enjoy inframarginal profits. There is no economic efficiency reason neither incentive argument that will justify paying more to a nuclear plant that operates without interruption or to a hydro unit that does not incur any no-load cost just because some peaking units on the margin need to be scheduled as a two-shift unit. An alternative approach that will also accomplish the revenue adequacy objective without distorting marginal cost, is to simply reimburse each generator for its incurred start-up and no-load costs on top of its energy revenues. The energy revenues will be based on incremental cost and adjusted for intertemporal constraints while the reimbursement will be based on the specific no-load and start-up cost parameter submitted by each generator and on its dispatch schedule. This approach is more generous toward generators since the fixed cost reimbursement is paid even when a generator makes a profit on its revenue sales. The choice between the two alternatives amounts to a transfer between consumers and producers which is a matter of equity with no efficiency implications.

**Recommendation:**

*If MAE prices are not based on DESSEM shadow prices then initial schedule prices in each period should be based only on incremental cost of the most expensive unit operating in the submarket or importing energy to the submarket after excluding units that operate due to intertemporal constraints such as ramp rate limit or inflexibility. Start-up and no-load cost parameters (fixed over weekly or monthly periods) should be*
used to calculate specific start-up and no-load costs of each operating unit. Generators should receive lump sum payments covered by uplift charges to either offset revenue shortfalls over a 24 hour period or to reimburse each generator for its incurred start-up and no-load costs. Excess cost of unit operating out of merit in M type periods due to ramp constraints should be loaded to the peak periods and reflected as a peak price addition to the initial schedule price (as proposed).

3 Settlement

Generators’ payments are made up by the following accounts:

1. MAE payments for energy (based on ex-post MAE price)
2. Constraint payments for ERM and non-ERM plants
3. Net capacity payments
4. Availability test compensation
5. Penalty payments
6. Peak price balancing item
7. Net consumption

3.1 MAE Payments for Energy:

MAE payments are determined by net generation, which is metered generation plus ERM adjustment (which may be negative) less contracted generation. The ex-post selling price is used to compensate generators who are net sellers.

3.2 ERM:

The ERM is an insurance type mechanism that ensures assured energy levels to generators provided total generation in the ERM was sufficient to cover assured energy levels. This ensures quantity certainty and settlements are based on ex-post prices. The process works by allocating energy from generators who produced more than their assured energy sales levels to those that produced less. Energy is allocated within a sub-market first and then allocations are made across sub-markets. The process takes place monthly, but is carried out for each settlement period in the relevant month.
Total metered generation for ERM plants above total assured energy levels is called secondary energy. Secondary generation is allocated to all generators in the ERM after all generators are brought up to their assured energy sales levels. If this is zero, i.e. if total metered generation is less than or equal to the sum of assured energy levels then assured energy levels for individual power stations are scaled down proportionally.

If there is no secondary energy, allocation is not difficult. Each power station will be allocated energy or donate energy to bring them to their scaled assured energy levels. This will leave no surplus and the allocation process is over.

The details of the allocation process become important because of the rules governing allocation of secondary energy levels. Any generation by a power station in excess of scaled assured energy levels that is not allocated to generators within its sub-market or in other sub-markets is deemed secondary energy produced by that particular power station. The power station will then have rights to 50% of that energy which is allocated to it directly. The remaining 50% is pooled and distributed among all power stations in proportion to their scaled assured energy levels.

Total variable costs of power stations who are net donors into the ERM are compensated by those who are net receivers over the period of a month. Costs to be used in this calculation are determined by ANEEL and include variable costs and royalty costs in the case of hydro plants.

3.3 Constraint Payments:

If metered generation is above ex-post unconstrained scheduled generation it means that the generator was constrained-on for purposes of congestion relief within its sub-market. If the generator's metered price, which is calculated in the same way as schedule price but using actual generation, is lower than the MAE selling price the generator is compensated for this added cost on the excess generation. If metered generation is lower than ex-post unconstrained scheduled generation and if metered price is lower than the MAE selling price then the generator is compensated for the missed
opportunity of selling at the MAE selling price for the lowered amount of energy. Separate accounts are kept for ERM generators and non-ERM generators.

3.4 Capacity Payments:

Capacity payments are made to generators on uncontracted, undispatched capacity. Ex-post this value is equal to the total availability less total generation less ERM adjustment multiplied by the ex-post capacity fee. Generators are considered to have recovered capacity payments on contracted generation and therefore only part of this is paid to the generators ex-post. If cumulative ex-post capacity payments are more than 80% of the cumulative ex-ante value on initial contracted generation for the year, the difference for that settlement period is paid to the generators, if it is less and ex-post capacity payments are positive, no capacity payments are made. Capacity payments can be negative as reference availability is based on an annual average and current capacity may be more. In this case generators are considered to have received excess capacity payments and these are recovered through the capacity payments made by generators.

The total ex-ante capacity fee is calculated by ASMAE using the ONS dispatch model. This is done on an annual basis and a total ex-ante capacity fee is calculated by sub-market and month. ASMAE builds scenarios based on information provided by ONS and forecasts shadow prices for energy under each scenario. The ONS model does not consider the peak price additions that are taken into account when calculating the derived MAE price. The capacity fee under any scenario is then calculated as the loss of load probability times the excess of the value of lost load over the forecast shadow price. To calculate the ex-ante capacity fee to be paid by demand for each month a smoothing program is run which ensures that the highest monthly capacity fee is no more than 2 times the lowest one. Total payments to generators are calculated in a similar manner. There is no smoothing done in this case. Generators are also paid a capacity fee if they provide stability to another sub-market.

Total retailer payments may not equal total payments to generators within a sub-market and surpluses are allocated proportionally into sub-markets with deficits.
Participation factors are calculated which are the proportion of retailers' payments from a particular sub-market is paid to generators in other sub-markets.

Reference availability used in the capacity payment determination is calculated for hydro plants, which takes into account seasonal variation and various hydrological scenarios when calculating an average expected availability for the year.

Ex-ante capacity fees by month and category type, i.e. peak or shoulder, are profiled using forecast retailers demand for each period to give an ex-ante per MWh capacity fee. The total ex-post capacity fee is calculated by multiplying the ex-ante per MWh rate by metered demand in each sub-market. The participation factors are then used to distribute this amount to various sub-markets supplying to this one. In any sub-market the total received capacity fee is divided by the available generation capacity in that sub-market to get a per MWh ex-post capacity fee. This is added to the MAE purchase price to arrive at the MAE selling price and also to calculate capacity payments to generators.

The per MWh ex-ante capacity fee deemed to be paid to generators in a sub-market is equal to the participation factor weighted per MWh ex-ante capacity fee for each sub-market supplying to the sub-market under consideration. The total ex-ante capacity fee assigned to a generator in a sub-market is this figure times the positive difference between reference availability and actual generation.

While the detailed allocation formulas used in the settlement seems reasonable at least from an equity perspective, there are two fundamental issues that need to be addressed. The first issue concerns the role of ex-ante vs. ex-post calculation in determining financial settlements. The second issue concerns the role of capacity payments and the implementation of such a scheme as a means of assuring supply adequacy. These issues will be addressed in separate sections below.
4 Ex-Ante vs. Ex-Post Pricing

A basic choice in any energy market is the settlement system. ASMAE has proposed a single-settlement system. However, the emerging trend in other countries and in particular the US is to adopt multi-settlement systems. We discuss the issues in both below.

4.1 Single-settlement system

In a single-settlement system, the day-ahead bids or input specification which include generator specified data load forecasts and hydro inflow forecasts are used for scheduling and indicative price calculations, but prices used for financial settlements are determined ex-post based on real-time dispatch.

A typical single-settlement system in the Brazilian context consists of the following steps:

- Bids (e.g., incremental costs, start-up and no-load costs, availability, flexibility, ramp constraints), contract information and forecasts are submitted day-ahead or before.

- ONS schedules units for the next day to minimize operating costs, given the bids, forecasts, operating and transmission constraints, hydro information from medium term planning models.

- MAE calculates indicative ex-ante prices based on day ahead optimal schedule and shadow prices from day ahead schedule optimization or using MAE pricing rules based on day ahead schedules, generators inputs and day ahead forecasts.

- MAE may inform generators and loads of these prices but they are not used in settlements.

- ONS may accept input changes up to an hour (or half hour) before real time.
• ONS dispatches units in real time at least cost, given all current information and forecasts for subsequent hours.

• MAE determines ex-post prices based on actual dispatch, loads, hydro inflows and redeclared availability

• Ex-post prices are used for all settlements to pay generators and charge load.

• Compliance penalties are assessed against those failing to perform as scheduled.

4.2 Multi-settlement system

In a multi-settlement system, the day-ahead bids are used for both scheduling and settling day-ahead transactions. Only deviations from the day-ahead schedule are priced ex-post. The steps are as follows:

• Bids (e.g., incremental costs, start-up and no-load costs, availability, flexibility, ramp constraints), contract information and forecasts are submitted day-ahead or before.

• ONS schedules units for the next day to minimize operating costs, given the bids, forecasts, operating and transmission constraints, hydro information from medium term planning models.

• MAE calculates ex-ante (day ahead) prices based on day ahead optimal schedule and shadow prices from day ahead schedule optimization or using MAE pricing rules based on day ahead schedules, generators inputs and day ahead forecasts.

• MAE uses the day-ahead prices and scheduled quantities in the first settlement.

• ONS may accept input changes up to an hour (or half hour) before real time.

• ONS dispatches units in real time at least cost, given all current information and forecasts for subsequent hours.
MAE determines ex-post prices based on actual dispatch, loads, hydro inflows and redeclared availability

MAE settles deviations from day-ahead schedules at ex-post prices (second settlement)

Compliance penalties are assessed against those failing to perform as scheduled.

A three-settlement system is the same, but with an hour-ahead settlement for deviations from the day-ahead schedules, and then a real-time (ex-post) settlement for deviations from the hour-ahead schedules.

4.3 Is a single-settlement system sufficient?

The single-settlement system may appear simpler than multi-settlement systems. First, it involves just a single set of hourly prices. Second, it is closer to the way the Brazilian system operated before restructuring. However, this simplicity is deceptive. The difficulty with the single ex-post settlement is that much is riding on the ex-post prices, since all earlier commitments and transactions are settled at the prices established in real time. After the day-ahead schedule is formed, agents have an incentive to make adjustments to influence the spot price in a favorable direction. Since the spot price is used for all trades, the incentive for manipulation of inputs, particularly redeclared availability, and for deviation from schedules may be large. While imposing penalties attempts to prevent such behavior it may be not an economically efficient approach since penalties do not price the deviations correctly.

For example if 95% percent of dispatched quantities are determined day ahead and only 5% are adjusted in real time to account for demand and inflow fluctuations, it does not make sense that price changes due to the 5% deviation will determine the settlement for the entire 95%. Agents with sufficient knowledge and control of resources can manipulate the inputs in order to affect the real time schedule so as to reap excess profits. Knowing how to do this is complex, and can be exploited best by large suppliers or buyers with sufficient scale to make the efforts worthwhile. The added complexity and risk tends to discourage entry and participation by small players whose net revenue might
be whipsawed by price volatility in the real-time market. This gaming can be mitigated by financial penalties for failures to perform as scheduled. But then the question is: How to set the penalties? Some flexibility is needed because of uncertainties in demand and supply. Setting the penalties too high leads to inefficient responses to this uncertainty, and setting the penalties too low leads to excessive gaming. The reliance on penalties is highly inefficient and problematic in its workings. It is a carryover from the tight power pool mentality and is unworkable on a sustained basis in a competitive market. The whole idea of relying on administered penalties is inefficient, subject to disputes and subject to continual pressure (as in Alberta Canada) to seek modifications and exceptions. Compliance is shown to be a problem in Victoria, where a supplier can and does curtail generation by claiming an operating problem, etc.

A multi-settlement system mitigates gaming on two fronts. First, the day-ahead bids are binding financial commitments. The bids and resulting schedules are credible precisely because they are financially binding. Second, bidders are unable to alter the day-ahead prices. These remain fixed for all transactions scheduled in the day-ahead market. Deviations from the day-ahead schedule affect the spot price, but the spot price is used only to price these deviations. Hence, in a multi-settlement system the incentive to manipulate the spot price is not magnified as it is in a single-settlement system. Penalties for non-performance are not needed in a multi-settlement system, since deviations from the schedule are priced correctly. If a generator fails to deliver as scheduled, then that generator is liable for the spot price for the quantity it was supposed to deliver.

The multi-settlement system reduces risk for the bidders, since the bidders can lock in the day-ahead prices. For the ONS, the multi-settlement system reduces scheduling uncertainty because it discourages changes, and it automatically sets the right penalties for non-performance. The system maintains the flexibility required to respond efficiently to fluctuations in demand and supply.

A difficulty with the multi-settlement system is that it involves multiple prices for energy. One might think that energy at a particular time (and place) should have one price. However, this is not correct. The price should be determined at the time resources
are committed. Hence, if there are two commitment points (day-ahead based on forecasts and real-time based on events), then there should be two prices, one a forward price for early commitments and a second that recognizes the effects of contingencies.

Despite the advantages of multi-settlement systems, single-settlement systems can perform adequately for at least a short period of time. The United Kingdom, Victoria, and Alberta provide examples of such systems. However, there is a strong tendency to move away from single-settlement systems. The United Kingdom reformed market design (NETA) that will become operational in October 2000 employs a multi settlement system. PJM and New England are also in the process of switching to a two settlement system while NYPP already has a two settlement system, California has a three settlement system and Spain has a twelve settlement system.

There is general consensus among academics and practitioners involved with electricity market design that multi-settlement systems offer significant advantages. The use of single-settlement systems in some markets is largely a historical artifact. Single-settlement systems are evolutionarily closer to the organization of power pools before competitive wholesale markets were introduced.

The differing incentive effects of the alternative settlement systems are illustrated in the Alberta (single-settlement) and California (multi-settlement) designs. The design of the California PX may seem awkward at first, and indeed it is awkward in terms of the software required for settlements, since each MWh of energy might be assigned any one of several prices. In the PX’s energy market, one clearing price is financially binding for trades completed in the day-ahead forward market, another clearing price is binding in the hour-ahead forward market, and the spot price in real time applies to ancillary services and supplemental energy purchased by the ISO. On the other hand, the advantage of this design is that traders have an incentive to bid seriously in each of the forward markets, since the trades concluded there are financially binding at the clearing price in that market.

Alberta uses the opposite design in which all settlements are made at the final spot price, calculated ex-post. That this design produces incentive problems can be seen in the
rules required to implement it. Traders were originally prohibited from altering their day-ahead commitments, but then pressures from suppliers led to a compromise in which each trader was allowed a single re-declaration, and lately the argument has been over whether the final time for all declarations should be moved to just two hours before dispatch. These developments reflect all suppliers’ preferences to delay commitments until close to the time at which prices for settlement are established, so that uncertainty is reduced, and each supplier’s advantage from committing last so that it can take maximal advantage of the likely pattern of prices thereby revealed. The Alberta design also invited a kind of gaming. Importers and exporters are allowed to submit multiple “virtual” declarations. They have used this opportunity to declare several alternatives on a day-ahead basis and then to withdraw all but one shortly before dispatch in order to obtain the best terms. Of course the other traders in Alberta now want the same privilege. The difficulties implementing the Alberta design are intrinsic to any design in which transactions are not financially binding at the clearing price in the market in which they are made. Having the day-ahead bids clear at the spot price, rather than the day-ahead price, introduces a basic conflict. One can argue that a sequence of binding forward prices might sacrifice some efficiency in coordinating the day-ahead and real-time markets, as compared to one in which settlements are based only on spot prices, but this sacrifice is necessary to ensure that bids are serious in the forward markets. While the need for viable forward markets is less crucial in a hydro rich system such as Brazil the introduction of a two settlement system will level the playing field for the thermal generators that are competing with the hydro generators. For example, it will reduce the exposure of thermal generators to precipitation uncertainty. Unexpected precipitation that creates excessive run-of river hydro and reduces thermal energy production will also reduce the ex-post price so the thermal generators who will have to "buy back" part of their scheduled energy will do so at a reduced price. On the other thermal generator will benefit less from a shortfall in run of river energy since the increased ex-post price applies only to the incremental energy above the scheduled day ahead energy production.

Since MAE is already planning to calculate indicative ex-ante prices it will be very easy to adopt a two settlement system in which the indicative prices are actually used in the first settlement and applied to the day ahead scheduled quantities whereas the
ex-post prices are applied to the differences between the actual dispatch and the day ahead scheduled dispatch.

**Recommendation:**

*MAE should switch to a multi-settlement system in which the ex-ante prices are applied to scheduled generation based on ONS day ahead schedule whereas ex-post prices based on actual dispatch and redeclared generator parameters will be applied to deviations of the actual dispatch from the day ahead schedule. In view of the high percentage of hydro in the Brazilian system a three-settlement system will probably not have much benefit beyond what a two settlement system can provide. Such an approach provides market incentives for participants to respond efficiently to uncertain demand and supply. Moreover, it mitigates incentives for gaming and reduces uncertainty for generators and buyers. The proposed approach also has the effect of replacing administrative penalties for deviation with market based penalties.*

5 **Capacity Payments**

Capacity payment is a supplemental payment to generators which is added to the energy price or paid for availability of undispatched generation capacity. The rationale for capacity payments is articulated in the first paragraph of the general description in Appendix G of the MAE rules. It states:

"The capacity fee is a means of ensuring that there is sufficient available generation in the system. This reduces the risk that there will be an interruption in supply that will cause the market price of electricity to rise to the value of loss of load (VLL)."

5.1 **Capacity fee as mandatory price insurance.**

Capacity payments is unique to the electric power industry and it is a highly controversial issue since according to economic theory there is no need for capacity payments to induce optimal investment in capacity. In theory, in a long run competitive equilibrium investment in capacity will converge to a level at which market clearing
prices and scarcity rents (i.e., payments to generators above their marginal cost of production) will exactly cover the amortized capacity cost. Furthermore, that level of capacity is the socially optimal level of capacity at which shortage costs equal amortized investment cost in additional capacity. The theory finds support in California which is one of the few electricity markets were producers are paid only for energy. Currently, nearly 20GW of capacity is ready to be built by private investors (the entire California system has about 50GW) suggesting that the fear of generation inadequacy in the absence of capacity payments is unfounded.

Theoretical rationale and practical experience thus suggest that energy-only markets with spot prices that are allowed to reflect scarcity rents will generate sufficient income to allow capacity cost recovery by generators. The massive influx of new planned generation capacity in California indicates that investors do indeed believe that they will be able to recover their investment and make a profit. Hence from a supply adequacy point of view a well functioning energy-only market can provide the correct incentives for efficient market based generation planning. Yet there may be good reasons for some form of capacity payment and even for regulatory intervention to ensure generation adequacy. Legitimate concerns for failure of the energy markets to reflect scarcity rents or failure of the capital market to produce proper levels of investment in response to such rents may justify some intervention. In some cases regulatory intervention in adequacy assurance is needed to compensate for regulatory interference in the energy market. The supply resource stack of electricity generation in systems with significant amounts of thermal generation exhibits an inherently steep rise in cost around the capacity limit. This phenomenon combined with the typically low short-term elasticity of electricity demand tends to produce high price volatility in fully competitive energy spot markets. Spot markets that clear on an hourly or half hourly basis tend to average out some of the volatility but even in such markets it may be politically infeasible to allow the energy spot prices to fully reflect scarcity rents. Consequently, energy prices are often suppressed through regulatory intervention (price caps) and by the market design, which in turn creates revenue deficiency. Price uncertainty, the prospects of regulatory interference and risk aversion of generators may indeed result in insufficient investment in generation capacity. Often the threat of regulatory interference to curb scarcity rents
will also inhibit capital formation and raise the capital cost for investment in generation capacity. Thus, capacity payments or capacity obligations that stimulate capacity markets are largely viewed as remedial measures needed to offset suppression of energy prices and to ensure generation adequacy.

One can also justify capacity payments, as a price smoothing mechanism. In order for generators to capture sufficient scarcity rents to cover their capacity costs energy prices should reflect the true price volatility needed to equilibrate supply and demand at any point in time. This implies that occasionally prices should be allowed to rise to value of lost load (VLL) which in a perfect market can be determined by demand side bids or in the absence of such bids (due to technological limitations) can be set administratively. As explained in the second paragraph of Appendix G of the MAE rules:

"If all agents had perfect knowledge then there would be no need for a capacity fee. This is because agents would build generators that would be utilised only in the price spike periods. However, agents find it difficult to plan for uncertain and infrequent events and so a market level optimism tends to occur. In this case, each generator believes that there will never be a large reduction in generation capacity due to break-down and so does not build capacity for that demand.

The capacity fee tries to smooth out the price spikes by paying for available capacity even when it is not despatched. The capacity fee is not a subsidy, the expected payments to generators is exactly the same with or without the capacity fee, it merely changes the profile of payments over time."

In other words, according to the above rationale capacity fee can be viewed as a form of mandatory insurance against price volatility that the load must purchase from generators. The capacity payment serves as an insurance premium that is paid to generators based on their available capacity on the grounds that such payment will induce them to build more generation capacity. Following this logic then according to Appendix G in MAE rules:
"The basic principle in calculating the capacity fee is to forecast the likely market price for the year ahead under ordinary circumstances, without any plant breakdowns or transmission outages. Then we model the probability of any plants breaking down and see if that has any effect on price i.e. does the price go to VLL. The capacity fee is then given as the \text{LOLP} \times (\text{VLL} \text{ – Forecast MAE price}) for a particular period."

The calculation of the Loss Of Load Probability (LOLP) is differentiated by submarkets, patamar, settlement month and hydrological scenario. As is the forecasted MAE price. For each combination of submarket, patamar, settlement month and hydrological scenario the LOLP is calculated by drawing a random samples of generating unit failures from multinomial distribution whose parameter are based on ONS data. For each drawn sample a DESSEM type optimization problem is solved whose objective is to minimize unserved load in each submarket (allowing for imports and exports subject to transmission constraints). The LOLP is then calculated as the fraction of runs in which unserved load could not be reduced to zero.

The unadjusted capacity fee for each submarket, month, patamar and hydrological scenario is calculated according to the formula given above and then averaged across all hydrological scenarios. A smoothing mechanism is used to spread the capacity fee across time period so that variation in fee across month does not exceed a factor of two between the highest and the lowest. The above calculation determines the actuarial value of the scarcity rent forgone by generators that receive the MAE purchase price for their energy sales. The remainder of the calculation are concerned with determination of payments to individual generators based on their availability for providing energy and stability and with the reconciliation of the capacity payments to generators and payments by demand.

It is important to emphasize that according to this capacity payment calculation generators availability declaration will only affect their proportional capacity fee but not the LOLP calculation. This eliminates the possibility of manipulation that was prevalent in the UK where the LOLP calculation was based on day ahead availability declarations by the generator and therefore subject to gaming.
In summary, the overall approach to the capacity fee calculation is methodologically correct in the sense that it captures the actuarial value and determines a fair premium for the mandatory price insurance provided to uncontracted loads through the MAE pricing rules. The distribution of capacity fees to the generators and charges to the load is also equitable with the exception of the treatment of demand side bids which will be addressed below. The more global issue which I will also address in details concerns the wisdom of a centralized uniform and mandatory insurance approach to achieving the goals of capacity payments and the possibility of market based alternatives that can achieve the same goals.

5.2 Demand side participation in generation adequacy insurance

Current MAE rules do not award capacity payments to demand side bids. Economic theory tells us that optimal generation expansion is at the level for which expected scarcity rents equal the expected social cost of lost load or the amortized capacity cost of the least expensive generation technology (e.g. CT). When demand can respond to prices, it will adjust to the optimal level. However, when scarcity rents are mitigated through a capacity payment scheme, load is not fully exposed to scarcity rents during the time of scarcity and over consumption will occur. This lead to excess capacity and inefficient use of resources. Figure 1 below illustrates a generation stack and three demand functions reflecting demand during three types of time periods. As shown at the optimal generation capacity the selling price will include a scarcity rent component (that equals to the amortized cost of the least cost generation capacity) above the incremental cost of the price setting generator. This scarcity rent is awarded to the generators in the form of a capacity fee. Hence the generators has an incentive to declare its true incremental cost. The demand side bids on the other hand who do not receive capacity payments will bid the reservation price based on the demand function which at the optimal capacity level includes the scarcity rent component. Consequently, DESSEM, which does not consider capacity payments in its dispatch, will see the demand side bid as noncompetitive and will instead dispatch Generator 6 shown in Figure 1, which will receive its capacity fee. In the long run such over dispatch results in excess capacity and
over consumption. In order to avoid such inefficiency demand side bids must be treated in the same way as generation i.e. receive a capacity payment as an adder. This will incent the load to reduce their energy bid by the amount of the adder so that they can be treated equitably in the dispatch decision. With such a capacity adder to the demand side bid the marginal load at the optimal capacity level will submit a bid that equals the incremental cost of the marginal generator leading to the socially efficient dispatch.

![Figure 1: Improper incentives to demand results in overcapacity](image)

As described above, in order to induce the proper demand side participation it is necessary to have a capacity type payment for the demand. Implementing such a payment under conditions of uncertainty can be done by buying back an energy call option from participating demand side bidders. A call option is a right but not an obligation to purchase energy at a specified strike price. A demand side bid can be viewed as an energy call option with a variable strike price that equals to the larger value of the bid price and the MAE purchase price. It is exercised whenever the MAE purchase price exceeds the bid price. This is equivalent to the standard exercise policy of a call option when the spot price (i.e. the MAE purchase price) reaches the option strike price. Of
course, in order to be eligible for such participation proper communication and control technology must exist that will allow selective curtailment i.e., the ability to exercise the call option. This call option, though, cannot be priced in the usual way as its payoff in any state of the system is zero. The value of the option comes from the real option of having dispatchable capacity available to the system operator to maintain system reliability. This value is equal to the capacity fee which is paid to undispatched capacity.

Capacity fee calculations will also need to incorporate a forecast of available demand side bids. The insurance interpretation of capacity payments dictates that load that is subjected to curtailment in the case of generation shortfall (demand side bid) should be paid the option value whenever it is not curtailed. If the demand side bid is not contracted then this payment amounts to exempting that load from the capacity fee portion of the MAE system service charge. In other words uncalled demand side bids should only be charged the MAE purchase price plus a system service charge without capacity fees included in it for their bid energy. When the option is called (i.e. the demand side bid is dispatched) the MAE purchase price is equal to or exceeds the demand side bid so amount that the load is charged for energy is exactly offset by the payment for its dispatched demand side bid so no payment needs to be made on either side. When a demand side bid is submitted against a contract which presumably included a capacity payment, then the bid should be paid the capacity fee for the call option it provides if it is not dispatched and should be paid the full MAE selling price if it is called. This approach can be rationalized on grounds that curtableable load is foregoing the protection bought by the capacity fee. In other words, the capacity portion of the MAE price is refunded to the demand side bid as payment for the call option they provide. This situation is symmetric to paying undispatched available generators a capacity fee which can also be viewed as payment for a call option provided through their availability.

**Recommendation:**

*Exempt eligible demand side bids (whether dispatched or not) from the capacity fee component of the MAE price. This will provide the proper incentive for demand side participation in shortage mitigation. Since demand side bids forgo the protection*
provided by the capacity fee they should be exempt from that fee. Since it is assumed in MAE pricing that contract already include capacity payments the above rationale suggests that contracted demand that is bid for curtailment at a specified incremental cost should be paid the capacity fee portion of the MAE price when it is not dispatched and should be paid the full MAE selling price when dispatched (unless the interruptibility provision is part of the contract in which case one should assume that such payments have been accounted for in the contract).

5.3 Market based capacity payments.

As described above a useful perspective in addressing the generation adequacy problem is to view the capacity fee as a form of insurance against price volatility. One view is that capacity fee represents a necessary regulatory intervention to compensates for the failure of the energy spot prices to properly reflect scarcity rents, i.e., it is a form of subsidy to generators. The more positive perspective, however, is to regard the capacity fee as a proactive measure in the form of a mandatory hedge or insurance that will assure that prices stay within a socially acceptable range. Such an insurance-based view recognizes the private good nature of generation adequacy. It lays the foundation for introducing customer choice in selecting the appropriate level of price protection and for establishing a relation between the capacity payment awarded to a generator and the responsibility that such payment entails.

5.3.1 Basic framework

The Basic questions that are raised once we accept the insurance framework are: why does it have to be uniform (one size fits all approach), why does it have to be mandatory and why does it have to be administered centrally. For instance, rather than setting a uniform capacity obligation or payment whose cost is evenly distributed among consumers, load serving entities, direct access customers and generators may be able to select their desired level of exposure to price risk and pay or receive an appropriate (negotiated) premium for the price insurance they buy or sell. Thus, generators receiving a capacity payment will guarantee the availability of their capacity to produce energy at a prespecified strike price so the capacity payment is interpreted as premium for a call
option on that capacity. The higher the payment the lower the strike price and vice versa. The centralized mandatory uniform insurance approach does not accommodate diversity of risk preference, financial capability and forecasts about the future. It results in inefficient risk sharing among agents and limits agents ability to make informed investment decisions that are consistent with their financial opportunities and subjective beliefs about the future. A market based system, on the other hand allows agents to bet their money on what they believe and permits a multiplicity of forecasts to guide the market prices. Such market prices for capacity then reflect the consensus of all the agents rather than the forecasts of the central authorities.

Forward markets and hedging instruments provide a market alternatives to mandatory capacity payments. The following features would characterize an idealized market based provision of generation adequacy:

- Customers (or their representative load serving entities) decide how much they want to pay for capacity according to the price risk they are willing and able to assume.
- Generators can diversify their investment risk through physical forward contracts or hedge their risk through other financial instruments.
- Generation gets built if and only if market value of capacity (as reflected by the contract markets) exceeds the cost of new generation.
- Equal opportunity for demand side participation in mitigating price risk.
- Administratively set VLL are replaced by demand side response to price signals
- Theoretical probabilistic models for calculating LLOP are replaced by empirically calibrated stochastic price models underlying the pricing of physical generation capacity and of hedging instruments.
5.3.2 Practical considerations:

An important concern that is often voiced in countries where there is no well developed institutional infrastructure that can enforce financial liability of corporations is that load serving entities or generators may assume more risk than they could handle reliably. So for instance, hydro generators may oversell their water in the present market and not be able to meet their generation adequacy obligations for which they collected capacity payments through premiums on private contracts. Likewise, load-serving entities left to their own devices may not hedge their supply sufficiently in order to reduce their capacity payments and may go out of business or default on their obligation to their customers if the spot prices for electricity skyrocket due to supply shortages. Such problems, however, face any commercial entity that is involved in underwriting risk. This is true for banks, savings and loans and insurance companies that require some form of regulation which will protect the customers from default. In the case of electricity it may be necessary to set some minimum contracting or hedging level on load serving entities and institute regulatory measures to insure that generators have the physical or financial resources to meet their contractual obligations.

The premium paid by load serving entities for meeting their hedging requirements through contracting with generators will produce the capacity payments that generators need to insure the stable income stream for financing adequate generation investment. In exchange for a stable source of income the generators will forgo some of the opportunity to collect high scarcity rents. However, there is no need for a "one size fits all" approach that awards a uniform capacity payment to all generators and imposes a uniform capacity charge on all the loads. A market based approach, which allows parties to trade energy price risk, and investment risk through different contractual arrangements can achieve better efficiency in risk sharing and investment. Regulatory intervention can then be limited to enforcement of minimal hedging requirements and oversight of commercial liability standards and adherence to contractual arrangements.
5.3.3 Capacity fee through energy call options.

To illustrate how such a market based scheme can be implemented let us assume that engineering consideration dictate the need for 20% planning reserves in the system. In a mature market the determination of the appropriate level of reserves can be left to market forces but let us assume that the market "cannot be trusted" to make the correct planning reserves choice. In that case, MAE could impose a requirement that all load serving entities cover all of their uncontracted load plus another 20% of their historical monthly peak load with energy call options which they can purchase from generators or interruptible customers (this can be viewed as a capacity obligation). Unlike a forward contract, a call option entitles but does not obligate its holder to purchase energy at a predetermined "strike price". Hence a seller of such an option must either have a physical cover such as available generation capacity or be able to settle the call financially when by paying the difference between the MAE purchase price and the contracted strike price. Naturally the owner of a call will only wish to exercise it if the MAE purchase price exceeds the strike price so from the MAE point of view the call option is a contract that is activated when the MAE purchase price reaches the strike price of the call option. Like any other contract the call option must be declared so it can be accounted for in the settlement. Furthermore, in case of a shortage any load that is not covered by a call option is subject to curtailment or payment of VLL.

In principle the value of an energy call option with a strike price K equals to the expected value of the (MAE purchase price - K). The income to generators from selling the call options on their available capacity plays the same role of a capacity payment except for the fact that it is not handled through MAE. Furthermore, the price of the call option will be determined between the buyers and sellers reflecting their beliefs about future prices. The role of the regulator in this approach is to enforce financial liability so that any party is liable for the risk they assume. The requirement for call option cover imposed on the load serving entities may be spread over multiple strike prices that could be standardized while the decision how to spread the obligation is left to the load serving entity.
An alternative way to implementing an energy call option is as a "one way contract for difference" with respect to the MAE purchase price. In such a setting the generators buy and sell their energy through MAE at the MAE purchase price (plus service charge) but do not pay or receive capacity payment. The generator will receive their capacity payment by selling the contract for difference to the load serving entity at a negotiated price and a specified strike price. If the MAE purchase price exceeds the strike price in the contract then the generator who sells power at the MAE purchase price will have to reimburse the load for the difference between the MAE purchase price and the strike price in the contract.

A system where the capacity payments represent a call option would require that generators who receive capacity payments be available to produce energy at the strike price, or purchase it and provide it at that price. On the other hand, generators that are not contracted and hence did not receive capacity payments should be allowed to collect up to the VLL for their power if dispatched. The short term inelasticity of demand and steep supply curve may necessitate the setting of a price cap at an administratively chosen VLL. That cap value will then serve as both, a penalty for unmet availability obligation and as a cap on the scarcity rents collected by generators who did not receive capacity payments.

5.3.4 Financing investment

Another problem that may arise in a market based capacity payment system concerns possible failure of the capital market to provide long term financing for generation investments at rates that commensurate with the associated risk. Such market failure may arise since supply contracts that will provide the equivalent capacity payments as option premiums are typically of short duration (no longer than five years) whereas generation investment requires fifteen to thirty years of financing. The practice of securitizing long term investment by rolling over short term contracts is prevalent in many industries (e.g. using short term savings to finance thirty year mortgages). However, lack of experience with commodity trading in the electricity industry and the perceived regulatory intervention risk (especially in developing countries) may raise the
cost of capital to levels that will reduce investment below the efficient adequacy level. Capacity payments are often viewed as a means of income stabilization that would enable generators to obtain financing for adequate investment level. If this indeed were the concern that capacity payments address a more appropriate mechanism would be some form of loan guarantees by the regulator. Since regulatory intervention is one of the important risks factors concerning investors in this business such loan guarantees may inspire confidence in the regulators commitment to uphold free market principles.

5.3.4 Prices vs. quantities

Setting contracting requirements to control the amount of planning reserves in the market enables us to invoke market based decision making while still ensuring that an acceptable level of generation adequacy is provided. This approach has some basic advantages over the traditional capacity payment approach proposed in the MAE rules. First, identifying the requirement imposed on load serving entities as energy call options helps to promote the market equilibrium between capacity prices and energy prices that stems from the fact that the intrinsic value of generation capacity stems from producing energy (with the exception of providing system stability). Second, by imposing a requirement on contract cover rather on the amount of operable capacity we allow some flexibility to substitute physical cover (i.e. operable capacity) with financial cover to settle contracts financially. Finally, imposing a requirement on quantity of contracted reserves rather than relying on generators' response to the capacity price is a more direct means to ensure the desired outcome. It has been argued\(^1\) that the supply function for capacity is relatively flat while the demand function is steep (driven primarily by economic growth) as illustrated in Figure 2 below. Hence a small error in the capacity price (due to inaccurate forecasts) may result in a large error in level of investment whereas controlling the quantity directly will produce a fairly accurate price. Thus, imposing contracting requirements whose prices are negotiated by the agents will ensure that the desired adequacy is achieved and the contract markets will produce the correct capacity price signal.

\(^1\) This argument and the supporting Figure 2 are due to Larry Ruff.
Figure 2: - Characteristic Shape of Supply and Demand Functions for Capacity

Recommendation:

Eliminate capacity payment component from MAE price and impose a requirement that load serving entities supplement their forward contract with energy call options or one sided contract for difference up to (1+X) of their monthly peak load, where X is the planning reserves requirement. The prices for the call option are negotiated directly between the load and the generators. Load serving entities can meet their contracting obligation outside their submarket but the reference MAE price for a call option and the exercise point correspond to submarket of the load. In other words the load can contracts with a generator in another submarket to deliver power in its submarket when the MAE price reaches the strike price or it can settle financially. This approach solves the problem of distribution of capacity payments across submarkets. Generators that cannot meet their contractual obligations during interruptions are liable for VLL while generators that generate above their contractual obligations during shortfalls should receive VLL for their surplus generation.