Diagnostic testing for earnings simulation engines in the Australian electricity market

Carl Chiarella
University of Technology Sydney

Jiri Svec
The University of Sydney

Max Stevenson
The University of Sydney

1 Introduction

The risk faced by electricity retailers stems from the purchase of wholesale electricity from the national wholesale market which is then on-sold to smaller retail customers. The electricity is purchased at a variable price that is dictated by demand and supply equilibrium. It is then on-sold to customers at a fixed price set by a designated government regulatory body. The difference between the fixed and floating price exposes the retailer to substantial earnings-at-risk (EaR). This exposure can be reduced by entering into financial hedging contracts that include electricity swaps, caps and other more exotic derivatives. These electricity hedging contracts go some way to offsetting the physical exposure of a retailer by effectively fixing part of the quantity demanded from them, as well as the wholesale electricity price at which they purchase the hedged quantity. Derivative contracts are purchased to keep risk levels within a retailer's board approved limits that are set in accordance with its risk appetite. The electricity market is unique with a large number of derivative products not found in other markets. Further, electricity derivative contracts comprise a market that is regarded as incomplete due to the inability to store and hold an inventory of the physical commodity. Compared to other markets, the pricing of these contracts is not trivial. The nature of the asset price processes underlying electricity often result in a poor fit to those existing pricing models that are based on diffusion processes and usually used to model and price equity, foreign exchange and fixed income. The retailer's exposure can be broadly quantified as a function of electricity demand, wholesale electricity price and fluctuations in temperature. A large academic literature contains a host of publications concerned with the modelling and evaluation of load and temperature forecasts across a wide spectrum of forecasting horizons. As most markets have different structures and are characterised by different climatic effects, models tend to be region-specific with no one model superior across all markets. On the other hand, extensive research into the comparison of price models is lacking to date and, as a consequence, there is little published research on the comparison of available price models. However, both load and price estimates are crucial inputs to the valuation of hedge contracts. As a consequence, the literature is devoid of Australian studies that have evaluated electricity hedging instruments using dollar values to calculate EaR. Electricity derivatives are written with payoff functions that settle at half hourly intervals. Historical load and price

	LOAD(MWh)	PRICE ($)
Mean	8614.56	35.80
Median	8731.18	20.77
Maximum	12598.07	9166.67
Minimum	5230.66	7.84
Std. Dev.	1303.46	249.97
Skewness	-0.05	27.46
Kurtosis	2.47	829.84

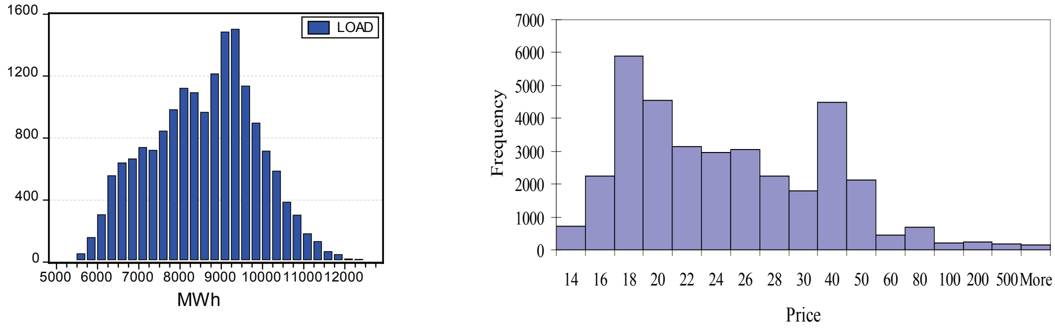
Table 1: Summary statistics for load and price for 2005

are positively correlated, and the density for price displays skewness in excess of a lognormal distribution. Furthermore, the underlying joint distribution for load and price is unobservable. Also, as hedges are generally written on only a portion of the day, then peak, off-peak and weekend valuation needs to be carried out separately. With most electricity markets characterised by unstable and non-parametric electricity price distributions, an efficient way to model price and load, and therefore the earnings distribution, is by incorporating simulation techniques rather than standard time series methods. This report is primarily concerned with evaluating the appropriateness of a number of risk simulation engines used by electricity retailers in NSW to accurately generate electricity price and load paths. These paths are then used as input in the determination of the distribution of a retailer's earnings and the assessment of their earnings-at-risk (EaR) measures. The EaR is an important metric associated with the risk management process. However, it is only when a simulation engine can be relied on to produce accurate future load and price estimates that are appropriate from a retailer's business perspective, that risk can be effectively managed. In §2, the characteristics of both load and price in the Australian National Electricity Market (NEM) are discussed, along with the argument for adopting a simulation approach for their modelling and future estimation. It is important to understand the characteristics of load and price in order to diagnose the ability of a risk simulation engine to produce load and price paths that accurately depict the historical data. The data from a number of risk simulation engines evaluated as part of this study are then outlined at the start of §3. The implementation of a series of diagnostic tests and analysis of the output from the various simulators then follows. Implications that can be drawn as to the suitability of the various simulation engines for an electricity retailer's risk management purposes are discussed in the concluding section.

2 Characteristics of electricity load and price

In order to describe the characteristics of typical NSW electricity load and prices (with Sydney as a reference node), we chose as an example the corresponding series for 2005. Table 1 contains the summary statistics for load and price aggregated over peak, off-peak and weekends where all observations were based on Eastern Standard Time[1]. The histograms for NSW

[1]In NSW, daylight saving starts at 2 a.m. Eastern Standard Time (EST) on the first Sunday in October and ends at 3 a.m. summer time on the first Sunday in April. During daylight saving, summer time in NSW is one hour in advance of EST.



(a) load

(b) spot price

Figure 1: Histograms of NSW state electricity load and spot prices in 2005

State load and prices for 2005 are shown in Figure 1. The histogram for load shows a tendency for bi-modality with a slight positive skew. On the other hand, it is clear from Figure 1 that the histogram for NSW state prices for 2005 is bi-modal and heavily skewed to the right. The majority of prices in 2005 were distributed between \$18 and \$40 per megawatt hour (MWh), but occasionally prices spiked close to the maximum permitted level of \$10,000. The maximum price for 2005 was more than 36 standard deviations away from the 2005 mean price. The average price during 2005 was approximately \$35 per MWh with over 92% of prices below \$40 per MWH. However, due to the excessive skewness, the load weighted average spot price was closer to \$40 per MWh. Both histograms have positive excess kurtosis, but only the price series is significantly higher than zero. Another interesting phenomenon can be observed from the graph of the half-hourly average electricity prices given in Figure 2. Although the price spikes and the episodes of "clustering" or volatility in 2005 were unpredictable, they were visibly more frequent during the warmer months between November and February. This pattern has been well documented by a report on various aspects of market behaviour compiled by NECA in 2003[2]. The report stated that after an initial transition period following market inception in December 1998, average NEM spot prices have exhibited a seasonal pattern with higher prices more likely to occur during summer and winter peak load periods.

Figure 3 shows the graph of NSW State load for 2005. The state demand oscillated between a minimum of 5,200 MWh and a maximum of 12,600 MWh. There is significantly more and variable consumption during the summer and winter months. When load is examined at the half-hour frequency, but over a shorter time interval as in Figure 4, diurnal and weekly effects becomes apparent. Intraday demand depends on many factors but the demand peaks are generally driven by domestic activity and temperature. Domestic activity typically occurs between seven and nine o'clock in the morning and between four and seven o'clock in the afternoon on weekdays. Note from Figure 4 that the 24th and the 25th of December fall on a holiday period with an associated smaller demand. Although there may be ample supply available in the system to meet peak demand levels, supply can come under extreme pressure on a few days of extremely high temperatures during the summer months. The high price spike on the 27th of December was one such example. During unexpectedly high temperatures

[2]National Electricity Code Administrator (NECA, 2003), "National Electricity Market Statistical Digest, October-December 2002". www.neca.com.au

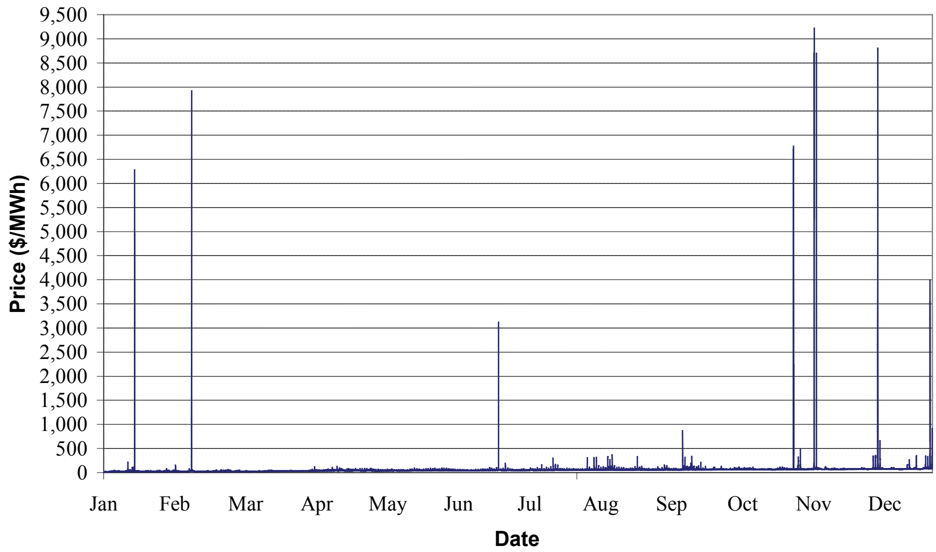


Figure 2: NSW State half-hourly electricity prices in 2005 ($/MWh)

it is not uncommon for the spot price to surge close to VoLL[3] for a brief period of time from a normal level of less than $40.

2.1 Load modelling

Electricity load models provide electricity retailers with long-term load forecasts for up to three years ahead in order to price derivative contracts and manage their market exposure. They display consumer-driven dynamics such as higher demand during morning and evening periods and different consumption patterns between workdays and weekends. Load modelling falls into two broad classes. The first of these classes comprises statistical time series models that produce load forecasts, while the second class of models generate expected loads using a simulation engine. Both classes of models can be represented by univariate and multivariate specifications. The essential differences between these two load forecasting approaches are now discussed. Both classes of load models usually involve decomposition of the data by de-trending with a deterministic exponential trend (or a linear trend if the logarithm transformation is imposed on the data), and then accounting for seasonal and daily cycles using periodic functions. The residual term after decomposition can then be bootstrapped to reconstruct multiple load paths that, when averaged at each tick, result in expected load values of the estimation (calibration) series. Rather than applying a bootstrap approach, an innovation distribution can be assigned by the modeller where random selections from the distribution can form the basis from which the original load series can be reconstructed. The alternative to a simulation approach is the statistical time series approach. A single forecast is generated from these models by reconstructing the signal on a n-period ahead basis using lagged values of the estimation sample and previously forecasted values, along with the inverse of the periodic functions that modelled the seasonal and daily cycles. Simulation can also be applied

[3]VoLL refers to Volume of Lost Load and is the NEM mandated maximum price per megawatt hour (MWh). It is presently set at $10,000.

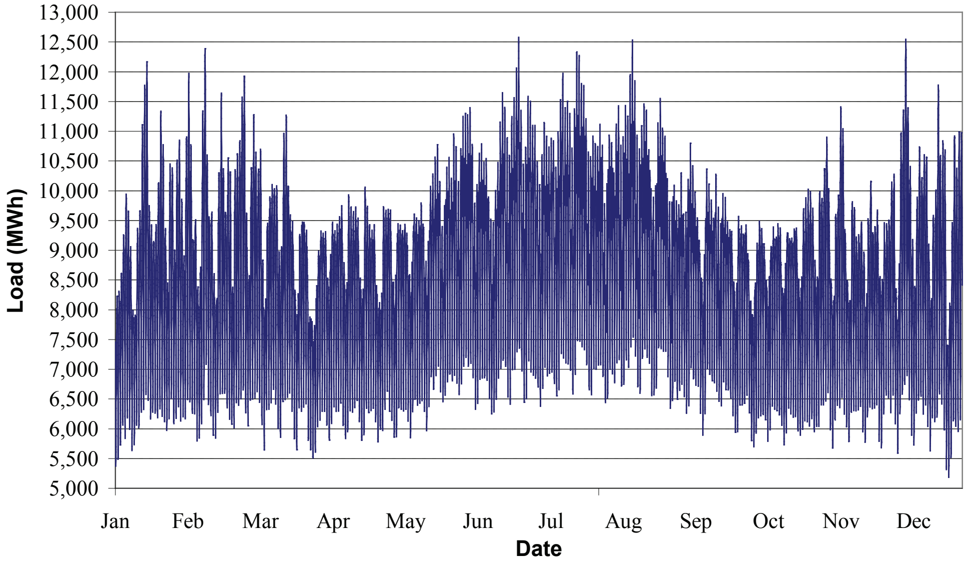


Figure 3: NSW State Half-hourly Electricity Demand in 2005 (MWh)

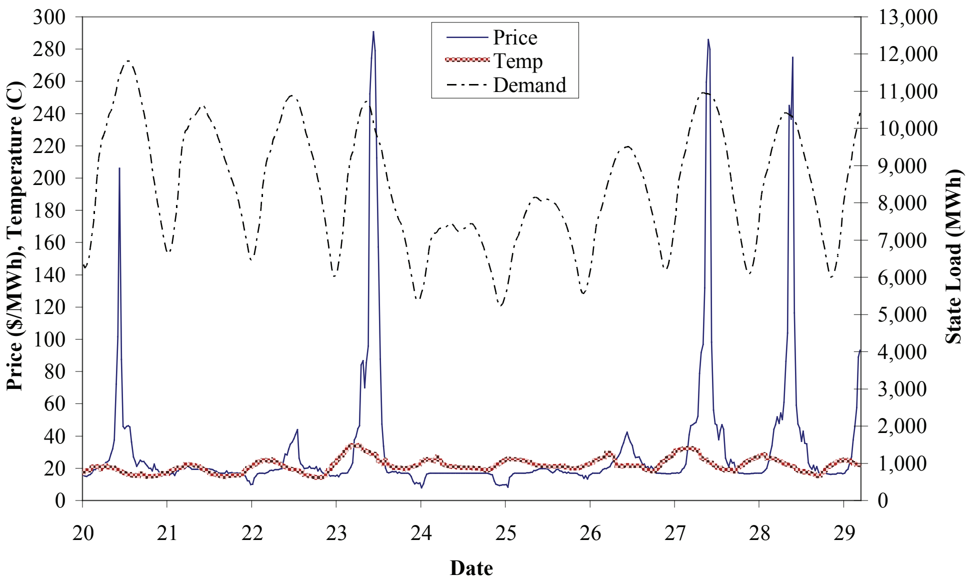


Figure 4: Intraday temperature, load and price from 20th to 29th December, 2005

to a statistical time series model. Random draws from the error distribution whose expected residual value is assumed to be zero, results in multiple forecast paths that can be averaged at each forecast step across the forecast horizon. The simulation models and the statistical time series models can be specified as univariate or multivariate models. Multivariate models divide the univariate electricity load series into daily or weekly vectors of half-hourly data. The univariate model consists of a one channel series while the multivariate model comprises of 336 individual load channels based on the 48 half-hour periods within a day over a whole week. The multivariate representation allows the vectors or 'channels' to be estimated by a seemingly unrelated regression (SUR) model in which a separate parametric regression is used for each intraday period with the errors correlated across equations. Further analysis has been carried out by [11] to examine whether modelling demand in a multivariate framework is more appropriate than as a univariate series in order to capture growth rates accurately. He found that the multivariate version of the model with a separate growth rate for each channel provides a much more accurate estimate of the NSW State load.

2.2 Price modelling

High electricity prices are driven by unpredictable and unforseen circumstances. These include episodes of higher than expected temperatures, infrastructure constraints and dependence on generation capacity. Electricity prices are capped at the VoLL, are usually not negative and have distributions that are highly skewed. While univariate forecasting models for electricity prices have been reported in the literature (for example, [9]), improvement in the accuracy of longer range forecasts from these models is required before they can be viewed as being suitable for generating an appropriate forward price curve. As a consequence, non-parametric pricing models based on simulations from empirical distributions are widely used in the electricity industry to generate forward prices. Knowledge of the specifications of three of the five risk simulation engines that form the basis of the analysis in this study have been withheld due to proprietary considerations. The remaining two engines are based on the historical empirical and returns bootstrap procedures applied on a channel by channel basis. As all the simulation engines will share similar characteristics in their specifications, it is instructive to discuss historical channel and returns bootstrap models in more detail.

2.3 Channel bootstrap model

Assuming that the process generating the historical data is stationary with independent increments, the prices can be simulated by re-sampling, with replacement, from the empirical distributions of the electricity prices constructed from the historical observations over the in-sample period. The procedure is then repeated N times to produce N sequences of the desired length. There are a number of ways historical bootstraps can be constructed. The simplest method is to re-sample from the actual distributions of each channel. This method is suitable for large stationary data series with fairly stable distributions. However, for the small weekend sample, re-sampling actual values would lead to frequent selection of identical extreme prices which is clearly not realistic. To overcome the problem the sampled series are further randomised, see [11], Chapter 5, for further details.

Once the historical observations are allocated into bins, the number of observations in each bin is divided by the total number of observations in each particular channel to obtain the cumulative probability distribution (CDF) of the prices in that channel. The benefit of obtaining the CDF is that all the transformed observations must lie between zero and one, inclusive. The electricity prices are then simulated by drawing a random number from a uniform distribution between zero and one which is then compared to the CDF. The random number generated between two points on the CDF is then used to isolate the appropriate bin from which to sample. Bins with a large number of observations have a longer interval within the CDF and are thus more likely to be drawn by the uniform distribution sampling process. In order to insure that the sampled prices are randomised and none of the historical observations replicated exactly, the log of the lower limit of the bin is added to another randomly generated number from a uniform distribution which is multiplied by the width of that particular bin. This ensures that the numbers simulated from each bin are evenly spaced within the bin and not systematically repeated.

2.4 Returns Bootstrap Model

The drawback of the random sampling approach from each channel is the lack of correlation structure between adjacent half-hours which significantly overstates the price volatility through time. This drawback can be addressed with the returns bootstrap. This procedure, developed by [8], converts raw prices into returns and divides them into blocks. The bootstrap then samples blocks of the returns data with replacement, rather than individual prices, to create a time series of returns. The advantage of such a process is that it relaxes the independent increment assumption of the previous bootstrap and retains the entire autocorrelation structure within each block. This produces volatility and intraday price dynamics comparable to the original series.

2.5 Earnings at Risk (EaR)

An electricity retailer has significant Earnings at Risk (EaR) due to the potential earnings variation of the combined retail portfolio, incorporating both wholesale and retail positions. The main risk for earnings is fluctuations in the wholesale pool price. Business overheads are quite predictable and are not viewed as a source of earnings risk. Unlike most financial and commodity markets where EaR is primarily concerned with an exposure to price volatility, participants in the complex electricity markets can also face exposure to volume variability, or combinations of both. A retailer can protect itself against high price volatility with the aid of derivative contracts, but only in the cases when the quantity of electricity specified within the derivative contract exactly matches the quantity of electricity that the customer consumed. Since the retailer has to put the derivatives in place before the fact, and only finds out how much electricity was actually sold to customers afterwards, the retailer is exposed to a volumetric risk. That is, the risk that the hedged quantity does not match the actual load. In Australia, the EaR risk metric is usually calculated by considering a committed customer base. The advantage is that the customer will be committed to being in the portfolio until the end of the specified length of their contract. The accurate forecasting of both load and price is therefore critical for obtaining an accurate measure of EaR for this portfolio. Due to the stochastic nature of the input variables, EaR analysis of the electricity portfolio usually consists of Monte Carlo simulations of forward retail load paths and forward price curves. The paths are based on observed historical characteristics of each of the series generated at a half-hourly resolution for the required reporting horizon. The half-hourly payoffs (in dollars) are calculated by multiplying the difference of the simulated and actual price by the contracted

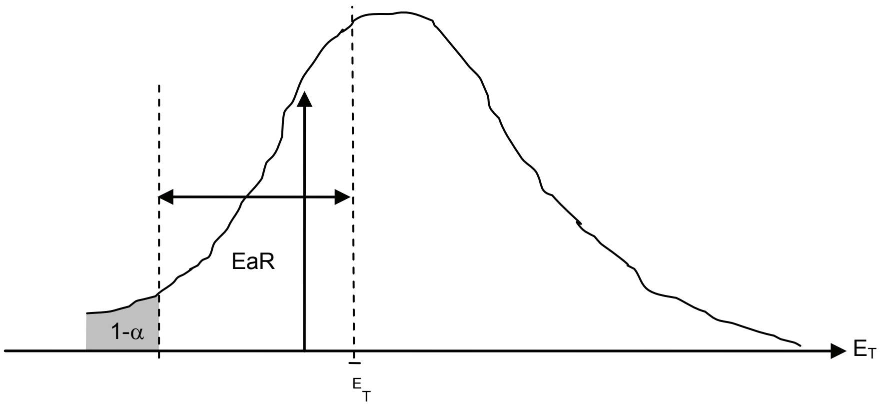


Figure 5: Earnings at Risk (EaR) and the earnings probability density function

load for each and every path that is generated, and for each of the retailer's contracts. The total payoff of the portfolio is then calculated by summing the half-hourly pay-offs across the full reporting horizon (usually three to five years), and at the defined calculation frequency (usually quarterly). The end result is a distribution of probabilistic outcomes, through which both the extent of the risk (in dollars) and the probability of occurrence can be determined. EaR outcomes are designed to simulate the potential economic loss created from a particular set of market variables. The EaR over a given time period (e.g. rolling quarters) is calculated according to

$$EaR = \bar{E}_r - E_{1-\alpha},$$

where $\bar{E}_r$ is the mean of a sum of half-hourly unhedged earnings and $E_{1-\alpha}$ is the earnings level corresponding to the $(1-\alpha)$ tail of the simulated earnings distribution depicted in Figure 5.

High prices can have devastating consequences on under-hedged retailers. After two long periods of sustained high prices in April and June 2007, two small non-franchise retailers, Momentum Energy and Energy One were forced to exit the retail electricity market and sell their portfolio of 20,000 customers to another retailer to avoid bankruptcy. Another small retailer, Jackgreen, also experienced large losses that forced the company to increase its hedging coverage at a significant market premium. Historically, June and April in 2007 were the first and third most expensive wholesale electricity months, respectively, since the inception of the NEM in 1998. The average market price for June, 2007 exceeded \$240 per MWh. This was three times higher than the previous highest ever recorded average price in a given month (Jackgreen Annual Report). The prices in June, 2007 reached a staggering eight times their monthly June average and ten times the average prices paid in the early months of 2007. With high prices having a major influence on EaR calculations, it is appropriate to report risk at a confidence interval that is reflective of the level of volatility associated with such prices. In calendar year 2006 for example, the top 1% of spot prices contributed approximately 20% of market value on a sample time-weighted basis. The 95th percentile is most commonly reported (relative to the 50th percentile) when considering the retailer's risk position. EaR is thus equal to the difference between the 50th percentile result and the 95th percentile result. However, the volatility within the NEM has increased significantly over the last couple of years. This has skewed the range of earning outcomes likely to be produced within a historically based EaR model. Accordingly, a diagnostic tool is introduced in a later section to identify the incidence of "high prices" in the simulated prices. Further the role played by price and volumetric risk in the determination of earnings and, therefore EaR, are discussed in the following two sections. Price risk is manifested by extremely high price volatility at times of peak demand and supply shortages. Price fluctuations may be caused by factors that impact on either the demand, or supply of electricity, or both. Electricity demand is highly variable (e.g. with time-of-day and weather), while the volatility in price is exacerbated by the way that consumers are charged a fixed retail price that is independent of the spot market price. There are few electricity substitutes in the short term. This, combined with its relative importance to the public and their expectation of an uninterrupted power supply, leads to price volatility rarely seen in other commodity markets [4, 5].

Research conducted by [3] calculated that instantaneous electricity price volatility readily exceeds 100%. This is significantly higher than volatility observed in other commodity markets such as oil and gas, with an instantaneous volatility of 40% and 60%, respectively. In comparison, equity markets have quite a low volatility of a mere 15%. [6] have shown that the Australian prices exhibit a greater frequency of extreme price spikes in comparison to the United States and Nordic pool markets. [2] calculated historical volatilities in the Australian market to be in excess of 900%. The particular characteristics of each electricity market are highly dependent on the physical fundamentals of that market. The local mix of generation supply plays an important role [7]. Load supply points usually come from one of three sectors. The first sector is coal-fired. Coal-fired power plants have low operating costs, high start-up costs and large start-lags. They need to be operated continuously to attain the maximum efficiency. The power generated from these power plants is known as 'base load'. The start-up time for a base load coal-powered unit can exceed a day depending on its size. The majority of the daytime load 'shoulder load' is supplied by closed-cycle or open-cycle gas turbines. Gas fired units have higher operating costs and will not start bidding into the market until the electricity price reaches their marginal operating costs. The typical start-up time of a gas turbine is around fifteen minutes.

Peak load covers the hours of highest demand. The leading provider of peak load for Sydney metropolitan area is Snowy Hydro Ltd through its hydroelectric and gas-fired units. Although hydro is inexpensive, due to limited water resources within the NEM, hydro generators are operated predominantly during the morning and evening demand peaks. Energy Australia, a leading retailer in NSW, defines the peak demand as a period between 2pm and 8pm on working weekdays, while the shoulder period occurs between 7am to 2pm and 8pm to 10pm working weekdays, and 7am to 10pm on weekends and public holidays. The remaining time between 10pm and 7am each day is classified as off-peak. Demand is usually characterized as highly inelastic because electricity is a necessary commodity [10]. During periods of low demand, generators supply electricity using base-load units with low marginal costs. However, during periods of elevated demand generators with higher marginal costs enter into the system. The elasticity of demand in the NEM has also been estimated by [**?**]. They divided the data into three different series (weekday peak, shoulder and off-peak), based on the total system load observed in that half-hour. Additionally, the peak series was also split into weekday and weekend for the pre-processing. The "weekday peak" series aggregated the top four half-hour periods for each day; "shoulder" covered the next 34 periods, while "off-peak" included the ten lowest periods. These observed elasticities conform to conventional wisdom. Excluding the operation of large base-load power plants, supply of electricity is generally elastic. The peak generation capacity has a very short start-up time and can quickly adapt to demand changes.[4]

If regional electricity demand exceeds supply during times of inadequate system reserves, plant failures or infrastructure constrains, electricity supply has the potential to become highly inelastic. According to [12] , prices were highly sensitive to quantity shifts during peak demand with an estimated elasticity of 0.06. During such periods the supply curve is quite steep because of the low utilisation of these assets to cover the short duration peaks. The relationship between demand and weather patterns during periods of high and low temperatures also impacts on electricity prices. For example, the proliferation of air conditioners plays a crucial role in the Australian markets but is largely irrelevant in colder climate countries like Norway. The volatility of electricity price is further compounded by generator bidding behaviour and influenced by economic and business activities. In the long-term, prices generally revert to the mean as peak periods are transitory and there is enough generation capacity to cover the desired demand. In order to understand the nature of the electricity markets it is essential to realise that the price risk is not symmetric. The distribution of prices is highly skewed and the risk is correspondingly one-sided. The largest risk is associated with high prices, with occasional peak pool prices close to VoLL. In comparison, the average peak demand level is around \$40 to \$50. The White Paper issued by the Australian Government [1] highlighted the high economic impact of price spikes. While peak demand periods were typically short in duration and lasted for approximately 3.2% of the annual duration, they accounted for 36% of total spot market costs. The typical combination of events that can lead to extreme prices, and yet are virtually impossible to forecast, are summarised using an actual example from the summer of 2004 and depicted in Figure 6.

Figure 6 shows the half-hourly NSW pool prices, NSW State load and the temperature as measured at Bankstown Airport between Tuesday 30th November and Wednesday 1st December, 2004. On both days the demand was higher than normal due to extreme weather conditions across NSW and it can be seen that as the temperature climbed towards the 40C mark on both days there was a corresponding increase in the electricity demanded. This was most likely associated with an increased usage of air conditioners. To compound the problem, the elasticity of supply was very low due to lower than anticipated power reserves from planned generator outages. This had a dramatic impact on the price of electricity which rose from less than \$100 per MWh on Tuesday morning to more that \$8,500 per MWh by Tuesday afternoon. Due to the increased demand and the reduced available supply, the generation availability in NSW, inclusive of all the imports from neighbouring regions, resulted in a surplus generation of less than 400 MW. The surplus was reduced further by a generating unit that tripped on Tuesday evening. Although this had little impact on Tuesday and is unobservable in the plot, it was a significant event that contributed to the substantially higher Wednesday prices. Originally this unit was due back early on Wednesday morning, however, delays in its recommissioning meant it did not start to generate power until late Wednesday afternoon. This led to an over-estimation in the reserve levels in most of the predispatch runs on 1st of December. NEMMCO's assessment of available reserves in NSW on

[4]For example, hydroelectric generators have a response time of less than sixty seconds. [12] estimated that the elasticity during the shoulder and off-peak periods was around 0.25 and 2, respectively. In the event of a supply shortage or a demand overload, Snowy Hydro has the capacity to quickly provide the required supply provided the load does not exceed the network constraints. [12] also investigated the amount of hedging at each of the three levels and found that peak periods were hedged to a significantly higher level compared to off-peak periods. On average, retailers hedged 90% of their peak load, 75% of their shoulder load and 25% of their off-peak load.

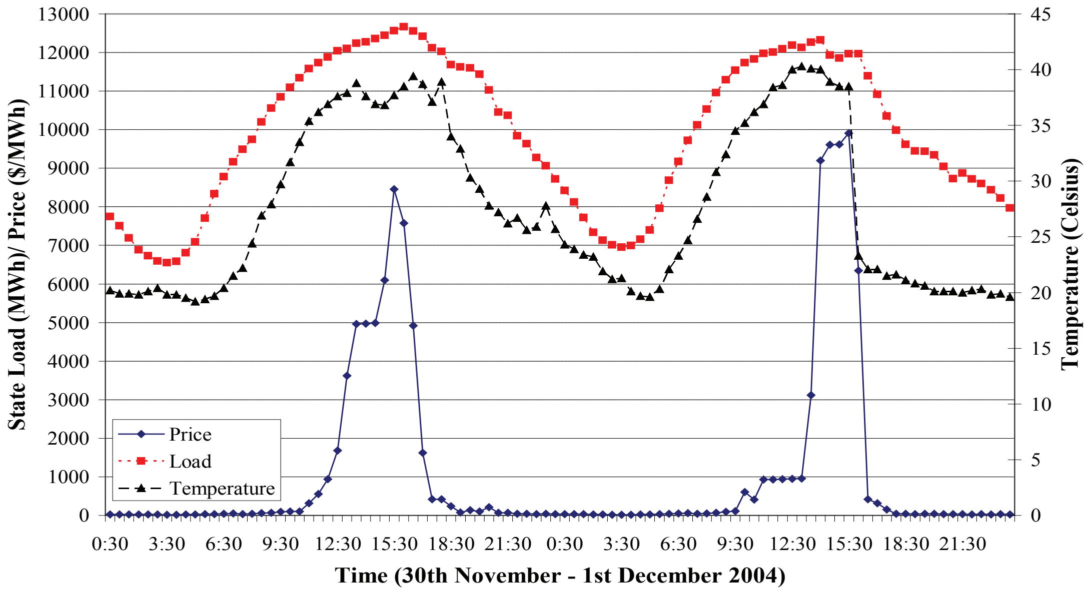


Figure 6: Price, Demand and Temperature from 30th Nov to 1st Dec, 2004

Wednesday afternoon indicated that there was insufficient reserve to cover the potential loss of the largest generator. The peak NSW demand was well over 12,000 MW with a generation surplus of 510 MW. Following another generating unit that tripped during this low reserve, the price of electricity jumped from $3000 per MWh to in excess of $9000 MWh. Regional interconnectors operated at capacity and further energy could not be sourced from outside the NSW region. NEMMCO was forced to declare a supply shortage in NSW and instructed an increase in off-line generation to restore the power system to a secure operating state. At the same time, NEMMCO also imposed voluntary load shedding and large industrial customers were encouraged by participants to voluntarily reduce load. Although the electricity price at this stage approached the $10,000 limit per MWh, the voluntary demand reductions alleviated the need for involuntary load shedding. Consequently, the VoLL flag was not set. Voluntary load shedding conditions lasted for twenty minutes until a stand-by generation facility began producing power and eased the demand pressure. Furthermore, whilst temperatures across Sydney still exceeded 40C, a cool change in the form of a severe weather pattern was passing over the transmission system on the south coast and the temperatures across Sydney on the 1st of December fell by 15C in less than one hour and there was a large decrease of 1000MW in NSW demand during this weather change. Simultaneously with this temperature and load drop, the price of electricity retreated from over $9,500$ per MWh to less than 500 per MW.

Extreme price events can generally be attributed to the following combination of factors; high demand during working days as most significant peaks occur during the week, extreme or higher than anticipated temperatures, supply shortages due to generator or transmission failure, reduced output from hydroelectric power stations due to low rainfall across the eastern states (reduced reserve margins have driven up NEM prices in 2007 to record levels) and limits to the ability to source power from interstate due to transmission and interconnector constraints.

Volumetric risk is often used to denote the phenomenon whereby electricity market participants have uncertain volumes or quantities of consumption or production. Energy retailers are obliged to sell uninterrupted electricity to their customers at a prearranged price without a consumption limit. This is stipulated by the government, with legislation placing stringent demand on market participants to meet load requirements. Failure to satisfy this demand can lead to a loss of licence to operate. The majority of the volumetric risk stems from the variability of energy demanded due to variability of weather. Under regulation, electricity retailers maintain excess capacity to protect themselves against weather related demand fluctuations. The cost associated with excess capacity is passed on to the consumers through higher rates. Under the new deregulated structure, supply and demand must be met in order to have the lowest common clearing price per unit delivered. This leads to a more efficient use of infrastructure, but the retailer must bear the financial impact of being at risk from the weather. Volumetric hedging is used to complement price hedges because it has the potential to dampen a portion of the effects of the price spikes. However, it does not eliminate the entire effect as spikes can also be caused by non-weather events. Weather hedges are generally written to cover a period of several months and are thus ideal to hedge against volume of energy transacted over a longer period rather than to cover large price spikes on a particular day.

3 Risk simulation engines and their output

Simulation paths from five individual risk engine simulators were supplied for diagnostic checking as part of the MISG2008 project. Three of the simulators are at present used by leading electricity retailers, while the remaining two are part of an ongoing research project in risk management in the electricity industry being undertaken at the University of Sydney. The price paths on which the analysis is undertaken in this report are based on paths generated in-house by either the retailer to which the risk simulation engine is licensed, or the research group at the University of Sydney. They were supplied to the MISG2008 project members. Price simulation paths for three of the simulators (C, D and E) were supplied for both the Financial Year 2007/08 (Fin07) and Calendar Year 2008 (Cal08). Paths for simulator A were supplied for Cal08 only, while paths for simulator B were supplied for Fin07 only. The origin of each of the simulators was identified alphabetically from A to E.[5] For consistency, 3,000 price paths were chosen from paths supplied from each simulator, and for each time period that it was represented in. As part of the scope of the analysis, no information was supplied as to the whether the simulations were based on a univariate or a multivariate model.

The usefulness of an earnings simulation engine has to be measured against how well it performs the task of providing inputs to the risk management process. These inputs include simulated price paths. Accordingly, we applied a number of diagnostic tests to the price paths from the simulation engines in order to determine their capability for generating appropriate price representations for the calculation of earnings measurements used in determining the EaR measure. The diagnostics are designed to benchmark a given set of simulations from a risk simulation engine against historical data to assess the quality of the output. The discussion of the results from applying the diagnostic tests to the price paths from the different simulation engines for the Calendar year 2008 (Cal08) and the Financial year (Fin07/08) are now given. The majority of the results from the diagnostic tests are represented either visually or in

[5] As mandated in the project guidelines, we provide no indication as to the licensed owners of the simulators diagnosed.

tabular form, while conclusions drawn from the results are left to discussion.

3.1 The first diagnostic test; means and medians

The first diagnostic test was to graphically represent the expected values (means) and the medians over the 3,000 price paths at the level of each of the 17,520 ticks (half-hour periods). Price paths from the simulators represented in Cal08 are diagnosed first. Diagnostics of output from simulators represented in Fin07/08 follow. For consistency in the discussion of the results for the first diagnostic test across all simulators, the following criteria are examined. First, the overall levels of the mean and median values for each of the 17,520 ticks for each set of price paths are compared and contrasted to realised prices for the period being analysed. Second, the degree of conformity of the shape of the mean and median average prices to what was expected and what was actually the case for both time periods analysed is discussed. It would be a mistake to think that the ability of the simulators to capture the jump process in the realised data could be assessed from this diagnostic tool. This is due to the expected prices being averaged at each tick across 3,000 prices paths. It is highly unlikely that jump behaviour will be repeated 3,000 times for a particular tick and, accordingly, being reflected in an expected price. By definition, the median over 3,000 observations will not pick up an outlier unless all observations are outliers. It follows that the identification of the incidence of jump behaviour in the price paths is not a realistic exercise using this diagnostic test. It will be incorporated in another diagnostic tool to be discussed later.

3.1.1 The calendar year 2008

The means and medians of the price paths for simulator A are graphed for Cal08 in Figures 7 and 8. From Figure 7, the expected (mean) prices indicate overpricing for levels across all quarters. Additionally, there is a marked discrepancy between simulated weekday and weekend pricing for all quarters. Given that the values are averages of 3,000 simulated paths, the increased magnitude of the levels for weekday prices as compared to those for a weekend are unrealistic. Apart from the shoulder months of March/April and September/October, this difference in the levels is approximately $150 across all weeks of the year. This deviation from what is a realistic weekday/weekend difference is more pronounced for expected prices in Figure 7 than for median prices given in Figure 8. Overpricing is indicated in Figure 8 during the summer and winter months. However, the overpricing in November/December can be explained by the abnormally cool weather experienced in quarter 4 of Cal08.

Simulator A provides price paths that result in means and medians that exhibit the appropriate shape on a seasonal basis. However, the shape of the expected prices during the winter months in Figure 7 is far too pronounced. On the other hand, median prices from April through to August are more reflective of the shape expected and are consistent with the realised prices for these months. Overall, the volatility of the expected prices is too great. This is a major issue for the use of this simulation engine as a price generator. The key risk metric, EaR, requires the input of an accurate expected price.

The means and medians of the price paths for simulator C are graphed for Cal08 in Figures 9 and 10. What is evident from both Figures 9 and 10 is how both the levels of the expected and median prices are significantly higher than the actual prices for all quarters in Cal08. Apart from the effect of price volatility during the winter months (May to August), as well as in January and parts of February, even the lowest simulated expected and median prices are

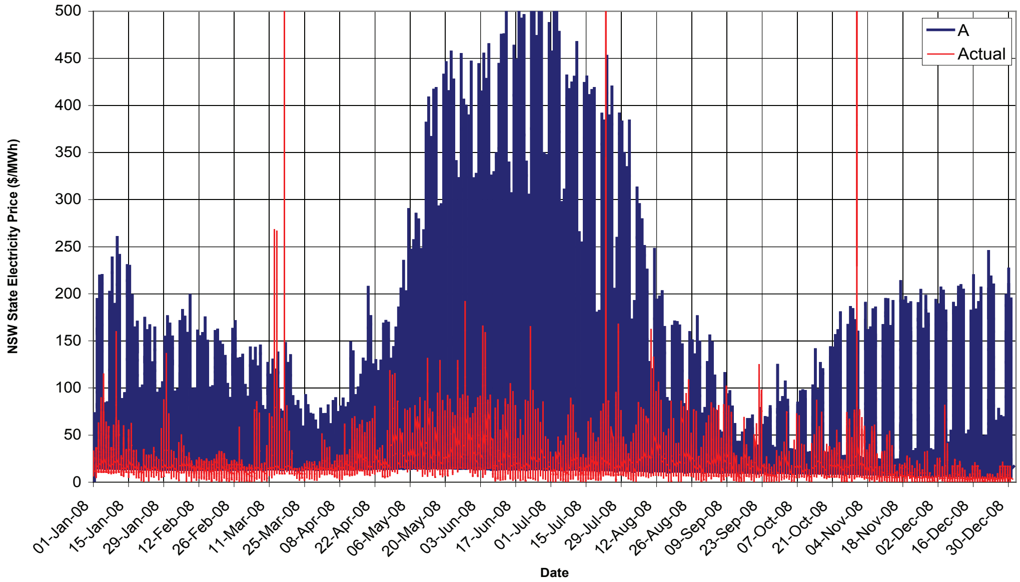


Figure 7: Mean values across 3,000 simulated price paths from simulator A at each half-hour tick during Cal08

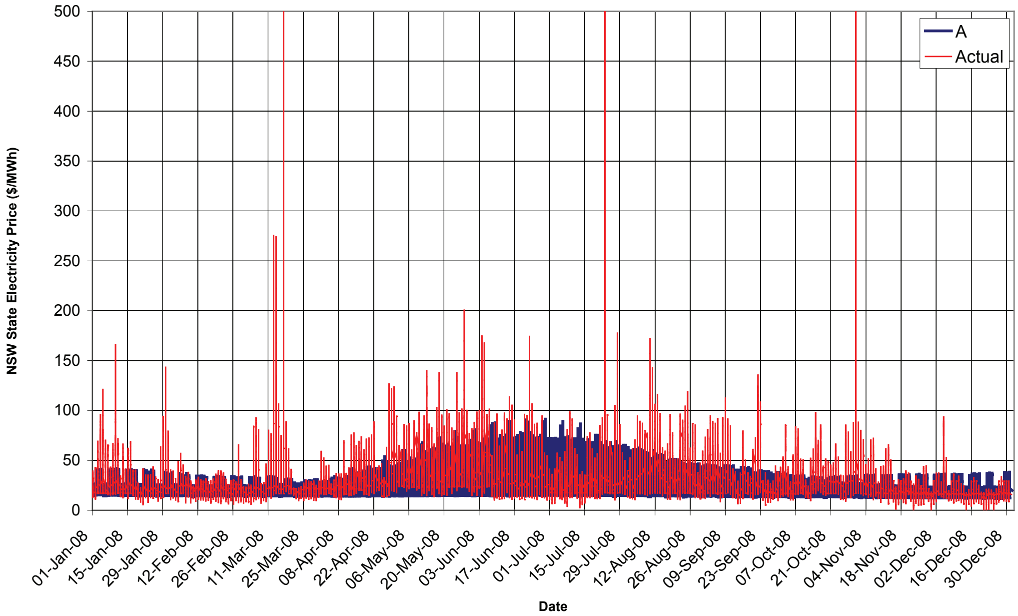


Figure 8: Median values across 3,000 simulated price paths from simulator A at each half-hour tick during Cal08

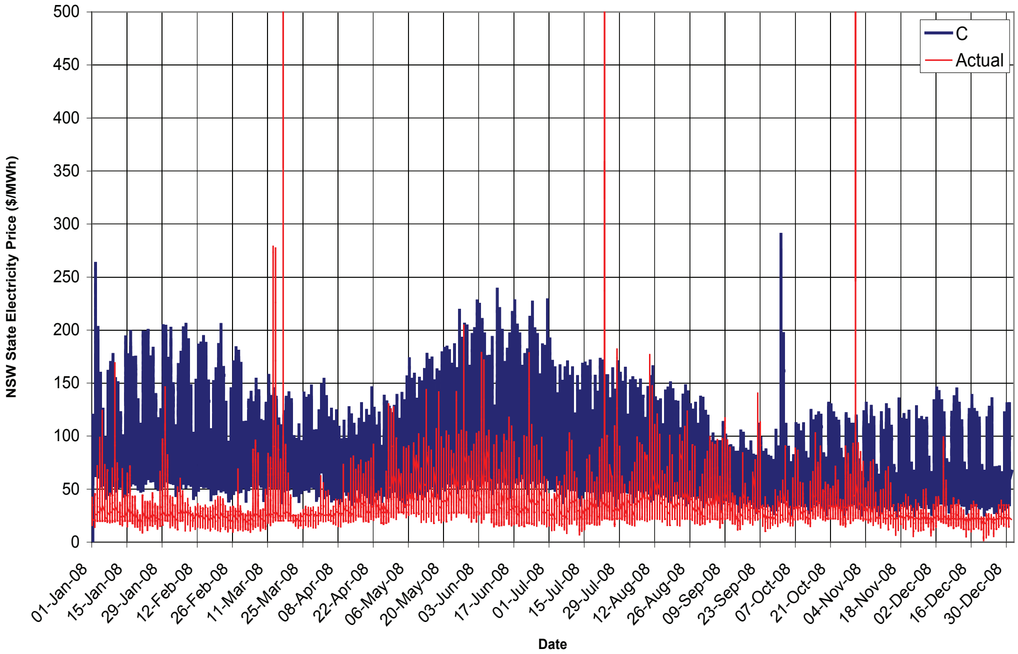


Figure 9: Mean values across 3,000 simulated price paths from simulator C at each half-hour tick during Cal08

greater than the actual prices. The degree of difference between the simulated expected and actual prices is a major problem for the use of this simulation engine as a risk management tool. The extreme level of overpricing evident in the expected values will result in an artificial and dangerously low level of EaR. If this was to result in a cut in the hedge portfolio to minimal levels, then any adverse upward movement in market prices could leave the business of the retailer seriously exposed to considerable losses. In addition, there are pronounced and unrealistic price differences between weekday and weekend simulated prices. This will affect the hedging strategy adopted by the company and will contribute to inaccuracy in the risk management metrics. A further reason for questioning the quality of simulator C concerns the single dominant spike in the means of the simulated paths in Figure 9 at the commencement of the fourth quarter. Over a three day period from 30/9/08 to 2/10/08 there were abnormally high averages throughout those days. The median prices between those times on those days ranged from $65 to $69. In order for this type of clustering of very high mean prices to occur based on 3,000 simulation runs immediately raises questions as to how simulation engine C deals with extreme prices.

This simulator generates price paths that, when averaged and graphed in Figure 9, produce appropriate seasonal shapes. This feature is also replicated for the median prices given in Figure 10. However, while the seasonal shape is appropriate, it is overstated for both expected and median prices during the summer and winter months.

The means and medians of the price paths for simulator D are graphed for Cal08 in Figures 11 and 12. The expected prices depicted in Figure 11 are unrealistically overstated across all half hours in Cal08. While by no means as serious a problem as was the case of the means in simulator C, the over-pricing of the simulated paths corresponding to simulator D carries

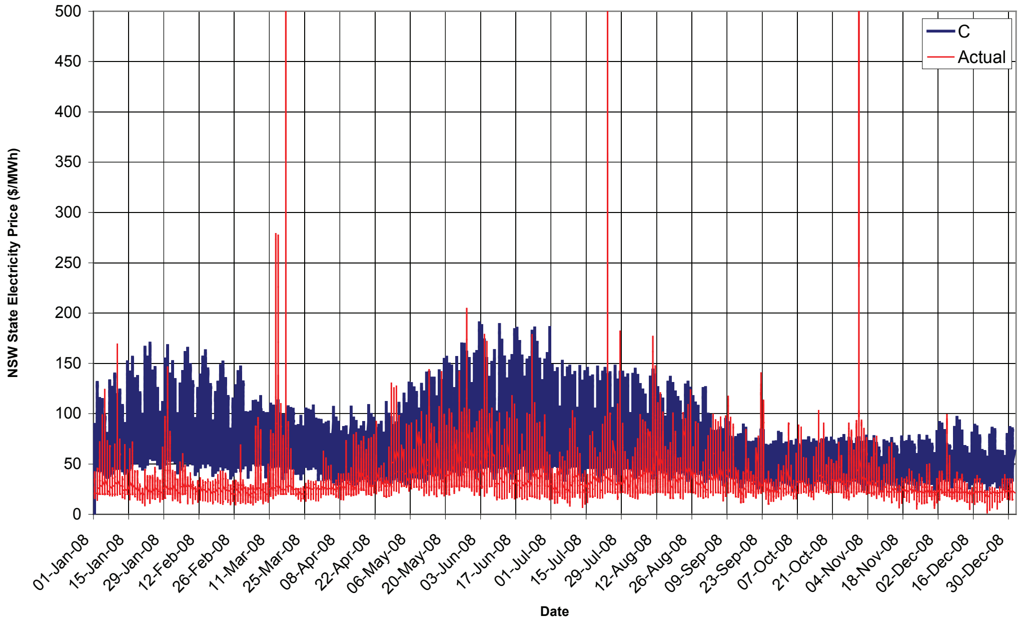


Figure 10: Median values across 3,000 simulated price paths from simulator C at each half-hour tick during Cal08

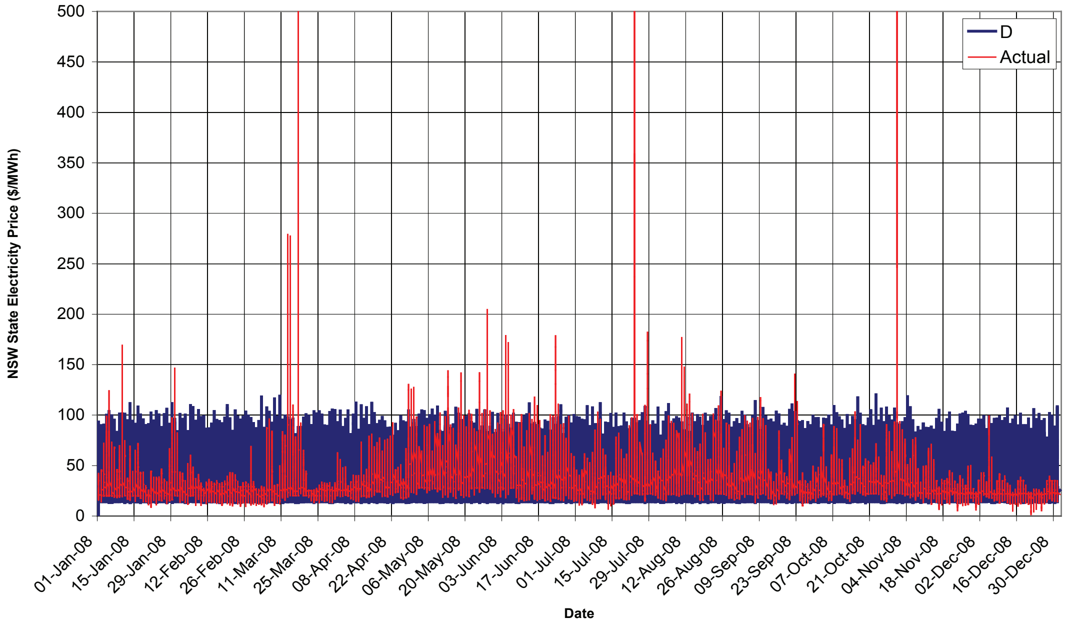


Figure 11: Mean values across 3,000 simulated price paths from simulator D at each half-hour tick during Cal08

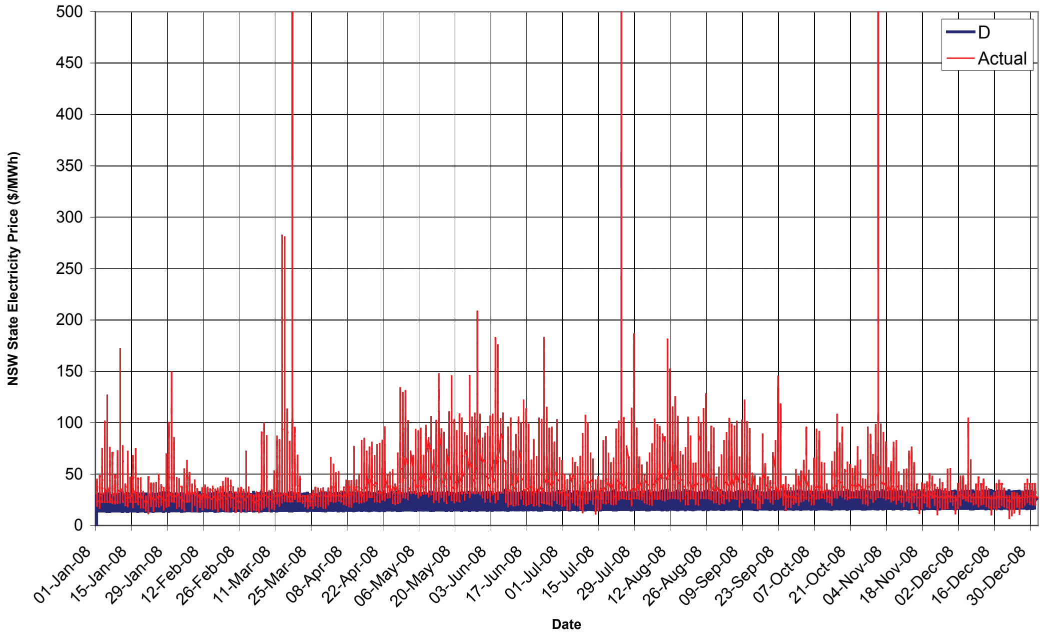


Figure 12: Median values across 3,000 simulated price paths from simulator D at each half-hour tick during Cal08

with it the same consequences of potential under-hedging of the derivative portfolio and an unrealistically low value of the EaR metric. The medians in Figure 12 are characterised by smaller levels compared to the means in Figure 11 and the overall levels implied by the realised prices. A difference between this simulator and the previous two simulators is that the mean and median prices are devoid of any seasonal shape and exhibit an extremely low level of volatility. This reflects a serious problem with the price paths generated by this simulator. The expected and median prices fail to reflect a fundamental seasonal pattern, with higher prices and levels of volatility in the summer and winter months when compared to the shoulder periods. These characteristics are implicit in electricity prices.

The means and medians of the price paths for simulator E are graphed for Cal08 in Figures 13 and 14. As was the case for simulator D, the expected prices depicted in Figure 13 are unrealistically overstated across all half hours in Cal08. This carries with it the same risk as that corresponding to all three previous simulators of an artificially low EaR and a likely under-hedged portfolio. The shape of the expected values, or the medians, by no means appropriately reflects the seasonal nature of electricity data. However, when compared to simulator D, the mean values corresponding to this simulator exhibit higher volatility in the generated price paths.

3.1.2 The financial year 2007/08

The means and medians of the price paths for simulator B are graphed for Fin07/08 in Figures 15 and 16. Figure 15 depicts the abnormally high realized winter prices in quarter 3 and early in quarter 4 of 2007. These extremely high realized prices are repeated in quarter 2 of 2008. Late quarter 3 of 2007, as well as in quarter 1 of 2008, there was the usual high

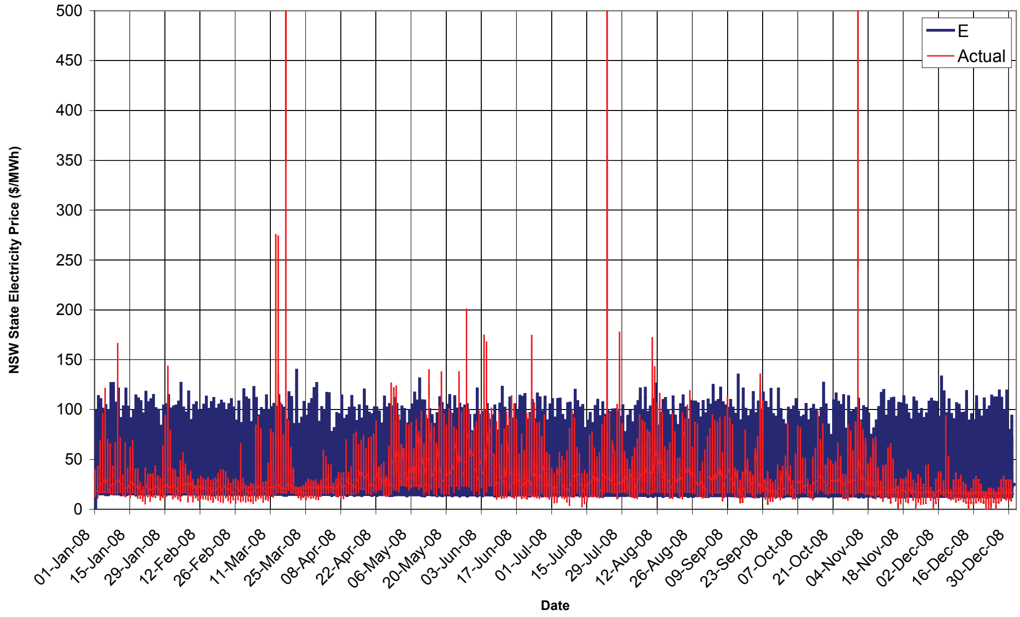


Figure 13: Mean values across 3,000 simulated price paths from simulator E at each half-hour tick during Cal08

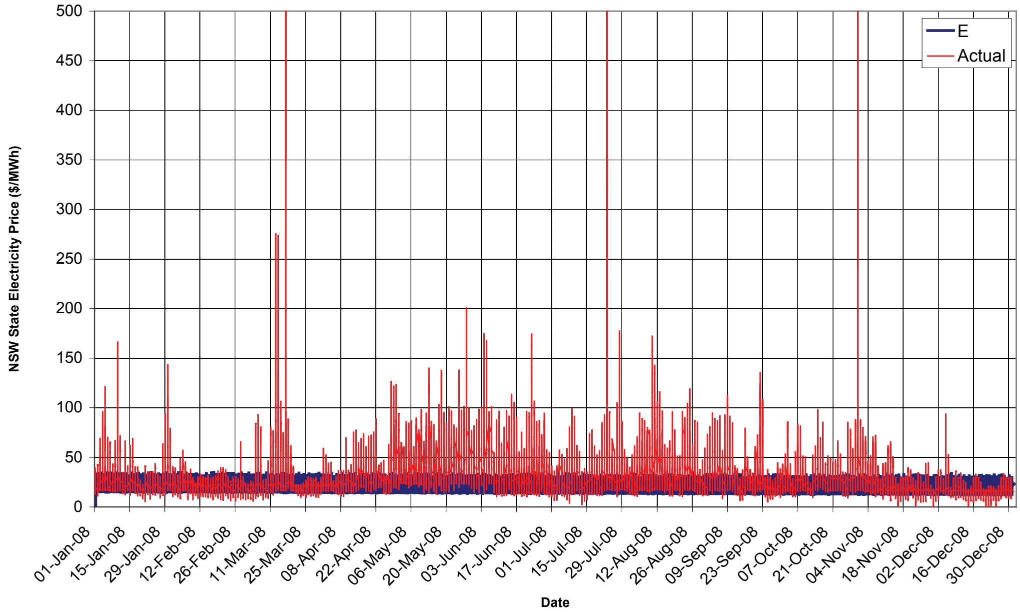


Figure 14: Median values across 3,000 simulated price paths from simulator E at each half-hour tick during Cal08

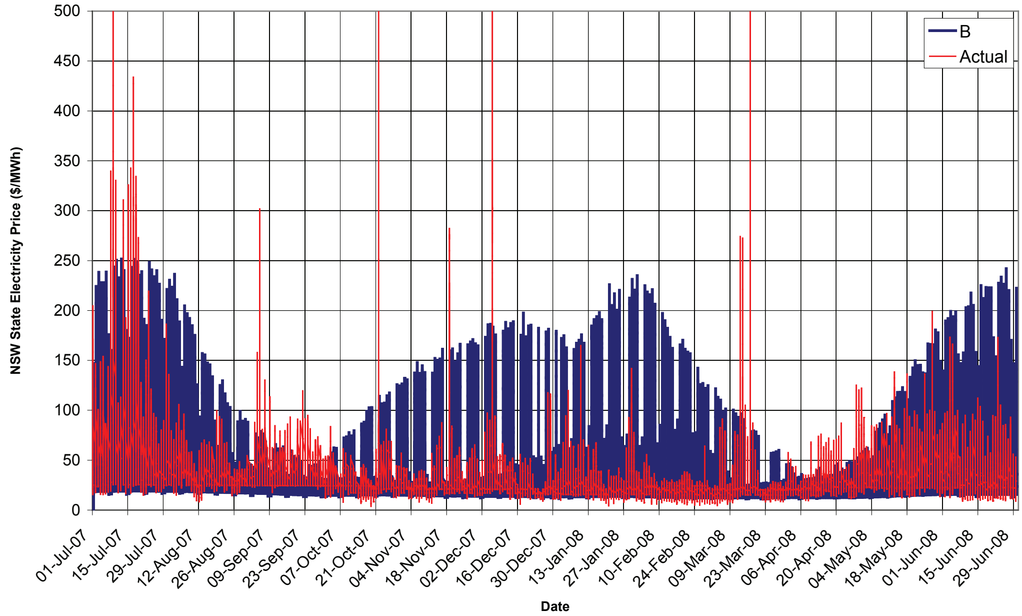


Figure 15: Mean values across 3,000 simulated price paths from simulator B at each half-hour tick during Fin07/08

realized summer prices. Simulator B does a credible job in anticipating these extreme price movements. While it can be argued that the simulator overprices during the summer and winter months of 2007 and 2008, it should be noted that many of the abnormally high prices during those periods are contained within the simulated mean prices. Given the extreme nature of pricing during 2007 and 2008 brought on by record abnormal drought conditions throughout Australia, and the fact that a risk engine is basing its pricing on past realised prices with no knowledge of future conditions likely to affect prices, the result produced by simulator B is most impressive. The medians also do a reasonable job of tracking the realised prices but, as expected, do not capture the extreme levels of the actual prices as well as the means.

The seasonal shape in the means of the paths generated by simulator B conforms to what is expected. Perhaps the volatility for the summer months is somewhat elevated, however, this is not so much problem when considering that these months historically are characterised by high weather-related pricing. What is evident from Figure 15 is a definite intraday pattern where peak and off-peak pricing can be easily observed. A weekly pattern is also transparent with the only criticism being an abnormal drop in the weekend means relative to the working week and visible by the clear slots between adjacent weekly profiles.

The means and medians of the price paths for simulator C are graphed for Fin07/08 in Figures 17 and 18. As was the case for this simulator during Cal08, both the levels of the expected and median prices are significantly higher than the actual prices for all quarters in Fin07/08. In fact, as is evident from Figures 9 and 10 for Cal08 and Figures 17 and 18, all the concerns expressed about the levels of the paths generated by this simulator in Cal08 are still concerns for Fin07/08. The seasonal, weekly and diurnal shapes are well captured by both the means and the medians. Indeed, if levels were more appropriately modelled and

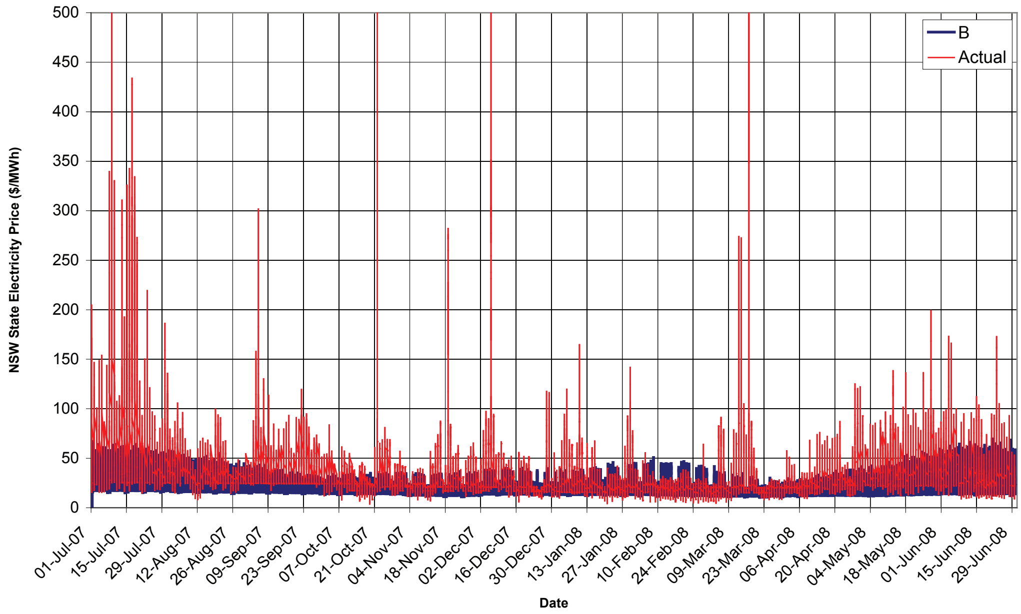


Figure 16: Median values across 3,000 simulated price paths from simulator B at each half-hour tick during Fin07/08

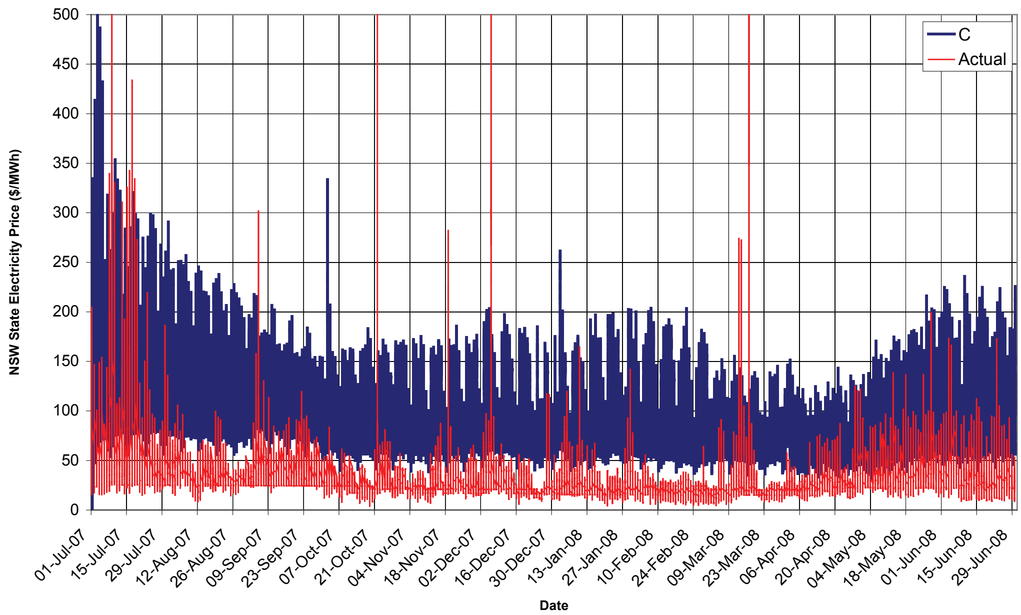


Figure 17: Mean values across 3,000 simulated price paths from simulator C at each half-hour tick during Fin07/08

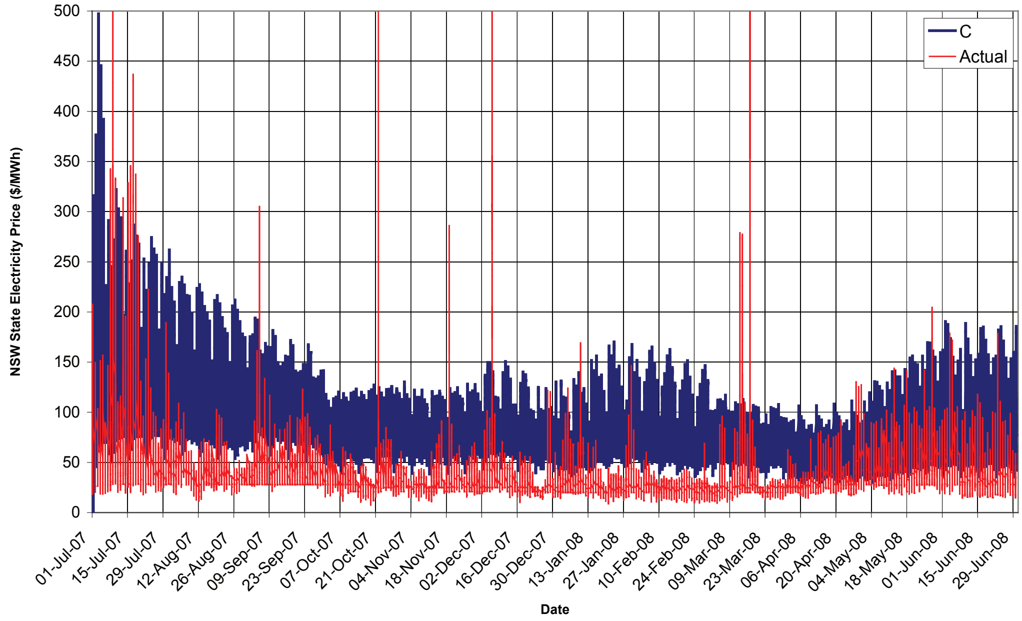


Figure 18: Median values across 3,000 simulated price paths from simulator C at each half-hour tick during Fin07/08

the shape superimposed over the revised levels, then this simulator would be regarded more highly as a useful tool for producing forward prices.

The means and medians of the price paths for simulator D are graphed for Fin07/08 in Figures 19 and 20. Comments concerning this characteristic of the price paths generated by this simulator are similar to that expressed for Cal08. What is worthwhile noting is how the mean values of the price paths capture the levels but not the volatility of the abnormally high winter prices in 2007 and 2008. As was the case for Cal08, this simulator fails to identify seasonal, weekly or diurnal patterns in either the means or the medians. The means and medians of the price paths for simulator E are graphed for Fin07/08 in Figures 21 and 22. Comments concerning levels and shape for this simulator are the same as for simulator D for Fin07/08, and for the discussion corresponding to Figures 13 and 14 for Cal08.

3.2 The second diagnostic test; outliers

A second diagnostic test centres on the percentage of times the means of simulation paths lie outside plus or minus three standard deviation units of the true mean. If the mean of a price path falls outside plus or minus three standard deviation units, then it is regarded as an outlier and an unsatisfactory representation of the series from which it is being generated. First, we consider the percentage of times the means lie outside these limits over the four quarters of the Calendar year 2008 (Cal08), then the months and, finally, the weeks. The same analysis is then repeated for Financial year 2007/08 (Fin07/08). In any quarter, we would expect the means to be outside these limits for only a small percentage of times. If the simulation engines cannot deliver a zero or small percentage of misses over all quarters then their quality has to be seriously questioned. The percentage of times the means lie outside the limits will most likely increase when moving from quarters to months, and then to weeks.

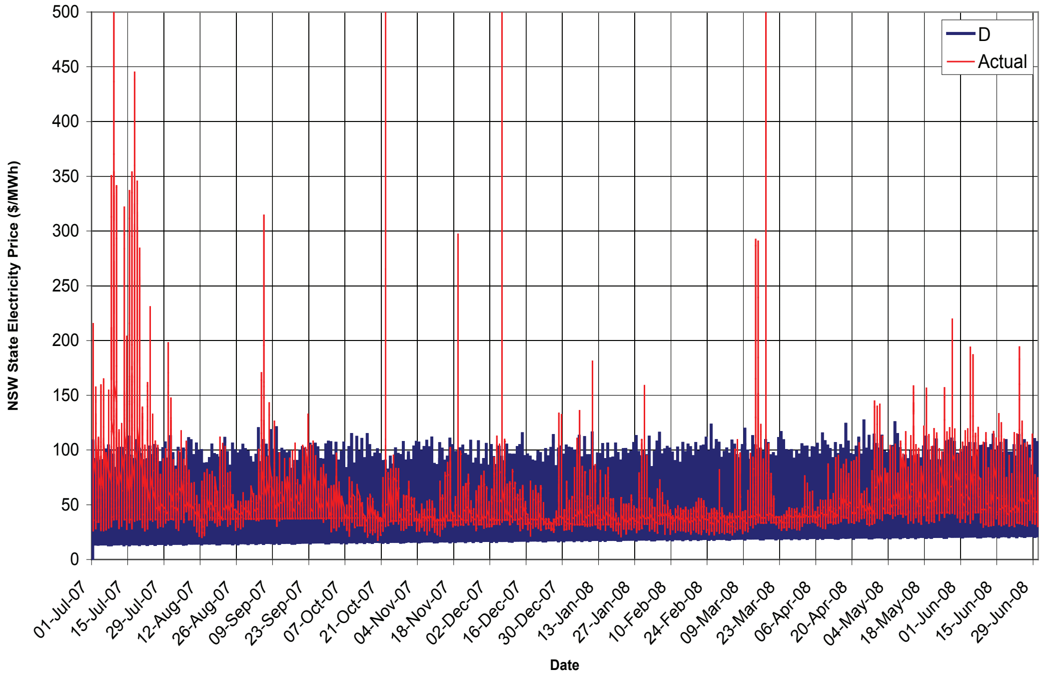


Figure 19: Mean values across 3,000 simulated price paths from simulator D at each half-hour tick during Fin07/08

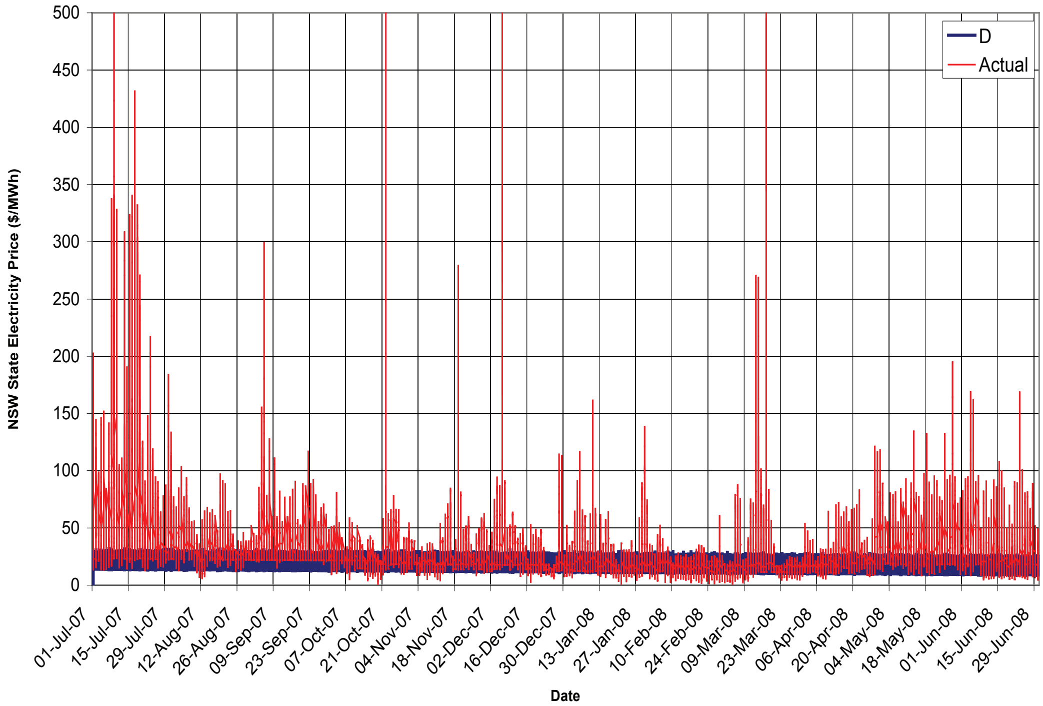


Figure 20: Median values across 3,000 simulated price paths from simulator D at each half-hour tick during Fin07/08

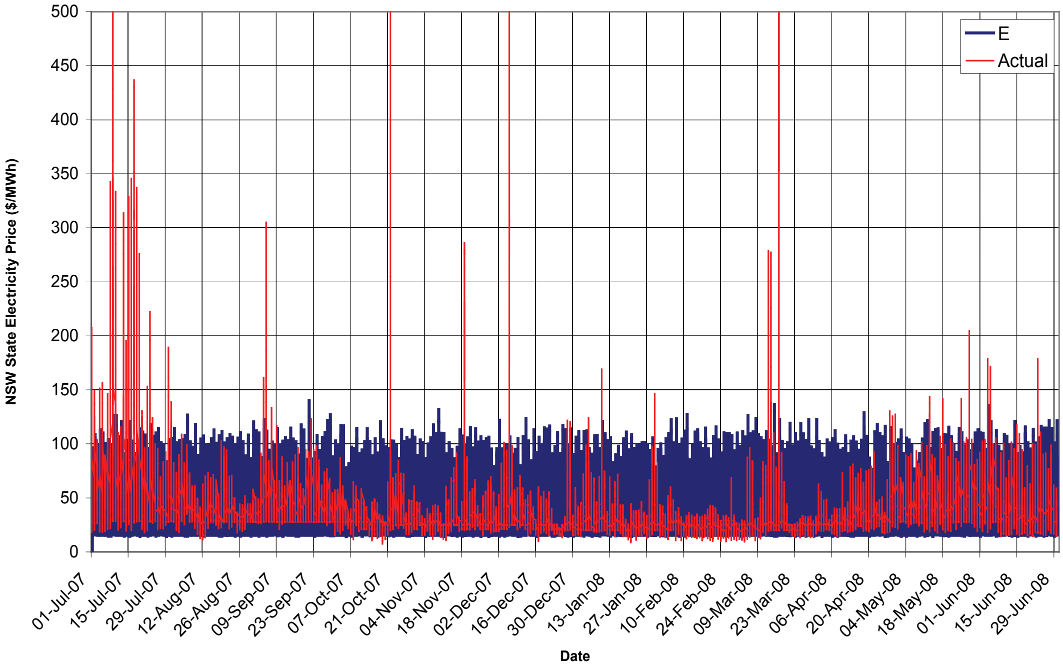


Figure 21: Mean values across 3,000 simulated price paths from simulator E at each half-hour tick during Fin07/08

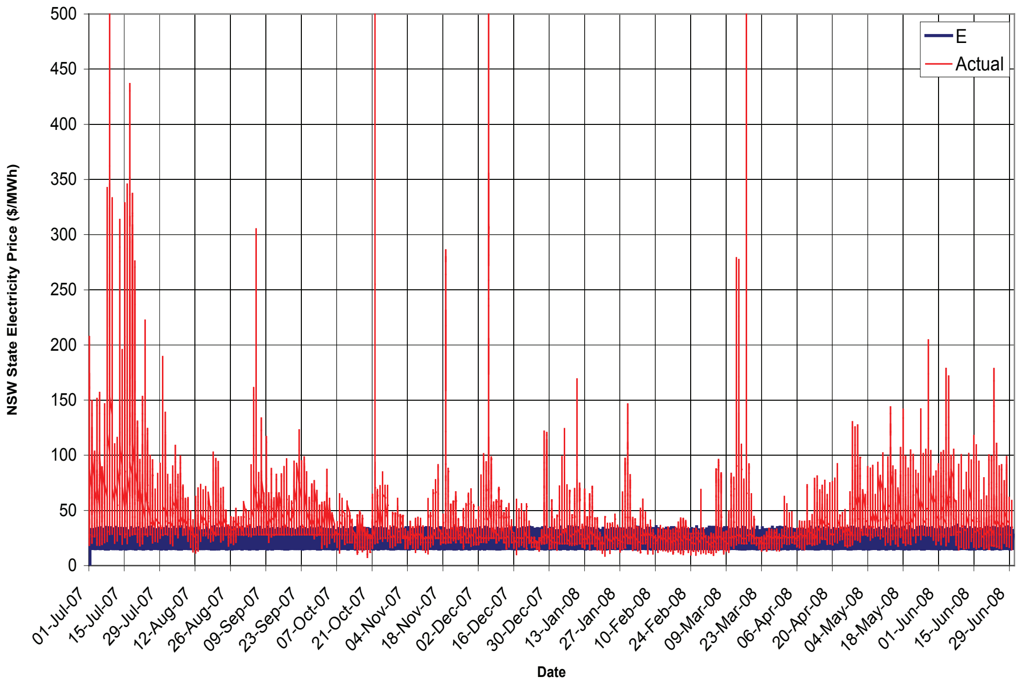


Figure 22: Median values across 3,000 simulated price paths from simulator E at each half-hour tick during Fin07/08

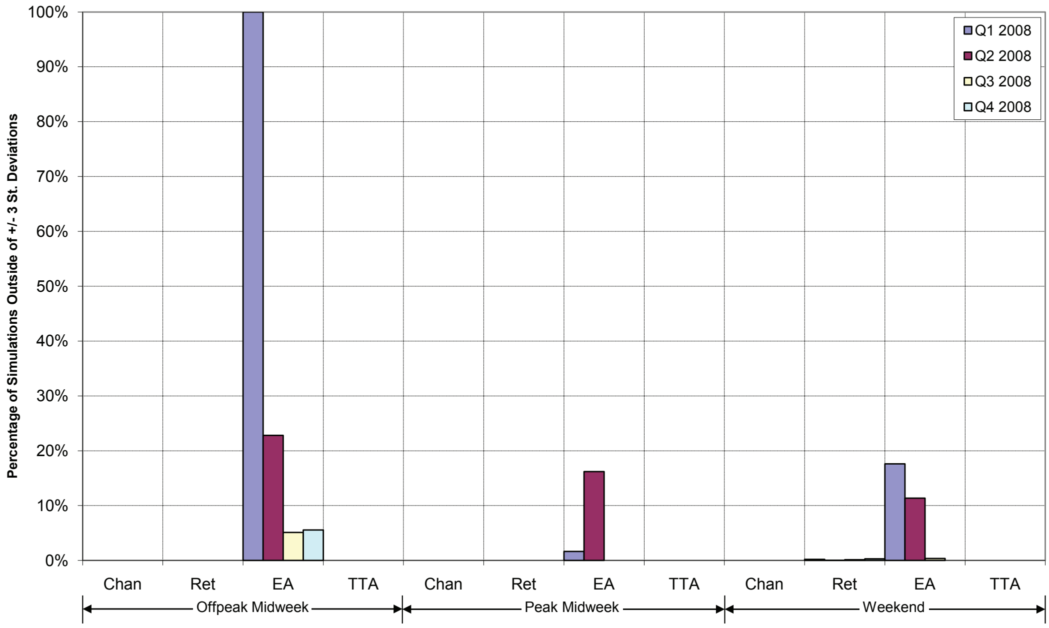


Figure 23: The percentage of times the quarterly means of the simulated data from simulators A, C, D and E were outside three standard deviation units of the true quarterly means for Cal08

However, the engines should deliver relatively small percentages even at these higher levels of resolution. If the price paths from a simulator are characterised by small means (or medians) relative to the mean of the actual prices, along with relatively small standard deviations, then it is likely that the mean price path levels will be consistently low. In this case, it is likely that the simulated means (or medians) will be consistently below three standard deviations of the true mean. When small mean or median values are less than three standard deviations from the true mean, then this is a breach in the other direction to what is usually expected. Usually, it is high volatility of the simulated prices that causes high mean levels to lie outside three standard deviations of the true mean.

3.2.1 The Calendar year 2008

Figure 23 shows the percentage of times the means of the 3,000 price simulations from simulators A, C, D and E breached three units of standard deviation from the true quarterly means for Cal08. The data was divided into midweek, peak and off-peak periods, as well as weekends, and graphed accordingly. What is noticeable is the poor performance of simulator C at the quarterly level of resolution. Especially of concern was quarter 1 in the off-peak period where the means breached three units of standard deviation 100% of the time. Breaches occurred to a lesser extent during quarter 2, and were minimal in quarters 3 and 4. Based on the means of the price paths, the numbers of outliers produced by this simulator in the peak period are not as pronounced as for the off-peak period and are predominantly contained to quarter 2. There were between 10% and 20% of breaches in the first two quarters of the weekend period. Overall, the unsatisfactory number of outliers produced by this simulator

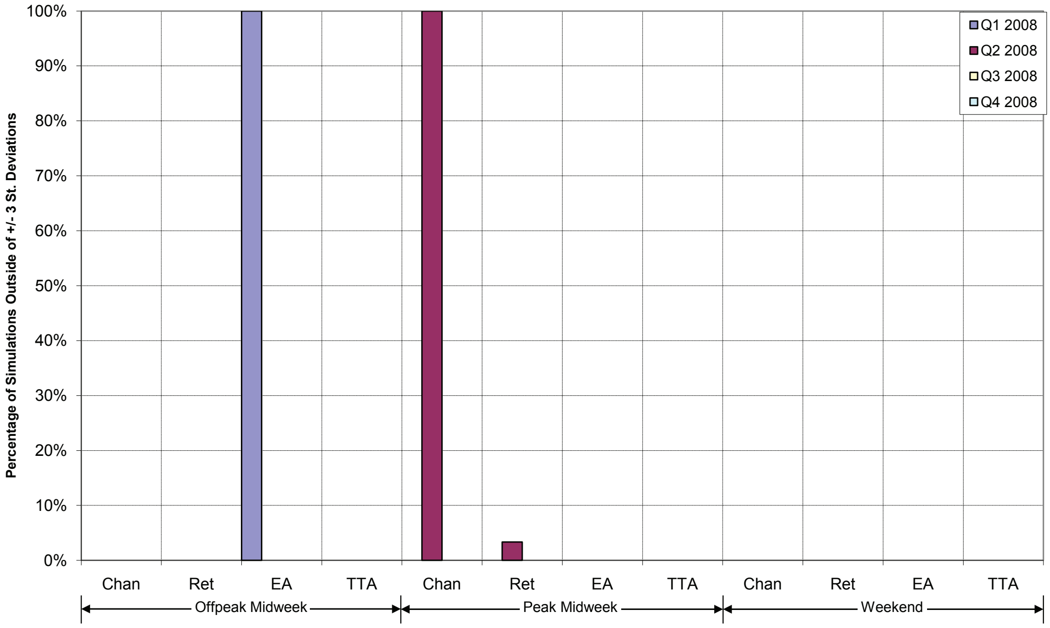


Figure 24: The percentage of times the quarterly medians of the simulated data from simulators A, C, D and E were outside three standard deviation units of the true quarterly means for Cal08

according to this diagnostic is consistent with the conclusion from Diagnostic 1, where the mean prices were unrealistically elevated above the actual. A conclusion that can be drawn from the results from the first two diagnostics is that simulator C is inappropriate technology to be generating price paths to determine earnings distributions and to calculate the EaR metric. The remaining three simulators showed no evidence of mean outlier behaviour in any of the partitioned periods. Figure 24 depicts the percentage of times the quarterly medians exceeded three standard deviation units from the true quarterly means for Cal08. The medians of the price paths for simulator D during the peak period breached the limits 100% of the time. The reason for this is the relatively low level of the medians and the fact that they fall outside, but below three units of standard deviation less than the mean of the actual prices. Additionally, and as expected, due to abnormally high level of the medians from the price paths generated by simulator C, the limits set by this diagnostic were breached 100% of the time. However, at the level of the medians of the price paths, simulators A, D and E pass this diagnostic test at the quarterly granularity for the peak, off-peak and weekend periods.

Figures 25 and 26 depict the percentage of times the means and the medians of the price paths generated by all simulators breach three units of standard deviations of the true monthly means during Cal08. There is likely to be an increase in the number of breaches of the limits set by the second diagnostic test due to an increase in the level of granularity from quarters to months. This follows with the recalculation of the means and medians of the price paths generated by all the simulators, and the true mean and standard deviation from the realised prices for Cal08. Figure 25 shows that the number of outliers identified from the price path means from simulator C has increased dramatically across all three periods. For the off-peak period, outliers are identified in most months with unacceptable levels in the summer

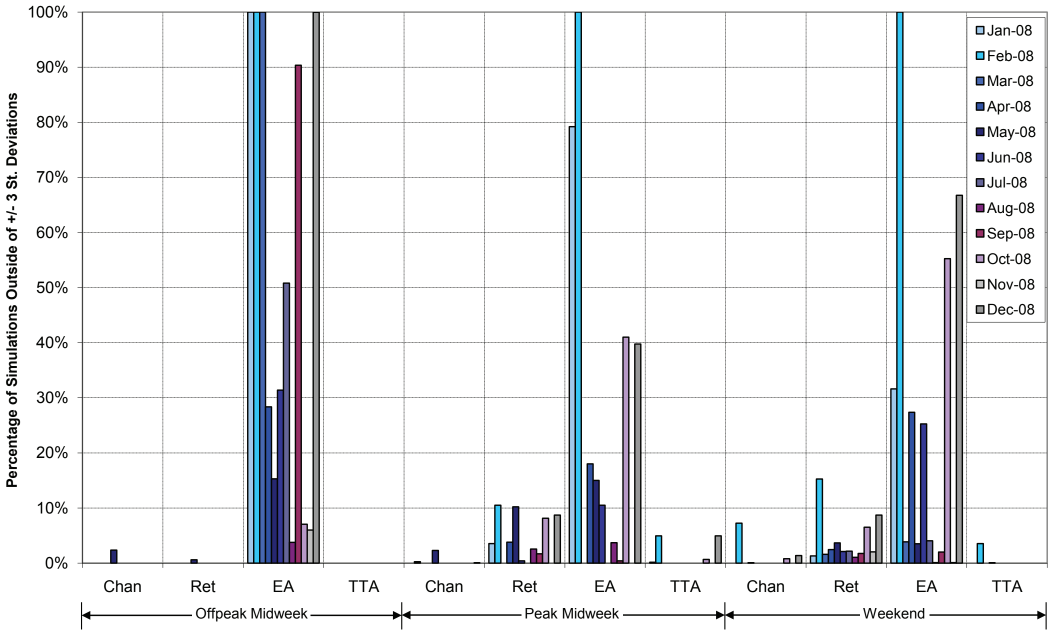


Figure 25: The percentage of times the monthly means of the simulated data from simulators A, C, D and E were outside three standard deviation units of the true quarterly means for Cal08

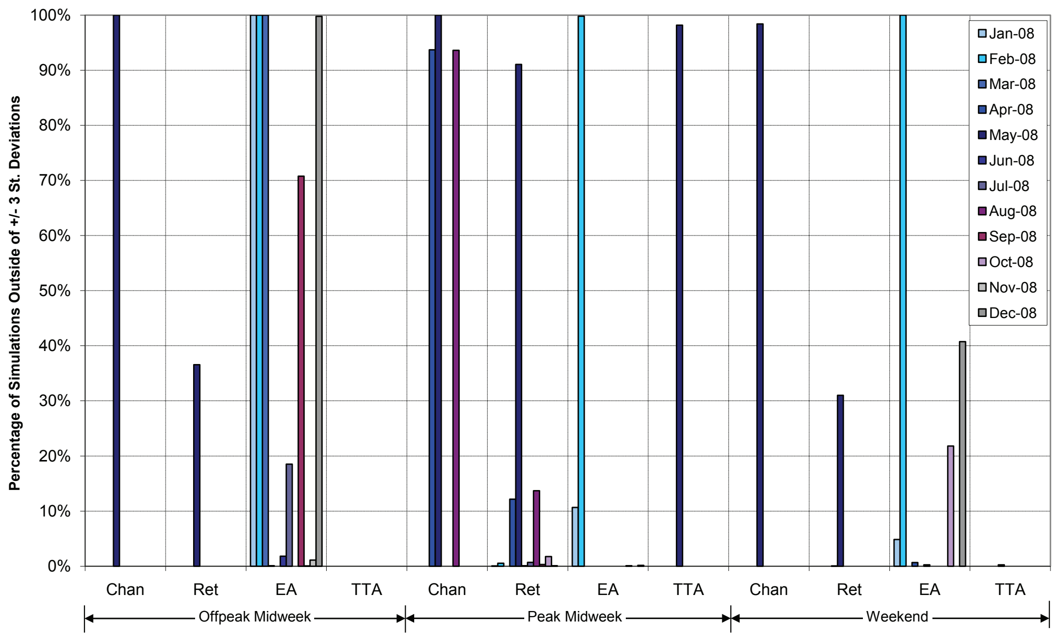


Figure 26: The percentage of times the monthly medians of the simulated data from simulators A, C, D and E were outside three standard deviation units of the true quarterly means for Cal08

months of January, February, March and December, and in June, July and September during the winter period. Outliers were at an unacceptable level in February for the peak period, while February and December produced an abnormal number of breaches for the weekend. The increased number of breaches is to be expected given the abnormally high level of mean pricing resulting from this simulator. When medians are used to identify outliers, the same problems identified from the means still exist, but not to the same extent. On the basis of these results this simulator should not be used to price risk management metrics. This diagnostic test identifies minimal evidence of mean outlier presence at the monthly level of resolution for simulators A and D. While there is some indication of outlier behaviour for simulator E, it is not problematical as breaches occur at most around 10% of the time. What is interesting from Figure 26 is the extreme level of breaches based on medians that occur for simulator D, and to a lesser extent for simulator E. This is caused by the very low level of the medians of the price paths from these simulators resulting in the values falling below three units of standard deviation from the true mean, which represents a breach of the diagnostic test.

As indicated in the introductory discussion to the second diagnostic test, there is a large increase in the number of breaches of the set limits at the weekly level of granularity, for both the weekly means and medians. All simulators were adversely represented by a high incidence of outliers, with perhaps an exception for simulator A. However, the pattern that emerged for the quarters, and then the months, is reinforced at the weekly level.

3.2.2 The Financial year 2007/08

As seen in Figure 27, simulator C again performs poorly at the quarterly level of granularity for Fin07/08. In the off-peak period, this diagnostic test shows that the price paths from this simulator that are more than three units of standard deviation away from the true mean 100% of the time during quarters 2 and 3. In quarter 4, outliers are identified over 20% of the time. While not as pronounced in the peak period, the percentage number of breaches is over 60% in quarter 2 and 50% in quarter 3. Quarter 4 during the weekends is also signals a problem with breaches 25% of the time. When breaches are determined using medians, Figure 28 points to breaches of the limits 100% of the time in quarters 2 and 3 in the off-peak period. As was the case for Cal08, simulator C yet again proves to be unsatisfactory technology for generating price paths for measuring EaR. According to Figures 27 and 28, the remaining simulators performed creditably across all quarters in Fin07/08 according to Diagnostic 2. At the monthly level of resolution, Figures 29 and 30 diagrammatically depict how simulator C continues to produce far too many outliers for most of the months in the off-peak period, and how it performs not much better in the other two periods. At the mean level, outliers produced by simulators D and E are abnormal during September during the off-peak period, and to a lesser extent during July and September at weekends. The only reason for concern with the price paths of simulator B at the monthly level of resolution are breaches for 25% of the time in September during the off-peak period. This pattern is repeated when using the medians to identify outliers, with the added effect of increased breaches for simulators D and E due to the low level of the medians relative to the true mean and its standard deviation. The breaches are more numerous at the weekly level of granularity, and follow the trend of the monthly results.

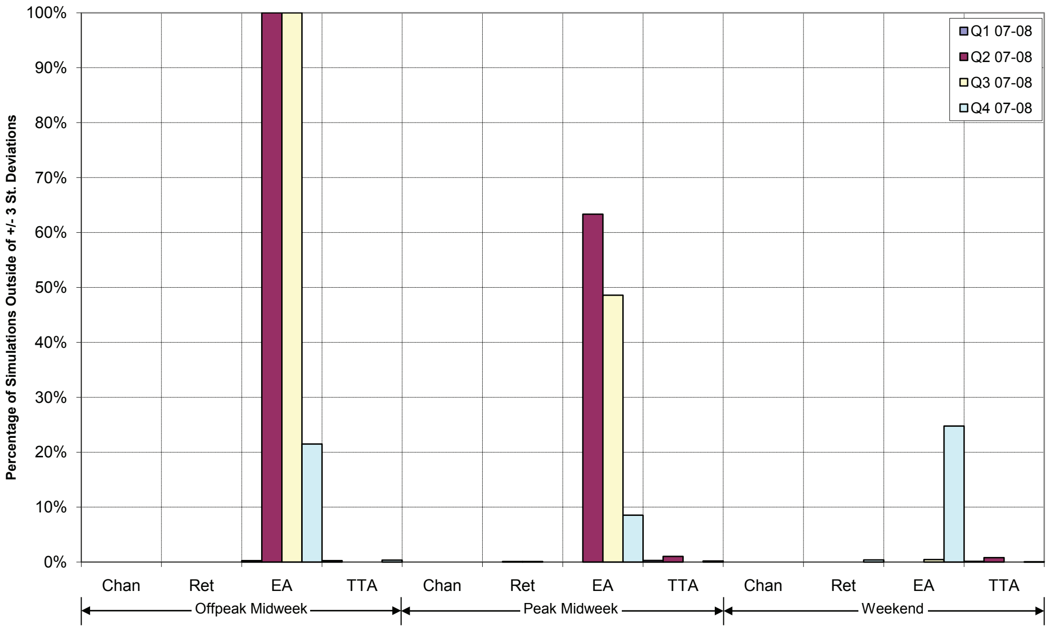


Figure 27: The percentage of times the quarterly means of the simulated data from simulators B, C, D and E were outside three standard deviation units of the true quarterly means for Fin07/08

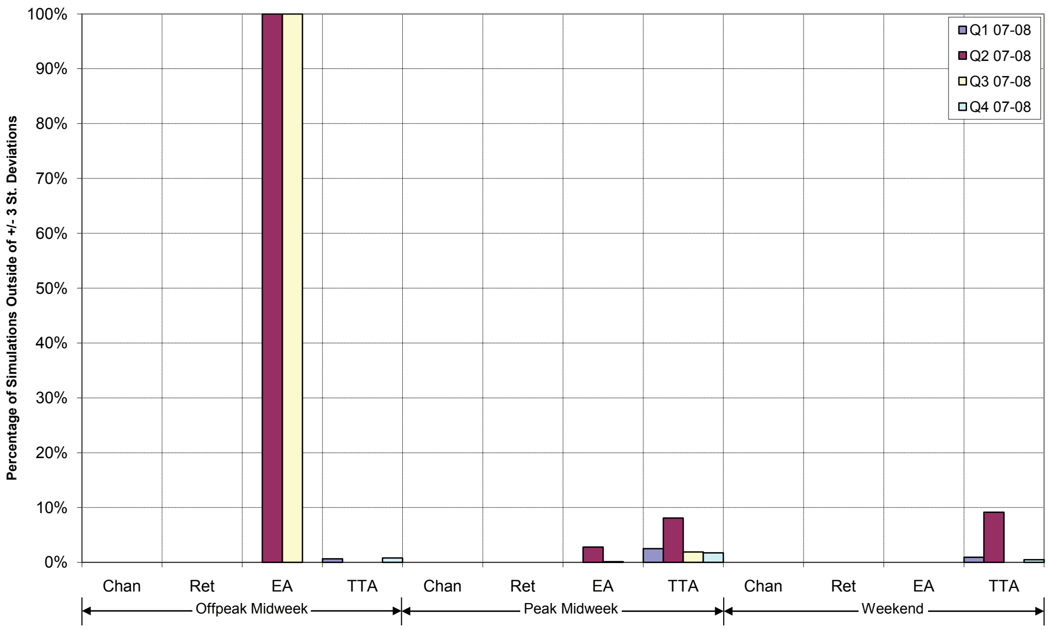


Figure 28: The percentage of times the quarterly medians of the simulated data from simulators B, C, D and E were outside three standard deviation units of the true quarterly means for Fin07/08

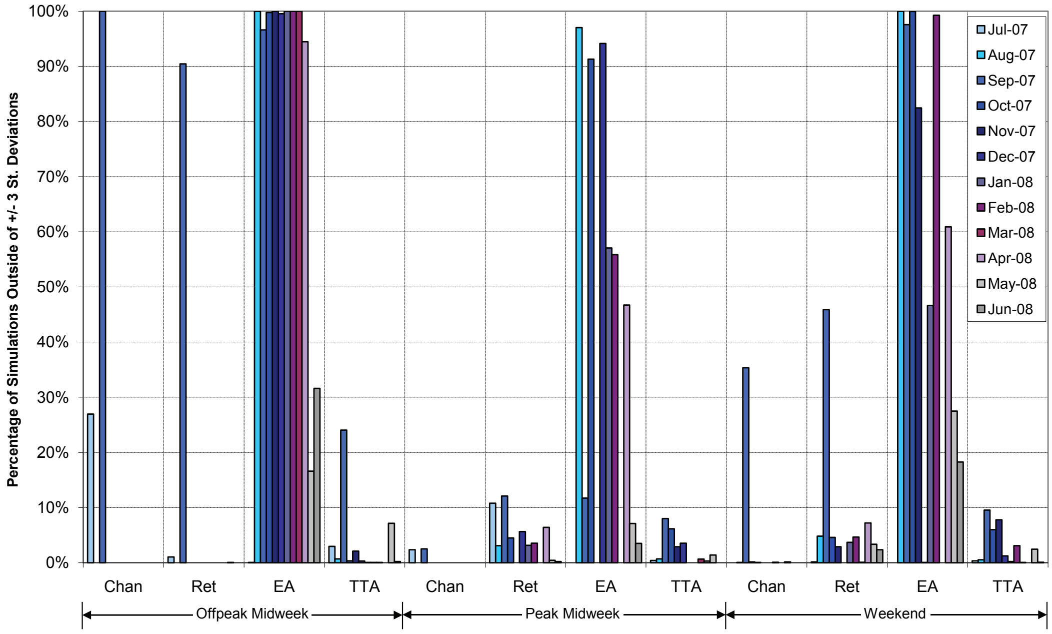


Figure 29: The percentage of times the monthly means of the simulated data from simulators B, C, D and E were outside three standard deviation units of the true quarterly means for Fin07/08

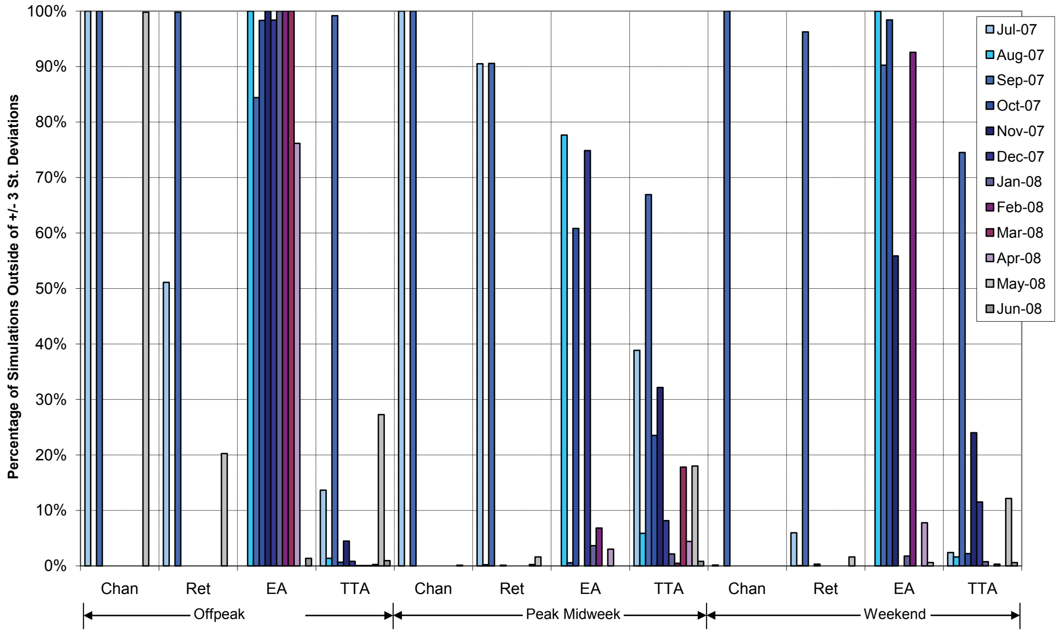


Figure 30: The percentage of times the monthly medians of the simulated data from simulators B, C, D and E were outside three standard deviation units of the true quarterly means for Fin07/08

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	All
Flat	35.63	33.27	39.76	26.40	45.14	35.83	31.01	67.07	39.14	39.25

Table 2: Historical flat average pool prices from 2000 to 2008

3.3 The third diagnostic test; volatility washing

The third diagnostic test was used to assess the quality of a risk simulation engine concerns its ability to avoid what is known as "volatility washing". Table 2 contains the historical flat average pool prices for the years from 2000-08. It follows that the average for any year over that time span can lie somewhere between \$26.40 and \$67.07. The underlying half-hour pool price generator that generates these price averages in any year is characterised by an average across the years of \$39.25 and a standard deviation of 11.74. We would expect a simulator that accurately generates price paths from this underlying pool price distribution to also have price paths over a year that have averages with a similar standard deviation.

3.3.1 The Calendar year 2008

The summary statistics of the averages for the 3,000 price paths generated by the simulators analysed in Cal08 can be found in Table 3. Figure 31 graphs the distributions of the average prices for all simulators and measured against the actual price for 2008 that is given by the yellow histogram bar. What is obvious at first glance from both Table 3 and Figure 31 is how inappropriate is the location of the distribution for the averages of the price paths generated by simulator C relative to the distributions of the other simulators and the true average price (yellow line). No further consideration is given to the merits (or otherwise) of simulator C in this discussion as it only serves to distract from the analysis of the remaining three. Volatility washing can be clearly observed for the averages from all three simulators. The term volatility washing refers a loss of information concerning average prices as a result of the simulation process. This loss is measured by a reduction in the standard deviation of the average prices generated by a simulator. A reduction in the spread of the average prices from a simulator could have potentially serious pricing implications. What is expected is the ability of a risk engine simulator to generate a distribution of prices that encompasses all the previous average prices given in Table 2. Reducing the spread, as measured by the standard deviation, could result in a distribution so tight that the probability of that distribution enclosing the true (realised) average value is very small. From Table 3 and Figure 31 it is clear that simulator D suffers from volatility washing. Additionally, the expected value of the price path averages is substantially less than the average of \$39.25 given in Table 2. A low value for standard deviation of 1.34, severely restricts the range of values comprising the distribution of the average prices. As shown in Figure 31, the tight distribution about an inaccurate mean has the distribution for Simulator D missing the true average price. Simulator A seems in Figure 31 to capture the true average rather well. Simulator E is not far behind. Simulator A has also been thoroughly washed with the result that the distribution of its average prices is tight with a standard deviation of 1.67. With such a tight distribution, it is important to have an accurate mean. However, this was the case for simulator A as can be seen from Table 3. On the other hand, Simulator E has the advantage of a larger standard deviation (less volatility washing) and therefore, a greater range of potential average prices. While all the simulators studied for Cal08 exhibited volatility washing, it should be noted that the abnormally high

	A	C	D	E
Mean	40.82	85.12	31.83	35.19
Median	40.78	84.56	31.76	34.81
St Dev	1.67	6.83	1.34	3.62
Kurtosis	0.24	0.05	-0.04	0.40
Min	35.71	69.39	28.41	27.10
Max	47.96	108.51	36.71	50.19

Table 3: Summary statistics for the averages of 3,000 simulated price paths for Cal08 from risk simulation engines A, C, D and E

	B	C	D	E
Mean	35.91	107.29	31.85	35.22
Median	35.29	106.46	31.76	34.84
St Dev	5.86	6.99	1.34	3.62
Kurtosis	12.69	0.00	-0.01	0.42
Min	20.98	89.14	28.42	27.10
Max	106.34	136.02	36.72	50.14

Table 4: Summary statistics for the averages of 3,000 simulated price paths for Fin07/08 from risk simulation engines B,C,D and E

(and record) mean flat price for 2007 will have inflated the overall standard deviation for the averages across 2000 to 2008. Without this outlier the impact of volatility washing would be reduced.

3.3.2 The Financial year 2007/08

The summary statistics of the averages for the 3,000 price paths generated by the simulators analysed in Fin07/08 can be found in Table 4. Figure 32 graphs the distributions of the average prices for all simulators and measured against the actual price for 2008. As previously noted in the discussion of this diagnostic for the Cal08, the location of the distribution of the average prices for simulator C is so remote compared to the other simulators and the true average value as to render futile any further analysis of this simulator. Furthermore, as evidenced from the standard deviations in Table 4, volatility washing is present in all the other simulators. However, it is not as pronounced as for the simulated price paths in Cal08. The standard deviations reported in Table 4 indicate that simulator B is less affected than the other two. With a standard deviation of 5.86 compared to 11.74 for the historical, this simulator has a volatility washing of more than 50%. The consequence of this is a reduced chance that the distribution of simulator B will include the historical expected value. As for Cal08, the distribution of the average prices of simulator D is so reduced due to volatility washing that it fails to include the expected historical mean. Alternatively, simulator E whose volatility is not as severely washed does include the true average. However, its distribution does not have the range of that of simulator B and, as such would not be preferred to that simulator according to this diagnostic. Again, it should be stressed that the historical standard deviation of 11.74 has been inflated by the record average flat price in 2007.

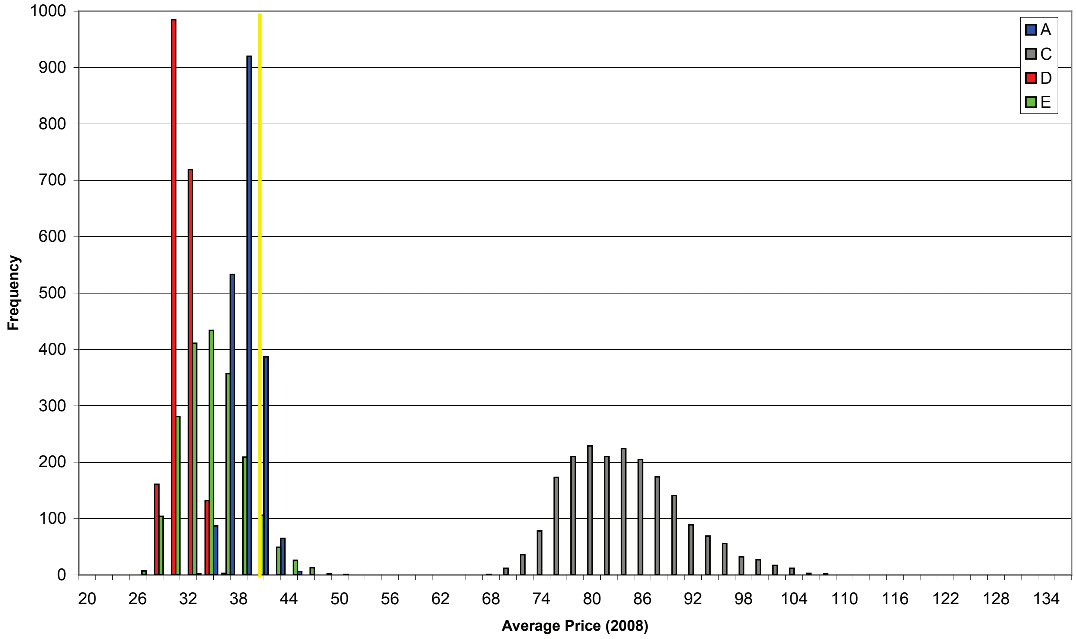


Figure 31: Distributions of the average prices for the 3,000 price paths during Cal08 for all simulators

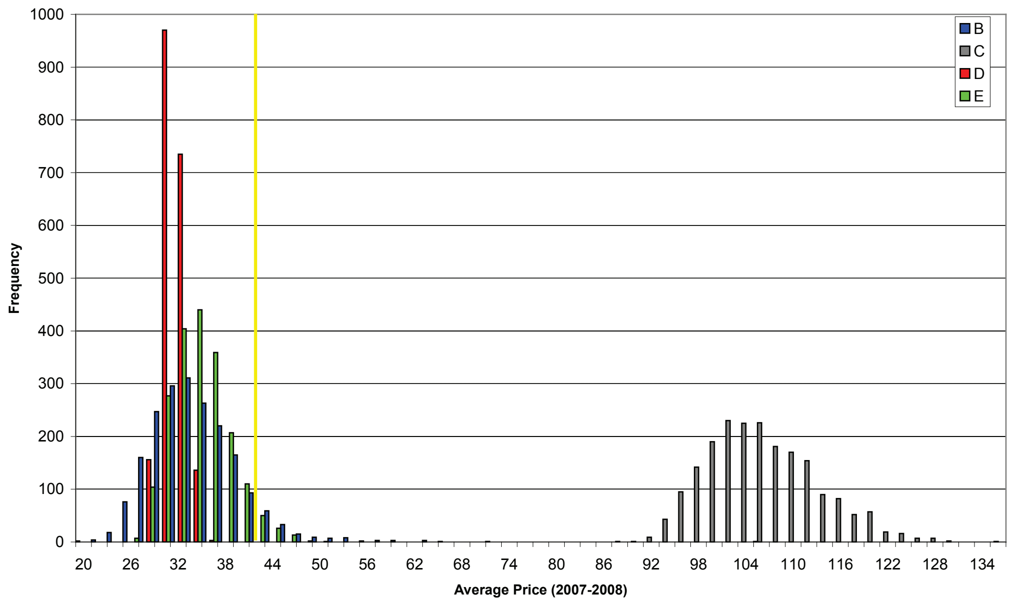


Figure 32: Distributions of the average prices for the 3,000 price paths during Fin07/0808 for all simulators

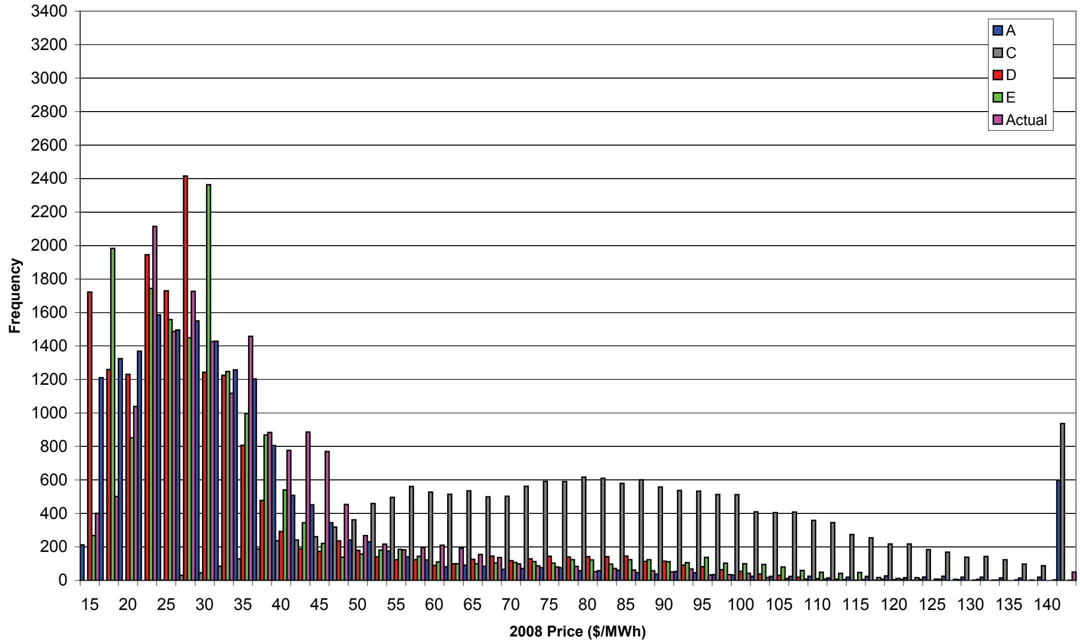


Figure 33: The distribution of a sample of the various price paths generated by simulators A,C, D and E, as well as the realised prices for Cal08

3.4 The fourth diagnostic test; extreme prices

The fourth diagnostic test determines the ability of each simulator to generate extreme prices. Bi-modal distributions of prices across both Cal08 and Fin07/08 are expected. In addition, a large positive skew indicates the incidence of high prices and possible price spikes. Kolmogorov-Smirnov tests were run to test the hypotheses that the price distributions from the simulators matched the historical realized prices. The p-values from these tests are reported in Table 5.

3.4.1 The Calendar year 2008

Figure 33 shows the distribution of prices from a randomly selected price path that was generated by simulators A, C, D and E, as well as the distribution of the realized prices for Cal08.

As can be observed from Figure 33, the distributions of the prices from all the simulators, as well as the realized prices, are bimodal. This is an accepted characteristic of the distribution of electricity prices and is due to the seasonal effect. Notwithstanding the distribution of prices from simulator C that we choose not to discuss further, the distributions from the other simulators all have a large positive skew. In particular, simulator A has a spike in the frequency of prices in the extreme positive bin in Figure 33. This spike indicates a large frequency of abnormally high prices and even price spikes. However, it is substantially large in comparison to the frequency of the realised prices in this bin. This would suggest an over-representation of high price regimes in the price paths generated by simulator A. For Cal08, the Kolmogorov-Smirnov tests reported in Table 5 found that there were significant

Simulator	CAL08 K-S statistic (p-value)	Fin07/08 K-S statistic (p-value)
A	0.0003	
B		0.9442
C	0.0000	0.0000
D	0.0048	0.0049
D	0.7054	0.7185

Table 5: Summary statistics for the averages of 3,000 simulated price paths for Fin07/08 from risk simulation engines B,C,D and E

differences between the simulated series generated by risk engines A, C and D when compared to the realized series. Additionally, this test found that there was no significant difference between the simulated series from risk engine E and the realized series. Given that these tests were based on 17,568 observations, it was not expected that a statistical test would have found that there was a match between the two distributions. In interpreting the results from this statistical test, it should be kept in mind that each risk engine was represented by only one price path.

3.4.2 The financial year 2007/08

Figure 34 shows the distribution of prices from a randomly selected price path that was generated by simulators B, C, D and E, as well as the distribution of the realized prices for Fin07/08. No further analysis is carried out for simulator C due to the unrealistically high prices associated with this risk engine. The expected bi-modality of the distribution of each randomly selected price path from simulators B, C, D and E is evident from Figure 34. As was the case for Cal08, all distributions exhibit a pronounced positive skew. However, what is of importance is the incidence of abnormally high realized prices as indicated by the frequency count in the extreme right-hand bin of the histogram given in Figure 34. Fin07/08 was a period where there were numerous price spikes and regimes of high prices that were the result of the effect of serious water shortages on base load generating cost and capacity caused by drought conditions throughout Australia. This effect was especially noticeable in NSW. Of the price paths examined by simulators B, D and E, it was only simulator B that matched the actual incidence of high prices.

This observation from Figure 34 concerning simulator B is confirmed by the result from the Kolmogorov-Smirnov test in Table 5. The null hypothesis of no statistical difference between the distribution of the actual prices and the price path generated from simulator B was not rejected. This result was also the case for simulator E and was a repeat of the result for this simulator in Cal08. However, as is evident from Figure 34, simulator E seems incapable of generating the high price regimes that are a characteristic of electricity prices.

3.5 The fifth diagnostic test; earnings

The fifth diagnostic test examines the distributions of earnings differences corresponding to the differences in earnings generated by each collection of price paths relative to the earnings calculated using the settlement price. The earnings difference distributions are analysed using half-hour simulated and actual prices for Cal08 and Fin07/08. Unhedged earnings difference distributions are graphed against each other on a quarterly basis. From these graphs, an

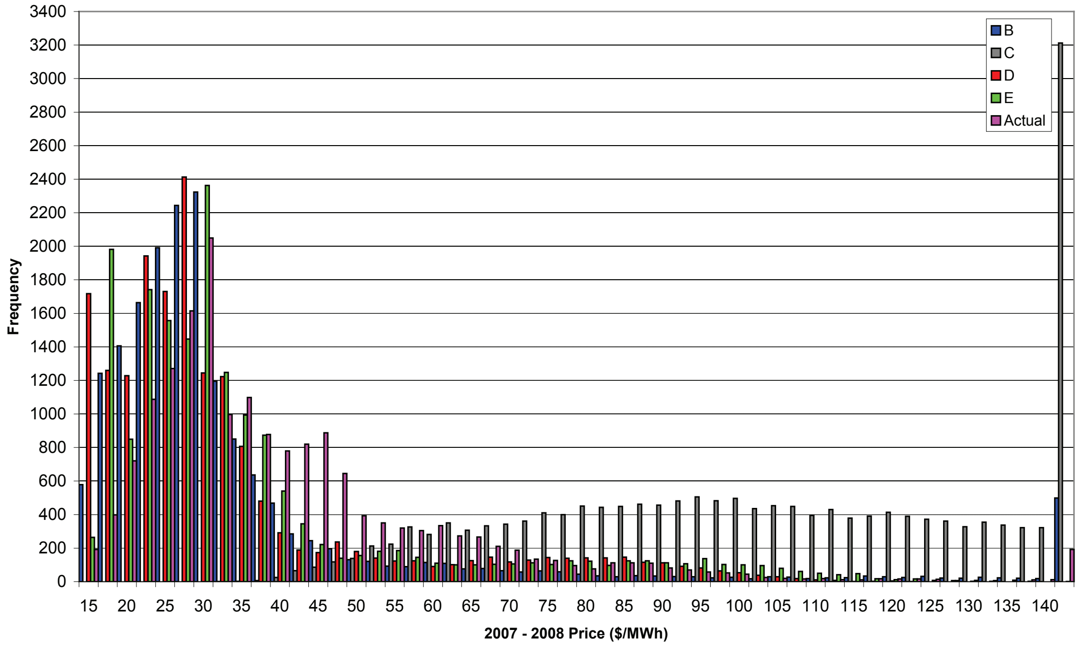


Figure 34: The distribution of the various price paths generated by simulators B,C, D and E, as well as the realised prices for Fin07/08

opinion is offered as to the suitability of each engine for generating an EaR measure to be used as a risk management tool by an electricity retailer. Additionally, the earnings difference distributions corresponding to peak periods are provided. This makes possible a contrast of the effect on the earnings difference distribution aggregated across all periods during the day from that of the peak period. The peak period is regarded as more likely to expose the business to increased risk.

EaR measures the amount of earnings volatility around the expected gross margin of an electricity retailer if the current hedge portfolio is held to maturity. It is a function of both the pool price outcomes and load volatility. It is a gross margin-based measure where an empirical distribution of the expected gross margin is simulated. To obviate the simulation process, data is imported from forward prices generated by a risk simulation engine, as well as forecasts of the retailer's committed portfolio load and trading portfolio. During iterations of the simulation process, options and swaps are marked against the simulated forward price curve. Expected gross margin estimates are then derived from cash flows calculated under the load forecasts. Given that no trading portfolios were part of the supplied data set, the expected gross margin distribution collapses to that of an unhedged earnings distribution. It is calculated as the difference between the revenue accruing to the retailer by on-selling electricity purchased from the pool, and the cost of purchasing the electricity to be sold at the variable pool price. Forward prices are used to estimate the future variable pool prices and are generated by the risk simulation engines that have been the subject of the previous diagnostic testing. Further, given that the data supplied for this study did not include the actual retailer's forecasted load, realized NSW State Load was substituted for the periods under consideration. After an empirical earnings distribution is simulated from 3000

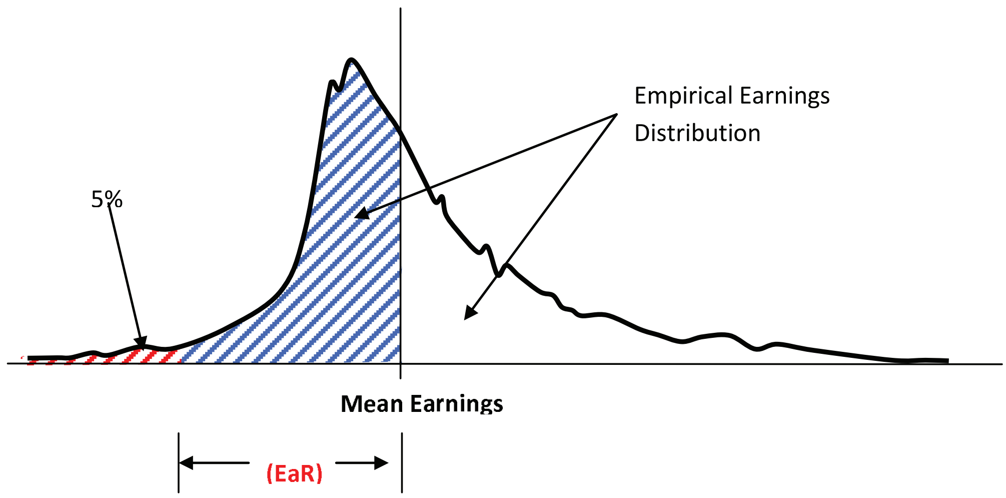


Figure 35: The empirical earnings distribution and earnings at risk (EaR) for an unhedged portfolio

iterations of this process, the EaR measure is calculated by subtracting the lower 5% level of the empirical distribution of the earnings distribution from its mean. This calculation is shown in Figure 35.

Because the price charged to customers is fixed, the retailer's revenue is fixed. Let the fixed revenue accruing to the retailer in a nominated period be R and the level of load purchased from the pool and on-sold to customers in that period be L. If the actual settlement price is p_A and the simulated price is p_S, then the half-hourly difference in earnings, ΔE, that results from that actually settled each half-hour over the period and that derived from the corresponding simulated price is given by

$$\Delta E = (R - p_S L) - (R - p_A L) = (p_A - p_S)L.$$

If the simulated prices are overpriced, then that implies there will be a positive difference between the estimated earnings using the overpriced simulated prices relative to the earnings derived from the settlement prices. In this case, the earnings difference, ΔE, reflects a cost to earnings resulting from using the simulated prices relative to the actual prices. This degree of inflation in the earnings figure can be interpreted as the opportunity cost associated with using a particular simulator to generate earnings numbers. If the realized and simulated prices were the same then the earnings difference under this definition would be zero. In other words, the earnings difference is a way of assessing the effect from simulated prices that systematically differ from the realized prices. The converse is the case for a simulator that underprices. It follows that when we talk about earnings differences and their distribution, we are referring to the effect that simulated prices have on overestimating or underestimating earnings and how this effect is distributed across the simulated prices. The effect of a simulator overpricing is to reduce the EaR figure. This could encourage an inappropriately low investment in a hedge portfolio to overcome the incidence of negative earnings. As a consequence, a low EaR figure could adversely affect the viability of the business by increasing

Cal08	A	C	D	E
Quarter 1	+	+	-	+
Quarter 2	+	+	-	-
Quarter 3	+	+	-	-
Quarter 4	-	+	-	-

Fin07/08	B	C	D	E
Quarter 1	-	+	-	-
Quarter 2	-	+	-	-
Quarter 3	+	+	-	+
Quarter 4	-	+	-	-

Table 6: The sign of the expected earnings differences corresponding to each simulation engine for each quarter of Cal08 and Fin07/08

the exposure to abnormally high and volatile pricing. This effect is to be avoided. By definition, earnings for an unhedged portfolio are exposed to higher realized (settlement) prices than to the corresponding prices that are the expected output from a simulator. Realized electricity prices are extremely volatile. Not only are they characterised by unexpectedly large price spikes, but also by regimes of high prices before reversion to significantly lower levels. Negative differences between the simulated prices and the prices that a retailer pays for electricity from the pool will also differ on a half-hour basis according to whether the data is partitioned into peak, off-peak or weekend periods. The uncertainty associated with this volatility results, more often than not, in negative earnings differences on a half-hourly basis. As a consequence, the expectation is that the mean earnings difference for an unhedged portfolio will be negative. Using this diagnostic, we assessed the suitability of the various pricing simulators being studied by determining the appropriateness of the distributions of earnings differences from the risk management perspective previously discussed. As described, the EaR metric is a simple calculation from an earnings distribution. Of critical importance in the calculation is how the expected (mean) earnings figure from using simulated prices differs from using that using the realized prices. Accordingly, by comparing and contrasting the distributions of earnings differences and their means, conclusions are arrived at concerning the merits of each risk simulation engine.

3.5.1 The Calendar year 2008

The left table in 6 indicates the sign of the expected earnings differences corresponding to each simulation engine for each quarter of Cal08, while the corresponding quarterly earnings difference distributions are shown for quarter 1 in Figure 36, quarter 2 in Figure 37 and quarter 4 in Figure 38. Note the results for quarter 3 are qualitatively similar to those shown in Figure 37. Simulator A has positive expected earnings differences in the first three quarters of Cal08. Recall the argument that, due to the volatility of electricity prices, we should expect negative earnings differences in an unhedged portfolio more often than not. It is then surprising that the expected (mean) earnings difference using this simulator is a positive value in the majority of quarters. In the case of simulator C, there is a positive expected earnings difference figure for all quarters. With the expected earnings difference figures all positive and greater than those of simulator A by a factor of ten, we conclude that this simulator results in expected earnings difference estimates that do not reflect industry expectations. Yet again, simulator C disqualifies itself as suitable risk management tool for calculating EaR estimates. Additionally, the use of both these simulators could result in an artificially low EaR number and, as a consequence, suggest a reduced need for investment in a hedge portfolio in order to protect earnings. This could leave an electricity retailer with an earnings book that is dangerously under-hedged. The remaining simulators (D and E) conform more

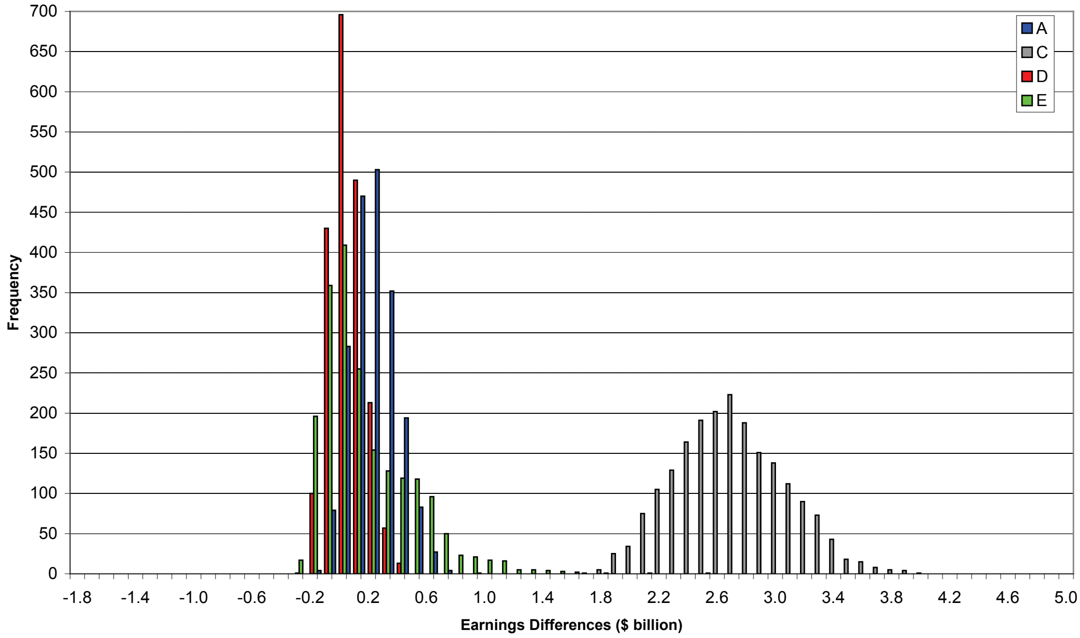


Figure 36: The earnings difference distributions using the various price paths generated by simulators A,C, D and E for quarter 1, Cal08

to industry expectations with a majority of negative expected earnings across the quarters. In the calculation of the EaR metric, what is of critical importance is the empirical earnings distribution and the corresponding distribution of earnings differences. A simulator ideally should have an earnings difference distribution that has a reasonable spread that allows for both positive and negative tail behaviour. While extreme positive and negative differences are not ideal, given the extreme volatility of electricity prices, they are to be expected and any expectation that a simulator could precisely capture their second moment (variance) behaviour is unrealistic. From Figure 36, this appears to be case for simulators A and E. While there is expected positive and negative tail behaviour for simulators A and E, the bulk of their earnings difference distributions are characterized by positive numbers that indicate a positive expected earnings figure. This latter characteristic counts against the usefulness of these two simulators. In the case of simulator C, its earnings difference distribution is located so far to the right in the positive domain that it appears that any incidence of negative earnings difference would be totally unexpected. While simulator D has positive and negative earnings difference tail behaviour, it is small relative to that of simulators A and E. This seriously detracts from its attractiveness as an appropriate simulator for generating future price paths. A small spread in the context of an earnings difference distribution can occur for a number of reasons. One reason would be for the case when the simulated half-hourly prices were very similar to the corresponding settled values. Of course, if the small variance was the result of accurate pricing by a simulator, then this would be a good attribute. However, given the high volatility of the actual prices relative to the simulated prices for simulator D, this is unlikely to be the case. At another extreme, it could be due to a higher than expected mean

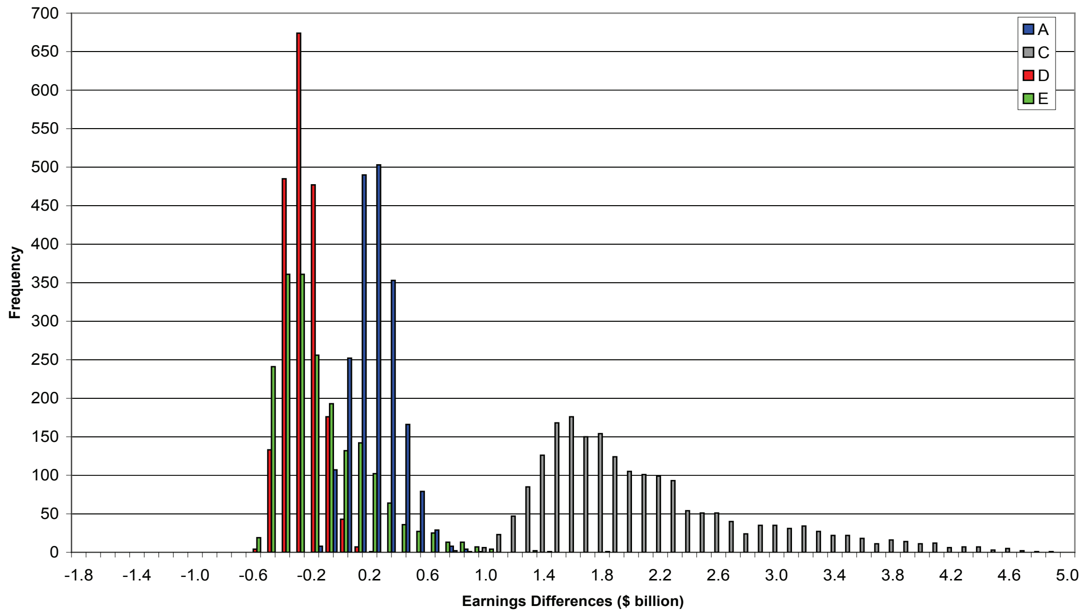


Figure 37: The earnings difference distributions using the various price paths generated by simulators A,C, D and E for quarter 2, Cal08

and low volatility in the simulated prices. If the expected mean of the simulated prices is greater than the mean of the settlement prices, in addition to a corresponding low volatility of the simulated price paths, then what is essentially being captured in the spread of the earnings difference distribution are relatively small positive and negative earnings differences due to the elevated mean of the simulated prices. Of course there will be some more extreme positive and negative tail behaviour representing the larger positive and negative differences resulting from high and low settlement prices. What is missing and unacceptable in the case of simulator D are simulated high prices that are an important characteristic of actual electricity prices. This is confirmed in what was previously reported for this simulator in Diagnostic 1. Recall how Figure 11 indicated unexpected low volatility for the half-hourly means for simulator D. A satisfactory simulator should be able to replicate the volatility in the settlement prices. The comparison of the diagrammatic representations of the earnings difference distributions for the first quarter of Cal08 corresponding to the four simulators rule out simulator C as a serious tool to be used in the risk management process.

As seen in Figure 37, the earnings difference distribution for simulator A in quarter 2 lies predominantly in the positive domain with very little negative tail behaviour. While not as pronounced as that of simulator C, the fact that the earnings difference distribution is positively located means that the expected value is positive. This implies that the expected value of the earnings distribution based on the simulated prices is much greater than the location of the settlement value. This will have an adverse effect on the EaR with likely under-hedging of the trading portfolio. The earnings difference distributions for quarter 2 corresponding to simulators D and E both have negative means in accordance with industry expectations. However, the more expansive tail behaviour associated with simulator E makes it a more preferable option to simulator D for generating price paths. The earnings difference

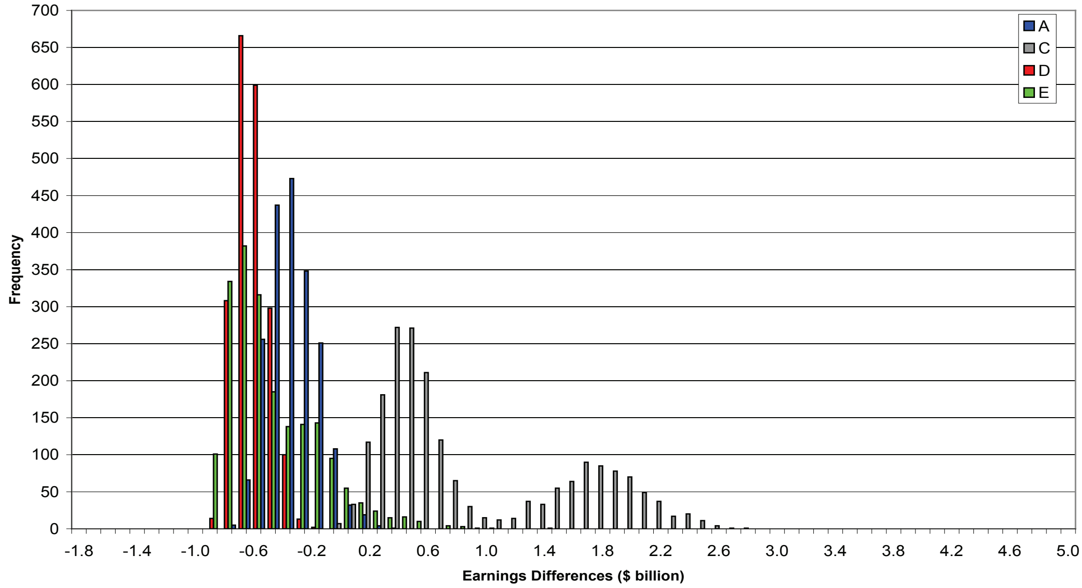


Figure 38: The earnings difference distributions using the various price paths generated by simulators A,C, D and E for quarter 4, Cal08

distributions for quarter 3, Cal08 are qualitatively the same as that for quarter 2.

Figure 38 depicts the earnings difference distributions calculated from the differences of the load weighted simulated and settlement prices for quarter 4, Cal08, for simulators A, C, D and E. The central location of the earnings difference distributions of the simulators A, D and E is more appropriate for quarter 4 than for the other quarters in Cal08. As was the case for quarter 3, simulator D has a low variance relative to simulators A and E. On the other hand, the variance of the earnings difference distributions corresponding to simulators A and E are more in keeping with expectations. The criticism offered concerning the low variance of the distribution of prices generated from simulator D again holds for this quarter. Simulator C continues to produce price paths that are inappropriate for risk management purposes.

3.5.2 The Financial year 2007/08

The right table in 6 indicates the sign of the expected earnings differences corresponding to each simulation engine for each quarter of Fin07/08. The corresponding quarterly earnings difference distributions are shown for quarter 1 in Figure 39, quarter 2 in Figure 40, quarter 3 in Figure 41 and quarter 4 in Figure 42. According to Table 6, simulator C is the only simulator during Fin07/08 to have positive expected earnings in all quarters. With the expectation of negative expected earnings differences, no improvement is observed in Fin07/08 for this simulator from that observed in Cal08. The other simulators studied during Fin07/08 generally all have negative earnings expectations as expected. The earnings difference distributions for Fin07/08 diagnosed correspond to price paths generated by three of the four simulators diagnosed in Cal08. The exception was simulator A that was substituted by simulator B during Fin07/08. It should also be recalled that the first two quarters of Fin07/08 were quarters where electricity prices took on levels not previously experienced. High price levels and volatility of the prices in these two quarters were generally attributed to the extreme

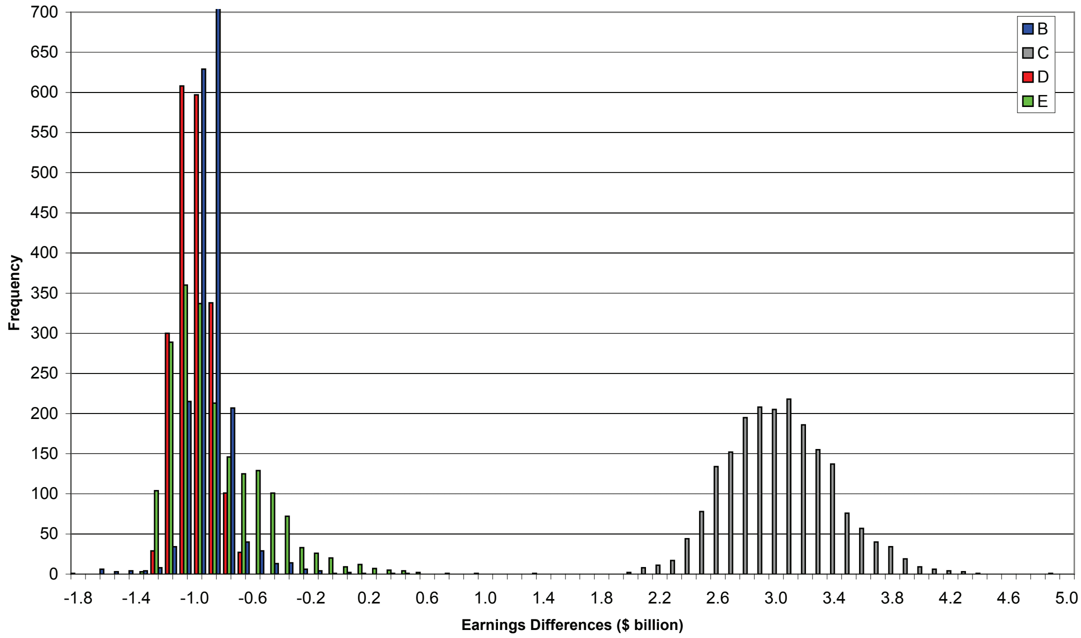


Figure 39: The earnings difference distributions using the various price paths generated by simulators B,C, D and E for quarter 1, Fin07/08

drought conditions experienced across the eastern seaboard of Australia. Figure 39 shows the earnings difference distributions for simulators B, C, D and E for quarter 1 of Fin07/08. All simulators, with the exception of simulator C, have earnings difference distributions that have negative expectations. On the other hand, the distribution corresponding to the price paths generated by simulator C lie far to the right in the positive domain of earnings differences relative to the other three. As such, it cannot be taken seriously as a price path generator for earnings distributions. The earnings difference distribution generated using the price paths from simulator B is characterised by a spread that is wide enough to suggest appropriate recognition of the volatility of electricity prices. However, the probability in the tails of this distribution is nowhere as pronounced as for simulator E. This indicates that its simulated price paths do not mismatch the volatility in the actual prices as frequently as does those generated from simulator E. As was the case for quarter 1, Cal08, the distribution associated with simulator D has too small a spread. The earnings difference distributions generated using the simulated price paths from simulators B, C, D and E for quarter 2 are shown in Figure 40. With the exception of simulator C, the remaining simulators produced earnings difference expectations that were negative. This was also the case for quarter 1. As far as spread is concerned, the weight of probability in the tails of the distributions corresponding to simulators B and E is reversed from what was the case for the first quarter, shown in Figure 39. The earnings differences show definite negative tail behaviour for simulator B, whereas it is far less pronounced for the differences generated from simulator E. This comparison is repeated when observing the positive tail behaviour for these two simulators. Recall that prices in quarter 2 of Fin07/08 were historically high and volatile. The extended tail behaviour of the earnings differences for simulator C during this quarter is evidence of a simulator performing

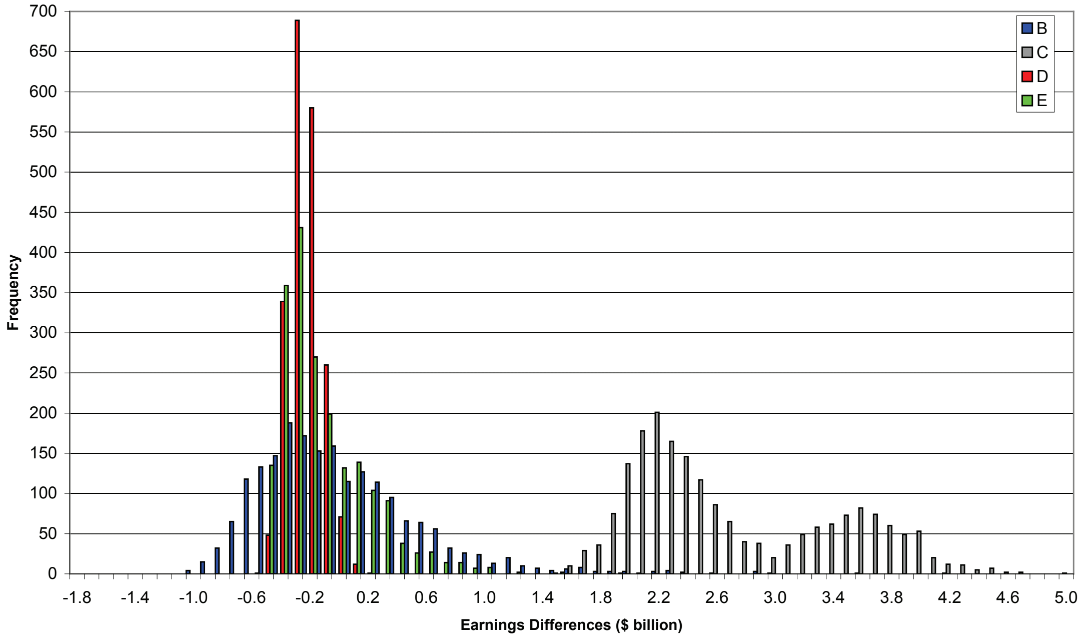


Figure 40: The earnings difference distributions using the various price paths generated by simulators B,C, D and E for quarter 2, Fin07/08

beyond feasible industry expectations, given that the simulated price paths were calibrated on historical data with relatively benign price behaviour. As was the case for quarter 1, the spread of the earnings differences calculated using price paths generated by simulator D appear to be too small to satisfy realistic expectations, while the earnings difference distribution corresponding to the prices generated from simulator C is totally unrealistic.

Figure 41 suggests that the only simulator to generate prices for quarter 3 that result in an anticipated negative expected earnings difference number is simulator D. However, as has been consistently the case, its spread is too small to satisfy what is realistically expected. The earnings distribution output from simulator C is, yet again, not to be taken seriously. The spreads of the earnings distributions based on the remaining two simulators are similar, with simulator B having less weight in both the positive and negative tails of the distribution. The earnings distributions for quarter 4 for the four simulators being studied are shown in Figure 42. The expected earnings difference number corresponding to simulator B prices is negative as expected. Further, the spread of the distribution of the earnings differences associated with this simulator fit neatly with industry expectations. Simulator E has a negative expected earnings difference number but has an earnings difference distribution whose spread appears to be too accentuated in the positive tail. The distribution calculated from simulator D has a negative mean earnings difference, but too tight a spread. The earnings difference distribution derived from prices generated by simulator C is again dismissed as approaching any sense of reality.

Figures 43 to 46 show the earnings difference distributions for all four quarters of the peak periods during Cal08. These distributions have been derived using the price paths generated by simulators A, C, D and E, which have been the focus of diagnosis for Cal08

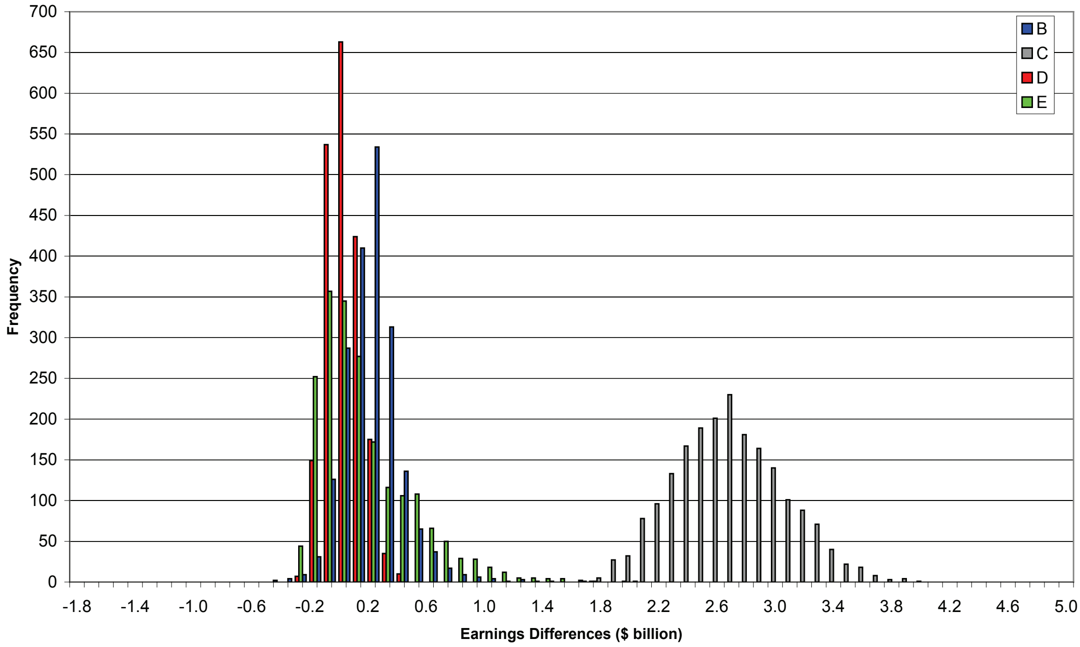


Figure 41: The earnings difference distributions using the various price paths generated by simulators B,C, D and E for quarter 3, Fin07/08

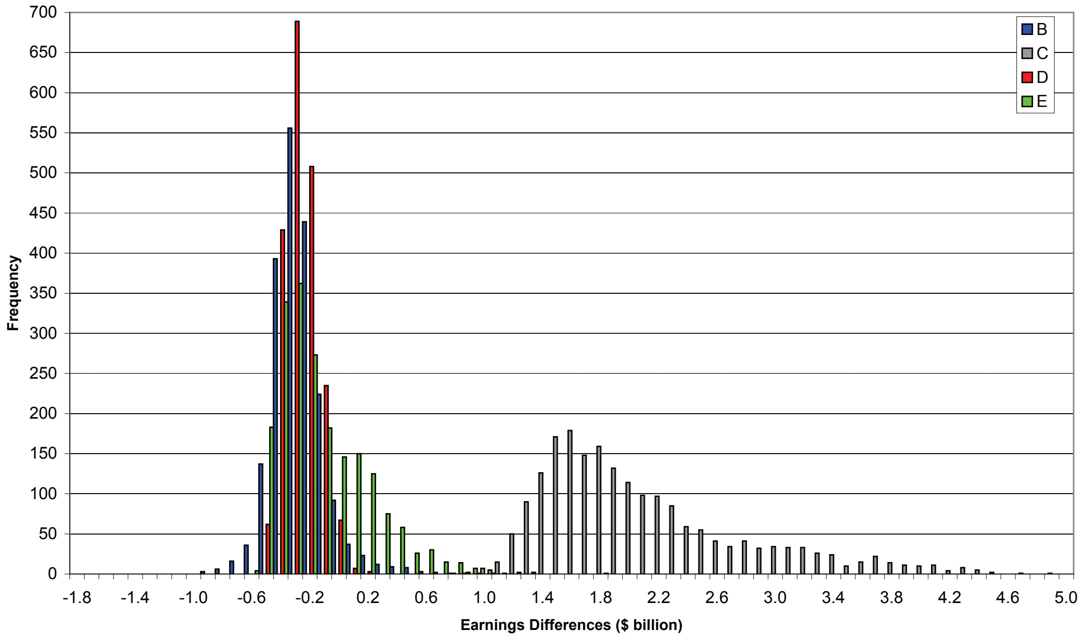


Figure 42: The earnings difference distributions using the various price paths generated by simulators B,C, D and E for quarter 4, Fin07/08

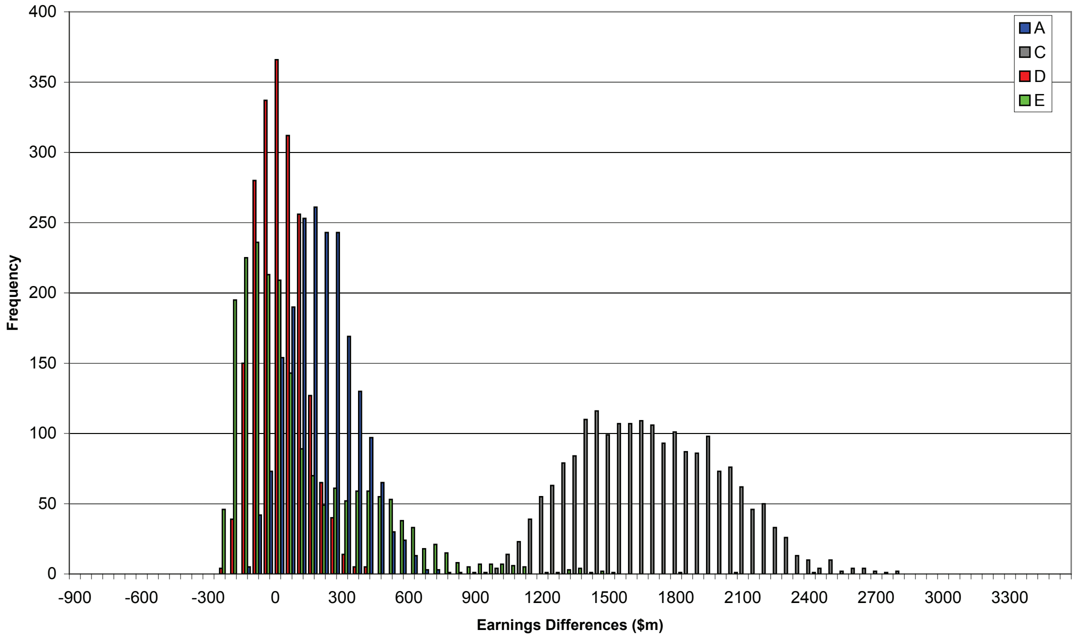


Figure 43: The earnings difference distributions for the quarter 1, Cal08 peak period using the price paths generated by simulators A,C, D and E

in this report. As previously noted, peak load covers the hours of highest demand. The peak is usually defined as the period between 2pm and 8pm on working weekdays. The leading provider of peak load for the Sydney metropolitan area is Snowy Hydro Limited through its hydroelectric and gas-fired units. Although hydro is inexpensive, due to limited water resources within the NEM, hydro generators are restricted to operating predominantly during demand peaks. When lack of water cancels out the use of hydroelectric generation, the more expensive gas-fired turbines are the usual alternative electricity generating option. It follows that an electricity retailer can be exposed to highly volatile prices in peak periods and, therefore, significant volumetric risk. The following discussion contrasts the aggregate earnings difference distributions taken across peak periods with those previously discussed for all periods during the day. The purpose of concentrating on peak periods is to determine whether the price paths from the simulators being studied do an appropriate and consistent job at generating earnings difference distributions for those periods when the business is likely to be exposed to increased risk. Given the totally unrealistic earnings difference distributions derived from price paths generated by simulator C, no discussion is entered into concerning the appropriateness of this simulator as a tool for peak earnings risk management.

The earnings difference distribution corresponding to simulator A in quarter 1 (Figure 43) has positive and inappropriate central tendency, but with acceptable positive and negative tail behaviour. However, positive differences for this simulator continue to be too pronounced in quarter 2 (Figure 44) and quarter 3 (Figure 45). In the case of simulator D, the spread of the peak earnings difference distribution is too narrow in the first three quarters. In the same periods, the earnings difference distribution for simulator E is characterised by negative differences that appear to be too accentuated, along with an over-extended positive tail. Figure 46 diagrammatically suggests that all three simulators have negative differences that

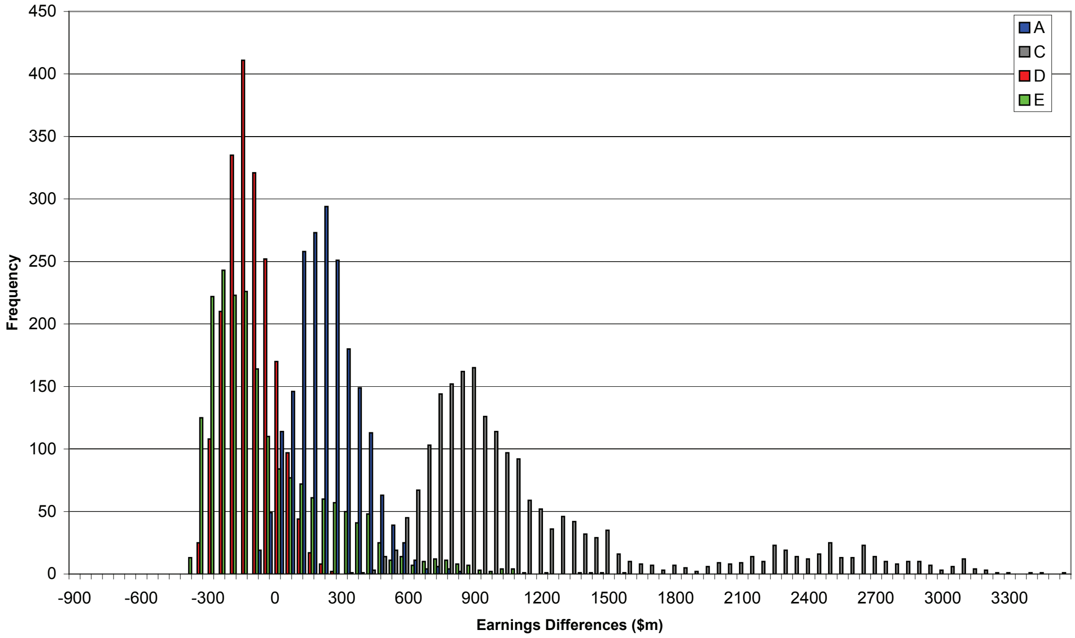


Figure 44: The earnings difference distributions for the quarter 2, Cal08 peak period using the price paths generated by simulators A,C, D and E

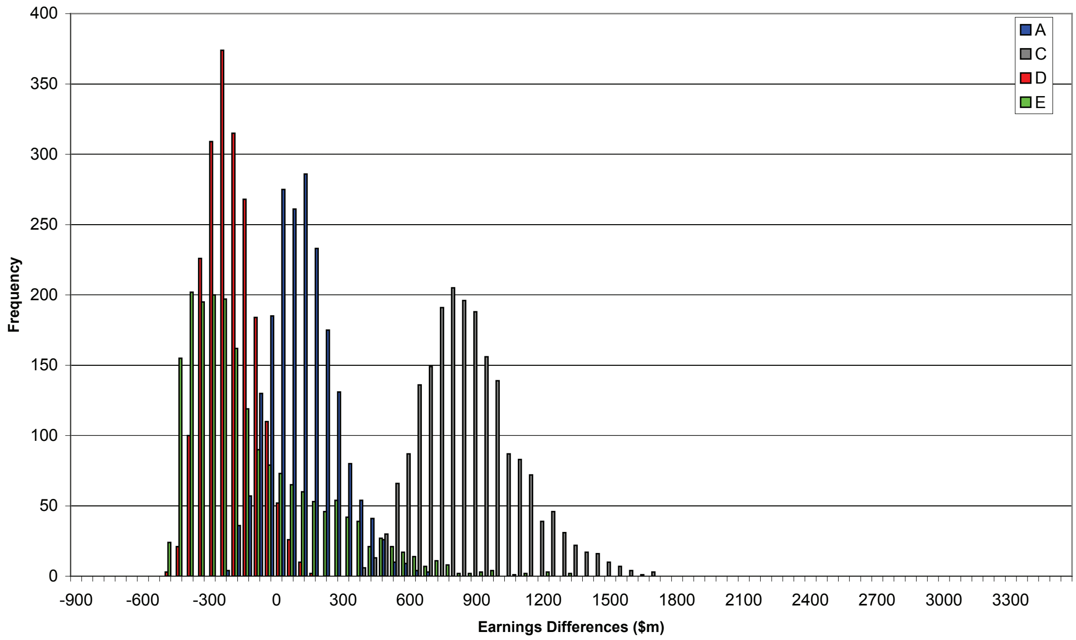


Figure 45: The earnings difference distributions for the quarter 3, Cal08 peak period using the price paths generated by simulators A,C, D and E

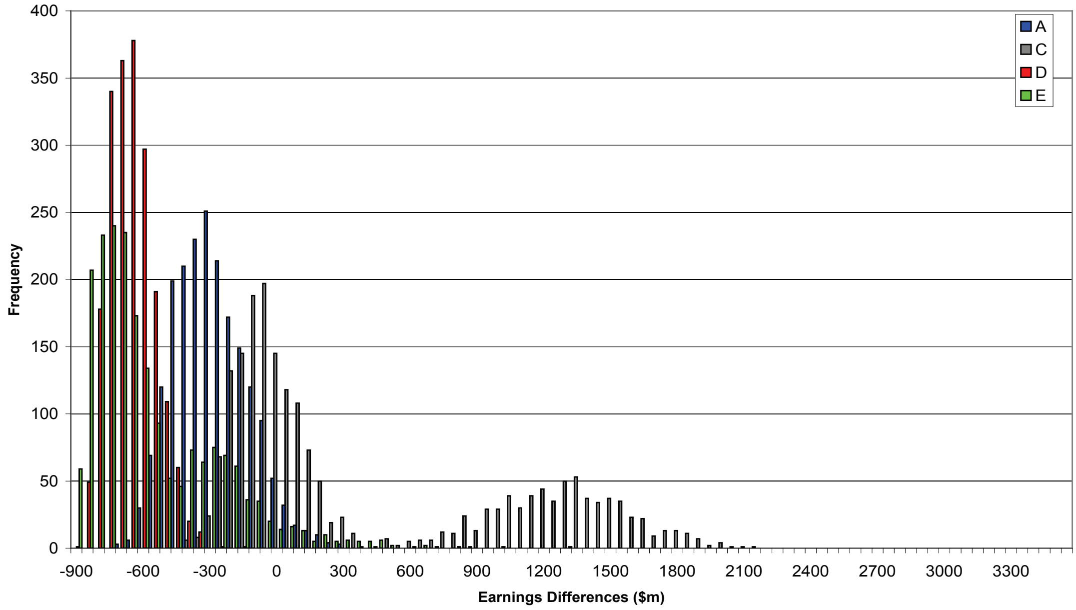


Figure 46: The earnings difference distributions for the quarter 4, Cal08 peak period using the price paths generated by simulators A,C, D and E

are too pronounced for quarter 4, indicating underestimation of the true earnings figure.

Figure 47-50 show the peak earnings difference distributions for the four quarters of Fin07/08. These distributions have been derived using peak price paths generated by simulators B, C, D and E. As was the case for the discussion involving peak periods during Cal08, simulator C is not discussed as this simulator is responsible for peak earnings distributions that have little relevance to industry standards. Figure 47 suggests that all three simulators exhibit pronounced negative tail behaviour in quarter 1 that is more dominant for simulators D and E when compared to that of simulator B. Negative differences imply underpricing by a simulator which can lead to an underestimation of the true earnings figure. For simulator B, this is consistent with the previous observation of the overestimation of prices using the first diagnostic test (Figure 15). For quarter 2 (Figure 48), all simulators are characterised by positive differences indicating overestimation of earnings. Simulator B has an extended positive tail which is consistent with high mean prices for this quarter (see Figure 15) and possible overestimation of the true earnings figure. While simulator D has only positive differences, the narrower spread for this simulator relative to the other two in this quarter can be seen as an attribute. The positive differences present in the earnings distribution for simulator E are greater than those for simulators B and D. However, the extended positive tail behaviour is not as obvious as it is for simulator B. Figure 49 indicates a further positive shift in the earnings difference distributions of all the simulators. The expected earnings differences are all positive. This suggests overestimation of the earnings differences, a low EaR and the potential for under-hedging of peak prices for all simulators for this quarter. From Figure 50 it can be concluded that all three simulators underestimate the true earnings figure in quarter 4. The earnings difference distributions corresponding to simulators B and D have a similar shape with the spread of D again too narrow. The distribution generated from price

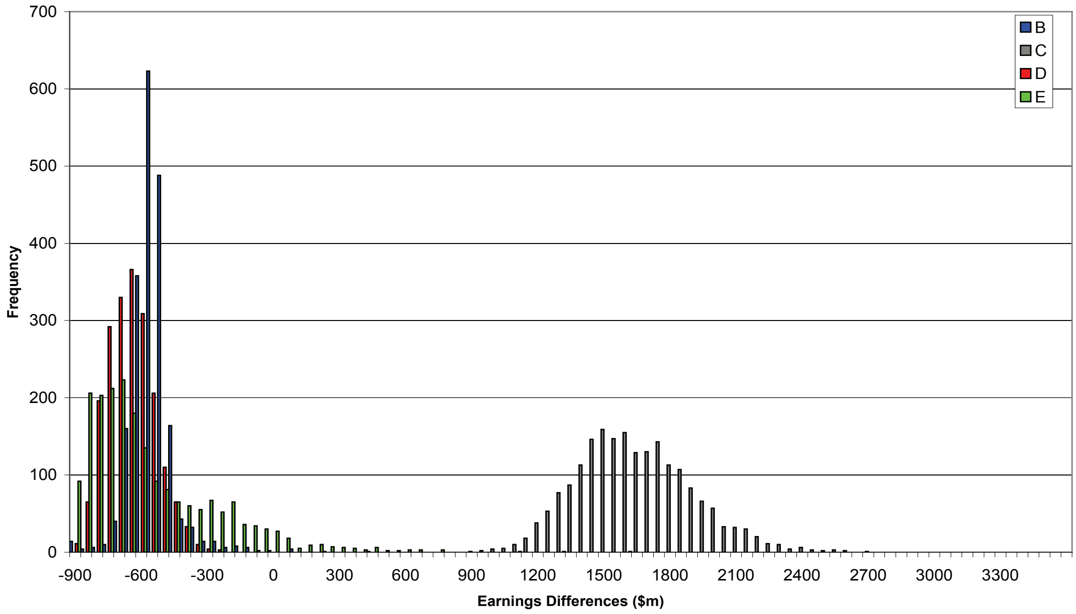


Figure 47: The earnings difference distributions for the quarter 1, Fin07/08 peak period using the price paths generated by simulators B,C, D and E

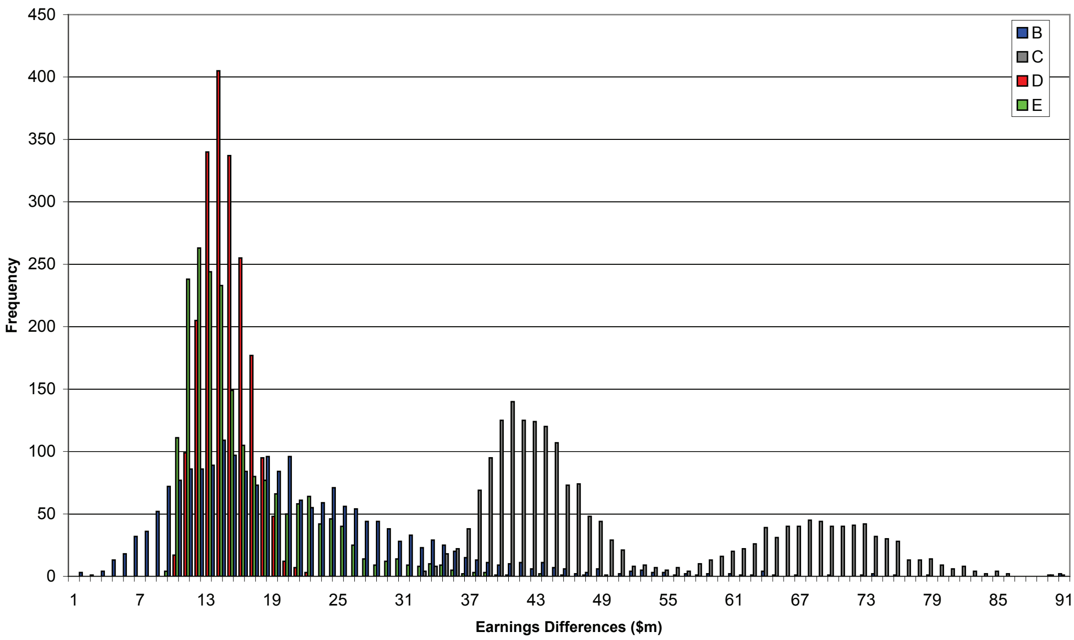


Figure 48: The earnings difference distributions for the quarter 2, Fin07/08 peak period using the price paths generated by simulators B,C, D and E

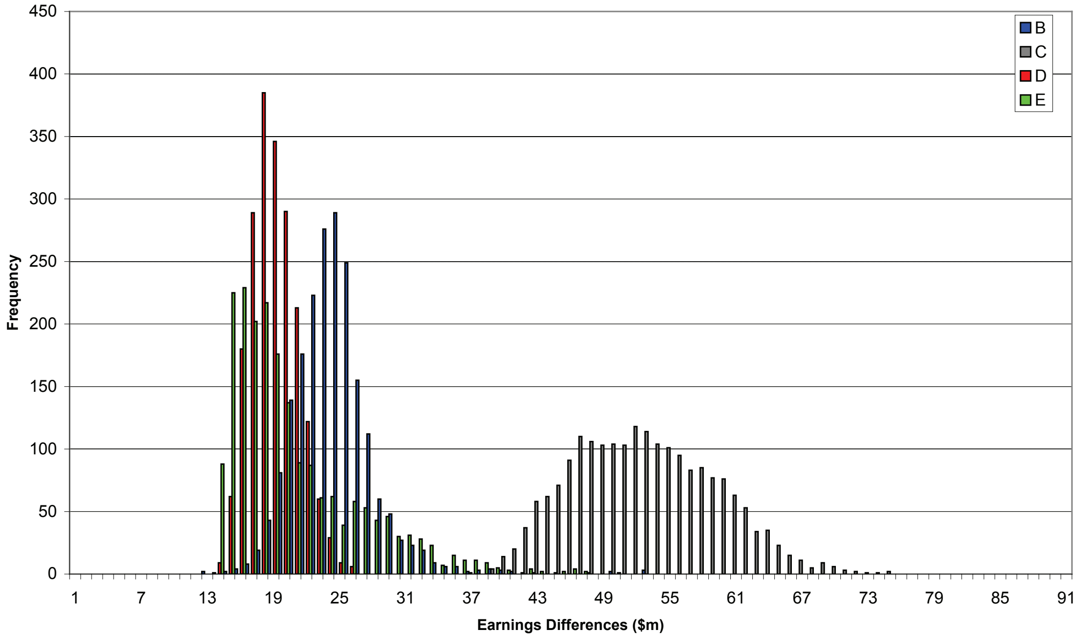


Figure 49: The earnings difference distributions for the quarter 3, Fin07/08 peak period using the price paths generated by simulators B,C, D and E

paths from simulator E has positive tail behaviour that is too accentuated. When comparing peak earnings difference distributions to the corresponding aggregate distributions diagrammatically, care needs to be taken to account for the scale differences between the earnings difference numbers. This scale differences make comparison of distributional characteristics between the aggregate and the peak earnings differences difficult and subject to potential misinterpretation. With this in mind, it is still fair to conclude that the volatility of the settlement prices during peak periods leads to more pronounced positive and negative tail behaviour than is the case for the distributions aggregated over the peak, off-peak and weekend periods. What is also noticeable for all peak periods in all quarters in Cal08, was the potential for simulator A to overprice relative to simulators D and E. This is consistent to overpricing by this simulator when the distributions were aggregated across all three periods in the day. This suggests that the overpricing by simulator A during the peak period may well explain its overpricing across the day. During the peak periods for Fin07/08, quite different distributional behaviour is noticeable from one quarter to the next. Quarter 1 and quarter 4 are characterised by underpricing, with a more acceptable earnings difference distribution the case for simulators B, D and E in quarter 4. On the other hand, overpricing seems to be the case in quarters 2 and 3 for all simulators. Recall that these two quarters were the quarters where the market experienced abnormally high prices across all three periods in the day, but especially in the off-peak and weekend periods. When the overpricing in the peak period was aggregated with the other two periods, the aggregate distributions for these two quarters appear closer to industry expectations.

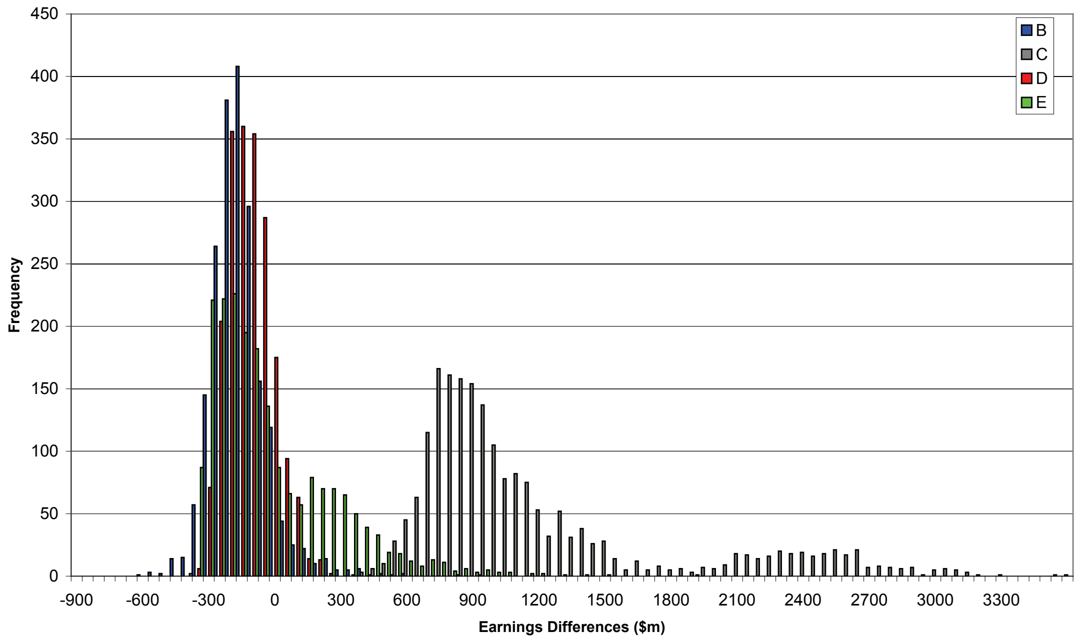


Figure 50: The earnings difference distributions for the quarter 4, Fin07/08 peak period using the price paths generated by simulators B,C, D and E

4 Conclusions

This study has endeavoured to propose and implement a series of diagnostic tests to determine the appropriateness of electricity simulation engines (ESEs) for generating electricity load and price paths to be used as input in the determination of a retailer's earnings distribution and the assessment of earnings-at-risk (EaR) measures. Additional diagnostic measures require development before a routine can be developed whereby a complete diagnostic report can be generated as output using simulated and historical data as input. This work includes:

(1) Further partitioning of output load and prices from an ESE into off-peak, peak and weekend periods to determine the subsequent effect on earnings.

(2) The diagnosis of simulated load paths. As simulated load was not supplied for all engines, the diagnostics developed in this report did not include an analysis of load.

(3) The building of a response surface to capture the interaction between temperature, load and price.

(4) Examination of the convergence behaviour of an ESE. Convergence in this context means the determination of the minimum number of load and price paths required from a simulator in order to return expected profiles that conform to industry expectations. This would involve the sequential testing of an increasing number of simulated paths from an ESE in order to determine the number required.

In conclusion, it is important to understand that each of the simulators that were diagnosed in this study were criticised according to industry expectations, and to the degree that the diagnostics employed here reflect those expectations. In fact, all simulators will attract criticism given that they are calibrated on historical data and are expected to generate future prices for market conditions that are unknown. The mark of an appropriate ESE is that the future load and pricing structure it generates is not too much at variance with industry expectations. A critical function of a simulator is for it not to overestimate or underestimate load and prices such that the risk metrics used to govern earnings risk faced by an electricity retailer are compromised to the extent that their book is either grossly over-hedged or under-hedged.

References

[1] Australian Government Energy White Paper (2004) Securing Australia's energy future, available at http://www.www.efa.com.au/Library/CthEnergyWhitePaper.pdf.

[2] Booth, R. (2004) Too much volatility is bad for you, available at http://www.bardak.com.au, accessed August 2007.

[3] Bossley, L. (2002) Trading power: science or art?, *CEPMLP Journal*, **17**, 8-11.

[4] Energy Information Administration Report (2002), Derivatives and risk management in the petroleum, natural gas, and electricity industries, U.S. Department of Energy, Washington D.C.

[5] Eydeland, A. & Wolyniec, K. (2003) *Energy and Power Risk Management-New Developments in Modelling, Pricing, and Hedging*, John-Wiley & Sons, New Jersey.

[6] Goto, M. & Karolyi, G. (2004) Working Paper No.12, Central Research Institute of the Electric Power Industry, Department of Finance, Ohio State University.

[7] Li, Y. & Flynn, P.C. (2004) Deregulated power prices: comparison of volatility, *Energy Policy*, **32** (14), 1591-1601.

[8] Politis, D.N. & Romano P.J. (1994) The Stationary Bootstrap, *Journal of the American Statistical Association*, **89** (428), 1303-1313.

[9] Stevenson, M., do Amaral, L.F.M. & Peat, M. (2006) Risk management and the role of spot price predictions in the Australian electricity market, *Studies in Nonlinear Dynamics and Econometrics*, **10**, No. 3, Article 4.

[10] Stoft, S. (2002) *Power System Economics*, John Wiley and Sons, New York.

[11] Svec, J. (2009) Risk Management in the Australian Electricity Industry, unpublished PhD thesis, Faculty of Economics and Business, The University of Sydney.

[12] Tipping, J., Read, G., Chattopadhyay, D. & McNickle, D. (2003) Can the shoe be made to fit? Cournot Modelling of Australian Electricity Prices, Energy Modelling Research Group, University of Canterbury, NZ.