Second Order Stochastic Dominance Portfolio Optimization for an Electric Energy Company

M.-P. Cheong, Student Member, IEEE, G. B. Sheble, Fellow, IEEE, D. Berleant, Senior Member, IEEE and C.-C. Teoh, Student Member, IEEE, J.-P. Argaud, M. Dancre, L. Andrieu, and F. Barjon

Abstract—This paper presents a framework of portfolio optimization for energy markets from an electric energy company’s perspective. The objective of this research is to determine the best possible investment plan by combining two potentially conflicting portfolio investment goals. First, given the general characteristics of the generating assets and forecast of market variables, the decision maker selects an efficient set of portfolios by optimizing the expected portfolio return. Secondly, an optimal portfolio is chosen based on company’s risk profile. This risk is controlled by guaranteeing that the portfolio model has second-order stochastic dominance (SSD) over the cumulative distribution of a minimum tolerable reference distribution. Decision criteria are then applied to obtain an optimal and robust portfolio.

The proposed approach is used to determine the amount of optimal market share value that maximizes the expected value of the profit. This is performed by treating risk as a distribution that represents the minimum expected profit acceptable by the energy company. Results show that different risk profile leads to different optimal portfolio. The optimal portfolio which gives the highest expected profit may not have the best robustness. This approach is also applicable to problems characterized by other sources of epistemic uncertainty besides unknown dependencies.

Index Terms—Portfolio optimization, second-order stochastic dominance, interval analysis, and epistemic uncertainty.

I. INTRODUCTION

The pioneering work in portfolio optimization began with the proposition of mean variance model by Harry Markowitz [17] in 1952. Since then, many related research work has been focused on finance and economics. In the recent years, the adoption of such concept has increased in the restructuring electric power sector. As more non-probabilistic factors affect the electricity market, not many research focused on solving the portfolio selection problem when dependency relationships between portfolio segments are unknown, in particular, when the market share is represented by a bounded variable with no known underlying distribution.

Because knowledge that supports investment decisions is often limited by what information is available, inference under conditions of epistemic uncertainty is important to do in order to support optimized portfolio. Epistemic uncertainty in the context of this paper refers to a lack of knowledge about the randomness of the decision variable. In this paper, such uncertainty is quantified as bounded families of probability distributions. Applying bounded families of distributions to electric power bidding problems exemplifies what may be expected in engineering from applying uncertainty quantification. There is a growth of interest in uncertainty quantification by journal special issues [1-2]. Applications to problems in electric power [3-4] follow naturally from the insights and investigations of other researchers, who have found that uncertainty quantification has applicability to power problems characterized by severe uncertainty. Prominent techniques include intervals [5-6] and fuzzy methods [7-8].

The well recognized need for decisions in the presence of severe uncertainty, coupled with the grounding of our approach in the mathematically well-founded theory of probability, support its use in addressing important problems in electric power. Guidance can then be obtained regarding investment decisions under conditions of uncertainty in which standard methods would require extra, unjustified assumptions.

The volatility of electricity prices caused the uncertain return (profit or loss) for an electricity portfolio. Often times, standard distributions may not accurately fit the distribution of returns. This calls for a methodology that does not depend on the type of distribution shape of the return but rather on the entire cumulative distribution function (CDF) of the return. This paper addresses this problem by optimizing a portfolio while satisfying the risk constraint which is represented by the second order stochastic dominance constraint.

In the proposed approach, generating units are modeled as a
series of options. The return is represented by spark (gas) or dark (coal) spread depending on the fuel used. There are generating units for this analysis. They include one nuclear unit and 3 thermal units (coal-fired and gas-fired). A two-step optimization approach is proposed to find the optimal financial portfolio for production. This problem is modeled from a viewpoint of a utility that services 4 different industries. The first model assumes that this utility company is obligated to serve the 4 sectors. The second model relaxes this constraint to allow the utility to sell to the forward market should it see profitable.

The following section describes the portfolio optimization model. Section III defines the robustness and stochastic dominance concepts. Background sections II and III are borrowed and modified from a recently submitted journal paper. The major differences in this paper as compared to the submitted journal paper include the assumption where the sum of all weights in the portfolio does not have to add up to 1, the return for each segment was generated using spark and dark spreads, and Anderson-Darling goodness-of-fit measure are used to fit distributions for each return segment. Section IV describes the interval algorithm. Section V and VI discuss conclusions and future work.

II. PORTFOLIO OPTIMIZATION MODEL

Optimal portfolios are often identified by finding the weights of the portfolio segments such that a mean-risk objective function is maximized [14]. Formally, the problem to be considered is to find such a portfolio given the constraints:

\[ R = \sum_{j} x(s_j) r(s_j) \geq \tilde{R} \]  

(1)

\[ x(s_j)_{lb} \leq x(s_j) \leq x(s_j)_{ub} \]  

(2)

where $R$: return distribution of the optimal portfolio  
$s_j$: portfolio segment $j$  
$r(s_j)$: weight of segment $s$ in the portfolio  
$r(s_j)$: return distribution of optimal portfolio segment $s$  
$\tilde{R}$: a given reference curve that represents the minimum tolerable return distribution (“risk limit”)

The symbol “$\geq$” designates stochastic dominance of $i$th order. As an additional constraint set, weights $x(s_{j1}), x(s_{j2}), x(s_{j3}), \ldots, x(s_{jn})$ has to be within its upper and lower bounds, $x(s)_{lb}$ and $x(s)_{ub}$. Each weight may be required to be within some interval in order to enforce a balance across segments, as might be specified by a company’s business model constraints and investment policies. This model also assumes that the firm has no restriction in borrowing money when it sees profitable to optimally invest in a particular sector.

The first step is to generate a set of optimal portfolios to search within for the best. A standard approach based on mean and risk and parameterized by risk position [17] is used. Let the desirability of a portfolio return random variable $r$ be determined by the following function of a parameter $z$ describing the importance of risk:

\[ f(z, r) = \text{mean}(r) - z \cdot \text{risk}(r) \]  

(3)

The following equation builds on the concept of Eq. 3, and states that given a risk position $z$, optimal return distribution is obtained by:

\[ OPT(z) = \sup_{y \in Y} \mu_y - z \sigma^2_y \]  

(4)

where  
$Y$: set of portfolios $y$ complying with constraints (1) and (2)  
$\mu_y$: expected return of a portfolio $y$  
$\sigma^2_y$: variance of the return of a portfolio $y$  
$z$: degree of risk aversion

It is possible to account for properties of portfolio variation besides variance [11], but $z\sigma^2_y$ is nevertheless widely used to model risk position.

III. STOCHASTIC DOMINANCE AND ROBUSTNESS

Stochastic dominance constraint can be thought as a ranking tool with no assumptions made on the shape of the return distribution. It can be used as a profit/risk indicator for any given portfolio. Thus, a given portfolio’s return distribution can be tested for compliance with stochastic dominance constraints. By definition, a return of a portfolio $X$ stochastically dominates another portfolio $Y$, to the first order, if the following is true:

\[ X \succ Y \Rightarrow P(X \leq A) \leq P(Y \leq A) \]  

(4)

where  
$P(X \leq A)$: Cumulative distribution function of $X$

Similar to the first order stochastic dominance, the condition for second order stochastic dominance constraint is given by:

\[ X \succ_2 Y \Rightarrow \int_{-\infty}^{A} P(X \leq B) dB \leq \int_{-\infty}^{B} P(Y \leq B) dB \]  

(5)

SSD constraint is computed using numerical integration or by summing areas of trapezoids under the curve. The size and number of trapezoids to sum is determined by the step size chosen for the integration process. An optimal portfolio might or might not comply with an additional requirement that it has stochastic dominance over a given reference return, $\tilde{R}$. The SSD constraint “$\succ_2$” ensures the dominant portfolio is preferred by any risk adverse player [13], and has the additional virtue of being less constraining than the FSD constraint. Thus, SSD constraint is used to be consistent with the degree of risk aversion specified in Eq. 3.

Robustness is defined as the amount by which a portfolio dominates a reference curve (robustness would be negative if it
does not dominate. By testing various optimal portfolios for robustness, one with the highest robustness can be identified. Alternatively, one with the highest expected return that also meets the SSD constraint could be found. In either case, the strategy is to search among a set of optimal portfolios provided by an under-constrained optimization problem for the one that is best according to a second criterion.

SSD is computed by the minimum horizontal distance between the integrals of two cumulative distributions. In other words, |SSD| measures how much one curve for the integral of a distribution can be moved toward another one along the x-axis before the two curves touch. |SSD| formalizes the amount of separation between the integrals of two distributions.

The amount of epistemic uncertainty is defined by \( \alpha \). The use of \( \alpha \) follows the convention in Info-Gap Theory [9]. In this paper, maximum of \( \alpha \) value is obtained by calling Eq. 6. This value determines the robustness of the optimal portfolio.

\[
\max_{\alpha} F_{\text{new}} = \alpha F_{\text{left}} + (1 - \alpha) F_{\text{best-guess}} \leq F_{\text{reference}} \tag{6}
\]

where
- \( F_{\text{new}} \): New curve obtained with max \( \alpha \). This is the \( \alpha \) curve shown in Fig. 6.
- \( F_{\text{left}} \): Left envelope based on the addition of 4 portfolio segments, given that their dependencies are unknown. This refers to the left curve in Fig. 6.
- \( F_{\text{best-guess}} \): The curve based on the addition of 4 independent portfolio segments. This is the best-guess curve shown in Fig. 6.
- \( F_{\text{reference}} \): This is the minimum tolerable risk level. This is the reference curve in Fig. 6.

IV. INTERVAL ALGORITHM

The unknown dependency among the portfolio segments is computed using Statool [19] that uses the Distribution Envelope Determination (DEnv) algorithm. This algorithm was developed by Berleant et al [20]. This section briefly introduces the algorithm. First, a discrete joint distribution tableau is constructed. Each input is discretized by representing its probability density function (PDF) with intervals (see figure below).

Figure 1. Discretization process.

These discretized inputs form the marginal of the joint distribution tableau, which sets the constraints for each interior cell. Each interior cell’s is represented by a range for its probability mass distribution. Next, the algorithm finds the bounds on the cumulative probability of the derived distribution. The left envelope is obtained by the derived distribution that rises the fastest and the right envelope is obtained by the derived distribution that rises the slowest. This is performed iteratively using linear programming to find the maximum and minimum values to form the left and right envelopes.

Fig. 2 illustrates the computation graphically. The operation is the addition of two random variables, \( X + Y = Z \). \( X \) and \( Y \) are shown in PDF form and \( Z \) is represented in both PDF and CDF forms.

\[
X_{\text{PDF}} + Y_{\text{PDF}} = Z_{\text{PDF}} = Z_{\text{CDF}}
\]

Figure 2. Illustration of the addition of two random variables, \( X \) and \( Y \), given their dependency is unknown.

V. ANALYSIS WITH INDUSTRIAL CUSTOMERS

The portfolio optimization model accepts two major classes of inputs: (i) the decision variables, which are the market segments and their profit distributions, and (ii) a reference distribution. The reference distribution refers to the cumulative distribution of a minimum tolerable return. Although a classic approach is to optimize the expected portfolio return subject to a risk aversion factor, this model has been modified to optimize based on a minimum tolerable “reference” distribution, the concepts of second-order stochastic dominance and Information-Gap Decision Theory, and market share constraints for different customers.

The decision variables here refer to portfolio segments, which are the market shares for customers within the following markets: the industry market (\( s_1 \)), middle market (\( s_2 \)), mass market (\( s_3 \)), and distribution market (\( s_4 \)). Each market share has an allowable range for level of investment determined by upper and lower limits. Suppose that the decision variable for each market is represented by \( x_i \). Then the market share is constrained by the following limits:

\[
x_1 \in [0.7, 1.0]; \quad x_2 \in [0.7, 1.0]; \quad x_3 \in [0.6, 0.8]; \quad x_4 \in [0.8, 1.0]
\]

Each number \( x_i \) represents a percentage of the entire European demand in market \( s_i \). For example, in the industry market (\( s_1 \)), the total European demand is 10767 MW/day, so the market share limits for this segment are \([0.7*10767, 1*10767] = [7536.90, 10767]\) MW/day. Therefore, there is no constraint indicating that the sum of all market shares has to add up to 1. This is true assuming that the European firm is able to sell short and would be able to borrow money when needed. Notice that this is different than most portfolio problem because it is commonly assumed that all the weights are fractions of the total budget that has to add up to 1.

Different customers pay different prices to the company.
The price charged to each market is given by:

\[ \pi_{s1} = 37\text{£} \]
\[ \pi_{s2} = \pi_{s} + 2\text{£} \]
\[ \pi_{s3} = 55\text{£} \]
\[ \pi_{s4} = \pi_{s} - 1\text{£} \]

where

\[ \pi_{s} \]: Daily spot price of electricity

The prices above are used to calculate the revenue for each market. For example, the revenue for industry market segment \( s_1 \) is in range \([37 \times 7,536.90, 37 \times 10,767] = [278,865.3, 398,379] \).

The cost of production is based on the average cost for all units if they are committed to sell. The units are committed based on the least cost dispatch order. There are 4 generating units. If the 3 units with the cheapest production costs can meet the demand for that day, then only 3 units are committed and the cost of production is the average cost of the 3 units times the demand for that day. Average cost is used because there is no one-to-one mapping from a specific unit’s output to a specific market. In other words, if the gas and nuclear units are committed to sell to the industry market (\( s_1 \)) and middle market (\( s_2 \)) segments, there is no information on whether \( s_1 \) is getting its electricity from the gas generating unit or the nuclear generating unit.

Profit distributions based on average cost of production for each segment is generated to obtain the expected return distribution. Since the demand data for the entire year exhibits seasonality, the profit distribution exhibits seasonality as well (see Fig. 3).

![Return Distribution](image)

**Figure 3.** Annual return distribution for industry market (\( s_1 \)), middle market (\( s_2 \)), mass market (\( s_3 \)), and distribution market (\( s_4 \)). Total return, which is the summation of all returns are also shown.

To use the yearly data without accounting for seasonality would produce inaccurate results. Since this paper’s focus is not on adjusting the data for seasonality, a snapshot of one month’s data is used for the analysis, specifically the month of July’s data. July data was chosen because it is typical of a peak period in the summer. This data is used as an input to the portfolio optimization model together with an arbitrarily selected reference distribution.

Table 1 shows the distributions for each demand sector based on the Anderson-Darling goodness-of-fit measure [21].

The Anderson-Darling statistic (\( A^2 \)) is a test that compares the fit of an observed CDF to an expected CDF. Formally, for \( n \) number of observations, it is defined as:

\[ A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \cdot (\ln F(X_i) + \ln(1 - F(X_{n-i+1}))) \]  

where

\[ F(X_i): \text{CDF of the ordered data, } X_i \]

This is a one-sided test where the critical values depend on the specific distribution that is being tested. The hypothesis for the distribution to follow a specific shape is rejected the test statistic is greater than the critical value.

From Table 1, the percentage spreads with respect to the corresponding mean value is on the order of 10%. Although they are in thousands, the values are relatively small if compared to the mean values.

<table>
<thead>
<tr>
<th>Demand Sector</th>
<th>Distribution</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>Beta</td>
<td>311,391.84</td>
<td>25,639.18</td>
<td>8.23%</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>Beta</td>
<td>368,804.96</td>
<td>36,930.41</td>
<td>10.01%</td>
</tr>
<tr>
<td>( s_3 )</td>
<td>Student T</td>
<td>411,659.98</td>
<td>13,752.44</td>
<td>3.34%</td>
</tr>
<tr>
<td>( s_4 )</td>
<td>Beta</td>
<td>191,379.79</td>
<td>17,155.56</td>
<td>8.96%</td>
</tr>
</tbody>
</table>

Fig. 4 shows the optimal value for market share for each demand segment for different degrees of risk aversion (\( z = 0 \) to \( z = 5 \)). The market shares converge to their lower bounds when \( z \) approaches 5. (Note: Although \( z \leq 0 \) is not generally used for the degree of risk aversion, the market shares for all negative \( z \)-values would be at their upper bounds because the variances for the demand sectors are high enough that the objective function would benefit from their consequent riskiness.)

![Market Share vs. Z Values](image)

**Figure 4.** Market share for different risk aversion values (\( z \)-values) for industry market (\( s_1 \)), middle market (\( s_2 \)), mass market (\( s_3 \)), and distribution market (\( s_4 \)).
However, the optimal weights change for very small positive values of $z$. Fig. 5 illustrates how the weights for each demand sector move from their upper limits to their lower limits from $z = 0$ to $z = 0.005$.

The weights change over a small range of $z$-values. Since the reference distribution was arbitrarily chosen, a few values for $z$ with different weights are selected to illustrate how one can use the $|\text{SSD}|$ and alpha metrics to select the $z$-value that gives the best results. The following tables give the optimal values for $x_1$, $x_2$, $x_3$, $x_4$, $|\text{SSD}|$, and max $\alpha$, and the expected returns for $z = 0$, 0.0003, 0.0006, 0.0008, 0.003, and 1 through 5.

### Table 2

<table>
<thead>
<tr>
<th>$z$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>0.0003</td>
<td>1</td>
<td>0.9014</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>0.0006</td>
<td>0.7895</td>
<td>0.7</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>0.0008</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7497</td>
<td>0.8128</td>
</tr>
<tr>
<td>0.003</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7497</td>
<td>0.8</td>
</tr>
<tr>
<td>1.00-5.00</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
<td>0.8</td>
</tr>
</tbody>
</table>

### Table 3

| $z$   | $|\text{SSD}|$ | Max $\alpha$ | $\mu$       |
|-------|----------------|--------------|-------------|
| 0     | 86,409.31      | 1.14         | 1,200,900.00|
| 0.0003| 56,112.53      | 0.42         | 1,164,500.00|
| 0.0006| -63,392.25     | 0.00         | 1,024,700.00|
| 0.0008| -118,980.54    | 0.00         | 961,020.00  |
| 0.003 | -141,069.29    | 0.00         | 937,870.00  |
| 1.00-5.00| -202,112.57 | 0.00         | 876,240.00  |

Under the specified assumptions and the reference distribution used, the best portfolio occurs when $z$ is 0. The weight for the middle market ($x_2$) decreases first followed by the industry market ($x_1$), distribution market ($x_4$) and mass market ($x_3$). As we will discuss later, these results are due to the fact that there is no scarcity.

Fig. 6 shows the left envelope of the unknown dependency curve, the best-guess curve, the reference curve and the max alpha curves ($\alpha=1.14$) for $z=0$.

The left envelope curve is the curve generated by the addition of all profit segments without making any assumption about their dependency relationships. Since the left envelope provides a worst case bound for portfolio performance under uncertainty due to unspecified dependencies among segments, a high value of 1 is assigned to $\alpha$ as the measure of uncertainty expressed by the left envelope, ignoring the right envelope henceforth. The best-guess curve is the curve using the optimal weights for $z=0$, where the weights are all at their upper limits. The reference curve is an arbitrary curve. The max alpha curve defines the robustness measure of the portfolio.

### VI. Future Work

Other constraints such as resource minimum capacity, ramp up rate and maintenance (or possible outages) such that the firm cannot sell the committed amount of generation, are not included in this model. If these are included, the results could be far more beneficial to a firm. In addition, startup costs and the forced outage rate for each unit should be included in order to be more complete.

Another extension to this work would be to look into identifying the coherent risk measures to use in the SSD constraint. Seasonality of the data has to be accounted for and more work needs to be done to adjust for this seasonal trend. Multi-period should be a natural extension to this paper as portfolio problem for a company should be solved in a multi-period manner.

### Acknowledgment

The authors would like to thank the sponsors, in particular, J.-P. Argaud, M. Dancre, L. Andrieu, and F. Barjon for their support and constructive comments in the completion of this research project.

### References

[3] Sheblé, G. and D. Berleant, Bounding the composite value at risk for energy service company operation with DEnv, an interval-based


M.-P. Cheong received her B.S. and M.S. in Computer Engineering from Iowa State University in year 2002 and 2004. She also received her M.S. in Economics in 2006 from Iowa State University. She is currently pursuing Ph.D. in Electrical Engineering at Portland State University. Her research interests include portfolio optimization and data mining techniques applied to the dynamic electricity market.

Gerald B. Sheblé (F ‘98) is the Maseeh Professor at Portland State University. He attended Purdue University, receiving BSEE (1971) and MSEE (1974). He received his Ph.D. in 1985 from Virginia Tech. Dr. Sheblé received his MBA from the University of Iowa, specializing in Economics and Finance in 2001. His research interests include artificial life techniques for optimization; process modeling of power system operations and markets; process modeling of dynamic decision and optimization support system; real option analysis of asset valuation; probabilistic analysis of contract compliance and ancillary service requirements; data mining of markets and of health care systems using artificial life techniques, supply chain management in the energy industries, bidding strategies based on economic microanalysis, time series forecasting, and financial engineering.

D. Berleant is an Associate Professor in the Dept. of Information Science, University of Arkansas at Little Rock. He received his Ph.D. from the University of Texas at Austin in 1991. His research interests focus on imprecise probabilities, technology foresight, and bioinformatics.

Chin-Chuen Teoh (S’07) is currently pursuing a Ph.D. degree in the Department of Electrical and Computer Engineering at Portland State University, Portland Oregon. He attended Iowa State University, receiving his M.S. degree in Electrical Engineering in 2004 and M.S. degree in Economics in 2006. His research interests include application of financial tool for energy system expansion, and real option analysis of asset valuation.

J.-P. Argaud is Researcher in the R&D Division of Electricité de France (EDF). He received his "Maitrise" in Mathematics in 1989, his Engineer degree from Ecole Centrale de Paris (french "Grande Ecole") in 1991 and his PhD from the University of Paris Dauphine in 1995. His research interest focuses on discrete and continuous optimization, including data assimilation, risk management and application of finance for electricity.

M. Dancr is a Risk Analyst in the Dept. of Energy Market Risk Control at EDF. He received his Ph.D. from the University of Computer Science at Aix-Marseille (France) in 1999. His research interest focuses on risk assessment and policy, energy portfolio management, energy market price modeling.

L. Andrieu is a research engineer with the Dept. of Optimization, Simulation, Risk and Statistics, R&D Division, Electricité de France. She received her Ph.D. from the Ecole Nationale des Ponts et Chaussées (France) in 2004. Her research interests focus on financial engineering applications (portfolio optimization), optimization risk management (risks in energy markets) and stochastic optimization (mostly on probabilistic constrained optimization and stochastic gradient algorithm).