IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 10, NO. 6, 2016, PP. 1029–1039

Data-driven Stochastic Pricing and Application to Electricity Market

Saahil Shenoy[1] and Dimitry Gorinevsky[2]

***Abstract*—This paper develops a novel approach to computation of the probability integrals encountered in derivative pricing using stochastic models estimated from historical data. First, nonparametric probability distribution models are built directly from the data as a solution of a convex optimization problem scalable to very big data sets. Second, these models are used for numerical calculus of probability integrals, where the quadrature includes long tails of the probability distributions. The application example is the procurement contract in the day-ahead bulk market for electricity. The data for PJM utility loads and prices in the day-ahead and spot markets were used to estimate the risk and to price the contract. The data-driven forward contract pricing allows to optimize the contract cost and reduce it by 2% compared to the baseline; this corresponds to about $0.6B/year in potential utility savings.**

I. INTRODUCTION

This paper presents a data-driven probability integral method for pricing fixed-term derivative contracts. The motivating problem is described next but the method is potentially applicable to many other derivative pricing problems. We consider a Load Serving Entity (LSE), such as an electrical distribution utility, that holds a service obligation to supply electricity at the predetermined hour without restriction on volume (a load following contract). In the day-ahead forward contract, the LSE procures the forecasted amount of the electricity from the wholesale day-ahead market and faces the opportunity cost of procuring an unknown remaining amount at the contract delivery time in a volatile spot market.

The stated problem has special features that call for use of modern signal processing methods, see [1]. First, the electrical load and the spot prices are non-Gaussian random variables. Their distributions are conditional on the known regressor variables and often have long tails. Second, this is a fixed term problem, which allows to use data-driven models trained on historical data for the past fixed term periods. Third, the stochastic pricing problem involves two interdependent random variables: the electricity load and the spot price. The electricity load influence means the contract cannot be priced by arbitrage relative to the spot price.

The established derivative pricing approaches, such as the celebrated Black-Scholes model, assume normal distributions of the variables, see [2]. Most derivative pricing approaches for the electricity market use normal distribution models as well, e.g., see [3], [4], and [5].

Closer to the subject of this paper is work on fixed term derivatives, starting from [6], [7], where nonparametric models of distributions can be estimated directly from the historical data. In [8], [9], nonparametric models are used for pricing electricity forward contracts.

The cited work does not include nonparametric modeling of conditional distributions with many regressor variables. The electricity markets are more predictable than stock markets. This allows using more complex models with regressors including seasonal variables (monthly, weekly, intra-day hour), weather variables, and others. An approach known as quantile regression (QR) is well suited for modeling of such nonparametric distributions, see [10], [11], [12], [13]. The QR distribution models have been used for derivative pricing, e.g., see [14] and [15]. The QR models have been also used for electricity data forecasting, see [16], [17], and [18]. The QR models are applied to risk estimation in [13], [19], [20], [21].

Most of the related prior QR work uses kernel regressors, e.g., see [22], [23], [24], [25], [26], and [27]. The finite-support or Gaussian kernels cannot model the long tails of the probability distribution.

Long tail models have been long used in the actuarial science approaches to insurance, see [28], [29], [30]. The use of the long tail model in finance has increased in the recent years, e.g., see [31], [32], [33], [34]. With the proliferation of variable wind and solar generation, the probability distributions in the electricity market are increasingly non-Gaussian. However, the only prior work using QR models with long tails in electricity market

[1]Saahil Shenoy is a PhD student in the Department of Physics, Stanford University, Stanford, CA 94305, USA, saahils@stanford.edu

[2]Dimitry Gorinevsky is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA, gorin@stanford.edu and Mitek Analytics LLC, Palo Alto, CA 94306, dimitry@mitekan.com

applications known to the authors is [16].

The contributions of this paper are as follows. First, it develops QR non-parametric models of the probability distributions based on historical data and including long tails. Second, these models are extended to two interdependent variables: the electrical load and the spot price. Third, the paper shows how to use the developed models for numerical computation of stochastic integrals in the derivative pricing. Finally, it demonstrates applications of the methodology to the motivating example of the load following contracts in the electricity market.

Sections III-A through III-E of this paper are based on the earlier conference papers [35], [36].

II. Problem Formulation

We consider the dataset

$$D = \{y_{P,i}, y_{\pi,i}, Z_i\}_{i=1}^{N}, \tag{1}$$

where scalars $y_{P,i}$ and $y_{\pi,i}$ are response variables, vectors $Z_i \in \Re^n$ are explanatory variables (regressors), i is the sample number, and N is the number of the samples.

In the motivating problem of the forward contract pricing, variable $y_{P,i}$ describes the amount of the electricity to be procured in the contract. Variable $y_{\pi,i}$ describes the future spot price of the electricity at the delivery time. In what follows, we assume that data (1) are i.i.d., and follow unknown conditional multivariate distribution $y_P, y_\pi | Z$. In forecasting applications, i is the time sample and the i.i.d. assumption means the underlying process is stationary.

The stochastic pricing requires to estimate the multivariate probability density $p(y_P, y_\pi | Z)$. This distribution is used to compute a stochastic integral of the form

$$\mathbf{E}[f(y_P, y_\pi)] = \iint_{-\infty}^{\infty} f(y_P, y_\pi) p(y_P, y_\pi | Z) dy_P dy_\pi. \tag{2}$$

We consider the class of probability integrals (2) for the functions of the form

$$f(y_P, y_\pi) = \phi(y_\pi) \cdot (A\chi(y_P) - B)_+, \tag{3}$$

where the notation $(\cdot)_+ = \max(\cdot, 0)$ is used; $\chi(\cdot)$ and $\phi(\cdot)$ are two monotonically increasing functions; $\chi(\cdot)$ is unbounded. Such integrals are often encountered in derivative pricing, e.g., for spread options, see [37].

Let y_* be the solution to $\chi(y_*) = B/A$; for the unbounded monotonically increasing function ψ, the solution always exists. The full stochastic integral (2), (3) is now approximated as

$$\mathbf{E}[f(y_P, y_\pi)] = A \cdot \Psi(y_* | \phi, \chi) - B \cdot \Psi(y_* | \phi, 1), \tag{4}$$

$$\Psi(y_*; \phi(\cdot), \chi(\cdot)) = \int_{-\infty}^{\infty} \int_{y_*}^{\infty} \phi(y_\pi)\chi(y_P) p(y_P, y_\pi | Z) dy_P dy_\pi, \tag{5}$$

where the dependence on the definite integral Ψ on the integrand functions χ and ϕ has been made explicit. The term $\Psi(y_* | \phi, 1)$ in (4) means that the integral is computed for $\chi(y_P) = 1$. The computation of integral (5) for specific functions χ and ϕ is discussed in Section IV-B.

For special case of $\phi(\cdot) = 1$, integral (2), (3) is encountered in the stochastic calculus of pricing options. If we take , $\phi(\cdot) = 1$, $A = 1$, and $\chi(\cdot) = \exp(\cdot)$ in (3), then $f(y_P, y_\pi) = (e^{y_P} - B)_+$ gives pricing of a European call option for the log-price variable y_P and strike price B. If, further, y_P follows a Gaussian distribution, we get the celebrated Black-Scholes model [2]. Most of the existing work related to derivatives pricing assumes Gaussian distributions of the prices.

This paper extends the stochastic integral (2) to a general non-parametric probability distribution model. Such model of the joint probability distribution for the amount and the price can be acquired directly from the data. The paper estimates the non-parametric model and uses it for computing (2).

Figure 1 overviews the proposed methodology for stochastic pricing. First, historical data is used for Model Estimation. The obtained Distribution Model is then used in Stochastic Pricing to compute Contract Pricing (2).

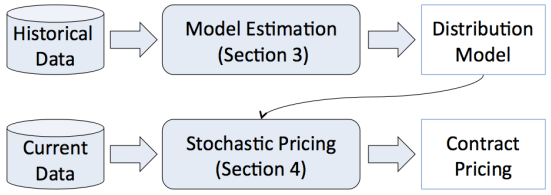

Fig. 1. Overview of the data-driven stochastic pricing approach.

The next sections describe the detail of the approach. The Model Estimation is discussed in Section III. Computation of Stochastic Pricing (2) is discussed in Section IV. Section V demonstrates example applications related to the electricity market.

III. Estimating the Probabilistic Model

The distribution $p(y_P, y_\pi | Z)$ in (2) is modeled from data (1) using multiple quantile regression. The non-parametric model provides a flexible description of the distribution shape. The constructed model can be then

used in the stochastic integral. This section describes the proposed data-driven modeling method.

A. Single Quantile Regression

As a starting point, consider a special case of the dataset D (1) with one scalar dependent variable y_i.

$$D_{qr} = \{Z_i, y_i\}_{i=1}^N, \tag{6}$$

where vectors Z_i are the same as in (1). It is assumed the data (6) are i.i.d. and follow an unknown conditional multivariate distribution with the probability density function $p(y|Z_i)$. We are trying to estimate the probability density $p(y|Z_i)$ from data (6) .

We assume that the distribution $p(y|Z)$ for (6) is described by the following quantile regression model

$$\begin{aligned} \mathbf{P}\left(y \le y(q)|Z\right) &= q, & (7)\\ y(q) &= Z\beta(q) + \alpha(q), & (8) \end{aligned}$$

where $q \in (0,1)$ is the quantile level; $\beta \in \Re^n$ and $\alpha \in \Re$ are the quantile regression hyperplane parameters.

For a given q, model (7) can be estimated by solving a LP (linear programming) problem, see [38]. This LP problem can be compactly written as

$$\begin{aligned} &\underset{\alpha,\beta}{\text{minimize}} \;\; h(Y - \mathbf{Z}\beta - \alpha\mathbf{1}_N; q),\\ &h(Y;q) = \frac{1}{2}\|Y\|_1 + \left(q - \frac{1}{2}\right)\mathbf{1}_N^T Y, \end{aligned} \tag{9}$$

where vector $Y = [y_1 \dots y_N]^T \in \Re^{N\times 1}$, matrix $\mathbf{Z} = [Z_1 \dots Z_N]^T \in \Re^{N\times n}$, and $\mathbf{1}_N \in \Re^N$ is a column vector of ones. In the case of $q = 1/2$, quantile regression is the well-known median regression. Quantile regression (9) uses pinball loss function $h(Y;q)$ and provides estimate that differs from the ordinary least squares (OLS) regression with squared loss $\|Y\|_2^2$.

The quantile regression can be used for estimation of the entire generating distribution in (6). The distribution can be estimated through a non-parametric model of the cumulative density function $\mathbf{P}(y|Z)$ in (7). The predictive power of such non-parametric model estimated using quantile regression is limited by two issues. The first issue is that if there are few data points on one side of the quantile hyperplane, solution (9) might have very large variance. This issue is with the distribution tails: the left tail quantile levels $q \ll 1$ and the right tail quantile levels, $1-q \ll 1$. This issue is discussed below in Subsection III-E.

The second issue is that the quantile regression hyperplanes for different quantiles are generally not parallel to each other. This means they intersect, at which point the estimated quantiles loose their ordering. This is known as the quantile crossing problem. The issue of the quantile crossing is discussed in Subsection III-D.

B. Multiple Quantile Regression

We need a model of the form (7), where $\alpha = \alpha(q)$ and $\beta = \beta(q)$ are smooth functions that can be differentiated to compute the probability density. Solving individual quantile regression problems (9) for many different values q might not give the desired result. To get a better solution, we solve a multiple quantile regression problem on the grid of m quantiles q_j

$$0 < q_1 < \ldots < q_m < 1. \tag{10}$$

These multiple quantile regression problems are solved as a joint optimization problem with a smoothing penalty

$$\begin{aligned} \underset{\{\alpha_i,\beta_i\}_{i=1}^m}{\text{minimize}} \;\; & \sum_{j=1}^m h(Y - \mathbf{Z}\beta_j - \alpha_j\mathbf{1}_N; q_j)\\ &+ \lambda\sum_{j=2}^m ||\beta_j - \beta_{j-1}||_2^2\\ &+ \mu\sum_{j=2}^{m-1}\left(\alpha_{j+1} + \alpha_{j-1} - 2\alpha_j\right)^2,\\ \text{subject to} \;\; & \beta_L = \beta_i, \quad (i = 1,\ldots,L),\\ & \beta_R = \beta_i, \quad (i = R,\ldots,m), \end{aligned} \tag{11}$$

where $h(Y,q)$, Y, $\mathbf{Z}$, and $\mathbf{1}_N$ are defined in (9); λ is a penalty on the first difference on β_j; μ, on the second difference of α_j. The reasoning behind the introduction of the smoothing penalties and selection of parameters λ and μ is presented in Subsection III-D. The constraints on β_j in (11) are introduced because for the low and the high quantiles there is not enough data on one side of the hyperplane to get accurate estimates of both regression slope β_j and its intercept α_j. The tail modeling is further discussed in Subsection III-E.

C. Solving the Multiple Quantile Regression

Convex optimization problem (11) is a quadratic programming (QP) problem. For small to moderate problem size, it can be solved with many available QP solvers. Large problem sizes for large training data sets (6) and large dimension of the regressor vector Z require a specialized convex optimization approach. The large problem sizes are encountered in the practical applications of the described quantile regression modeling methods. A scalable approach to solving problem (11) is discussed in our paper [35], where more detail can be found. This convex optimization problem can be solved by using the block splitting formulation of the alternating

direction method of multipliers (ADMM) [39]. The ADMM solution is scalable and parallelizable in the number of data points and the number of regressors. It has been demonstrated for $N = 21,696$ data points with 4,455 decision parameters in [35]. For a problem with $N = 10^6$ data points and 10^5 decision parameters the solution took 15 minutes to compute on a quad-core PC.

D. Quantile Crossing Problem

The quantile crossing issue is brought up in the end of Subsection III-A. It is examined here in more depth.

The multi-quantile model is obtained by solving (11). The solution is the set of the slopes β_j and intercepts α_j (11) on the quantile grid (10). This set describes the functions $\alpha(q)$ and $\beta(q)$ in (7). The probability density $p(y) = \frac{dq}{dy}$ can be obtained by differentiating $y = Z\beta + \alpha$. In practice, a secant method of differentiation can be employed using β_j and α_j.

The quantile crossing is avoided if $\frac{dq}{dy} > 0$. This is equivalent to $\frac{dy}{dq} > 0$, which can be expressed as

$$Z\frac{d\beta(q)}{dq} + \frac{d\alpha(q)}{dq} > 0. \tag{12}$$

We will use an equivalent form of (12)

$$-\left(ZM^{-1}\right) \cdot \frac{d\left(M\beta(q)\right)}{dq} \cdot \left[\frac{d\alpha(q)}{dq}\right]^{-1} \leq 1, \tag{13}$$

where M is a preconditioner matrix. Consider the regressor scatter matrix $\mathbf{Z}^T\mathbf{Z}$, where $\mathbf{Z} = [Z_1 \dots Z_N]$, for data set (6). Preconditioner M can be selected such that the matrix $M^{-1}(\mathbf{Z}^T\mathbf{Z})M^{-1}$ is well conditioned.

Consider the following use case. Model (7) is estimated (trained) for a set of past data (6). It is then exploited in some on-line algorithm for new data points. Will quantile crossing be an issue? The quantile hyperplanes intersect somewhere in the regressor space, unless $\beta(q)$ is constant. This means one can always find a regressor Z_* such that model has quantile crossing.

Below is a sufficient condition that there is no quantile crossing for a given regressor Z. Taking matrix norms of ZM^{-1} and the remaining multiplier in (13), maximizing over q, and dividing by the second norm yields

$$\|ZM^{-1}\|_2 \leq \Xi, \tag{14}$$

$$\Xi = \max_q \left\| M\frac{d\beta(q)}{dq} \cdot \left[\frac{d\alpha(q)}{dq}\right]^{-1} \right\|_2^{-1}, \tag{15}$$

where Ξ is the radius of the scaled regressors ball where the model is guaranteed to have no crossing.

The derivatives $d\beta(q)/dq$ and $d\alpha(q)/dq$ in (15) can be numerically estimated by the secant method for a solution of (11) computed on the grid (10) to evaluate Ξ in (15). The smoothing parameters λ and μ in (11) should be tuned such that condition (14), holds for all (or almost all) data points. The numerical analysis of these quantile crossing conditions and of the smoothing parameter tuning for the problem closely related to the examples in Section V can be found in [35].

E. Tail Modeling

The described smoothed quantile regression model (7), (8), (10), (11) interpolates the available data. The distribution tails, i.e., the extreme values of response variables y have to be modeled separately. The stochastic model for the last (or the first) quantile can be extrapolated beyond the data range if its parametric form is known. Extreme Value Theory (EVT) predicts that in many cases the distribution tails follow a Pareto (power law) or exponential distribution [40]. The EVT methods for tail estimation are peaks over threshold (POT) and the block maxima. This paper uses the POT method because it employs all data in the last (or the first) quantile.

The application examples of Section V, use log-variables. In log-coordinates, the Pareto distribution becomes an exponential distribution. The exponential tails can be estimated as a separate step, after the solution to (11) has been computed. The first $q_L = q_1$ and the last $q_R = q_m$ quantile levels on the quantile grid in (10) are used as the tail thresholds. The POT exceedances are

$$e_{L,j} = y_j - Z_j\beta_1 - \alpha_1, \quad j \in J_L, \tag{16}$$

$$e_{R,k} = y_k - Z_k\beta_m - \alpha_m, \quad k \in J_R, \tag{17}$$

where $J_L \equiv \{j : y_j < Z_j\beta_1 + \alpha_1\}$ and $J_R \equiv \{k : y_k > Z_k\beta_m + \alpha_m\}$.

We assume the tail probability distributions to be

$$Z\beta_1 + \alpha_1 - y | y < Z\beta_1 + \alpha_1 \quad \sim \quad \text{Exp}(\theta_L), \tag{18}$$

$$y - Z\beta_m - \alpha_m | y > Z\beta_m + \alpha_m \quad \sim \quad \text{Exp}(\theta_R). \tag{19}$$

These distributions are conditional on the POT exceedance events in (18) and (19), which have probabilities $q_L = \mathbf{P}(y < Z\beta_1 + \alpha_1)$ and $q_R = \mathbf{P}(y < Z\beta_m + \alpha_m)$. The Maximum Likelihood Estimates of the tail rate parameters θ_L in (18) and θ_R in (19) are

$$\theta_L^{-1} = -\text{mean}(e_{L,j} | j \in J_L), \tag{20}$$

$$\theta_R^{-1} = \text{mean}(e_{R,k} | k \in J_R). \tag{21}$$

F. Quantile Model of the Distribution

The numerical calculation of the stochastic integral (2) requires expressing the probability density $p(y|Z)$ through the data-driven probabilistic model of the form (7), (8), (10), (11), (18), (19).

Model (7) defines the cumulative density function (CDF): $\mathbf{P}\left(y \le y(q)|Z\right) = F_{y|Z}(y) = q$. Model (8) can be then considered as the inverse CDF (quantile function) $y = F^{-1}_{y|Z}(q)$, where q is the quantile level.

The quantile regression model (8), (11) is estimated at the quantile grid points (10). We extended this model to the entire $(0, 1)$ quantile interval as

$$y(s|Z) = \sum_{j=1}^{m} (Z\beta_j + \alpha_j) K_j(s), \tag{22}$$

where kernel functions $K_j(s)$ are such that $K_j(q_j) = 1$, and $K_j(q_k) = 0$ for $k \neq j$. Model (22) does linear interpolation between the grid nodes (10) and extrapolation for the tails in accordance with (18), (19). We use

$$K_j(s) = B_{j,2}(s) \qquad \text{for } q_L \le s \le q_R, \tag{23}$$

$$K_1(s) = \theta_L^{-1} \log \frac{s}{q_L} \qquad \text{for } 0 < s < q_L, \tag{24}$$

$$K_m(s) = -\theta_R^{-1} \log \frac{1-s}{1-q_R} \quad \text{for } q_R < s < 1, \tag{25}$$

where $B_{j,2}(s)$ are the second-order (triangular) B-spline kernels with knots (10). The log terms in (24) and (25) come from the quantile functions (inverse CDFs) of the tail exponential distributions (18) and (19).

G. Bivariate Distribution in Quantile Variables

The bivariate distribution $p(y_P, y_\pi|Z)$ in (2) can be expressed using the conditional probability formula

$$p(y_P, y_\pi|Z) = p(y_\pi|Z, y_P) \cdot p(y_P|Z). \tag{26}$$

We model $p(y_\pi|Z, y_P)$ and $p(y_P|Z)$ through the respective inverse CDFs of the form (22)–(25). The inverse CDF model for $p(y_P|Z)$ can be expressed as

$$y_P(s|Z) = \sum_{j=1}^{m} (Z\beta_{P,j} + \alpha_{P,j}) K_j(s), \tag{27}$$

where the model parameters for y_P, such as $\beta_{P,j}$ and $\alpha_{P,j}$, are obtained according to (8), (10), (11), (18), (19).

The inverse CDF model for $p(y_\pi|Z, y_P)$ in (26) explicitly depends on y_P, which is an additional regressor parameter. Similar to quantile model (27), we have

$$y_\pi(r|Z, y_P) = \sum_{j=1}^{m} (Z\beta_{\pi,j} + \gamma_{\pi,j} y_P + \alpha_{\pi,j}) K_j(r). \tag{28}$$

IV. Numerical Stochastic Integration

This section presents a method for numerical computation of the stochastic integral (5) using the nonparametric probability distribution model of the form (27), (28). In this section the regression variable vector Z is given and fixed. Therefore, the explicit dependence on Z is omitted to make the explanation clearer.

A. Probability Integral Transform

Models (27), (28) expressed in terms of the quantile levels allow to formulate integral (5) using the *probability integral transform*. The joint pdf in (5) can be expressed through quantile models (27), (28) by using (26), where in accordance with (7)

$$p(y_P(q)) = \frac{dF_{y_P}}{dy_P} = \left[\frac{\partial y_P(q)}{\partial q}\right]^{-1}, \tag{29}$$

$$p(y_\pi(r)|y_P) = \frac{dF_{y_\pi|y_P}}{dy_\pi} = \left[\frac{\partial y_\pi(r|y_P)}{\partial r}\right]^{-1}. \tag{30}$$

To compute integral (5), consider the transformation of integration variables to the quantile level variables

$$\left(\begin{array}{rcl} y_P &=& y_P(q) \\ y_\pi &=& y_\pi(r|y_P(q)) \end{array}\right) \to \left(\begin{array}{c} q \\ r \end{array}\right), \tag{31}$$

where $y_P(q)$ is given by (27) and $y_\pi(r|y_P(q))$ is given by (28) with $y_P(q)$ substituted for y_P. In the (q, r) variables, the first two integrand multipliers in (5) become

$$\psi(q) = \chi(y_P(q)), \tag{32}$$

$$\varphi(r, q) = \phi(y_\pi(r|y_P(q))). \tag{33}$$

Using (26), (29), and (30), the probability density multiplier in (5) can be expressed as

$$p(y_P(q), y_\pi(r|y_P(q))) = \left[\frac{\partial y_P(q)}{\partial q}\right]^{-1} \left[\frac{\partial y_\pi(r|y_P)}{\partial r}\right]^{-1}.$$

The integration variable change formula requires to include an additional transformation determinant multiplier

$$\det\left[\frac{\partial(y_P, y_\pi)}{\partial(q, r)}\right] = \frac{\partial y_P}{\partial q} \cdot \frac{\partial y_\pi}{\partial r}.$$

This holds because $y_P(q)$ does not depend on r.

The last two multipliers cancel each other, their product is unity. Thus, integral (5) can be computed as

$$\Psi(y_P(q_*); \varphi, \psi) = \int_0^1 \int_{q_*}^1 \varphi(r, q) \cdot \psi(q)\, dq dr, \tag{34}$$

where q_* solves $y_P(q_*) = y_*$ with y_* from (5), $\varphi(\cdot)$ is given by (27), (32) and $\psi(\cdot)$ by (27), (28), and (33).

B. Numerical Integration Approach

The evaluation of $\Psi(q_*)$ in (5) is illustrated in Table I that shows partitioning of the 2-D $[0, 1] \times [0, 1]$ integration domain into nine subdomains. The double integral $\Psi(q_*)$ is a sum of the integrals over these subdomains, which are annotated in Table I.

	$[0, r_L]$	$[r_L, r_R]$	$[r_R, 1]$
$[q_R, 1]$	$F_{L,R}(q_*)$	$\rho_R(q_*)$	$F_{R,L}(q_*)$
$[q_L, q_R]$	$\Gamma_L(q_*)$	$\Gamma(q_*)$	$\Gamma_R(q_*)$
$[0, q_L]$	$F_{L,L}(q_*)$	$\rho_L(q_*)$	$F_{R,L}(q_*)$

TABLE I
INTEGRATION DOMAIN PARTITIONING FOR INTEGRAL (5)

The middle block in Table I does not include singularities and can be computed using numerical quadrature. For simplicity, assume that the same uniform quantile grid of the form (10) is used for building both models (27) and (28), where $q_j - q_{j-1} = \Delta q$. The quadrature using linear interpolation (trapezoidal rule) is, then (assuming that $q_* < q_R$)

$$\Gamma(q_*) = (\Delta q)^2 \sum_{i=1, j=j_*}^{m} w_i^{(1)} w_j^{(j_*)} \varphi(r_i, q_j)\psi(q_j), \quad (35)$$

$$j_* = \arg\min_j \{|q_* - q_j|\}, \quad (36)$$

$$w_j^{(j_*)} = \begin{cases} 0, \ j < j_* \\ 1/2, \ j = j_*, j = m \\ 1, \ \text{otherwise} \end{cases}. \quad (37)$$

For $q_* \geq q_R$, we have $\Gamma(q_*) = 0$.

When computing integral (4), special care should be taken of the distribution tails that correspond to the eight edge blocks in Table I. Functions ψ (27) and ϕ (28) include logarithms at the integration domain edges that have singularities at 0 and 1 respectively. These logarithmic singularities describe the quantile behavior of the exponential distribution tails. Because of the singularities, the integrals for all eight edge blocks in Table I are not amenable to the numerical quadrature and must be computed analytically. The obtained analytical expressions for the eight edge block integrals given below can be used in the numerical computation of integral (4). Each of the eight tail integrals depends on the tail behaviors for the two the integration variables, q and r, and on the inner integral limit q_* in (34).

The tail integrals for the eight edge blocks can be expressed through the following auxiliary functions

$$\Omega_L(r, q_*) = \int_{\min(q_*, q_L)}^{q_L} \varphi(r, q) \cdot \psi(q) dq, \quad (38)$$

$$\Omega_R(r, q_*) = \int_{\max(q_*, q_R)}^{1} \varphi(r, q) \cdot \psi(q) dq, \quad (39)$$

$$g_L(q) = \int_0^{r_L} \varphi(r, q) dr, \quad (40)$$

$$g_R(q) = \int_{r_R}^{1} \varphi(r, q) dr. \quad (41)$$

Using notation (38)–(41) we have

$$F_{L,L}(q_*) = \int_0^{r_L} \Omega_L(r, q_*) dr, \quad (42)$$

$$F_{L,R}(q_*) = \int_0^{r_L} \Omega_R(r, q_*) dr, \quad (43)$$

$$F_{R,L}(q_*) = \int_{r_R}^{1} \Omega_L(r, q_*) dr, \quad (44)$$

$$F_{R,R}(q_*) = \int_{r_R}^{1} \Omega_R(r, q_*) dr, \quad (45)$$

$$\Gamma_L(q_*) = \Delta q \sum_{i=1}^{m} w_i^{(1)} \Omega_L(r_i, q_*), \quad (46)$$

$$\Gamma_R(q_*) = \Delta q \sum_{i=1}^{m} w_i^{(1)} \Omega_R(r_i, q_*), \quad (47)$$

$$\rho_L(q_*) = \Delta q \sum_{j=j_*}^{m} w_j^{(j_*)} g_L(q_j)\psi(q_j), \quad (48)$$

$$\rho_R(q_*) = \Delta q \sum_{j=j_*}^{m} w_j^{(j_*)} g_R(q_j)\psi(q_j). \quad (49)$$

Integrals (38)–(45) can be evaluated for specific functions $\varphi(r, q)$ and $\psi(q)$. This is done below.

C. Tail Integrals for Exponential Function

In the motivating example of Section V, logarithms of the variables y_P and y_π are used in the estimated model. Computing the expectation of the forward contract price in (3) requires the inverse, exponential, transformations. This means in (32), (33) we have

$$\phi(y_\pi) = \exp(\nu \cdot y_\pi), \quad (50)$$

$$\chi(y_P) = \exp(\kappa \cdot y_P), \quad (51)$$

where ν and κ are fixed exponent parameters.

Auxiliary functions (38)–(41) are then computed as

$$\Omega_L(r, q_*) = \Omega_e(r, q_L, \min(q_*, q_L), q_L, \sigma_L), \quad (52)$$

$$\Omega_R(r, q_*) = \Omega_e(r, q_R, \max(q_R, q_*), 1, -\sigma_R), \quad (53)$$

$$g_L(q) = \varphi(r_L, q) h(r_L, \theta_{\pi L}), \quad (54)$$

$$g_R(q) = \varphi(r_R, q) h(1 - r_R, -\theta_{\pi R}), \quad (55)$$

where

$$\sigma_L = (\nu\gamma_\pi(r) + \kappa)/\theta_{PL}, \quad (56)$$

$$\sigma_R = (\nu\gamma_\pi(r) + \kappa)/\theta_{PR}, \quad (57)$$

$$\Omega_e(r, q, q_*, a, \sigma) = \varphi(r, q) \cdot \psi(q) \cdot m(q)^{-\sigma} \cdot |m(a)^{\sigma+1} - m(q_*)^{\sigma+1}|/(\sigma + 1), \quad (58)$$

$$h(r, \theta) = r / (1 + \nu/\theta), \quad (59)$$

$$m(x) = \min(x, 1 - x), \quad (60)$$

and $\gamma_\pi(r)$ in (56), (57) is defined through (23), (25) as

$$\gamma_\pi(r) \quad = \quad \sum_{j=1}^{m} \gamma_{\pi,j} \cdot K_j(r). \tag{61}$$

The tail integrals (42)–(45) in Table I are

$$F_{L,L}(q_*) \quad = \quad \Omega_L(r_L, q_*) \cdot h(r_L, \theta_{\pi L}), \tag{62}$$
$$F_{L,R}(q_*) \quad = \quad \Omega_L(r_L, q_*) \cdot h(1 - r_R, -\theta_{\pi R}), \tag{63}$$
$$F_{R,L}(q_*) \quad = \quad \Omega_R(r_L, q_*) \cdot h(r_L, \theta_{\pi L}), \tag{64}$$
$$F_{R,R}(q_*) \quad = \quad \Omega_R(r_L, q_*) \cdot h(1 - r_R, -\theta_{\pi R}). \tag{65}$$

Integrals (46)–(49) in Table I are expressed through Ω_L in (52) and Ω_R in (53). To evaluate $\Psi(q_*; \varphi, 1)$ in the case of $\chi = 1$ one can set $\kappa = 0$ in (51), (56), and (57).

V. Application Examples

The motivating example for this paper is the stochastic pricing of the day-ahead electricity procurement contracts entered by a distribution utility. We consider the following stylized problem. The distribution utility is buying electrical power at the advance bulk electricity market. This happens every midnight, when the utility enters into 24 forward contracts for each hour of the next 24-hour day. As the day goes on, at each hour, the utility must fulfill the aggregate demand of the retail customers. This power load demand is unknown in advance. Possible shortages of the forward contract are filled by procuring the deficit amount at the spot electrical power market, which can be rather volatile. If the forward contract provides a surplus of the power, the overpayment cannot be recovered.

This section applies the developed methodology to a few problems related to the described day-ahead electricity procurement setup. Subsection V-A applies the modeling approach of Section III to data sets for a major US utility, PJM. The result is the data-driven stochastic models for the power loads and spot prices. The obtained models allow using the numerical stochastic pricing approach of Section IV for the forward contract pricing. As a simple example of the method usage, Subsection V-C evaluates the risk of the forward contract in fulfilling the power load demand. Subsection V-D provides forward contract pricing for a given hour of the day. This contract pricing is then applied to demonstrate procurement policy optimization in Subsection V-E.

A. Model for Day-ahead Forward Contract

Let P_t be the load demand and π_t the spot price at hour t. The response variables $y_{P,t}$ and $y_{\pi,t}$ in the dataset (1) are indexed by the number of hours t elapsed since the start of the data collection and are

$$y_{P,t} = \log\left(P_t / P_0\right), \tag{66}$$
$$y_{\pi,t} = \log\left(\pi_t / \pi_0\right), \tag{67}$$

where the log-load is normalized by $P_0 = 1$ GW and the log-price by $\pi_0 = \$1$/MWh. Regressors Z_t in (66), (67) include 6 indicators for week days, 11 for calendar months, and 3 for holidays, as well as 24-hour lagged log-load and log-spot price for a total of 23 regressors. For more discussion on the regressors see [35], [36].

This work uses a data set from the PJM utility, which is described in [41], [42]. The total system load P_t is used in the examples below. The range of the loads is 51 to 158 GW, with the average load of 88 GW. The data covers a time range from January 2011 to December 2013 with sampling interval of one hour, $N = 26,280$ samples at all. The dataset also includes the spot electricity prices. The prices range from \$0/MWh to \$768/MWh, with the average around \$37/MWh. The price is sampled at an hourly rate, the same as the load. We split up the dataset into two parts, the training and test set. The training set includes the data from January 2011 to December 2012, the total of 17,520 samples. It is used to estimate the data-driven stochastic models of Section III for the log power loads and log spot prices. The test set from January 2013 to December 2013, which contains 8,760 time samples, is used for validating the model Section V-B and the model-based policies in Sections V-C and V-E.

The data-driven methods of Section III were applied to estimate two probabilistic models from the data: the model of the form (27) for the log load $y_{P,t}$ (66) and the model (28) for the log price $y_{\pi,t}$ (67). The models use quantile grid (10) with $m = 99$ and spacing $\Delta q = 0.01$ spanning from $q_L = r_L = 0.01$ to $q_R = r_R = 0.99$.

Selection of smoothing parameters λ and μ is discussed in Section III-D. More detail on parameter tuning for a closely related application example can be found in [35]. In both data fit problems (11), for (66) and (67), the smoothing parameters were set to $\lambda = 5 \cdot 10^5$, $\mu = 5 \cdot 10^5$ for 2, 3, and 6-hour ahead forecasts and $\lambda = 10^6$, $\mu = 5 \cdot 10^5$ for all other hours. Cross validation of these parameters based on Pearson's chi square test statistics is discussed in Section V-B.

Figure 2 illustrates overall modeling logic. The model training is based on Data Set (6), which is the input into the Optimal Fit of Smoothed QR procedure (11). The output is Distribution Model on quantile grid (10) described by Quantile Regression parameters $\{\beta_j, \alpha_j\}_{j=1}^m$ in (11), and tail rate parameters (20) and (21).

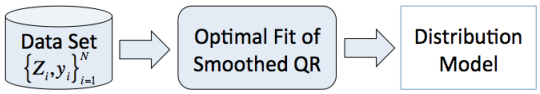

Fig. 2. Quantile regression model building logic.

The fit of the estimated models (27), (28) is illustrated by using them for forecasting the data in the training dataset. Overall, 24 models of the form (27) for the log loads at each of the 24 hours of the next day were estimated along with the 24 models of the form (28) for the log prices. At each midnight point, forecasts for the next 24 hours were computed and then compared to the actually realized values.

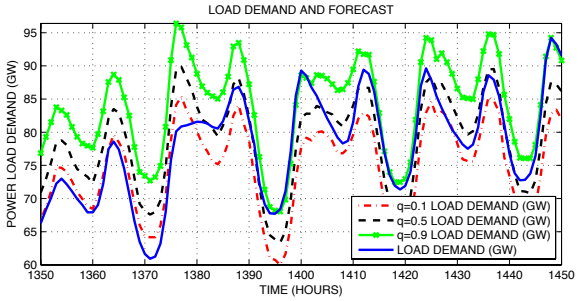

Fig. 3. Plot of quantile regression load forecast.

Figure 3 illustrates the load forecasts for a 101 hour segment (just over 4 days) computed for quantile levels $q = 0.1, 0.5, 0.9$ in the models of the form (27), (66). The 24 hourly forecasts for the 24 trained models are stitched together in the figure. The actual load demand curve is plotted along with the quantile level forecasts.

Figure 4 illustrates the quantile forecasts for the spot prices. The format is the same as in Figure 3.

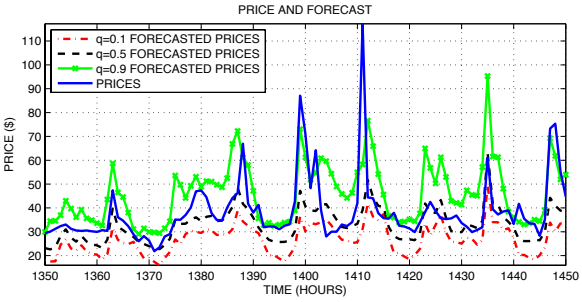

Fig. 4. Plot of quantile regression price forecast.

B. Statistical Validation of the Model

After the probability integral transform (31), the original distribution (26) becomes a uniform distribution in $[0,1] \times [0,1]$ square in the quantile variables (q, r), see [43]. Following [43], [44], the statistical validation of the estimate of distribution (26) is equivalent to testing the uniform distribution hypothesis in the transformed variables (q, r). This is done by dividing the $[0,1] \times [0,1]$ square into 100 cells of 0.1×0.1 size. Each cell has a probability mass of 0.01.

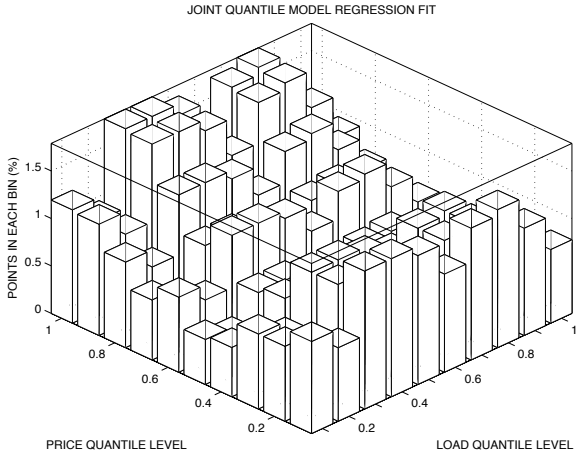

Fig. 5. Joint distribution histogram.

Figure 5 shows percentages of the test set data points in the 100 quantile cells for the estimated 20-hours ahead forecast model. As expected, the results are around 1% (the average of 3.65 data points). The Pearson's chi-squared test of the goodness of fit can be computed for the binned data in Figure 5 with 99 degrees of freedom, e.g., see [43], [44]. The test does not reject the uniform distribution hypothesis with significance of 99%.

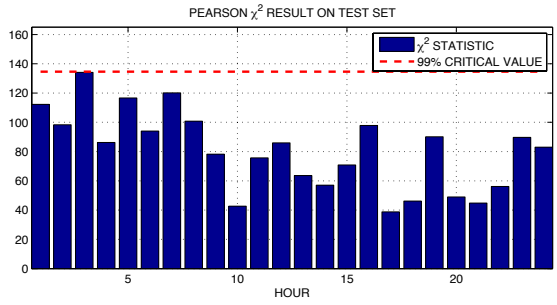

Fig. 6. Chi squared test value compared to the 99% critical value.

The chi-squared test results for all estimated models are shown in Figure 6. The dashed line shows the 99% critical value of the distribution. The bars show the chi-squared statistics for the 24 forecast hours. They are all

below the dashed line, which means the null hypothesis holds. These are results for the test set with the model estimated from the training set. The chi-squared statistics for the training set are substantially smaller.

Figures 5 and 6 characterize the goodness of fit of models (27) and (28) for the distribution body. The distribution tails describe rare extreme events that are not seen in these figures and not covered by the tests. The exponential model fit for the right tails of quantile functions (27) and (28) in log-variables (66), (67) is illustrated by the QQ plots in Figure 7. The errors of the 8am (8-hour ahead) forecast for the training set are plotted. The QQ plots show that the tail models fit the data very well. These log-exponential (Pareto) models have the best fit by far compared to the alternatives. More discussion on selecting and estimating tail distribution models for electrical load and price data, as well as further references can be found in [33], [35], [45].

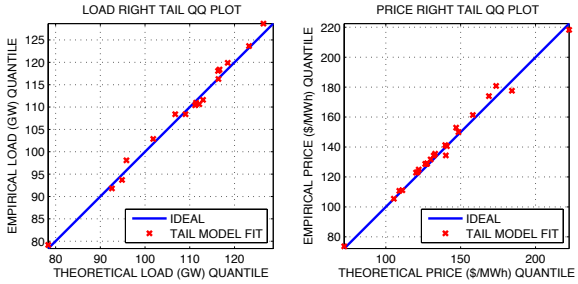

Fig. 7. QQ plots for the right tails: load (left), spot price (right).

C. Forward Contract Risk Estimation

The models built in Subsection V-A can be used for estimating the risk of the forward contract for electricity delivery. The probabilistic evaluation of risk is important because load volatility is increasing with the on-going proliferation of the renewable generation. The most important risk factor is that the required amount of the electricity cannot be procured at the spot market, because the stand-by generation capacity is insufficient.

Consider a stylized formulation, which resembles empirical approach used by some ISOs (Independent Systems Operators) for operational provisioning of the generating capacity, see [46]. Assume that power amount in the day-ahead electricity procurement contract is determined by utility using the Ordinary Least Squares (OLS) forecast of the form similar to (8)

$$\hat{y}_{P,t} = Z_t\beta_{OLS} + \alpha_{OLS}, \tag{68}$$

where α_{OLS} and β_{OLS} are determined by solving OLS data fit problem. Using the notation in (9), the OLS problem for dataset (6) is

$$\{\alpha_{OLS}, \beta_{OLS}\} = \arg\min_{\alpha,\beta} \|Y - \mathbf{Z}\beta - \alpha\mathbf{1}_N\|^2. \tag{69}$$

The ISO ensures that the generation capacity for the power procured by the utility is available with a 10% operating reserve margin. If the actual demand exceeds the operating reserve, it cannot be fully satisfied, no matter the price. This means some hard load must be shed, which is called a Loss of Load (LOL) event. The risk of the LOL events is controlled by the regulator, NERC. The LOL expectation (LOLE) is required to be less than of 1 day in 10 years, which corresponds to LOL probability (LOLP) of less than 0.0027.

In accordance with (66), the reserved capacity is $P_t^* = 1.1 \cdot P_0 \exp(\hat{y}_{P,t})$. The risk of the LOL event is $R_t = \mathbf{P}(y_{P,t} > \log 1.1 + \hat{y}_{P,t}|Z_t)$. It can be estimated using formula (4) and stochastic integral (5), where $\varphi(r,q) = 1, \psi(q) = 1, A = 1, B = 0$, and $y_* = \log 1.1 + \hat{y}_{P,t}$. The parts of integral (34) shown in Table I can be computed in accordance with (35), (42)–(45), and (46)–(49), (52), (53), where $\phi = 1$ is obtained by setting $\nu = 0$ in (50) and $\chi = 1$ for $\kappa = 0$ in (51). Alternatively, the risk R_t can be obtained from the numerical quantile model $y_P(s|Z)$ (27); $R_t = 1 - s_t$, where s_t is a solution to $y_P(s_t|Z_t) = \log 1.1 + \hat{y}_{P,t}$.

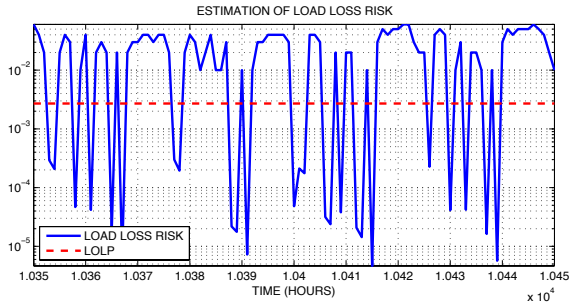

Fig. 8. Risk estimation of the current load.

To illustrate the risk estimation, 24 risk values are computed at each midnight for the 24 OLS forecasts for each hour of the next day. Forecast for each hour is using its own OLS model (68), (69) for the log loads y_P. The model is trained based on the data set described in Subsection V-A. The risk estimates for 24 hours of each day in the training dataset have been stitched together to provide coverage for each hour of the data. Figure 8 shows risk R_t computed for a 101 hour (about 4-day) segment of the data. The minimum of the risk estimate is less than 10^{-5} and mostly less

than 0.01. This means evaluation of the risk relies on the tail part of the data-driven probabilistic model of Subsection V-A. The horizontal dashed line in Figure 8 is the LOLP=0.0027 mandated by NERC. One can see that the common approach of using the OLS forecast plus 10% margin does not provide the required LOLP at all times. Figure 8 shows that using the OLS forecast provides average log-risk that looks reasonable; however the average risk is much higher than the required LOLP.

An alternative approach is to use the data-driven probabilistic model (27), (28) described in Subsection V-A to achieve the risk $R_t = r_* = 0.0027$ at all times by using quantile regression forecast $\hat{y}_{P,t} = y_P(r_*|Z_t)$. This risk corresponds to the horizontal dashed line in Figure 8.

D. Forward Contract Pricing

In this subsection, the data-driven probabilistic model of Subsection V-A is used for forward pricing of the electricity procurement contract.

We consider the day-ahead forward contract entered at the midnight for the electricity delivery at given hour t of the next day. With the regressor vector Z_t known, the conditional quantile models of the form $y_P(q)$ (27) for log-load (66) and of the form $y_\pi(r)$ (28) for the log-prices (67) can be used. The quantile model (27) and inverse of (66) can be used to parametrize the advance power order $P_a(t)$ though a quantile level variable s as $P_a(t) = P_0 e^{y_P(s)}$. Note that $s \to P_a(t)$ is a one-to-one transformation since both (27) and (66) are monotonic functions.

The forward contract cost C_t depends on the future log-load $y_{P,t}$ and log-spot price $y_{\pi,t}$, which are both random variables. Its expectation can be expressed as

$$\mathbf{E}[C_t(s)] = \mathbf{E}\left(\left[P_0 e^{y_{P,t}} - P_0 e^{y_P(s)}\right]_+ \pi_0 e^{y_{\pi,t}}\right). \tag{70}$$

To compute expected forward contract cost (70), we need to evaluate stochastic integral (2) with exponential functions (50), (51) in integrand function $f(y_{P,t}, y_{\pi,t})$ (3). Integral (70) is obtained by setting $A = P_0$, $B = P_0 e^{y_P(s)}$ in (3) and $\nu = \kappa = 1$ in (50), (51). We then used formulas (4) and (5). The parts of integral (34) shown in Table I were evaluated in accordance with (35), (42)–(45), (46)–(49), (52), (53).

E. Contract Cost Optimization

The forward contract pricing in Subsection V-D can be used as a basis for optimizing the procurement policy. This subsection considers optimization of the total expected cost that is the sum of the advance cost and the expected forward contract price. These costs depend on the advance power order $P_a(t)$.

The advance cost is the deterministic value $\pi_{adv,t-1} P_a(t)$. Using the same quantile parametrization as in Subsection V-D, we have

$$A_t(s) = \pi_{adv,t-1} P_0 e^{y_{P,t}(s)}, \tag{71}$$

where $\pi_{adv,t}$ is the advance price at time t. The advance price data from PJM utility is described in [42]. The advance prices range between \$0/MWh and \$313/MWh, with the average of \$37/MWh.

The total cost $T_t(s)$ at a given time t is

$$T_t(s) = A_t(s) + \mathbf{E}[C_t(s)], \tag{72}$$

where the expected forward contract cost $\mathbf{E}[C_t(s)]$ (70) is computed as the stochastic integral, see Subsection V-D. Based on (71), the advance cost $A_t(s)$ is a non-decreasing function of s. The expected spot cost $\mathbf{E}[C_t(s)]$ is a non-increasing positive function of s. The optimal trade-off between the advance cost and the spot cost that minimizes total cost $T_t(s)$ (72) can be found numerically by computing $T_t(s)$ on a grid of quantile levels $s \in (0, 1)$.

Figure 9 illustrates the total cost $T_t(s)$ for 12 pm on July 19th, 2013 computed from $\mathbf{E}[C_t(s)]$ (70), $A_t(s)$ (71), and (72) on the grid of the quantile levels s. Its minimum is achieved near $s_o = 0.81$. For $s < s_o$, $T_t(s)$ is dominated by $\mathbf{E}[C_t(s)]$ and for $s > s_*$, by $A_t(s)$.

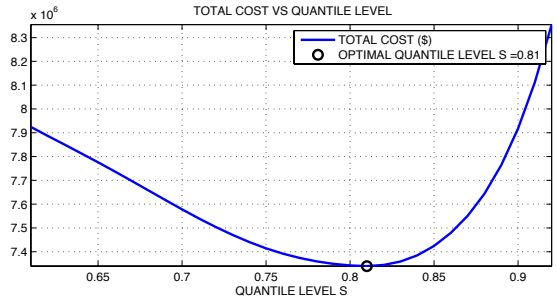

Fig. 9. The total cost and the minimizer quantile level s_o.

In backtesting, the historical data ahead in time of the contract allows computing the actual realized cost for an advance log-power procurement $\bar{y}_{P,t}$ as

$$U_t(\bar{y}_{P,t}) = \pi_{adv,t} P_0 e^{\bar{y}_{P,t}} + \pi_t P_0 \left(e^{y_{P,t}} - e^{\bar{y}_{P,t}}\right)_+, \tag{73}$$

where $y_{P,t}$ is the actual log power load that realized at the contract exercise time. We compute the actual realized cost U_t for the procured log power $\bar{y}_{P,t} = y_P(s_o(t))$, where $s_o(t)$ minimizes the total cost $T_t(s)$ in (72). We compare it with the results for the baseline

advance procurement policy $\bar{y}_{P,t} = \hat{y}_{P,t}$ that is based on the OLS forecast (68), (69). The comparison of these two actual realized costs is summarized in Table II. The optimized actual realized cost $U_t(y_P(s_o))$ shows 1.11% savings, which is \$153,000 for the chosen hour.

TABLE II
COST RESULTS (\$/MILLION)

Strategy/Model	Optimized	OLS Forecast
1-hour Cost	13.60	13.75
Optimal s	0.81	n/a

We backtest the optimized procurement policy $\bar{y}_{P,t} = y_P(s_o(t))$ and the baseline OLS forecast procurement policy $\bar{y}_{P,t} = \hat{y}_{P,t}$. The two actual realized costs (73) are computed for each hour of the test set (the year of 2013). Figure 10 illustrates the backtesting logic.

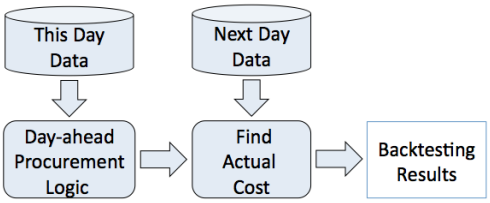

Fig. 10. Backtesting logic.

The backtesting results are shown in Figure 11 for a selected 61 hour period (about 1.5 days). It can be seen that most of the time the optimized policy has lower cost than the OLS forecast policy. Figure 11 shows that main savings come at the peak load times. This is because the developed non-parametric model describes the tails much more accurately than the normal distribution model implied in the OLS. The full year actual cost results are summarized in Table III. The yearly savings are 2.08% or \$641 million for the utility.

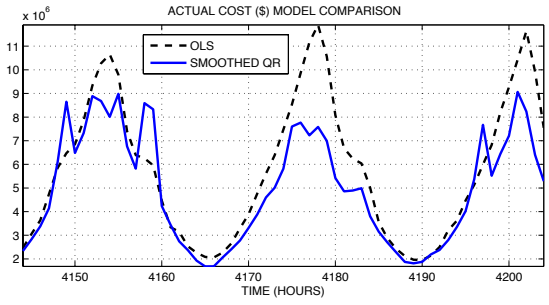

Fig. 11. Hourly cost comparison for a 61 hour period.

TABLE III
BACKTESTING RESULTS: TOTAL ACTUAL COST FOR A FULL YEAR

Strategy/Model	Optimized	OLS Forecast
1-year Cost (\$B)	30.1	30.8
Percentage Savings	2.08%	0%

VI. CONCLUSION

This paper has developed a numerical stochastic pricing method for forward procurement contracts using non-parametric long tail probabilistic models acquired directly from historical data. The method has been demonstrated in application to electricity procurement in day-ahead bulk markets. One example is risk estimation, another example is forward contract pricing and optimization. Backtesting of the optimized policy based on the developed data-driven method using electric utility data has demonstrated savings of 2.08% or \$641 million.

REFERENCES

[1] K. Varshney and A. Mojsilovic, "Business analytics based on financial time series," *IEEE Signal Process. Magazine*, vol. 28, no. 5, pp. 83–93, 2011.

[2] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *The Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.

[3] M. Carrión, A. B. Philpott, A. J. Conejo, and J. M. Arroyo, "A stochastic programming approach to electric energy procurement for large consumers," *IEEE Trans. on Power Systems*, vol. 22, no. 2, pp. 744–754, 2007.

[4] N. Secomandi and S. Kekre, "Optimal energy procurement in spot and forward markets," *Manufacturing & Service Operations Management*, vol. 16, no. 2, pp. 270–282, 2014.

[5] F. E. Benth, J. S. Benth, and S. Koekebakker, *Stochastic Modelling of Electricity and Related Markets*, vol. 11. World Scientific, 2008.

[6] M. Stutzer, "A simple nonparametric approach to derivative security valuation," *The Journal of Finance*, vol. 51, no. 5, pp. 1633–1652, 1996.

[7] Y. Aït-Sahalia and A. W. Lo, "Nonparametric estimation of state-price densities implicit in financial asset prices," *The Journal of Finance*, vol. 53, no. 2, pp. 499–547, 1998.

[8] P. Diko, S. Lawford, and V. Limpens, "Risk premia in electricity forward prices," *Studies in Nonlinear Dynamics & Econometrics*, vol. 10, no. 3, 2006.

[9] E. Hjalmarsson, "Does the Black-Scholes formula work for electricity markets? A nonparametric approach," *Working Papers in Econ.*, no. 101, 2003.

[10] L. Meligkotsidou, I. D. Vrontos, and S. D. Vrontos, "Quantile regression analysis of hedge fund strategies," *Journal of Empirical Finance*, vol. 16, no. 2, pp. 264–279, 2009.

[11] M. L. Barnes and A. T. W. Hughes, "A quantile regression analysis of the cross section of stock market returns," 2002.

[12] I.-C. Tsai, "The relationship between stock price index and exchange rate in Asian markets: A quantile regression approach," *Journal of Int. Financial Markets, Institutions and Money*, vol. 22, no. 3, pp. 609–621, 2012.

[13] K. Yu, Z. Lu, and J. Stander, "Quantile regression: Applications and current research areas," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 52, no. 3, pp. 331–350, 2003.

[14] I. Takeuchi, K. Nomura, and T. Kanamori, "Nonparametric conditional density estimation using piecewise-linear solution path of kernel quantile regression," *Neural Computation*, vol. 21, no. 2, pp. 533–559, 2009.
[15] J. W. Taylor, "A quantile regression neural network approach to estimating the conditional density of multiperiod returns," *Journal of Forecasting*, vol. 19, no. 4, pp. 299–311, 2000.
[16] F. Asnicar, *Optimization of Electricity Reserved Capacity*. PhD thesis, Technical University of Denmark, DTU, DK-2800, Denmark, 2006.
[17] S. Arora and J. W. Taylor, "Forecasting electricity smart meter data using conditional kernel density estimation," *Omega*, vol. 59, pp. 47–59, 2016.
[18] S. BenTaieb, R. Huser, R. J. Hyndman, and M. G. Genton, "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression," *IEEE Trans. on Smart Grid*, 2015.
[19] S.-K. Chao, W. K. Härdle, and W. Wang, "Quantile regression in risk calibration," in *Handbook of Financial Econometrics and Statistics*, pp. 1467–1489, Springer, 2015.
[20] Y. Fan, W. K. Härdle, W. Wang, and L. Zhu, "Composite quantile regression for the single-index model," tech. rep., Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, 2013.
[21] M. Bernardi *et al.*, "Bayesian tail risk interdependence using quantile regression," *Bayesian Analysis*, vol. 10, no. 3, pp. 553–603, 2015.
[22] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Probability & Its Applications*, vol. 14, no. 1, pp. 153–158, 1969.
[23] R. Koenker, P. Ng, and S. Portnoy, "Quantile smoothing splines," *Biometrika*, vol. 81, no. 4, pp. 673–680, 1994.
[24] I. Takeuchi, Q. V. Le, T. D. Sears, and A. J. Smola, "Nonparametric quantile estimation," *The Journal of Machine Learning Research*, vol. 7, pp. 1231–1264, 2006.
[25] X. He, "Quantile curves without crossing," *The American Statistician*, vol. 51, no. 2, pp. 186–192, 1997.
[26] J. Zhuang, I. W. Tsang, and S. C. Hoi, "A family of simple non-parametric kernel learning algorithms," *The Journal of Machine Learning Research*, vol. 12, pp. 1313–1347, 2011.
[27] J. V. Rosenberg, "Nonparametric pricing of multivariate contingent claims," *Journal of Derivatives*, no. 10, pp. 9–26, 2003.
[28] E. Jondeau, S.-H. Poon, and M. Rockinger, *Financial Modeling Under Non-Gaussian Distributions*. Springer Science & Business Media, 2007.
[29] P. Embrechts and H. Schmidli, "Modelling of extremal events in insurance and finance," *Zeitschrift für Operations Research*, vol. 39, no. 1, pp. 1–34, 1994.
[30] A. J. McNeil, "Estimating the tails of loss severity distributions using extreme value theory," *Astin Bulletin*, vol. 27, no. 01, pp. 117–137, 1997.
[31] L. K. Hotta, E. C. Lucas, and H. P. Palaro, "Estimation of VaR using copula and extreme value theory," *Multinational Finance Journal*, vol. 12, no. 3/4, pp. 205–218, 2008.
[32] A. J. McNeil, "Extreme value theory for risk managers," *Departement Mathematik ETH Zentrum*, 1999.
[33] S. Shenoy and D. Gorinevsky, "Estimating long tail models for risk trends," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 968–972, 2015.
[34] H. N. Byström, "Extreme value theory and extremely large electricity price changes," *Int. Review of Economics & Finance*, vol. 14, no. 1, pp. 41–55, 2005.
[35] S. Shenoy, D. Gorinevsky, and S. Boyd, "Non-parametric regression modeling for stochastic optimization of power market forecast," in *American Control Conference*, (Chicago, IL), July 2015.
[36] S. Shenoy and D. Gorinevsky, "Stochastic optimization of power market forecast using non-parametric regression models," in *Power & Energy Systems*, pp. 1–5, 2015.
[37] R. Carmona and V. Durrleman, "Pricing and hedging spread options," *Siam Review*, vol. 45, no. 4, pp. 627–685, 2003.
[38] R. Koenker, *Quantile Regression*. No. 38 in Econ. Society Monographs, Cambridge University Press, 2005.
[39] N. Parikh and S. Boyd, "Block splitting for distributed optimization," *Mathematical Programming Computation*, vol. 6, no. 1, pp. 77–102, 2014.
[40] L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction*. New York: Springer, 2006.
[41] PJM, "Metered load data." Available: http://www.pjm.com/markets-and-operations/ops-analysis/historical-load-data.aspx.
[42] PJM, "Energy pricing." Available: https://dataminer.pjm.com/dataminerui/pages/public/energypricing.jsf.
[43] M. Rosenblatt, "Remarks on a multivariate transformation," *The annals of mathematical statistics*, vol. 23, no. 3, pp. 470–472, 1952.
[44] J. Geweke and G. Amisano, "Comparing and evaluating Bayesian predictive distributions of asset returns," *Int. Journal of Forecasting*, vol. 26, no. 2, pp. 216–230, 2010.
[45] S. Shenoy and D. Gorinevsky, "Risk adjusted forecasting of electric power load," in *American Control Conference*, (Portland, OR), pp. 914–919, 2014.
[46] NERC, "2013 Summer Reliability Assessment," tech. rep., May 2013.

Saahil Shenoy (SM14) is a graduating Ph.D student in Physics and an ARCS Foundation Scholar at Stanford University. He completed the MS requirements in Particle Physics at the age of 20. His Stanford research is in big data analytics for extreme events, with applications in energy, finance, climate change, cloud computing, and risk analysis. He worked on portfolio optimization with Charles Schwab. He was Best Student Paper Award Finalist at 2015 American Control Conference and authored several papers in IEEE journals and conferences.

Dimitry Gorinevsky (M'91–SM'98–F'06) received a Ph.D. from Department of Mechanics and Mathematics at Moscow (Lomonosov) University in 1986, and a M.Sc. from Moscow Institute of Physics and Technology (Phystech) in 1982. He is a Consulting Professor in Electrical Engineering at Stanford University and a founder of Mitek Analytics, an Industrial IoT Analytics company in Palo Alto, CA. In the past, he he spent 10 years with Honeywell. He worked on many analytics applications across several industries. Dr. Gorinevsky was an Associate Editor of the IEEE Transactions on Control Systems Technology, 2001-2008. He received Control Systems Technology Award, 2002, and Transactions on Control Systems Technology Outstanding Paper Award, 2004, of the IEEE Control Systems Society. He received Best Paper Award (Senior Award), 2013, of the IEEE Signal Processing Society. He is a Fellow of IEEE.