Short-term hedging for an electricity retailer: online appendices

Debbie J. Dupuis, Geneviève Gauthier, Frédéric Godin

February 25, 2015

Appendices

A.1 Basis ratio

The weekly average price S_t^* paid for electricity differs from the weekly arithmetic average price S_t, which is the underlying asset of weekly futures. The extent to which S_t^* and S_t differ is represented by basis ratio η_t in (5). Figure 1 shows the observed ratio over the January 1, 2007 and July 23, 2012 period.

Figure 1: Basis ratio
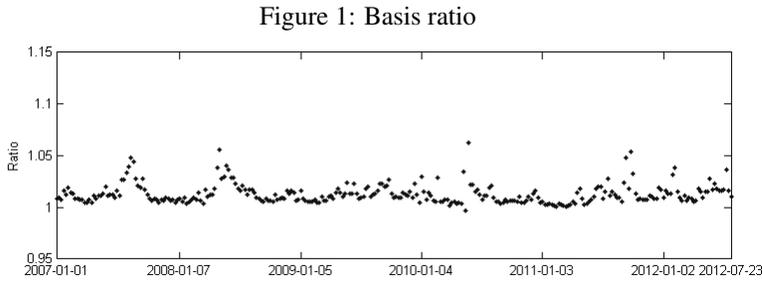
Notes. Observed ratio of the load weighted mean spot price to the arithmetic mean spot price as defined by (5). Data between January 1, 2007 and July 23, 2012.

As η_t is larger than one in all but one instance, S_t underestimates S_t^*. Such a departure has not yet been considered in the literature. This departure is due to the fact that more electricity is consumed during peak hours when its price is higher.

A.2 Delta-hedging with futures

If transaction costs are disregarded, the terminal value of the self-financing hedging portfolio with an initial value of 0 is given by $V_T = \sum_{j=t_0+1}^{T} \theta_j (B_T/B_j)(F_{j,T} - F_{j-1,T})$. Setting $\theta_j = B_j/B_T$, the terminal value of the portfolio becomes

$$V_T = \sum_{j=t_0+1}^{T} (F_{j,T} - F_{j-1,T}) = F_{T,T} - F_{T,t_0} = S_T - F_{T,t_0}.$$

Therefore, holding one unit of this portfolio for each unit of load sold (in the case where the load to serve is known with certainty) permits to lock in the price of electricity to F_{T,t_0}.

A.3 Solving problem (8)

The optimal trading strategy $(\theta^*_{T-2}, \theta^*_{T-1}, \theta^*_T)$ solving problem (8) with the semi-quadratic penalty (9) is obtained through the Bellman equation of dynamic programming (Bertsekas 1995):

$$\psi_{t,T} = \min_{\theta_{t+1}} \mathbb{E}\left[\psi_{t+1,T}|\mathcal{G}_t\right] \text{ with the terminal condition } \psi_{T,T} = G(\Psi_T - V_T) \quad (25)$$

$$\theta^*_{t+1} = \arg\min_{\theta_{t+1}} \mathbb{E}\left[\psi_{t+1,T}|\mathcal{G}_t\right]. \quad (26)$$

This optimization problem is tackled using backward induction over time. The traditional approach used for solving (25) is based on a lattice which includes all state variables of the problem; these include the current value of the load-basis and futures prices, current futures return volatilities, the current hedging portfolio value, lagged futures returns and the past portfolio composition. Such an approach is not viable due to its large dimension. Our approach is a stochastic tree which is feasible because the hedging portfolio is only rebalanced three times. The optimization of the trading position θ_t is performed numerically by discretizing its possible values.

If daily rebalancing was used instead of weekly rebalancing, the numerical complexity of the global hedging problem would dramatically increase. The current numerical scheme would not be viable. Either considerable simplifications to the problem would be necessary, see literature review in Section 1, or alternative dynamic programming schemes such as spectral interpolation (see Breton and de Frutos (2010)) or simulation and regression (see Denault et al. (2013)) would be required. Attempting the use of these procedures is beyond the scope of the current work. Moreover, using a daily resolution would (i) entail modeling the intra-week dynamics of the state variables which would add another layer of complexity to the problem, and (ii) lead to more rebalancing and consequently larger transaction costs. This justifies the weekly granularity with weekly rebalancing. The downside of considering weekly data is that the retailer cannot react immediately to intra-week information.

A.3.1 Simulation of the stochastic tree

Since the terminal condition $\psi_{T,T} = G(\Psi_T - V_T) = G(\mathcal{L}_T (F_{T,T} - F_{T-3,T}) - V_T)$ depends on the state variables (the load-basis $\mathcal{L}$ and the futures contracts related variables) and some endogenous variables (the portfolio value V_T and consequently the corresponding portfolio positions θ_{T-1}, θ_{T-2}, and θ_{T-3}), the random tree must account for all these dimensions.

At time $T-3$, M_{T-3} scenarios for the state variables are simulated from Equations (10)-(12) and (15)-(16).[1] These scenarios are combined with all the possible portfolio positions[2] $\theta_{T-2} \in \Theta_{T-2}$ to generate $N_{T-3} = M_{T-3}\text{Card}\{\Theta_{T-2}\}$ simulated values for endogenous variables (V_{T-2}, θ_{T-2}).

At time $T-2$, each of the N_{T-3} scenarios for the state and endogenous variables are subdivided into $N_{T-2} = M_{T-2}\text{Card}\{\Theta_{T-1}\}$ branches corresponding to all combinations of simulated state variables and possible portfolio positions. A similar iteration occurs at time $T-1$, leading to $N_{T-3} \times N_{T-2} \times N_{T-1}$ terminal nodes.

---

[1] Simulating a scenario at time t involves simulating the values of the load-basis and futures price innovations, respectively ε_{t+1} and $\omega_{t+1,t+j}$, $j = 1, 2, 3$.

[2] A discretized subset Θ_{T-2} of the possible positions is considered. $\text{Card}\{\Theta_{T-2}\}$ represents the number of elements it contains.

A.3.2 Backward induction

The algorithm solving (25) starts by computing the final hedging penalty at each terminal node[3] of the tree:

$$
\begin{aligned}
\hat{\psi}_T \quad & \binom{m_{T-3},m_{T-2},m_{T-1}}{\theta_{T-2},\theta_{T-1},\theta_T} \\
= \quad & G\bigg(\mathscr{L}_T\left(m_{T-3},m_{T-2},m_{T-1}\right)\left(F_{T,T}(m_{T-3},m_{T-2},m_{T-1})-F_{T-3,T}\right) \\
& \quad -V_T\binom{m_{T-3},m_{T-2},m_{T-1}}{\theta_{T-2},\theta_{T-1},\theta_T}\bigg).
\end{aligned}
$$

Equations (25)-(26) are then approximated using the following backward recursion for each node of the tree:

$$
\begin{aligned}
\hat{\theta}_T^*\binom{m_{T-3},m_{T-2}}{\theta_{T-2},\theta_{T-1}} &= \underset{\theta\in\Theta_T}{\arg\min}\,\frac{1}{M_{T-1}}\sum_{m=1}^{M_{T-1}}\hat{\psi}_T\binom{m_{T-3},m_{T-2},m}{\theta_{T-2},\theta_{T-1},\theta}, \\
\hat{\psi}_{T-1}\binom{m_{T-3},m_{T-2}}{\theta_{T-2},\theta_{T-1}} &= \underset{\theta\in\Theta_T}{\min}\,\frac{1}{M_{T-1}}\sum_{m=1}^{M_{T-1}}\hat{\psi}_T\binom{m_{T-3},m_{T-2},m}{\theta_{T-2},\theta_{T-1},\theta}, \\
\hat{\theta}_{T-1}^*\binom{m_{T-3}}{\theta_{T-2}} &= \underset{\theta\in\Theta_{T-1}}{\arg\min}\,\frac{1}{M_{T-2}}\sum_{m=1}^{M_{T-2}}\hat{\psi}_{T-1}\binom{m_{T-3},m}{\theta_{T-2},\theta}, \\
\hat{\psi}_{T-2}\binom{m_{T-3}}{\theta_{T-2}} &= \underset{\theta\in\Theta_{T-1}}{\min}\,\frac{1}{M_{T-2}}\sum_{m=1}^{M_{T-2}}\hat{\psi}_{T-1}\binom{m_{T-3},m}{\theta_{T-2},\theta}, \\
\hat{\theta}_{T-2}^* &= \underset{\theta\in\Theta_{T-2}}{\arg\min}\,\frac{1}{M_{T-3}}\sum_{m=1}^{M_{T-3}}\hat{\psi}_{T-2}\binom{m}{\theta}, \\
\hat{\psi}_{T-3} &= \underset{\theta\in\Theta_{T-2}}{\min}\,\frac{1}{M_{T-3}}\sum_{m=1}^{M_{T-3}}\hat{\psi}_{T-2}\binom{m}{\theta}.
\end{aligned}
$$

where[4]

$$
\begin{aligned}
m_{T-1} &= (\varepsilon_T, z_{0,T}), \\
m_{T-2} &= (\varepsilon_{T-1}, z_{1,T-1}, z_{0,T-1}), \\
m_{T-3} &= (\varepsilon_{T-2}, z_{2,T-2}, z_{1,T-2}, z_{0,T-2}).
\end{aligned}
$$

In the experiments of Section 4, when computing the initial optimal futures position θ_{T-2}^*, the number of scenarios are $M_{T-3}=M_{T-2}=M_{T-1}=500$ when the simulation at a given time point involves dependence between the load-basis and futures innovations, and $M_{T-3}=M_{T-2}=1000$ and $M_{T-1}=100$ otherwise. When the load-basis and futures innovations are independent, fewer scenarios are required at the final step since the conditional expectations can partially be solved analytically. More precisely, Equations (25)-(26) involve double integrals (one over the load innovation and the other over the futures return innovation with a one-week maturity). Fortunately, the load innovation is Gaussian, so the first integral can be computed analytically. Therefore, instead of using a regular Monte-Carlo simulation for the futures innovation, a quadrature in a single dimension is applied.

---

[3]The terminal nodes are identified with the set of indices corresponding to the branches constituting the path:

$$\binom{m_{T-3},m_{T-2},m_{T-1}}{\theta_{T-2},\theta_{T-1},\theta_T}.$$

[4]Although a single futures is used for hedging, returns must also be simulated for all futures that have a smaller time to maturity; these returns will affect the lagged returns and volatility variables of the futures that are used for hedging at subsequent time steps.

The discrete sets of portfolio positions are $\Theta_{T-2} = \{0.96, 0.965, \ldots, 1.04\}$ and $\Theta_{T-1} = \Theta_T = \{0.93, 0.94 \ldots, 1.07\}$, implying that $\text{Card}\{\Theta_{T-2}\} = 17$ and $\text{Card}\{\Theta_{T-1}\} = \text{Card}\{\Theta_T\} = 15$.

At time steps where load-basis and futures innovations are independent, variance reduction techniques improve the precision of the Monte Carlo estimates and compensate for small sample sizes. Antithetic variables are used in the simulation for load-basis innovations ε. The first half of scenarios are simulated by regular Monte-Carlo methods. In the last half of scenarios, the futures return innovations are identical to the ones in the first half. Load innovations are however set equal to their antithetic counterparts.

A.3.3 Re-simulation

The previous algorithm determines the optimal hedging strategy $\left(\theta^*_{T-2}, \theta^*_{T-1}, \theta^*_T\right)$ as seen from time $T - 3$. At time $T - 2$, the retailer holds θ^*_{T-2} long futures positions and has to select θ^*_{T-1} to perform the rebalancing. The realization of the state variables at time $T - 2$ will not exactly fall on one particular node of the random tree. The standard approach used to solve this issue is to interpolate between the nodes of the tree to determine the optimal hedging position θ^*_{T-1}. We opted for a re-simulation to obtain simulated data which incorporates the newly observed realization of state variables. More precisely, a two-period random tree is simulated from time $T - 2$ up to time T to update the optimal hedging strategy $\left(\theta^*_{T-1|T-2}, \theta^*_{T|T-2}\right)$. Since this tree is smaller than the previous one, we opted for a thinner discretization of the portfolio positions: $\Theta_{T-1} = \{0.93, 0.9325 \ldots, 1.07\}$, and $\Theta_T = \{0.93, 0.94 \ldots, 1.07\}$. Furthermore, at time steps where load-basis and futures innovations are independent, $M_{T-2} = 1000$ and $M_{T-1} = 100$. Otherwise, $M_{T-2} = M_{T-1} = 500$.

Finally, at time $T - 1$, a one-period random tree with

$$\Theta_T = \{0.900, 0.901 \ldots, 1.100\}$$

and $M_{T-1} = 20,000$ is simulated to update the final hedging position $\theta^*_{T|T-1}$.

A.4 Load-basis model estimation

A.4.1 Cross-validation procedure for load model selection

To determine the number P of Fourier terms in step 1 of the load-basis model estimation (or Q in step 2), a cross-validation procedure is implemented. The load-basis data are from 2007 to 2012. Data from year y are removed and retained as out-of-sample, while remaining data are in-sample. For each value of P (or Q), the model is estimated in-sample. Denote $\mathscr{J}^y_{1,P} = (\gamma, \beta_0, \ldots \beta_{2P+1})$ and $\mathscr{J}^y_{2,Q} = (\alpha_0, \ldots \alpha_{2Q+1})$. f denotes the pdf function.

$$\hat{\mathscr{J}}^y_{1,P} = \underset{\mathscr{J}^y_{1,P}}{\text{argmax}} \sum_{t, year(t) \neq y} \log f_{\mathscr{L}_t | \mathscr{L}_{t-1}}(\mathscr{L}_t | \mathscr{L}_{t-1})$$

$$\text{(under assumption that } v(t) \text{ is constant)}$$

$$\hat{\mathscr{J}}^y_{2,Q} = \underset{\mathscr{J}^y_{2,Q}}{\text{argmax}} \sum_{t, year(t) \neq y} \log f_{v_t(\mathscr{J}^y_{2,Q})\varepsilon_t}(\sqrt{\tilde{v}(t)}\tilde{\varepsilon}_t)$$

where $\tilde{g}(t)$ and $\tilde{v}(t)$ are obtained by respectively plugging $\hat{\mathscr{J}}^y_{1,P}$ in (11) and $\mathscr{J}^y_{2,Q}$ in (12). The $\tilde{\varepsilon}_t$ are calculated by replacing $g(t)$ and $v(t)$ by $\tilde{g}(t)$ and $\tilde{v}(t)$ in (10).

Then, a test statistic assessing the goodness-of-fit (MSE for P, log-likelihood for Q) is calculated out-of-sample:

$$\text{MSE}_y^P = \frac{1}{n_y} \sum_{t,\,year(t)=y} (\mathscr{L}_t - \text{Pred}(\mathscr{L}_t, \hat{\mathscr{J}}_{1,P}^y))^2$$

$$\text{log-l}_y^Q = \sum_{t,\,year(t)=y} \log f_{\sqrt{v_t(\hat{\mathscr{J}}_{2,Q}^y)}\varepsilon_t}(\sqrt{\hat{v}(t)}\hat{\varepsilon})$$

where n_y is the number of observations in year y. $\hat{g}(t)$ and $\hat{v}(t)$ are obtained by respectively plugging $\hat{\mathscr{J}}_{1,P}^y$ in (11) and $\hat{\mathscr{J}}_{2,Q}^y$ in (12). The $\hat{\varepsilon}_t$ are calculated by replacing $g(t)$ and $v(t)$ by $\hat{g}(t)$ and $\hat{v}(t)$ in (10). The predicted load-basis is $\text{Pred}(\mathscr{L}_t, \hat{\mathscr{J}}_{1,P}^y) = \hat{g}(t) + \hat{\gamma}(\mathscr{L}_{t-1} - \hat{g}(t-1))$ where $\hat{g}$ is calculated by plugging $\hat{\mathscr{J}}_{1,P}^y$ in (11) and $\hat{\gamma}$ is the first component of $\hat{\mathscr{J}}_{1,P}^y$. The prediction is obtained by applying a conditional expectation on (10). This operation is repeated for all years y and the test statistic is aggregated across all years:

$$\text{RMSE}_{total}^P = \sqrt{\frac{\sum_{y=2007}^{2012} n_y \text{MSE}_y^P}{\sum_{2007}^{2012} n_y}} \quad \text{or} \quad \text{log-l}_{total}^Q = \sum_{y=2007}^{2012} \text{log-l}_y^Q.$$

Parameters $\hat{P}$ and $\hat{Q}$ are selected to optimize the corresponding test statistic

$$\hat{P} = \underset{P}{\text{argmin}}\ \text{RMSE}_{total}^P \quad \text{and} \quad \hat{Q} = \underset{Q}{\text{argmax}}\ \text{log-l}_{total}^Q.$$

Results are shown in Tables 1 and 2 and suggest $\hat{P} = 3$ and $\hat{Q} = 2$.

Table 1: Cross-validation test results for the load-basis seasonality trend

Value for P	Cross-validation RMSE $(\times 10^5)$
1	2.386
2	2.376
3	**2.360**
4	2.364
5	2.367

Notes. Out-of-sample cross-validation prediction root-mean-square-error for the load-basis model with different numbers of Fourier terms P in the load-basis seasonality trend g defined by (11).

Table 2: Cross-validation test results for the load-basis variance trend

Value for Q	Cross-validation log-likelihood $(\times 10^{-3})$
1	-3.973
2	**-3.967**
3	-3.973
4	-3.974
5	-3.976

Notes. Out-of-sample cross-validation log-likelihood for the load-basis model with different numbers of Fourier terms Q in the load-basis variance trend v defined by (12).

A.4.2 Goodness-of-fit for the load model

In this section, the properties of the standardized residuals $\hat{\varepsilon}_t$ are analyzed to determine the adequacy of the load-basis model (10)-(12). Figure 2 shows a boxplot

of residuals by quarter of the year, a QQ-plot and a kernel density plot. Residuals look reasonably uniform across quarters; there is thus no obvious evidence that the seasonal trend is not properly being captured. The Gaussian distribution seems to be a suitable candidate for residuals, even if the empirical left tail of the load residuals is slightly heavier. A bootstrap Cramer-Von-Mises goodness-of-fit test for the adequacy of the Gaussian distribution is applied to the residuals and the p-value is 27%, not rejecting the Gaussian distribution. A Ljung-Box test for autocorrelation of residuals has a p-value of 92% and does not reject $\hat{\varepsilon}_t$ as white noise. The presence of a GARCH effect in the residuals is tested through the McLeod-Li test. The p-value is 18% and there is no significant presence of a GARCH effect and the ε load-basis innovations are modeled by a strong Gaussian white noise.

Figure 2: Load-basis model residuals
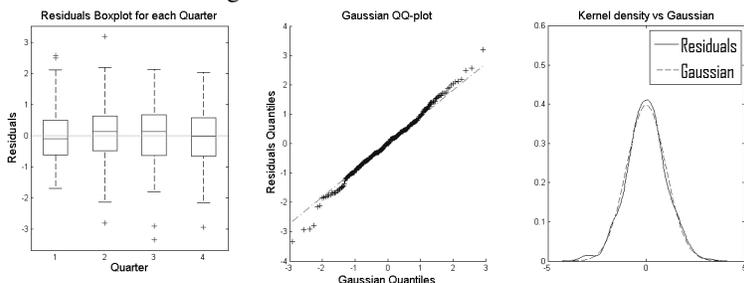
Notes. Boxplot, Gaussian QQ-plot and kernel plot for residuals $\hat{\varepsilon}$ of load-basis model (10)-(12). Data from January 1, 2007 and July 29, 2012.

A.4.3 Goodness-of-fit of futures return model

Ljung-Box and McLeod-Li tests for strong white noise are carried out on the scaled residuals $\hat{z}_{j,t}$, $j = 0, 1, 2$. *P*-values are obtained through simulation (usual *p*-value formulas incorrectly assume Gaussianity). *P*-values are given in Table 3 and none of the tests reject the white noise hypothesis.

Table 3: Autocorrelation tests for futures return innovations

Series	$z_{0,t}$	$z_{1,t}$	$z_{2,t}$
Ljung-Box *p*-value	0.36	0.35	0.41
McLeod-Li *p*-value	0.97	0.33	0.72

Notes. Bootstrapped $p-$values for the Ljung-Box and McLeod-Li tests applied on futures return innovations. Data between January 1, 2007 and July 29, 2012 for futures with 1, 2 and 3 weeks to maturity, respectively.

The choice of the NIG distribution for the innovations must be validated. Figure 3 compares the kernel density of the $\hat{z}_{i,t}$, its fitted NIG distribution and a corresponding Gaussian distribution. The NIG distribution represents more adequately the shape of the empirical residuals distribution than the Gaussian distribution, the latter is unable to capture the peakedness of the empirical futures returns distribution. Cramer-Von-Mises tests (with simulated *p*-value) are applied to assess the adequacy of the fit of the NIG distribution for the $z_{i,t}$ innovations. *P*-values are found in Table 4 for each univariate $z_{i,t}$, $i = 0, 1, 2$ series. The *p*-value for the joint trivariate series is 0.82. The NIG distribution thus provides an acceptable fit.

To validate the choice of the copula, Cramer-Von-Mises goodness-of-fit tests are applied for the Gaussian copula on the three following pairs of processes:

6

Figure 3: Futures return innovations distribution
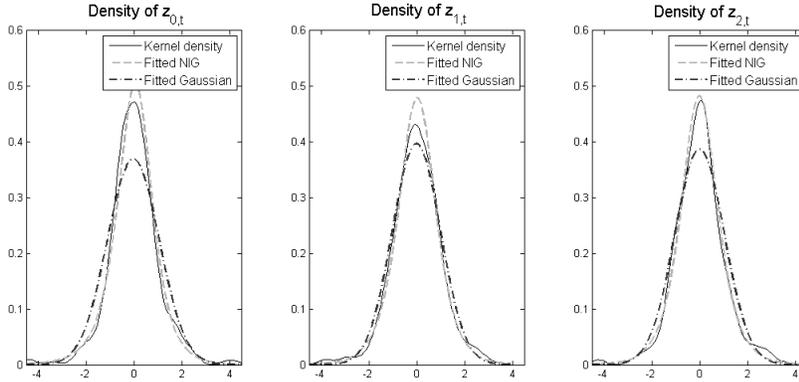
Notes. Kernel density plots of futures return scaled residuals with estimated parameters from Table 5, and fitted NIG and Gaussian distributions. Data between January 1, 2007 and July 29, 2012 for futures with 1, 2 and 3 weeks to maturity.

Table 4: Goodness-of-fit of the futures return distribution

Series	$z_{0,t}$	$z_{1,t}$	$z_{2,t}$
p-value	0.09	0.88	0.65

Notes. Bootstrapped $p-$values for the Cramer-Von-Mises goodness-of-fit test on the NIG distribution for futures return. Data between January 1, 2007 and July 29, 2012 for futures with 1, 2 and 3 weeks to maturity, respectively.

$(z_{0,t}, z_{1,t})$, $(z_{0,t}, z_{2,t})$ and $(z_{1,t}, z_{2,t})$.[5] The *p*-values for the three tests are given in Table 5. Since *p*-values are all high, the Gaussian copula provides an acceptable fit.

Table 5: Goodness-of-fit of the futures return copula

Innovation Pair	$(z_{0,t}, z_{1,t})$	$(z_{0,t}, z_{2,t})$	$(z_{1,t}, z_{2,t})$
p-value	0.90	0.67	0.97

Notes. Bootstrapped $p-$values for the Cramer-Von-Mises goodness-of-fit test applied to the Gaussian copula linking futures returns. Tests are applied on pairs of returns instead of the triplet $(z_{0,t}, z_{1,t}, z_{2,t})$. Data between January 1, 2007 and July 29, 2012 for futures with 1, 2 and 3 weeks to maturity.

---

[5]The test was not carried on the triplet $(z_{0,t}, z_{1,t}, z_{2,t})$. The numerical burden associated with such a test is very high.