Data-Driven Optimization for Commodity Procurement under Price Uncertainty: Online Appendix

Christian Mandl, Stefan Minner
Logistics & Supply Chain Management, TUM School of Management, Technische Universität München, Arcisstraße 21, 80333 Munich, Germany, christian.mandl@tum.de, stefan.minner@tum.de, http://www.log.wi.tum.de

Electronic Companion

EC.1. Optimal Procurement Policy Structure

In the following, we show that a price threshold $P_t^\tau(x_t)$ fully characterizes the optimal procurement policy, which is of a bang-bang type and using a particular option in a period, i.e., we either procure all uncovered demand or nothing with a contract $\tau \in \mathcal{F}^+$.

However, the state $z_t \in Z_t$ is not fully known. We only know that z_t contains the firm's position in the forward market $\vec{I} = (I_t^\tau)$ and the current forward curve $\vec{F}_t = (p_t^\tau : \tau \geq 0)$. However, without price model specifications, we do not know the drivers of the evolution of $\vec{F}$. Therefore, we introduce $x_t \in \mathcal{X}$ as the unknown parts (features) of the state space $z_t = (\vec{I}_t, \vec{F}_t, x_t)$ that drives the evolution of $\vec{F}$ and must be learned from the data $\mathbf{X}$ for which we formulate DDA models.

In order to provide the policy structure, we formulate the problem as a standard SDP using a Bellman equation with the endogenous state transition $I_{t+1}^{\tau-1} = I_t^\tau + y_t^\tau$ and the exogenous price evolution $\vec{F}_{t+1} = g(\vec{F}_t, x_t)$. This (unknown) exogenous price transition replaces typical stochastic price processes in models where they are assumed.

$$C_t(\vec{I}_t, \vec{F}_t, x_t) = \min_{\substack{y_t^\tau \geq 0 \\ I_t^0 + y_t^0 \geq d_t}} \left\{ \sum_{\tau \in \mathcal{F}} p_t^\tau y_t^\tau + \mathbb{E}_t \left[C_{t+1}(\vec{I}_{t+1}, \vec{F}_{t+1}, x_{t+1}) \right] \right\} \qquad \forall t = 0, ..., n. \tag{EC.1}$$

For every period t, we prove that for all $\tau \in \mathcal{F}^+$,

$$y_t^\tau(x_t) = \begin{cases} d_{t+\tau} & \text{if } p_t^\tau \leq P_t^\tau(x_t) \text{ and } I_t^\tau = 0, \\ 0 & \text{if } p_t^\tau > P_t^\tau(x_t). \end{cases} \tag{EC.2}$$

The proof exploits two properties. The value function of a given period is separable in the procurement instruments and linear with regard to the quantities y_t^τ.

Period t=n. We start with the last period $t = n$, which contains a non-stochastic purchase decision y_n^0 with

$$C_n(\vec{I}_n, \vec{F}_n, x_n) = C_n(\vec{I}_n, p_n^0) = p_n^0 y_n^0, \tag{EC.3}$$

that yields

$$y_n^0 = [d_n - I_n^0]^+. \tag{EC.4}$$

Period t=n-1. At the second to last stage $t = n - 1$, the problem becomes stochastic, since procurement decisions y_{n-1}^τ affect the cost-to-go C_n of period $t = n$:

$$C_{n-1}(\vec{I}_{n-1}, \vec{F}_{n-1}, x_{n-1}) = \min_{\substack{y_{n-1}^1 \geq 0 \\ I_{n-1}^0 + y_{n-1}^0 \geq d_{n-1}}} \left\{ p_{n-1}^0 y_{n-1}^0 + p_{n-1}^1 y_{n-1}^1 + (d_n - y_{n-1}^1 - I_{n-1}^1) \mathbb{E}_{n-1}[p_n^0 | x_{n-1}] \right\}. \tag{EC.5}$$

This function is linear and separable in y^0_{n-1} and y^1_{n-1}. The two decisions to be taken in $t=n-1$ are (i) the spot purchase decision

$$y^0_{n-1} = [d_{n-1} - I^0_{n-1}]^+ \tag{EC.6}$$

and (ii) the forward purchase decision

$$y^1_{n-1}(x_{n-1}) = \begin{cases} [d_n - I^1_{n-1}]^+ & \text{if } p^1_{n-1} \leq \mathbb{E}_{n-1}[p^0_n|x_{n-1}], \\ 0 & \text{if } p^1_{n-1} > \mathbb{E}_{n-1}[p^0_n|x_{n-1}], \end{cases} \tag{EC.7}$$

which is driven by whether the derivative with regard to y^1_{n-1} is positive or negative. Consequently, the implied threshold is

$$P^1_{n-1}(x_{n-1}) = \mathbb{E}_{n-1}[p^0_n|x_{n-1}], \tag{EC.8}$$

and the functional value is again separable and linear in the remaining uncovered demands:

$$\begin{aligned} C_{n-1}(\vec{I}_{n-1}, \vec{F}_{n-1}, x_{n-1}) = (d_{n-1} - I^0_{n-1}) p^0_{n-1} \\ + (d_n - I^1_{n-1}) p^1_{n-1} \cdot 1_{p^1_{n-1} \leq P^1_{n-1}} \\ + (d_n - I^1_{n-1}) \mathbb{E}_{n-1}[p^0_n|x_{n-1}] \cdot 1_{p^1_{n-1} > P^1_{n-1}} \end{aligned} \tag{EC.9}$$

with 1 as the indicator function.

Periods t=0,...,n-2. For all stages $t = 0, ..., n-2$, the linearity of the period cost and the cost-to-go in the decision variables reinforces the all-or-nothing decisions for using a procurement instrument depending on the state-dependent thresholds $P^\tau_t(x_t)$ and again results in the linearity of the resulting value function:

$$P^1_{n-1}(x_{n-1}) = \mathbb{E}_{n-1}[p^0_n|x_{n-1}], \tag{EC.10}$$

and the functional value is again separable and linear in the remaining uncovered demands:

$$\begin{aligned} C_{n-2}(\vec{I}_{n-2}, \vec{F}_{n-2}, x_{n-2}) = (d_{n-2} - I^0_{n-2}) p^0_{n-2} \\ + (d_{n-1} - I^1_{n-2}) p^1_{n-2} \cdot 1_{p^1_{n-2} \leq P^1_{n-2}} \\ + (d_{n-1} - I^1_{n-2}) \mathbb{E}_{n-2}[p^0_{n-1}|x_{n-2}] \cdot 1_{p^1_{n-2} > P^1_{n-2}} \\ + (d_n - I^2_{n-2}) p^2_{n-2} \cdot 1_{p^2_{n-2} \leq P^2_{n-2}} \\ + (d_n - I^2_{n-2}) \mathbb{E}_{n-2}[p^1_{n-1}|x_{n-2}] \cdot 1_{p^2_{n-2} > P^2_{n-2}, p^1_{n-1} \leq P^1_{n-1}} \\ + (d_n - I^2_{n-2}) \mathbb{E}_{n-2}[p^0_n|x_{n-2}] \cdot 1_{p^2_{n-2} > P^2_{n-2}, p^1_{n-1} > P^1_{n-1}}. \end{aligned} \tag{EC.11}$$

The two decisions to be taken in $t = 0, ..., n-2$ are again (i) the spot purchase decision

$$y^0_t = [d_t - I^0_t]^+ \tag{EC.12}$$

and (ii) the forward purchase decision

$$y_t^\tau(x_t) = \begin{cases} [d_{t+\tau} - I_t^\tau]^+ & \text{if } p_t^\tau \le P_t^\tau(x_t), \\ 0 & \text{if } p_t^\tau > P_t^\tau(x_t). \end{cases} \tag{EC.13}$$

The resulting policy is a bang-bang-type policy. The optimal hedging policy uses the expected cheapest source, which is in accordance with Smith, Stulz (1985), who show that partial hedging that leaves the hedger exposed to some residual price risk is not cost-optimal in our problems setting.

EC.2. Performance Bounds

In §EC.2.1, we derive in-sample cost bounds for DDA-BD, DDA-SD and DDA-ML that enhance our general understanding of the relationship between the special cases of DDA. In §EC.2.2, we derive probabilistic guarantees for the out-of-sample performance that justify the use of regularization for DDA.

EC.2.1. Bounds on the In-Sample Performance

Proposition EC.1 **(In-Sample Bounds)**.

(i) $C^{\text{UB}} \ge \hat{C}^{\text{BD}} \ge C^{\text{PF}}$. *$C^{\text{UB}}$ describes the upper cost bound (worst case procurement policy) as obtained by solving the perfect foresight (PF) optimization model as a maximization problem.*

(ii) $\hat{C}^{\text{BD}} \to C^{\text{PF}}$ *for* $N \to \infty$. *DDA is trained with respect to the loss function of the prescription problem, i.e.,* $\ell_{\text{DDA}} = \hat{C} - C^{\text{PF}}$. *If there is no multicollinearity between features* $i = 1, ..., N$, *then with increasing number of features* N, *the data-driven solution converges to the perfect foresight optimum (PF) as the model can fit the underlying function more accurately.*

(iii) *For* $N \ge T$, $\hat{C}^{\text{BD}} = C^{\text{PF}}$. *If the number of features* N *is greater than or equal to the number of demand periods* T, *the model is able to give an individual price threshold* $P_t^\tau(\boldsymbol{X})$ *to each in-sample period* t *(overfitting). Consequently,* $\hat{C}^{\text{BD}}$ *is equal to* C^{PF}. *Again, this is only true if multicollinearity between features* $i = 1, ..., N$ *is sufficiently small.*

(iv) $\hat{C}^{\text{BD}} \le \hat{C}^{\text{SD}}$. *This follows from (ii).*

(v) $\hat{C}^{\text{BD}} \le \hat{C}^{ML}$. *This is true as* $\lambda \ge 0$ *and for all* $i = 1, ..., N$ *and* $\tau \in \mathcal{F}^+$, $w_i^\tau \ge 0$ *(DDA-ML1) respectively* $|\beta_i^\tau| \ge 0$ *(DDA-ML2). The idea of ML in data-driven optimization is to overcome overfitting and therefore accept an in-sample performance loss.*

(vi) $\hat{C}^{\text{BD}} \le C^{\text{REO}}$. *The popular reoptimization approach (REO) (see §6.2) uses deterministic forward curve information from* t, *i.e.,* $\vec{F}_t$, *as predictor for future periods. REO is a special case of DDA-BD if* p_t^τ $\forall \tau \in \mathcal{F}$ *is used as feature information within equation (3).*

EC.2.2. Probabilistic Bounds on the Out-of-Sample Performance

Theorem EC.1 **(Generalization Bounds Based on Rademacher Complexity)**. *Let $\beta \in \mathscr{B}$ be the hypotheses for the learning algorithm, i.e., solutions to our prescription problem (regression-like coefficients β_i^τ). Then, for any $\delta > 0$, with probability $1-\delta$, the following inequality holds for all $\beta \in \mathscr{B}$:*

$$C^{\text{out}}(\beta) \leq \frac{1}{T}\sum_{t=1}^{T}\hat{C}_t(\beta) + \bar{c}\sqrt{\frac{log\frac{1}{\delta}}{2T}} + L\mathfrak{R}_T(\mathscr{B})$$

and

$$C^{\text{out}}(\beta) \leq \frac{1}{T}\sum_{t=1}^{T}\hat{C}_t(\beta) + 3\bar{c}\sqrt{\frac{log\frac{2}{\delta}}{2T}} + L\hat{\mathfrak{R}}_T(\mathscr{B})$$

where L is the Lipschitz constant, $\bar{c}$ is the upper bound of the loss function $\ell_{DDA} \in [0,\bar{c}]$ and $\mathfrak{R}(\mathscr{B})$ ($\hat{\mathfrak{R}}(\mathscr{B})$) is the (empirical) Rademacher complexity of hypothesis class $\mathscr{B}$.

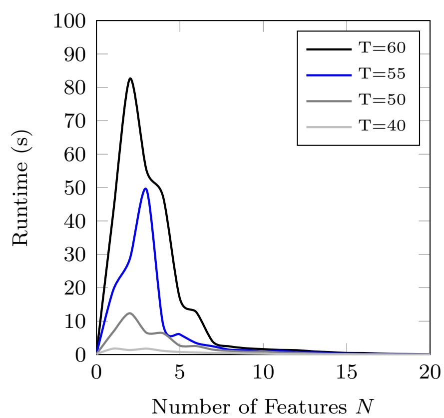
Proof. The proof is straightforward and similar to the proof for generic multivariate data-driven optimization problems as presented by Bertsimas and Kallus (2016). □

The generalization error Ω is a function of the model complexity $\mathscr{B}$, the sample size T and the confidence level δ. Ω decreases in δ and T and increases in $\mathscr{B}$ as the Rademacher function $\mathfrak{R}_T(\mathscr{B})$ increases with an increasing hypothesis space $\mathscr{B}$, i.e., with an increasing number of covariates N. For bounds on the Rademacher complexity $\mathfrak{R}_T(\mathscr{B})$ and Lipschitz constant L, we refer the reader to Mohri et al. (2012) and Bertsimas and Kallus (2016).

EC.3. Additional Plots and Tables

EC.3.1. Computation times

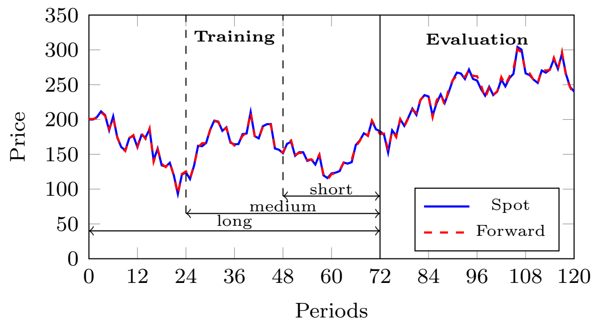
Figure EC.1 Computation times to train DDA-BD for different model sizes (Average across 100 runs)



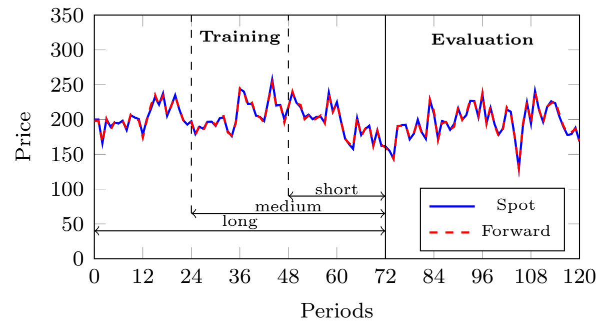
Note. Spot and forward ($\tau \in \{1,2,3,4\}$) prices randomly sampled from $N(100,20)$ $\forall t = 1,...,T$. Features $i = 1,...,N$ randomly sampled from $N(20,10)$. Main observation: Runtime-overfitting trade-off, i.e., by strongly increasing the number of features N (overfitting), computation times decrease (alternate optimal solutions). Computation times refer to the implementation with indicator constraints to avoid setting Big M.

EC.3.2. Results of the Numerical Study

Figure EC.2 Exemplary sample paths of spot (solid line) and forward (dashed line) prices under random walk (RW) and mean reversion (MR) assumptions and price process noise $\sigma_{\epsilon_t} = 10$

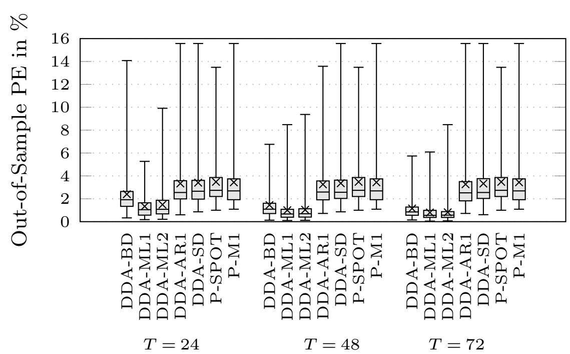


(a) RW

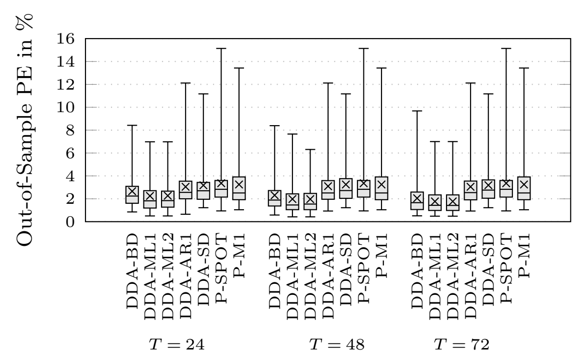


(b) MR

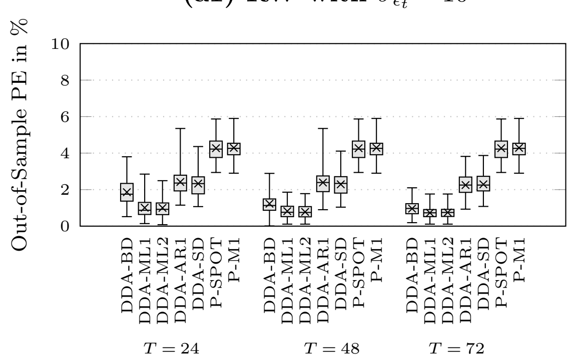
Figure EC.3 Out-of-sample prescription error (PE) of different procurement policies for $\sigma_{\epsilon_t} \in \{10, 20\}$ across 100 simulations conditional on training set size T and price process types (RW, MR)



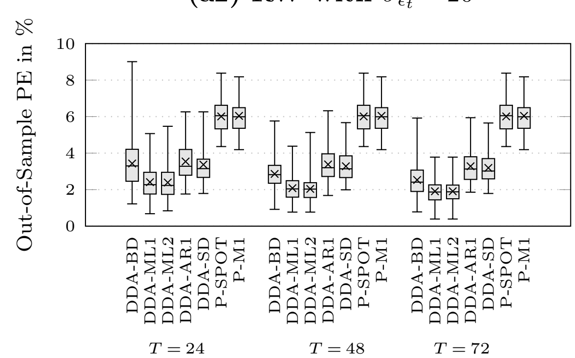
(a1) RW with $\sigma_{\epsilon_t} = 10$



(a2) RW with $\sigma_{\epsilon_t} = 20$

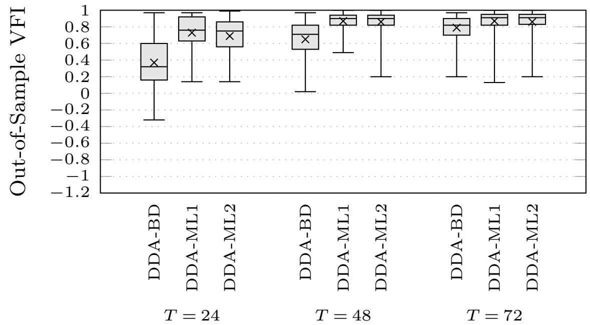


(b1) MR with $\sigma_{\epsilon_t} = 10$

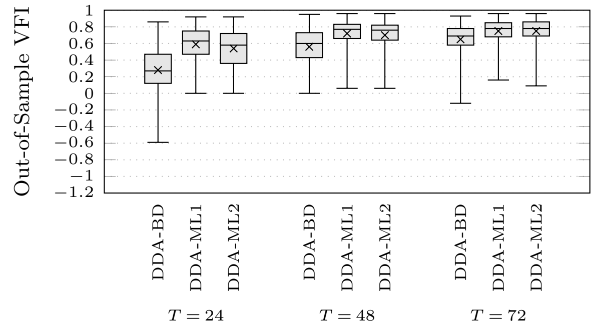


(b2) MR with $\sigma_{\epsilon_t} = 20$

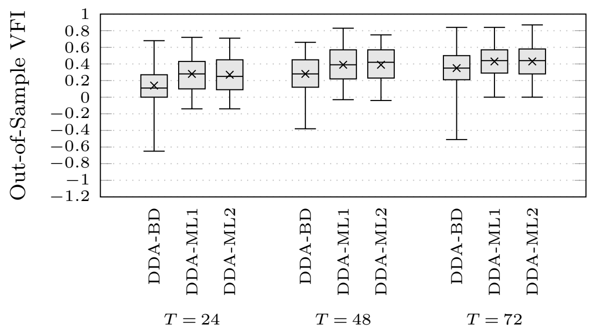
Figure EC.4 Out-of-sample value of feature information (VFI) of DDA-BD, DDA-ML1 and DDA-ML2 across 100 simulations for different training sets T, price process types (RW, MR) and noise levels σ_{ϵ_t}



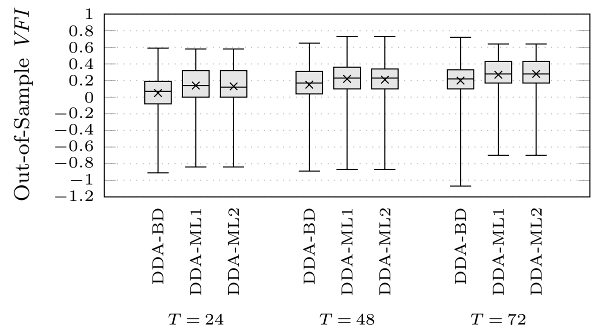
(a1) RW with $\sigma_{\epsilon_t}=5$



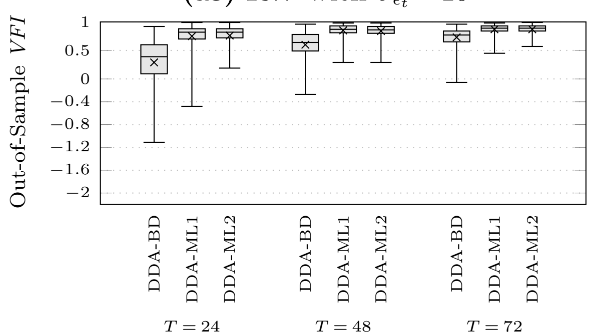
(a2) RW with $\sigma_{\epsilon_t}=10$



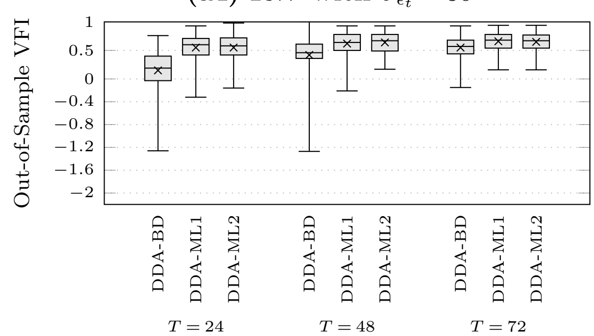
(a3) RW with $\sigma_{\epsilon_t}=20$



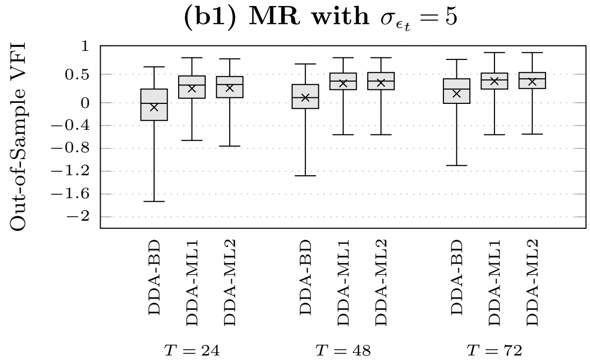
(a4) RW with $\sigma_{\epsilon_t}=30$



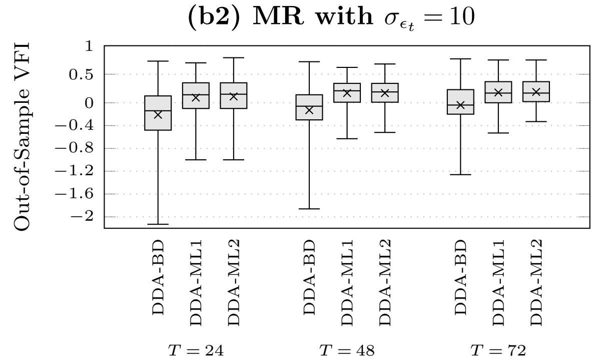
(b1) MR with $\sigma_{\epsilon_t}=5$



(b2) MR with $\sigma_{\epsilon_t}=10$



(b3) MR with $\sigma_{\epsilon_t}=20$



(b4) MR with $\sigma_{\epsilon_t}=30$

EC.3.3. Results on Empirical Data: Procurement of Natural Gas

Table EC.1 Historical monthly price and feature data (07/2007 - 06/2017 and 07/2017 - 12/2018) retrieved from Thomson Reuters Datastream and Eikon as of August 2017 and February 2019

Time Series	Unit	Data Source
Prices		
TTF Spot Price (Day-Ahead)	Euros/MWh	Eikon: TRNLTTFDA
TTF Futures Prices	Euros/MWh	Eikon: TRNLTTFMc1-4
Features		
Spot Price Returns	%	Eikon: TRNLTTFDA
TTF Spot Price Lags (t-1,t-2,t-3)	Euros/MWh	Eikon: TRNLTTFDA
Gaspool Spot Price (Day-ahead)	Euros/MWh	Eikon: TRDEGSPD1
Henry Hub Spot Price	USD/mmbtu	Datastream: NATLGAS
Henry Hub NYMEX Front Month	USD/mmbtu	Eikon: NGc1
Coal Price Front Month	USD/tonne	Eikon: TRAPI2Mc1
Brent Oil Spot Price	USD/barrel	Datastream: EIACRBR
Gas Production Germany	TWh	Datastream: BDELPRNGP
Domestic Gas Demand Germany	Mio Tons of Oil Equivalent	Datastream: BDXDGAS.P
EUR/USD	-	Eikon: EUR=
EUR/GBP	-	Eikon: UKEURSP
USD Index	-	Datastream: BOEUSA$
Federal Fund Rate	-	Datastream: USFDFUND
S&P 500 Index	-	Datastream: S&PCOMP
Bloomberg Commodity Index	USD	Datastream: DJUBSTR
PPI Energy Germany	-	Datastream: BDENERGYF
PPI Energy UK	-	Datastream: UKOPIEN2F
Temperature (Paris)	Celsius	Eikon: PARIS-OBS
Trading volumes (Spot, M1-M4)	MWh	www.powernext.com

Table EC.2 Procurement of natural gas: Average unit purchase cost

	Rolling In-Sample (07-2007 to 06-2016)				Rolling Out-of-Sample (07-2008 to 06-2017)			
	Mean	StDev	Min	Max	Mean	StDev	Min	Max
DDA-BD	19.58	4.41	10.47	25.17	20.32	4.13	12.31	25.65
DDA-SD	19.82	4.39	10.76	25.22	20.79	4.78	11.45	26.05
DDA-AR1	19.74	4.42	10.60	25.17	20.77	4.58	12.34	25.95
DDA-ML1	19.69	4.44	10.47	25.17	20.06	4.41	11.45	25.65
DDA-ML2	19.70	4.45	10.47	25.17	20.10	4.42	11.45	25.65
REO	20.80	4.48	11.59	25.68	20.28	4.69	11.59	25.68
1/N	21.62	4.57	12.41	26.02	20.97	4.92	12.41	26.02
P-SPOT	20.93	4.18	12.74	26.05	20.14	4.33	12.74	26.05
P-M1	21.09	4.32	12.40	26.09	20.43	4.60	12.40	26.09
P-M2	21.61	4.68	12.27	26.28	21.01	5.03	12.27	26.28
P-M3	22.08	5.03	12.24	28.47	21.46	5.43	12.24	28.47
P-M4	22.44	5.54	11.75	30.75	21.74	5.94	11.75	30.75
PF (=LB)	19.58	4.41	10.47	25.17	18.98	4.71	10.47	25.17
UB	23.83	5.12	14.46	32.02	23.07	5.47	14.46	32.02

Table EC.3 Dominance matrix: Percent of out-of-sample sub-periods in which a policy (row) performs strictly better than another (column)

	DDA-BD	DDA-SD	DDA-AR1	DDA-ML1	DDA-ML2	REO	1/N	P-SPOT	P-M1	P-M2	P-M3	P-M4
DDA-BD	-	44.4	66.7	0.0	0.0	55.6	66.7	22.2	55.6	55.6	55.6	55.6
DDA-SD	33.3	-	44.4	0.0	0.0	33.3	44.4	33.3	44.4	44.4	66.7	66.7
DDA-AR1	22.2	44.4	-	11.1	11.1	22.2	55.6	33.3	33.3	33.3	55.6	66.7
DDA-ML1	44.4	55.6	66.7	-	11.1	66.7	66.7	44.4	77.8	66.7	66.7	77.8
DDA-ML2	44.4	44.4	66.7	0.0	-	66.7	66.7	44.4	66.7	66.7	66.7	77.8
REO	44.4	66.7	77.8	33.3	33.3	-	66.7	55.6	66.7	77.8	77.8	77.8
1/N	33.3	55.6	44.4	33.3	33.3	33.3	-	44.4	33.3	66.7	66.7	66.7
P-SPOT	33.3	33.3	55.6	11.1	11.1	44.4	55.6	-	55.6	55.6	66.7	66.7
P-M1	44.4	55.6	66.7	22.2	33.3	33.3	66.7	44.4	-	55.6	66.7	66.7
P-M2	44.4	55.6	66.7	33.3	33.3	22.2	33.3	44.4	44.4	-	66.7	66.7
P-M3	44.4	33.3	44.4	33.3	33.3	22.2	33.3	33.3	33.3	33.3	-	66.7
P-M4	44.4	33.3	33.3	22.2	22.2	22.2	33.3	33.3	33.3	33.3	33.3	-

Table EC.4 Feature models of DDA-ML1 in backtests (07/2007 - 06/2017)

Features	07/07-06/08	07/08-06/09	07/09-06/10	07/10-06/11	07/11-06/12	07/12-06/13	07/13-06/14	07/14-06/15	07/15-06/16
Intercept		○◐◑●	○◐◑●	○◐◑●			○	○◐◑	○◐◑●
Spot Price Returns					○	○◐◑			
Gaspool Spot Price (Day-ahead)	●								
Henry Hub Spot Price	◐								
Henry Hub NYMEX Front Month	●				●		○		
Coal Price Front Month	○◐◑●								
Gas Production Germany	○◐◑				◑				
Domestic Gas Demand Germany						○◐◑●			
EUR/USD					○◐	●			
Federal Fund Rate	○◐◑●			●				○	
S&P 500 Index			○						
PPI Energy Germany					○	○◐◑●			
PPI Energy UK					◑				
Temperature (Paris)					○◑				

Note. ○: M1 futures contract, ◐: M2 futures contract, ◑: M3 futures contract, ●: M4 futures contract.

Table EC.5 Feature models of DDA-ML1 after its implementation (07/2017 - 12/2018)

Features	07/17	08/17	09/17	10/17	11/17	12/17	01/18	02/18	03/18
Intercept	○◐◑●		○◐◑●	○◐◑●		○◐◑●			○◐◑●
TTF Spot Price Lags (t-1)		○◐◑●							
TTF Spot Price Lags (t-4)					○		○		
Gaspool Spot Price (Day-ahead)	◑								
Domestic Gas Demand Germany		○			○		○	○●	
USD Index					○		○◐	○	
S&P 500 Index		○			◐◑		◐◑●	○◐◑●	
PPI Energy UK					●		●	●	
Temperature (Paris)					○			○	

Features	04/18	05/18	06/18	07/18	08/18	09/18	10/18	11/18	12/18
Intercept									
TTF Spot Price Lags (t-1)									
TTF Spot Price Lags (t-4)				●	●	●	●	●	●
Gaspool Spot Price (Day-ahead)									
Domestic Gas Demand Germany	●	●			○●	●	●	●	●
USD Index									
S&P 500 Index	○◐◑●	○◐◑●	○◐◑●	○◐◑●	○◐◑●	○◐◑●	○◐◑●	○◐◑●	○◐◑●
PPI Energy UK	●	●							
Temperature (Paris)			●	◐◑●	◐◑	◐	◐●	◐●	◐●

Note. ○: M1 futures contract, ◐: M2 futures contract, ◑: M3 futures contract, ●: M4 futures contract.

References

Bertsimas, D., N. Kallus. 2016. From predictive to prescriptive analytics. *Working Paper* .

Mohri, M., A. Rostamizadeh, A. Talwalkar. 2012. *Foundations of Machine Learning*. MIT Press.