Online Appendix

Optimal Management of Renewable Energy Certificates: A Reinforcement Learning Approach

Daeho Kim, Dong Gu Choi, Michael K. Lim

A. Proof of the Main Results

Proof of Lemma 1

To decompose the aggregator's total profit from all RECs into the value function of each REC, we expand the function $f(X_t, S_t, A_t)$ within the value function $V_t(X_t, S_t)$. By using equations (2) and (3), the value function $V_t(X_t, S_t)$ in (4) can be expressed as below:

$$V_t(X_t, S_t) = \max_{\pi \in \Pi} \Bigg\{ \alpha \sum_{n=1}^{N} \sum_{\tau=0}^{k_n} a_{n,\tau,t} P_{t,Y(t-\tau)} - \sum_{n=1}^{N} (i_{n,k_n,t} - a_{n,k_n,t}) \bar{P}_{t,Y(t-k_n)} + \sum_{\tau=k_N+1}^{k_0} a_{0,\tau,t} P_{t,Y(t-\tau)} + \delta_0 \mathbb{E}_t[V_{t+1}(X_{t+1}, S_{t+1}) | X_t, S_t] \Bigg\}. \tag{A.1}$$

We now express the above using the individual value functions defined below. First, we define the aggregator's value function $v_{0,\tau,t}(i_{0,\tau,t}, P_t)$, where it represents the aggregator's expected profit under the policy that maximizes the value of its own RECs with vintage τ. Hence, we can express $v_{0,\tau,t}(i_{0,\tau,t}, P_t)$ as follows:

$$v_{0,\tau,t}(i_{0,\tau,t}, P_t) := \max_{a_{0,\tau,t} \in [0, i_{0,\tau,t}]} \Big\{ a_{0,\tau,t} P_{t,Y(t-\tau)} + \delta_0 \mathbb{E}_t[v_{0,\tau+1,t+1}(i_{0,\tau+1,t+1}, P_{t+1}) | P_t, a_{0,\tau,t}] \Big\} \; \forall t, \forall \tau, \tag{A.2}$$

$$v_{0,k_0+1,t}(i_{0,k_0+1,t}, P_t) := 0 \;\; \forall t. \tag{A.3}$$

Further, we define the aggregator's value function $v_{n,\tau,t}(i_{n,\tau,t}, R_{n,t}, P_t, \bar{P}_t)$ for generator n's RECs with vintage τ as below.

$$\begin{aligned} & v_{n,\tau,t}(i_{n,\tau,t}, R_{n,t}, P_t, \bar{P}_t) \\ := & \max_{a_{n,\tau,t} \in [R_{n,\tau,t} i_{n,\tau,t}, i_{n,\tau,t}]} \Big\{ \alpha a_{n,\tau,t} P_{t,Y(t-\tau)} + \delta_0 \mathbb{E}_t[v_{n,\tau+1,t+1}(i_{n,\tau+1,t+1}, R_{n,t+1}, P_{t+1}, \bar{P}_{t+1}) | R_{n,t}, P_t, a_{n,\tau,t}] \Big\} \\ & \qquad \forall n, \forall t, \forall \tau \in \{1, \cdots, k_n - 1\}, \end{aligned} \tag{A.4}$$

$$\begin{aligned} & v_{n,k_n,t}(i_{n,t,k_n}, R_{n,t}, P_t, \bar{P}_t) \\ := & \max_{a_{n,k_n,t} \in [R_{n,\tau,t} i_{n,\tau,t}, i_{n,k_n,t}]} \Big\{ \alpha a_{n,k_n,t} P_{t,Y(t-k_n)} - (i_{n,k_n,t} - a_{n,k_n,t}) \bar{P}_{t,Y(t-k_n)} \\ & \qquad + \delta_0 \mathbb{E}_t[v_{0,k_n+1,t+1}(i_{n,k_n,t} - a_{n,k_n,t}, P_{t+1}) | R_{n,t}, P_t, a_{n,k_n,t}] \Big\} \; \forall n, \forall t. \end{aligned} \tag{A.5}$$

Finally, to express the value function of RECs yet to be issued, we use $V_t(X_t, \vec{0})$ by changing the inventory S_t to zeros $\vec{0}$ in $V_t(X_t, S_t)$.

Next, using the property that the maximum value of a sum of elements is less than or equal to the sum of the maximum of each element, it follows that the expected profit under the policy for the all RECs is less than or equal to the sum of the expected profit under the policy for each generator's or aggregator's RECs. Thus, we have the following:

$$V_t(X_t, S_t) \leq V_t(X_t, \vec{0}) + \sum_{n=1}^{N} \sum_{\tau=0}^{k_n} v_{n,\tau,t}(i_{n,\tau,t}, R_{n,t}, P_t, \bar{P}_t) + \sum_{\tau=k_N+1}^{k_0} v_{0,\tau,t}(i_{0,\tau,t}, P_t).$$

We note that the policy optimizing the management of separate RECs is a feasible policy for the aggregator. In addition, $V_t(X_t, S_t)$ in (A.1) represents the outcome under the optimal policy, which maximizes the expected profit across all feasible policies. Then, $V_t(X_t, S_t)$ is at least as large as the value under any feasible policy. This implies:

$$V_t(X_t, S_t) \geq V_t(X_t, \vec{0}) + \sum_{n=1}^{N} \sum_{\tau=0}^{k_n} v_{n,\tau,t}(i_{n,\tau,t}, R_{n,t}, P_t, \bar{P}_t) + \sum_{\tau=k_N+1}^{k_0} v_{0,\tau,t}(i_{0,\tau,t}, P_t).$$

From the above inequalities, we therefore conclude that

$$V_t(X_t, S_t) = V_t(X_t, \vec{0}) + \sum_{n=1}^{N} \sum_{\tau=0}^{k_n} v_{n,\tau,t}(i_{n,\tau,t}, R_{n,t}, P_t, \bar{P}_t) + \sum_{\tau=k_N+1}^{k_0} v_{0,\tau,t}(i_{0,\tau,t}, P_t). \tag{A.6}$$

This implies that the value function $V_t(X_t, S_t)$ at any day t in (4) can be decomposed into two main components: the value of RECs yet to be issued, $V_t(X_t, \vec{0})$, and the values of RECs already issued, which includes both the generators' individual RECs, $v_{n,\tau,t}(i_{n,\tau,t}, R_{n,t}, P_t, \bar{P}_t)$ and the aggregator's own RECs, $v_{0,\tau,t}(i_{0,\tau,t}, P_t)$. □

Proof of Lemma 2

As mentioned in Lemma 1, the value function $V_t(X_t, S_t)$ in (4) can be decomposed into two parts: the value associated with RECs yet to be issued and the value corresponding to RECs already issued at day t.

Now, we show that there exists an optimal policy for $v_{0,\tau,t}(i_{0,\tau,t}, P_t)$ and $v_{n,\tau,t}(i_{n,\tau,t}, R_{n,t}, P_t, \bar{P}_t)$, where the optimal action $a^*_{j,\tau,t}$ for all τ and t is either 0 or $i_{j,\tau,t}$ where $j \in \{0\} \cup \{1, \cdots, N\}$. We prove this by induction.

We start with the aggregator's own RECs. If $\tau = k_0$, it is straightforward to see that $v_{0,k_0,t}(i_{0,k_0,t}, P_t)$ is $i_{0,k_0,t} P_{t,Y(t-k_0)}$ by the optimal action $a^*_{0,k_0,t} = i_{0,k_0,t}$ for all t.

If $\tau = k_0 - 1$, then it follows that

$$\begin{aligned} v_{0,k_0-1,t}(i_{0,k_0-1,t}, P_t) &= \max_{a_{0,k_0-1,t} \in [0, i_{0,k_0-1,t}]} \left\{ a_{0,k_0-1,t} P_{t,Y(t-k_0+1)} + \delta_0 \mathbb{E}_t[i_{0,k_0,t+1} P_{t+1,Y(t-k_0+1)} | P_t, a_{0,\tau,t}] \right\} \\ &= \max_{a_{0,k_0-1,t} \in [0, i_{0,k_0-1,t}]} \Big\{ a_{0,k_0-1,t} \big(P_{t,Y(t-k_0+1)} - \delta_0 \mathbb{E}_t[P_{t+1,Y(t-k_0+1)} | P_t] \big) \\ &\qquad + i_{0,k_0-1,t} \delta_0 \mathbb{E}_t[P_{t+1,Y(t-k_0+1)} | P_t] \Big\}. \end{aligned} \tag{A.7}$$

We note that the function to be maximized in (A.7) is linear in $a_{0,k_0-1,t}$, and thus the optimal action $a^*_{0,k_0-1,t} = 0$ if $(P_{t,Y(t-k_0+1)} - \delta_0 \mathbb{E}_t[P_{t+1,Y(t-k_0+1)} | P_t]) < 0$; otherwise, $a^*_{0,k_0-1,t} = i_{0,k_0-1,t}$. Therefore, we have

$$\begin{aligned} &v_{0,k_0-1,t}(i_{0,k_0-1,t}, P_t) \\ &= \begin{cases} i_{0,k_0-1,t} \delta_0 \mathbb{E}_t[P_{t+1,Y(t-k_0+1)} | P_t], & \text{if } P_{t,Y(t-k_0+1)} - \delta_0 \mathbb{E}_t[P_{t+1,Y(t-k_0+1)} | P_t] < 0 \\ i_{0,k_0-1,t} P_{t,Y(t-k_0+1)}, & \text{if } P_{t,Y(t-k_0+1)} - \delta_0 \mathbb{E}_t[P_{t+1,Y(t-k_0+1)} | P_t] \geq 0 \end{cases}. \end{aligned} \tag{A.8}$$

It is important to note that the value function (A.8) is linear in inventory level $i_{0,k_0-1,t}$ regardless of the action. Hence, it follows that $v_{0,k_0-1,t}(i_{0,k_0-1,t}, P_t) = C_{0,k_0-1,t} \cdot i_{0,k_0-1,t}$, where $C_{0,k_0-1,t}$ and $C_{0,k_0,t}$ are respectively defined as

$$C_{0,k_0-1,t} := \max\{P_{t,Y(t-k_0+1)}, \delta_0 \mathbb{E}_t[P_{t+1,Y(t-k_0+1)} | P_t]\}, \tag{A.9}$$

and

$$C_{0,k_0,t} := P_{t,Y(t-k_0)}. \tag{A.10}$$

Lastly, we consider τ' that lies in $1 \leq \tau' \leq k_0 - 2$. Assume that, for $\tau' + 1$, the optimal action $a^*_{0,\tau'+1,t+1}$ is either 0 or $i_{0,\tau'+1,t+1}$, and that the value function $v_{0,\tau'+1,t+1}(i_{0,\tau'+1,t+1}, P_{t+1})$ is linear in inventory level given by $C_{0,\tau'+1,t+1} \cdot i_{0,\tau'+1,t+1}$. In addition, we can express $C_{0,\tau'+1,t+1}$ in a recursive form as

$$C_{0,\tau'+1,t+1} = \max\{P_{t+1,Y(t-\tau')}, \delta_0 \mathbb{E}[C_{0,\tau'+2,t+2} | P_{t+1}]\}, \tag{A.11}$$

and confirm that this does not depend on $a^*_{0,\tau'+1,t+1}$ nor $i_{0,\tau'+1,t+1}$. Hence, the value function $v_{0,\tau',t}(i_{0,\tau',t}, P_t)$ can be expressed as follows:

$$\begin{aligned} &v_{0,\tau',t}(i_{0,\tau',t}, P_t) \\ &= \max_{a_{0,\tau',t} \in [0, i_{0,\tau',t}]} \{a_{0,\tau',t} P_{t,Y(t-\tau')} + \delta_0 \mathbb{E}_t[C_{0,\tau'+1,t+1} \cdot (i_{0,\tau',t} - a_{0,\tau',t}) | P_t, a_{0,\tau',t}]\} \\ &= \max_{a_{0,\tau',t} \in [0, i_{0,\tau',t}]} \{a_{0,\tau',t} \big(P_{t,Y(t-\tau')} - \delta_0 \mathbb{E}_t[C_{0,\tau'+1,t+1} | P_t] \big) + i_{0,\tau',t} \delta_0 \mathbb{E}_t[C_{0,\tau'+1,t+1} | P_t]\}. \end{aligned} \tag{A.12}$$

In (A.12), the optimal action $a^*_{0,\tau',t}$ is given by either 0 or $i_{0,\tau',t}$ depending on the signage of $(P_{t,Y(t-\tau')} - \delta_0 \mathbb{E}_t[C_{0,\tau'+1,t+1} | P_t])$. In addition, we verify that the value function $v_{0,\tau',t}(i_{0,\tau',t}, P_t)$ can

be expressed as $C_{0,\tau',t} \cdot i_{0,\tau',t}$ where $C_{0,\tau',t} = \max\left\{P_{t,Y(t-\tau')}, \delta_0 \mathbb{E}[C_{0,\tau'+1,t+1}|P_t]\right\}$. Hence, using induction, we have shown that there exists an optimal policy in $v_{0,\tau,t}$ where the action $a^*_{0,\tau,t}$ for all τ and t are either 0 or $i_{0,\tau,t}$.

We next consider the generator's RECs. We show the existence of optimal policy in $v_{n,\tau,t}(i_{n,\tau,t}, R_{n,\tau,t}, P_t, \bar{P}_t)$, where the actions are set to either 0 or $i_{n,\tau,t}$. We have already showed that $v_{0,\tau,t}(i_{0,\tau,t}, P_t) = i_{0,\tau,t} \cdot C_{0,\tau,t}$. Then, using equation (A.5), we express $v_{n,k_n,t}(i_{n,k_n,t}, R_{n,t}, P_t, \bar{P}_t)$ as below:

$$\begin{aligned} &v_{n,k_n,t}(i_{n,k_n,t}, R_{n,t}, P_t, \bar{P}_t) \\ &\quad = \max_{a_{n,k_n,t} \in [R_{n,\tau,t} i_{n,\tau,t}, i_{n,k_n,t}]} \Big\{ \quad a_{n,k_n,t}\big(\alpha P_{t,Y(t-k_n)} + \bar{P}_{t,Y(t-k_n)} - \delta_0 \mathbb{E}_t[C_{0,k_n+1,t+1}|P_t]\big) \\ &\qquad\qquad + i_{n,k_n,t}\big(\delta_0 \mathbb{E}_t[C_{0,k_n+1,t+1}|P_t] - \bar{P}_{t,Y(t-k_n)}\big)\Big\}. \end{aligned} \tag{A.13}$$

In equation (A.13), we can see that $v_{n,k_n,t}(i_{n,k_n,t}, R_{n,t}, P_t, \bar{P}_t)$ determines the action between $R_{n,k_n,t} i_{n,k_n,t}$ or $i_{n,k_n,t}$ depending on the signage of $(\alpha P_{t,Y(t-k_n)} + \bar{P}_{t,Y(t-k_n)} - \delta_0 \mathbb{E}_t[C_{0,k_n+1,t+1}|P_t])$ in (A.13). Since $R_{n,k_n,t}$ is a binary variable, it is straightforward to see that the optimal action $a^*_{n,k_n,t}$ is set to 0 or $i_{n,k_n,t}$. Furthermore, since the optimal action $a^*_{n,k_n,t}$ is $R_{n,k_n,t} i_{n,k_n,t}$ or $i_{n,k_n,t}$, we note that $v_{n,k_n,t}(i_{n,k_n,t}, R_{n,t}, P_t, \bar{P}_t)$ is linear in $i_{n,k_n,t}$. Then, defining $Z_{n,k_n,t}$ as

$$Z_{n,t,k_n} := \max\{\alpha P_{t,Y(t-k_n)}, (1 - R_{n,k_n,t})(\delta_0 \mathbb{E}_t[C_{0,k_n+1,t+1}|P_t] - \bar{P}_{t,Y(t-k_n)})\}, \tag{A.14}$$

we have

$$v_{n,k_n,t}(i_{n,k_n,t}, R_{n,t}, P_t, \bar{P}_t) = Z_{n,k_n,t} \cdot i_{n,k_n,t}.$$

We now verify that the optimal actions in $v_{n,\tau,t}(i_{n,\tau,t}, R_{n,t}, P_t, \bar{P}_t)$ is set to either 0 or $i_{n,\tau,t}$ using induction. Consider some τ' that lies in $1 \le \tau' \le k_n - 1$. Assume that, for $\tau' + 1$, the optimal action $a^*_{n,\tau'+1,t+1}$ is either 0 or $i_{n,\tau'+1,t+1}$ and the value function $v_{n,\tau'+1,t+1}(i_{n,\tau'+1,t+1}, R_{n,t+1}, P_{t+1}, \bar{P}_{t+1})$ is linear in inventory level as $Z_{n,\tau'+1,t+1} \cdot i_{n,\tau'+1,t+1}$ where $Z_{n,\tau'+1,t+1}$ does not depend on $a_{n,\tau'+1,t+1}$ and $i_{n,\tau'+1,t+1}$.

Using equation (A.5), we can express $v_{n,\tau',t}(i_{n,\tau',t}, R_{n,t}, P_t, \bar{P}_t)$ as below:

$$\begin{aligned} &v_{n,\tau',t}(i_{n,\tau',t}, R_{n,t}, P_t, \bar{P}_t) \\ &\quad = \max_{a_{n,\tau',t} \in [R_{n,\tau',t} i_{n,\tau',t}, i_{n,\tau',t}]} \Big\{ \alpha a_{n,\tau',t} P_{t,Y(t-\tau')} + \delta_0 \mathbb{E}_t[Z_{n,\tau'+1,t+1} \cdot (i_{n,\tau',t} - a_{n,\tau',t})|R_{n,t}, P_t, a_{n,\tau',t}] \Big\} \\ &\quad = \max_{a_{n,\tau',t} \in [R_{n,\tau',t} i_{n,\tau',t}, i_{n,\tau',t}]} \Big\{ a_{n,\tau',t}\big(\alpha P_{t,Y(t-\tau')} - \delta_0 \mathbb{E}_t[Z_{n,\tau'+1,t+1}|R_{n,t}, P_t]\big) \\ &\qquad\qquad + i_{n,\tau',t} \delta_0 \mathbb{E}_t[Z_{n,\tau'+1,t+1}|R_{n,t}, P_t] \Big\} \end{aligned} \tag{A.15}$$

From this, we find that $v_{n,\tau',t}(i_{n,\tau',t}, R_{n,t}, P_t, \bar{P}_t)$ determines the optimal action between $R_{n,\tau',t}$ and $i_{n,\tau',t}$. That is, since $R_{n,\tau',t}$ is a binary variable, the optimal action is either 0 or $i_{n,\tau',t}$ for any n, τ and t. Hence, we have shown that there exists an optimal policy in $v_{n,\tau,t}$ where the action $a^*_{n,\tau,t}$ for all n, τ, and t are either 0 or $i_{n,\tau,t}$. Furthermore, from equation (A.15), we obtain that the value function $v_{n,\tau,t}(i_{n,\tau,t}, R_{n,t}, P_t, \bar{P}_t)$ can be expressed as $Z_{n,\tau,t} \cdot i_{n,\tau,t}$, where $Z_{n,\tau,t} = \max\{\alpha P_{t,Y(t-\tau)}, (1 - R_{n,\tau,t})\delta_0 \mathbb{E}_t[Z_{n,\tau+1,t+1}|R_{n,t}, P_t]\}$.

Therefore, there exists an optimal policy such that the actions $a^*_{j,\tau,t}$ for all τ and t are set to either 0 or $i_{j,\tau,t}$ where $j \in \{0\} \cup \{1, \cdots, N\}$. □

Proof of Proposition 1

Under the Immediate Transfer service, the aggregator only manages the transferred RECs. Consequently, the value function can be expressed as $V_t(X_t, S_t) = V_t(X_t, \vec{0}) + \sum_{\tau=0}^{k_0} v_{0,\tau,t}(i_{0,\tau,t}, P_t)$. In the proof of Lemma 2, we establish that, for all t and τ, the optimal value function is given by $v_{0,\tau,t}(i_{0,\tau,t}, P_t) = i_{0,\tau,t} C_{0,\tau,t}$, where $C_{0,\tau,t}$ is defined in (A.11). The optimal action $a^*_{0,\tau,t}$ is determined by the signage of the term $P_t - \delta_0 \mathbb{E}[C_{0,\tau+1,t+1}|P_t]$.

Next, we prove that $C_{0,1,t} \geq C_{0,2,t} \geq \cdots \geq C_{0,k-1,t} \geq C_{n,k_0,t}$ holds for any given t, using induction.

For any t', if $\tau = k_0$, we know that $C_{0,k_0,t'} = P_{t',Y(t'-k_0)}$ from the proof of Lemma 2. If $\tau = k_0 - 1$, then we have

$$\begin{aligned} C_{0,k_0-1,t'} &= \max\{P_{t',Y(t'-k_0+1)}, \delta_0 \mathbb{E}_t[P_{t'+1,Y(t'-k_0+1)}|P_{t'}]\} \geq \max\{P_{t',Y(t'-k_0)}, \delta_0 \mathbb{E}_t[P_{t'+1,Y(t'-k_0+1)}|P_{t'}]\} \\ &\geq \max\{C_{0,k_0,t'}, \delta_0 \mathbb{E}_t[P_{t'+1,Y(t'-k_0+1)}|P_{t'}]\} \geq C_{0,k_0,t'}. \end{aligned} \tag{A.16}$$

Suppose that, for some τ' that lies in $1 \leq \tau' \leq k_0 - 2$ for any t', $C_{0,\tau'+1,t'} \geq C_{0,\tau'+2,t'} \geq \cdots \geq C_{0,k_0,t'}$ holds. From (A.11), we have $C_{0,\tau',t'} = \max\{P_{t',Y(t'-\tau')}, \delta_0 \mathbb{E}[C_{0,\tau'+1,t'+1}|P_{t'}]\}$ and $C_{0,\tau'+1,t'} = \max\{P_{t',Y(t'-\tau'-1)}, \delta_0 \mathbb{E}[C_{0,\tau'+2,t'+1}|P_{t'}]\}$. Since $P_{t',Y(t'-\tau')} \geq P_{t',Y(t'-\tau'-1)}$ and $C_{0,\tau'+1,t'+1} \geq C_{0,\tau'+2,t'+1}$, it follows that $\mathbb{E}[C_{0,\tau'+1,t'+1}|P_{t'}] \geq \mathbb{E}[C_{0,\tau'+2,t'+1}|P_{t'}]$ and $C_{0,\tau',t'} \geq C_{0,\tau'+1,t'}$ holds. Therefore, by induction, we conclude that $C_{0,1,t'} \geq C_{0,2,t'} \geq \cdots \geq C_{0,k_n,t'}$ holds for any given t'.

In the proof of Lemma 2, the coefficient of $a_{0,\tau,t}$ in the value function $v_{0,\tau,t}(i_{0,\tau,t}, P_t)$ in (A.12) is given by $(P_{t,Y(t-\tau)} - \delta_0 \mathbb{E}_t[C_{0,\tau+1,t+1}|P_t])$. Since $C_{0,\tau,t}$ is non-increasing in τ, as established above, the coefficient $(P_{t,Y(t-\tau)} - \delta_0 \mathbb{E}_t[C_{0,\tau+1,t+1}|P_t])$ is non-decreasing in $\tau \in \Omega_{0,t,y}$.

To determine the optimal bidding policy for RECs in $\Omega_{0,t,y}$ under the Immediate Transfer service, we define $\tau^*_{0,t,y}$ as the minimum τ with compliance year y such that this coefficient becomes non-negative. Specifically, we define $\tau^{A*}_{0,t,y}$ as

$$\tau^{A*}_{0,t,y} := \min\{\tau | P_{t,y} - \delta_0 \mathbb{E}_t[C_{0,\tau+1,t+1}|P_t] \geq 0, \tau \in \Omega_{0,t,y}\}$$

. Given that the coefficient is non-decreasing in τ, the value for RECs with compliance year y does not take negative values once it exceeds $\tau^{A*}_{0,t,y}$.

Therefore, for the compliance year y and threshold $\tau^{*}_{0,t,y}$, the optimal action is given as $a^{*}_{0,\tau,t}=0$ if $\tau<\tau^{A*}_{0,t,y}$, and $a^{*}_{0,\tau,t}=i_{n,\tau,t}$ if $\tau\geq\tau^{A*}_{0,t,y}$. □

Proof of Corollary 1

The aggregator does not provide a transfer service under the Spot Sales service; i.e. $v_{0,\tau,t}(i_{0,\tau,t},P_t)=0$ and $\bar{P}_t=\vec{0}$. Hence, we can re-express the value function $v^{B}_{n,\tau,t}(i_{n,\tau,t},R_{n,t},P_t,\bar{P}_t)$ as $v^{B}_{n,\tau,t}(i_{n,\tau,t},R_{n,t},P_t)$ for the Spot Sales service. Further, $v^{B}_{n,k_n+1,t}(i_{n,k_n+1,t},R_{n,t},P_t)=0$.

For $\tau=k_n$, the optimal action $a^{*}_{n,k_n,t}$ is equal to $i_{n,k_n,t}$ regardless of the generator's request $R_{n,k_n,t}$. As a result, the value function $v^{B}_{n,k_n,t}(i_{n,k_n,t},R_{n,t},P_t)$ is given by $\alpha P_{t,Y(t-k_n)}i_{n,k_n,t}$.

Next, we show that there exists a threshold $\tau^{B*}_{n,t}$ in the REC vintage such that the optimal action $a^{*}_{n,\tau,t}=0$ if $\tau<\tau^{B*}_{n,t}$, and $a^{*}_{n,\tau,t}=i_{n,\tau,t}$ otherwise. Using equation (A.4), the value function $v^{B}_{n,k_n-1,t}(i_{n,k_n-1,t},R_{n,t},P_t)$ can be expressed as follows:

$$\begin{aligned}
&v^{B}_{n,k_n-1,t}(i_{n,k_n-1,t},R_{n,t},P_t)\\
&\quad=\max_{a_{n,k_n-1,t}\in\{R_{n,k_n-1,t}i_{n,k_n-1,t},i_{n,k_n-1,t}\}}\Big\{\alpha a_{n,k_n-1,t}(p_{t,Y(t-k_n+1)}-\delta_0\mathbb{E}_t[P_{t+1,Y(t-k_n+1)}|P_t])\\
&\qquad\qquad+\alpha i_{n,k_n-1,t}\delta_0\mathbb{E}_t[P_{t+1,Y(t-k_n+1)}|P_t]\Big\} \qquad \text{(A.17)}
\end{aligned}$$

Since the optimal action is either 0 or $i_{n,k_n-1,t}$ as shown in Lemma 2, we note that the value function (A.17) is linear in the inventory $i_{n,k_n-1,t}$ regardless of the action. Consequently, it follows that $v^{B}_{n,k_n-1,t}(i_{n,k_n-1,t},R_{n,t},P_t)=\alpha C_{n,k_n-1,t}\cdot i_{n,k_n-1,t}$, where

$$C_{n,k_n-1,t}:=\max\{P_{t,Y(t-k_n+1)},(1-R_{n,k_n-1,t})\delta_0\mathbb{E}_t[P_{t+1,Y(t-k_n+1)}|P_t]\}, \qquad \text{(A.18)}$$

and

$$C_{n,k_n,t}:=P_{t,Y(t-k_n)}. \qquad \text{(A.19)}$$

We now verify that $v^{B}_{n,\tau,t}(i_{n,\tau,t},R_{n,\tau},P_t)=\alpha C_{n,\tau,t}i_{n,\tau,t}$ for all τ under the Spot Sales service using induction. We consider τ' that lies in $1\leq\tau'\leq k_n-2$ such that the value function $v^{B}_{n,\tau'+1,t+1}(i_{n,\tau'+1,t+1},R_{n,t+1},P_{t+1})$ is $\alpha C_{n,\tau'+1,t+1}\cdot i_{n,\tau+1,t+1}$. By using equations (A.18) and (A.19), we denote $C_{n,\tau'+1,t+1}$ in a recursive form as

$$C_{n,\tau'+1,t+1}=\max\{P_{t+1,Y(t-\tau')},(1-R_{n,\tau'+1,t+1})\delta_0\mathbb{E}_t[C_{n,\tau'+2,t+2}|R_{n,t+1},P_{t+1}]\}. \qquad \text{(A.20)}$$

From the above and equation (A.4), the value function $v^B_{n,\tau',t}(i_{n,\tau',t}, R_{n,\tau'}, P_t)$ can be expressed as follows:

$$\begin{aligned} &v^B_{n,\tau',t}(i_{n,\tau',t}, R_{n,\tau'}, P_t) \\ &\quad = \max_{a_{n,\tau',t}\in\{R_{n,\tau',t} i_{n,\tau',t}, i_{n,\tau',t}\}} \Big\{ \alpha a_{n,\tau',t}(P_{t,Y(t-\tau')} - \delta_0 \mathbb{E}_t[C_{n,\tau'+1,t+1}|R_{n,t}, P_t]) \\ &\qquad\qquad + \alpha i_{n,\tau',t}\delta_0 \mathbb{E}_t[C_{n,\tau'+1,t+1}|R_{n,t}, P_t] \Big\} = \alpha C_{n,\tau',t} \cdot i_{n,\tau',t}, \end{aligned} \tag{A.21}$$

where

$$C_{n,\tau',t} = \max\{P_{t,Y(t-\tau')}, (1 - R_{n,\tau',t})\delta_0 \mathbb{E}_t[C_{n,\tau'+1,t+1}|R_{n,t}, P_t]\}. \tag{A.22}$$

Hence, we have verified that $v^B_{n,\tau,t}(i_{n,\tau,t}, R_{n,\tau}, P_t)$ is $\alpha C_{n,\tau,t} i_{n,\tau,t}$ for all τ under the Spot Sales service.

Finally, we show that $C_{n,1,t} \geq C_{n,2,t} \geq \cdots \geq C_{n,k_n,t}$. Since $C_{n,k_n,t} = P_{t,Y(t-\tau)}$ in Spot Sales service, we know that $C_{n,k_n-1,t} \geq P_{t,Y(t-k_n+1)} \geq P_{t,Y(t-k_n)} = C_{n,k_n,t}$ holds for all n and t. Since $R_{n,\tau,t} \leq R_{n,\tau+1,t}$ and $P_{t,Y(t-\tau)} \geq P_{t,Y(t-\tau+1)}$, one can show that $C_{n,\tau,t}$ is monotone decreasing in τ, i.e., $C_{n,1,t} \geq C_{n,2,t} \geq \cdots \geq C_{n,k_n,t}$, by induction as shown in the proof of Proposition 1.

Using the above monotonicity of $C_{n,\tau,t}$, we know that the coefficient $(P_{t,Y(t-\tau)} - (1 - R_{n,\tau,t})\delta_0 \mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t}, P_t])$ in (A.21) is non-decreasing in $\tau \in \Omega_{n,t,y}$. Then, we define $\tau^{B*}_{n,\tau,t}$ as

$$\tau^{B*}_{n,\tau,t} := \min\{\tau | P_{t,Y(t-\tau)} - (1 - R_{n,\tau,t})\delta_0 \mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t}, P_t] \geq 0, \tau \in \Omega_{n,t,y}\}.$$

Since the coefficient is non-decreasing in τ, the optimal action is $a^{B*}_{n,\tau,t} = 0$ if $\tau < \tau^{B*}_{n,t,y}$ and $a^{B*}_{n,\tau,t} = i_{n,\tau,t}$ otherwise.

Further, since the generator n can request the sales of RECs $\mathbb{P}(R_{n,\tau,t} = 1) \geq 0$ for any τ and k_n is less than or equal to k_0, it follows that $C_{n,\tau,t} \leq C_{0,\tau,t}$ for all t and τ. Given that $C_{n,\tau,t}$ is non-negative, it reaches its maximum when $R_{n,\tau,t} = 0$ for all τ and t. Additionally, in the case where k_n approaches k_0, then $C_{n,\tau,t}$ approaches $C_{0,\tau,t}$. Therefore, we conclude that $C_{n,\tau,t} \leq C_{0,\tau,t}$ holds for all t and τ, which implies that $\tau^{B*}_{n,t,y} \leq \tau^{A*}_{0,t,y}$ holds for all t and y. □

Proof of Proposition 2

Under two scenarios s and s', where $\mathbb{P}(R_{n(s),\tau,t'} = 1) \geq \mathbb{P}(R_{n(s'),\tau,t'} = 1)$ for all $t' \in \{t+1, \ldots, T\}$ and τ, we denote $C_{n,\tau,t}$ for each scenario s and s' as $C_{n(s),\tau,t}$ and $C_{n(s'),\tau,t}$ respectively.

Using equations (A.18), (A.19) and the fact that $R_{n(s),t} = R_{n(s'),t}$, it follows that $C_{n(s),\tau,t}$ and $C_{n(s'),\tau,t}$ are identical when τ is $k_n - 1$ or k_n. Suppose that, for some τ' that lies in $1 \leq \tau \leq k_n - 2$ and $t' \in \{t+1, \cdots, T\}$, we have $\mathbb{E}_{t'}[C_{n(s),\tau'+1,t'+1}|R_{n,t'}, P_{t'}] \leq \mathbb{E}_{t'}[C_{n(s'),\tau+1,t'+1}|R_{n,t'}, P_{t'}]$. Then, using equation (A.22), $\mathbb{E}_t[C_{n,\tau',t+1}|R_{n,t}, P_t]$ can be expressed as below:

$$\begin{aligned} &\mathbb{E}_t[C_{n,\tau',t+1}|R_{n,t}, P_t] \\ &= \mathbb{E}_t[\max\{P_{t+1,Y(t+1-\tau')}, (1 - R_{n,\tau',t+1})\delta_0 \mathbb{E}_{t+1}[C_{n,\tau'+1,t+2}|R_{n,t+1}, P_{t+1}]\}|R_{n,t}, P_t]. \end{aligned} \tag{A.23}$$

Since $P_{t+1,Y(t+1-\tau')}$ is identical under both scenarios and $\mathbb{E}_{t'}[C_{n(s),\tau'+1,t'+2}|R_{n,t'+1},P_{t'+1}] \leq \mathbb{E}_{t'}[C_{n(s'),\tau+1,t'+2}|R_{n,t'+1},P_{t'+1}]$ in equation (A.23), the value of $\mathbb{E}_t[C_{n,\tau',t+1}|P_t,R_{n,t}]$ is determined by $R_{n,\tau,t+1}$. Because $\mathbb{P}(R_{n(s),\tau,t'}=1) \geq \mathbb{P}(R_{n(s'),\tau,t'}=1)$ for all τ and $t' \in \{t+1,\ldots,T\}$, it follows that $\mathbb{E}_{t'}[C_{n(s),\tau',t'+1}|R_{n,t'},P_{t'}] \leq \mathbb{E}_{t'}[C_{n(s'),\tau',t'+1,}|R_{n,t'},P_{t'}]$ for all $t' \in \{t+1,\ldots,T\}$ and τ.

For the Spot Sales service, the threshold $\tau^{B*}_{n,t,y}$ is determined as:

$$\min\{\tau|P_{t,Y(t-\tau)} - (1-R_{n,\tau,t})\delta_0\mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t},P_t] \geq 0, \tau \in \Omega_{n,t,y}\}.$$

Hence, $\tau^{B*}_{n(s),t,y} \leq \tau^{B*}_{n(s'),t,y}$ holds for two scenarios s and s' where $\mathbb{P}(R_{n(s),\tau,t'}=1) \geq \mathbb{P}(R_{n(s'),\tau,t'}=1)$ for all τ and $t' \in \{t+1,\ldots,T\}$. □

Proof of Proposition 3

Recall from equation (A.15) in the proof Lemma 2 that the value function $v^C_{n,\tau,t}(i_{n,\tau,t},R_{n,t},P_t,\bar{P}_t)$ can be expressed as follows:

$$\begin{aligned} &v^C_{n,\tau,t}(i_{n,\tau,t},R_{n,\tau,t},P_t,\bar{P}_t) \\ &= \max_{a_{n,\tau,t}\in\{R_{n,\tau,t}i_{n,\tau,t},i_{n,\tau,t}\}} \{a_{n,\tau,t}\alpha P_{t,Y(t-\tau)} + (i_{n,\tau,t}-a_{n,\tau,t})\delta_0\mathbb{E}_t[Z_{n,\tau+1,t+1}|R_{n,t},P_t,\bar{P}_t]\}. \end{aligned} \tag{A.24}$$

Since the optimal action is either $R_{n,\tau,t}i_{n,\tau,t}$ or $i_{n,\tau,t}$, it follows that $v^C_{n,\tau,t}(i_{n,\tau,t},R_{n,\tau,t},P_t,\bar{P}_t)$ is given by

$$\max\{\alpha P_{t,Y(t-\tau)},(1-R_{n,\tau,t})\delta_0\mathbb{E}_t[Z_{n,\tau+1,t+1}|R_{n,t},P_t,\bar{P}_t]\}\cdot i_{n,\tau,t} = Z_{n,\tau,t}\cdot i_{n,\tau,t}.$$

Thus, the value function $v^C_{n,\tau,t}(i_{n,\tau,t},R_{n,\tau,t},P_t,\bar{P}_t)$ for the RECs in the pre-transfer phase can be written as $Z_{n,\tau,t}\cdot i_{n,\tau,t}$ for all n, τ, and t. Here, $Z_{n,\tau,t}$ includes both the immediate sale profit $\alpha P_{t,Y(t-\tau)}$ and the expected future profit from holding the RECs, given by $(1-R_{n,\tau,t})\delta_0\mathbb{E}_t[Z_{n,\tau'+1,t+1}|R_{n,t},P_t,\bar{P}_t]$.

Because $Z_{n,\tau,t}$ and $C_{n,\tau,t}$ have same recursive form and $Z_{n,k_n,t}$ is greater or equal to $\alpha C_{n,k_n,t}$ in equation (A.14), it is straightforward to show that $\delta_0\mathbb{E}_t[Z_{n,\tau+1,t+1}|R_{n,t},P_t,\bar{P}_t] \geq \delta_0\alpha\mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t},P_t]$ holds. Therefore, the optimal action is to hold if $P_{t,Y(t-\tau)} < (1-R_{n,t})\delta_0\mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t},P_t]$. Further, since $C_{n,\tau,t}$ is non-increasing in τ, as shown in Proposition 1, we define the threshold $\tau^{C*}_{n,t,y}$ for the compliance year y as follows:

$$\tau^{C*}_{n,t,y} := \min\{\tau|P_{t,y} - (1-R_{n,\tau,t})\delta_0\mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t},P_t] \geq 0, \tau \in \Omega_{n,t,y}\}.$$

Since $P_{t,Y(t-\tau)} - (1-R_{n,t})\delta_0\mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t},P_t]$ is non-decreasing in $\tau \in \Omega_{n,t,y}$, the optimal action $a^*_{n,\tau,t}$ is 0 for $\tau < \tau^{C*}_{n,t,y}$ and $\tau \in \Omega_{n,t,y}$.

Now, we show the optimal action of the RECs with vintage τ such that $P_{t,Y(t-\tau)} \geq (1 - R_{n,t})\delta_0 \mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t}, P_t]$. From equation (A.24), we obtain that the aggregator's optimal action is to sell, $a^*_{n,\tau,t} = i_{n,\tau,t}$, if and only if

$$\alpha P_{t,Y(t-\tau)} \geq (1 - R_{n,\tau,t})\delta_0 \mathbb{E}[Z_{n,\tau+1,t+1}|R_{n,t}, P_t, \bar{P}_t].$$

Otherwise, the optimal action is to hold, $a^*_{n,\tau,t} = 0$.

In addition, we show that $Z_{n,\tau,t}$ is not monotone in τ. We show this by contradiction. First, assume $Z_{n,\tau,t}$ is non-decreasing in τ, which implies that $Z_{n,k_n-1,t} \leq Z_{n,k_n,t}$ must hold. If the transfer price $\bar{P}_{t,Y(t-k_n)}$ is high such that the aggregator's expected profit after the transfer is negative ($\bar{P}_{t,Y(t-k_n)} > \delta_0 \mathbb{E}_t[C_{0,k_n+1,t+1}|P_t]$), then $Z_{n,k_n,t}$ is equal to $\alpha p_{t,Y(t-k_n)} = \alpha C_{n,k_n,t}$. Since $Z_{n,k_n-1,t} \geq \alpha C_{n,k_n-1,t} \geq \alpha C_{n,k_n,t}$, we find that $Z_{n,k_n-1,t} \leq Z_{n,k_n,t}$ does not hold if there exists a situation where $\delta_0 \mathbb{E}_t[Z_{n,k_n,t+1}|R_{n,t}, P_t] > \alpha p_{t,Y(t-k_n)}$. This can be occur when $\delta_0 \mathbb{E}_t[P_{t+1,Y(t-k_n+1)}|P_t] > P_{t,Y(t-k_n)}$, which contradicts the above assumption. Therefore, $Z_{n,\tau,t}$ is not non-decreasing in τ.

Second, we assume $Z_{n,\tau,t}$ is non-increasing in τ and similarly construct a counterexample. Consider a case where the RECs with vintage $k_n - 1$ and k_n have the same energy year (i.e., $k_n - 1, k_n \in \Omega_{n,t,y}$), the market price at time t reaches the SACP. Further, the transfer price $\bar{P}_{t,y(t-k_n)}$ is less than $\delta_0 E[C_{0,k_n+1,t+1}|R_{n,t}, P_t] - \alpha P_{t,y(t-k_n)}$, while $\bar{P}_{t+1,y(t-k_n+1)}$ is high enough to make the transfer profitless. In this case, the aggregator expects to earn more from the transfer than from the immediate sales of RECs with vintage k_n, i.e., $\alpha P_{t,Y(t-k_n)} < \delta_0 \mathbb{E}_t[C_{0,k_n+1,t+1}|R_{n,t}, P_t] - \bar{P}_{t,Y(t-k_n)}$. Conversely, for RECs with vintage $k_n - 1$, the immediate sales profit is the largest, i.e., $Z_{n,k_n-1,t} = \alpha P_{t,Y(t-k_n+1)}$. Hence, we see that $Z_{n,k_n-1,t} \geq Z_{n,k_n,t}$ does not hold in this case. Hence, $Z_{n,\tau,t}$ is not monotone in τ. All in all, we conclude that $Z_{n,\tau,t}$ is not monotone in τ. □

Proof of Corollary 2

As shown in Lemma 2, the optimal actions under each service are determined by comparing the immediate sales profit with the expected future profit. Specifically, under Spot Sales service, the optimal action $a^{B*}_{n,\tau,t}$ is $i_{n,\tau,t}$ if $P_{t,Y(t-\tau)} \geq (1 - R_{n,\tau,t})\mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t}, P_t]$. Otherwise, $a^{B*}_{n,\tau,t}$ is 0. Similarly, in the pre-transfer phase of the Hybrid Agreement service, the optimal action for generator n's RECs, $a^{C*}_{n,\tau,t}$, is $i_{n,\tau,t}$ if $\alpha p_{t,Y(t-\tau)} \geq (1 - R_{n,\tau,t})\mathbb{E}_t[Z_{n,\tau+1,n+1}|R_{n,t}, P_t, \bar{P}_t]$.

In the proof of Proposition 3, we have shown that $\alpha C_{n,\tau,t} \leq Z_{n,\tau,t}$ under the same contract conditions. This implies that $\alpha(1 - R_{n,\tau,t})\mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t}, P_t] \leq (1 - R_{n,\tau,t})\mathbb{E}_t[Z_{n,\tau+1,t+1}|R_{n,t}, P_t, \bar{P}_t]$. Consequently, if the optimal action for RECs with vintage τ is to hold ($a^{B*}_{n,\tau,t} = 0$) in the pre-transfer phase of the Hybrid Agreement service, then the optimal action is also to hold ($a^{B*}_{n,\tau,t} = 0$) in the Spot Sales service under the same conditions. □

Proof of Corollary 3

As shown in the proof of Proposition 1, the threshold for the Hybrid Agreement service $\tau^{C*}_{n,t,y}$ is determined as

$$\min\{\tau | P_{t,Y(t-\tau)} - (1-R_{n,\tau,t})\delta_0 \mathbb{E}_t[C_{n,\tau+1,t+1}|R_{n,t}, P_t] \geq 0, \tau \in \Omega_{n,t,y}\}.$$

Furthermore, in Proposition 2, we have already shown that the value of $\mathbb{E}_{t'}[C_{n,\tau,t+1}|R_{n,t}, P_t]$ is determined by $R_{n,\tau,t+1}$ under both scenarios. Given that the probability $\mathbb{P}(R_{n(s),\tau,t'}=1) \geq \mathbb{P}(R_{n(s'),\tau,t'}=1)$ for all τ and $t' \in \{t+1,\ldots,T\}$ under the scenarios s and s', it follows that

$$\mathbb{E}_{t'}[C_{n(s),\tau,t'+1}|R_{n,t'}, P_{t'}] \leq \mathbb{E}_{t'}[C_{n(s'),\tau,t'+1}|R_{n,t'}, P_{t'}]$$

for all τ and $t' \in \{t+1,\ldots,T\}$.

Thus, we conclude that $\tau^{C*}_{n(s),t,y} \leq \tau^{C*}_{n(s'),t,y}$ under the two scenarios s and s', where $\mathbb{P}(R_{n(s),\tau,t'}=1) \geq \mathbb{P}(R_{n(s'),\tau,t'}=1)$ holds for all τ and $t' \in \{t+1,\ldots,T\}$. □

Proof of Corollary 4

For the post-transfer phase in the Hybrid Agreement service, the value functions (A.2) and (A.3) are identical to the value function $v^A_{0,\tau,t}(i_{0,\tau,t}, P_t)$ for the Immediate Transfer service. Therefore, the value function for the post-transfer phase in the Hybrid Agreement service can also be expressed as $i_{0,\tau,t} \cdot C_{0,\tau,t}$, where $C_{0,\tau,t} = \max\{P_{t,Y(t-\tau)}, \delta_0 \mathbb{E}_t[P_{t+1,Y(t-\tau)}|P_t]\}$.

We define the threshold $\tau^{C*}_{0,t,y}$ as

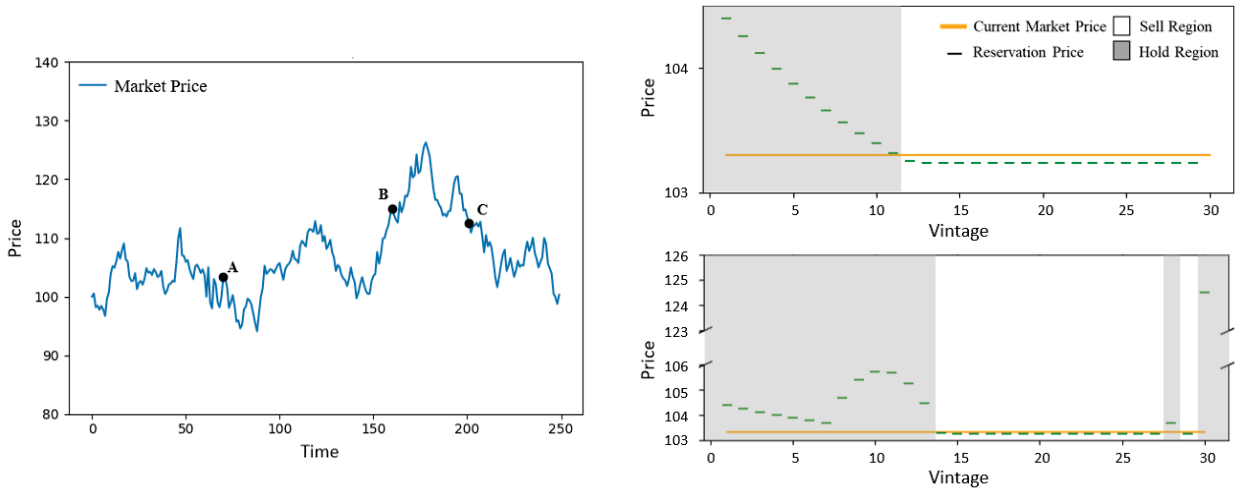
$$\tau^{C*}_{0,t,y} := \min\{\tau | P_{t,y} - \delta_0 \mathbb{E}_t[C_{0,\tau+1,t+1}|P_t] \geq 0, \tau \in \Omega_{0,t,y}\}.$$

Since $C_{0,\tau,t}$ is non-increasing in τ, we observe that for the compliance year y, there exists the threshold $\tau^{C*}_{0,t,y}$. The optimal action is $a^{C*}_{0,\tau,t} = 0$ if $\tau < \tau^{C*}_{0,t,y}$ and $a^{C*}_{0,\tau,t} = i_{0,\tau,t}$ otherwise.

B. Illustration of the Aggregator's Optimal Policy

Figure B.1 provides an example illustrating the aggregator's optimal policy. Figure B.1(a) shows a sample path of REC market prices, while Figure B.1(b) compares the aggregator's optimal actions with and without the REC transfer option.

The aggregator uses current and historical market data to forecast price trends. During an upward price trend, as at point B, the aggregator tends to hold RECs expecting higher future profits. Conversely, in a downward trend, as seen at point C, the aggregator opts for immediate REC sales. Under highly volatile conditions, such as at point A, the transfer price becomes unstable,



(a) Sample Path of an REC Market Price (following a Geometric Brownian Motion with $\mu = 0$ and $\sigma = 0.25$)

(b) Optimal Policy at Point A: (**Top**) Spot Sales Service; (**Bottom**) Hybrid Agreement Service (δ=0.9997, k=50, k_n=30, $\bar{p}_t = \frac{1}{10}\sum_{\Delta=0}^{9} p_{t-\Delta}$, α=0.02, $R_{n,\tau,t} = 0$)

Figure B.1 Sample Path of an REC Market Price and the Aggregator's Optimal Policies at Point A

resulting in fluctuations in the expected profits from exercising the transfer option. This leads to a mixed strategies of conditional sell, as depicted in the bottom of Figure B.1(b). The figure also demonstrates that the optimal policy under the Hybrid Agreement service has a larger hold region than that of the Spot Sales service. As discussed in Corollary 2, the transfer option in the Hybrid Agreement enhances the value of the RECs, increasing their reservation price.

C. Additional Information on the Numerical Experiment

C.1. Details of Simulation Experiment

The training and testing periods for our simulation experiment are presented in Table C.1.

Table C.1 Train, Test, and REC Issuance Period in each Market

NJ SREC Market	Train	Test
Simulation Period	Jun. 2013 – May. 2018	Jun. 2018 – May. 2023
REC Issuance Period	Jun. 2013 – May. 2015	Jun. 2018 – May. 2020
DC SREC market	Train	Test
Simulation Period	Jan. 2013 – Dec. 2017	Jan. 2018 – Dec. 2022
REC Issuance Period	Jan. 2013 – Dec. 2014	Jan. 2018 – Dec. 2019

Hyper-parameters for our proposed DRL algorithm are given in Table C.2.

Table C.2 Summary of Hyper-parameters for the DRL Algorithm

Hyper-parameter	Value
Number of Layers	4
Dimension of Hidden Layer	256
Activation Function	ELU with slope 0.2*
Learning Rate of the network for RECs in pre- and post-transfer phases	$10^{-4}, 10^{-3}$
Initial Rate for Soft Target Update ($\hat{\epsilon}_t$, ϵ_t)	0.01
Rate Decay for Soft Target Update	0.001
Initial Rate for Exploration ($\hat{\varepsilon}_t$, ε_t)	1.0
Rate Decay for Exploration	0.0001, 0.0010
Number of Episodes	150

*Note: ELU represents Exponential Linear Unit. See Clevert et al. (2016) that introduces the ELU activation function to accelerate the learning while alleviating the vanishing gradient issues.

C.2. Understanding the DRL Algorithm

We utilize two notable techniques from eXplainable Reinforcement Learning (XRL) to understand why our customized DRL algorithm outperforms the standard DRL algorithm. First, we conduct a Feature Importance (FI) analysis to highlight the significance of the structural properties incorporated in our proposed DRL algorithm. This helps us identify the most useful input features for making decisions in the algorithm. In addition, we conduct a correlation analysis between the policies generated by our algorithm and those derived under full information.

The FI analysis assigns a score to each input feature in the value function approximation model of the DRL algorithm. These scores indicate the 'importance' of each feature, with larger scores indicating a stronger contribution to approximating the value function in the model. Since the value-based DRL algorithm makes decisions based on action values at a given state, this analysis helps us understand the influence of each feature on the algorithm's decision-making process. We utilize the widely used SHapley Additive exPlanations (SHAP) method for this analysis. We assign a contribution value to each feature in the model, referred to as the SHAP values, indicating its impact on the model output based on the marginal contribution of each feature when considering all possible feature combinations. For more detail on the SHAP method, please see Lundberg and Lee (2017) and Lundberg et al. (2020).

Figure C.2 illustrates the FI analysis results for the NJ SREC markets, showcasing the five most important features used in the decision making. The analysis involves two GNN models: our proposed algorithm, *DDDQN with GNN*(Q'), and the baseline algorithm, *DDDQN with GNN*(Q), presented in Figures C.2(a) and C.2(b) respectively. The FI analysis results in the DC SREC market exhibit similar trends to those in the NJ SREC market, as depicted in Figure C.3.

The proposed DRL algorithm employs two distinct GNN models for value function approximation depending on the status of REC transfer: REC pre- and post-transfer phases. In each phase, the

(a) FI analysis on *DDDQN w/ GNN*(Q')

(b) FI analysis on *DDDQN w/ GNN*(Q)

Figure C.2 Feature importance (FI) analysis results for NJ market

features are ranked based on their average SHAP values, with the most important features listed at the top. Each data point represents the SHAP value of a feature for a specific REC on a particular day. Thus, features with a wider spread of data points generally indicate a higher level of importance in the decision-making process. The color of each data point corresponds to the input feature value, with red representing high and blue representing low, which allows us to discern the relationship between the SHAP value and its corresponding input feature value. For example, in the REC post-transfer phase under both GNN models, the market price stands out as the most important feature with the widest spread. This indicates that the market price plays the most critical role in calculating the profit difference of the RECs between the *hold* and *sell* actions. This is because the market price directly determines the value of the *sell* action, and the GNN models estimate the relative value of the *hold* action. Therefore, we observe that a data point colored red (blue) in the market price feature, which indicates a higher (lower) market price, corresponds to a negative (positive) SHAP value in the REC post-transfer phase.

(a) FI analysis on *DDDQN w/ GNN*(Q')

(b) FI analysis on *DDDQN w/ GNN*(Q)

Figure C.3 Feature importance (FI) analysis results for DC market

The FI analysis results in Figure C.2 highlight the significant role of the structural properties employed in our proposed DRL algorithm in shaping the optimal policy. In particular, the notable distinction in the important feature composition between pre- and post-transfer phases indicates a clear difference in the decision-making process between the two phases. This validates our decision to utilize separate GNN models based on the REC transfer status, as illustrated in Figure 2(a), as an effective approach to designing the algorithm. In addition, our analysis reveals that the 'vintage' of RECs is the second most important feature in the REC post-transfer phase in our proposed model. This can be attributed to the monotone threshold policy identified in §3.2. Moreover, we observe a clear correlation between the color gradient scheme in the vintage feature and the vintage-based monotonicity property outlined in (10). Specifically, older RECs with smaller vintage feature values ($\frac{k_0-\tau}{k_0}$) have lower expected future profit, resulting in lower SHAP values. In Figure C.2, we can clearly observe this relationship, where blue data points in the vintage feature are more likely to have negative SHAP values, while red data points are more likely to have positive SHAP values. In contrast to Figure C.2(a), the baseline GNN model in Figure C.2(b) does not attribute significant

importance to vintage, which suggests that the baseline algorithm does not adhere to the optimal policy characterized by the structural property based on the vintage of RECs. This highlights the difficulty for a DRL algorithm to autonomously identify relevant structural properties in complex problems without explicitly incorporating them, as we have done in our approach.

Next, we discuss the correlation analysis using the approach proposed by Guan and Liu (2021), which was used to validate the effectiveness of DRL in the domain of financial portfolio management. This approach examines the correlation between the policy of the DRL algorithm and the optimal policy obtained under full information on future market conditions. A high correlation indicates that the DRL algorithm closely aligns with the true optimal policy. In our study, we similarly examine the correlation between the action values in the optimal policy under full information and those derived from three DRL algorithms: our proposed DRL algorithm, *DDDQN with GNN*(Q'), and two baseline DRL algorithms, the standard *DDDQN* and *DDDQN with GNN*(Q).

Table C.3 summarizes the results of correlation analysis on DRL algorithms in NJ and DC SREC markets. The trained algorithms, which were based on the data from the training period described in Section 5.1, are used to derive the action values during both the training and test periods. We observe that the degree of correlation increases progressively as the algorithms become more sophisticated. Among the algorithms with GNN, the correlations of our proposed algorithm are higher than the other (mostly higher than 0.7), indicating that its actions align closely with those derived from the optimal policy with full information on the market prices.

Table C.3 Correlation Analysis on DRL algorithms

		NJ Market	DC Market
Train period	*DDDQN w/ GNN*(Q)	0.4877	0.6610
	DDDQN w/ GNN(Q')	0.6539	0.8647
Test period	*DDDQN w/ GNN*(Q)	0.2020	0.6048
	DDDQN w/ GNN(Q')	0.8639	0.7041

References

Clevert, D.-A., Unterthiner, T., Hochreiter, S. (2016). Fast and accurate deep network learning by exponential linear units (ELUs) *International Conference on Learning Representations (ICLR 2016)*.

Lundberg, S., Erion, G., Hugh, C., DeGrave, A., Prutkin, J., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2:56–67.

Lundberg, S. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*.