Online Appendix to "When Gray Markets Have Silver Linings: All-Unit Discounts, Gray Markets and Channel Management". PROOFS.

Proof of Proposition 1. To satisfy the demand $d \geqslant 0$, the order size is $q \geqslant d$ and the diversion size is $q - d \geqslant 0$. If $d \geqslant \eta$, then $q \geqslant d \geqslant \eta$; the order cost $c(q,d) = w_\eta q - s(q-d)$ that is linear in q with a slope $w_\eta - s > 0$, thus it is minimized at $q^*(d) = d$. If $d < \eta$, the order cost is $c(q,d) = w_o q - s(q-d)$ when $d \leqslant q < \eta$ and $c(q,d) = w_\eta q - s(q-d)$ when $q \geqslant \eta$. On the above two regions the cost-minimizing solutions are respectively $q = d$ and $q = \eta$. To find the optimum $c^*(d)$ when $d < \eta$, it suffices to compare the cost at the two solutions, i.e., $c^*(d) = \min\{w_o d, (w_\eta - s)\eta + sd\}$: if $0 \leqslant d \leqslant \hat{q}$, $c^*(d) = w_o d$ and if $\hat{q} < d < \eta$, $c^*(d) = (w_\eta - s)\eta + sd$. ▫

Proof of the claim that the all-unit discount generates a unique incentive for diversion. First, consider an incremental discount: $C(q) = w_o q$ if $0 \leqslant q < \eta$ and $C(q) = w_\eta(q-\eta) + w_o\eta$ if $q \geqslant \eta$. To satisfy the demand $d \geqslant 0$, the order size is $q \geqslant d$ and the diversion size is $q - d \geqslant 0$. If $d \geqslant \eta$, then $q \geqslant d \geqslant \eta$; the order cost $c(q,d) = w_\eta(q-\eta) + w_o\eta - s(q-d)$ that is linear in q with a slope $w_\eta - s > 0$, thus it is minimized at $q^*(d) = d$. If $d < \eta$, the order cost is $c(q,d) = w_o q - s(q-d)$ if $d \leqslant q < \eta$ and $w_\eta(q-\eta) + w_o\eta - s(q-d)$ if $q \geqslant \eta$. On the above two regions the cost-minimizing solutions are respectively $q = d$ and $q = \eta$. To find the optimum $c^*(d)$ when $d < \eta$, it suffices to compare the cost at the two solutions, i.e., $c^*(d) = \min\{w_o d, w_o\eta - s(\eta - d)\} = w_o d$. For both cases of $d \geqslant \eta$ and $d < \eta$, the optimal order size is $q^*(d) = d$.

Second, consider a two-part tariff: $C(q) = F + w_o q$, $q > 0$. To satisfy the demand $d \geqslant 0$, the order size is $q \geqslant d$ and the diversion size is $q - d \geqslant 0$. The order cost is $c(q,d) = F + w_o q - s(q-d)$ for $q \geqslant d$ and is linear in q with a slope $w_o - s > 0$. Thus it is minimized at $q^*(d) = d$. ▫

Proof of Lemma 1. At any time t the reseller can change his inventory position from the current position $I(t)$ to a new position $I(t) + \Delta I(t)$ by a combination of order $q(t)$ from the supplier and gray market diversion $g(t)$. Fix any time t. We first argue that since replenishment is instantaneous, it is suboptimal for $\Delta I(t) = i > 0$ if $I(t) > 0$. The action of changing the current inventory position $I(t)$ to $I(t) + i$ could be profitably delayed to the time $t_0 \equiv \inf\{x > t : I(x) = 0\}$. Letting $\Delta I(t) = 0$ and $\Delta I(t_0) = \Delta I(t_0) + i$ has an improvement $h(t_0 - t)i > 0$ in the holding costs up to time t_0. Recursively applying this process results in an improved set of orders where $\Delta I(t) > 0$ only when $I(t) = 0$. We now argue that it is suboptimal for disposal of goods $\Delta I(t) = j < 0$ if $I(t) > 0$ since this action could have been profitably performed at an earlier time $t_{-1} \equiv \sup\{x < t : I(x) = 0\}$, which has a holding cost improvement $h(t - t_{-1}) \mid j \mid > 0$. Therefore, in an efficient inventory policy any ordering or gray market diversion occurs only at times when $I(t) = 0$. The set of times when $I(t) = 0$ represents a set of renewal points. Since the demand rate is stationary, the optimal action is identical at each of these times. To complete the proof let I be equal to the optimal inventory adjustment when $I(t) = 0$ and then $q^*(I)$ and $g^*(I)$ correspond to the optimal order and gray market diversion quantities respectively. ▫

Proof of Proposition 2. To solve for the optimal inventory policy, the reseller selects the cycle inventory level I that minimizes the total costs. Given the optimal zero-inventory policy characterized by Lemma 1, the total costs for each order cycle of length I/λ consist of order cost $c^*(I)$ given by Proposition 1 and holding cost $hI^2/(2\lambda)$. We can then calculate the long-run average cost per unit time, denoted by $g(p, I, s)$ with dependence on p and s suppressed in this proof, as $c^*(I)\lambda/I + hI/2$. By substituting the reseller's optimal cost function given by Proposition 1 (where $d = I$), we obtain the expression for $g(I)$ as follows: $g(I) = w_o m/(p - \gamma p_s)^\alpha + hI/2$ if $0 \leqslant I < \hat{q}$, $g(I) = (w_\eta - s)\eta m/[I(p - \gamma p_s)^\alpha] + sm/(p - \gamma p_s)^\alpha + hI/2$ if $\hat{q} \leqslant I < \eta$ and $g(I) = w_\eta m/(p - \gamma p_s)^\alpha + hI/2$ otherwise.

Recall that in the first and third cases, the reseller orders up to the desired cycle inventory level I and sells the entire order through the authorized channel over time; no goods are diverted to the gray market in these two cases. However, in the second case, the reseller orders up to the quantity of η to enjoy the quantity discount and sells the excess amount $\eta - I$ to the gray market. Within this range ($\hat{q} \leqslant I < \eta$), the reseller will choose a locally optimal cycle inventory level I^o that minimizes the cost $g(I)$, where $I^o \equiv \sqrt{[2(w_\eta - s)\eta m]/[h(p-\gamma p_s)^\alpha]} = \eta\sqrt{[w_\eta - s]/[H(p-\gamma p_s)^\alpha]}$.

The reseller selects the optimal cycle inventory and gray market diversion by comparing the minimum cost $g(I)$ in each of the three regions.

Since the demand and resale price is fixed for the reseller, the reseller's revenue is fixed. The reseller is aiming at minimizing cost $g(I)$. To find the minimum of $g(I)$ we compare the optimal solutions for each region of $I \in [0, \hat{q}]$, $I \in (\hat{q}, \eta)$ and $I \in [\eta, \infty)$. Over the first and third regions, $g(I)$ is a linearly increasing function and is minimized at $I = 0$ and $I = \eta$ respectively. Over the second region $I \in (\hat{q}, \eta)$, $g(I)$ is convex and minimized at an interior point I^o if it is indeed in $(\hat{q}, \eta)$. Otherwise, $g(I)$ is minimized at one of the boundary points $\hat{q}$ or η.

The necessary and sufficient condition for $I^* = I^o$ is $I^o \in (\hat{q}, \eta)$, $g(I^o) < g(0)$ and $g(I^o) < g(\eta)$. The feasibility condition $I^o > \hat{q}$ holds if and only if $H(p-\gamma p_s)^\alpha < (w_o - s)^2/(w_\eta - s)$; The other feasibility condition $I^o < \eta$ holds if and only if $H(p-\gamma p_s)^\alpha > (w_\eta - s)$. The optimality condition $g(I^o) \leqslant g(\eta)$ always holds since $g(I)$ is continuous at $I = \eta$. Finally, the other optimality condition $g(I^o) < g(0)$ holds if and only if $H(p-\gamma p_s)^\alpha < (w_o - s)^2/[4(w_\eta - s)]$. Taking the intersection of regions defined by the feasibility and optimality conditions yields that $I^* = I^o$ if and only if $(w_\eta - s) < H(p-\gamma p_s)^\alpha < (w_o - s)^2/[4(w_\eta - s)]$. Such holding costs exist only if $w_\eta \in (s, (w_o + s)/2]$.

$I^* = \eta$ if and only if $I^* \neq I^o$ and $g(\eta) \leqslant g(0)$. $g(\eta) \leqslant g(0)$ holds if and only if $H(p-\gamma p_s)^\alpha \leqslant w_o - w_\eta$ with equality holding at $H = w_o - w_\eta$. Intersecting $H(p-\gamma p_s)^\alpha \leqslant w_o - w_\eta$ with the region where $I^* \neq I^o$, i.e., $H(p-\gamma p_s)^\alpha \notin (w_\eta - s, (w_o - s)^2/[4(w_\eta - s)])$, results in $H(p-\gamma p_s)^\alpha \leqslant \min\{w_\eta - s, w_o - w_\eta\}$. Note that $w_\eta - s < w_o - w_\eta$ if and only if $w_\eta < (w_o + s)/2$. The necessary and sufficient condition for $I^* = \eta$ follows immediately.

The remaining possible holding cost regions are $H(p-\gamma p_s)^\alpha > (w_o - s)^2/[4(w_\eta - s)]$ if $w_\eta < (w_o + s)/2$, $H(p-\gamma p_s)^\alpha > w_o - w_\eta$ otherwise, which correspond to $I^* = 0$. □

Lemma 2. (Supplier's Profit Function under Exogenous Resale Price). *Given that the reseller employs the optimal inventory policy in response to a discount wholesale price* w_η, *the supplier receives the following profit per unit of time: if* $H < (w_o - s)/[2(p-\gamma p_s)^\alpha]$, *then* $\Pi(w_\eta) = m(w_\eta - c_\eta)\sqrt{H/[(w_\eta - s)(p-\gamma p_s)^\alpha]}$ *when* $s < w_\eta \leqslant s + H(p-\gamma p_s)^\alpha$, $\Pi(w_\eta) = [m(w_\eta - c_\eta)]/[(p-\gamma p_s)^\alpha]$ *when* $s + H(p-\gamma p_s)^\alpha < w_\eta \leqslant w_o - H(p-\gamma p_s)^\alpha$, $\Pi(w_\eta) = [m(w_o - c_o)]/[(p-\gamma p_s)^\alpha]$ *when* $w_o - H(p-\gamma p_s)^\alpha < w_\eta \leqslant w_o$; *if* $H \geqslant (w_o - s)/[2(p-\gamma p_s)^\alpha]$, *then* $\Pi(w_\eta) = m(w_\eta - c_\eta)\sqrt{H/[(w_\eta - s)(p-\gamma p_s)^\alpha]}$ *when* $s < w_\eta \leqslant s + [(w_o - s)^2]/[4H(p-\gamma p_s)^\alpha]$, $\Pi(w_\eta) = m(w_o - c_o)/(p-\gamma p_s)^\alpha$ *when* $s + (w_o - s)^2/[4H(p-\gamma p_s)^\alpha] < w_\eta \leqslant w_o$.

Proof of Lemma 2. The supplier's profit depends on the inventory strategy the reseller optimally selects according to conditions shown in Proposition 2. The supplier's profit per unit of time is equal to the rate the supplier supplies the reseller multiplied by the profit margin per unit. The margin is $w_o - c_o$ per unit if $I^* = 0$ and $w_\eta - c_\eta$ per unit otherwise. If no goods are sold to the gray market (i.e. $I = 0$ or $I = \eta$), the supplier's demand rate is equivalent to the reseller's demand rate λ. However, if $I = I^o$ then the supplier's demand rate is $\eta\lambda/I^o = \lambda\sqrt{H(p-\gamma p_s)^\alpha/(w_\eta - s)}$. The supplier's profit function is therefore $\Pi(w_\eta) = \lambda(w_o - c_o)$ if $I^* = 0$, $\Pi(w_\eta) = \lambda(w_\eta - c_\eta)\sqrt{[H(p-\gamma p_s)^\alpha]/[(w_\eta - s)]}$ if $I^* = I^o$ and $\lambda(w_\eta - c_\eta)$ if $I^* = \eta$. It remains to show that the conditions denoted in the proposition are sufficient to entail the appropriate reseller's best response.

We begin by considering the conditions on the discount which imply $I^* = \eta$. The conditions where $I^* = \eta$ denoted in Proposition 2 are equivalent to $w_\eta \in (s + H(p-\gamma p_s)^\alpha, (w_o + s)/2] \bigcup [(w_o + s)/2, w_o - H(p-\gamma p_s)^\alpha)$. Hence $I^* = \eta$ if and only if $w_\eta \in (s + H(p-\gamma p_s)^\alpha, w_o - H(p-\gamma p_s)^\alpha)$ and such a w_η exists if and only if $w_o - s > 2H(p-\gamma p_s)^\alpha$. Therefore, as stated in the proposition, $\Pi(w_\eta) = \lambda(w_\eta - c_\eta)$ if and only if $w_\eta \in (s + H(p-\gamma p_s)^\alpha, w_o -$

$H(p-\gamma p_s)^\alpha)$ which is non-empty only if $w_o - s > 2H(p-\gamma p_s)^\alpha$.

Similarly the conditions from Proposition 2 which imply $I^* = I^o$ can be summarized as $w_\eta \in (s, s + H(p-\gamma p_s)^\alpha) \bigcap (s, s + (w_o - s)^2/(4H(p-\gamma p_s)^\alpha)) \bigcap (s, (w_o + s)/2)$, which can be simplified by considering whether $w_o - s \leqslant h\eta/\lambda$. If $w_o - s \leqslant 2H(p-\gamma p_s)^\alpha$ then $I^* = I^o \Leftrightarrow w_\eta \in (s, s + (w_o - s)^2/(4H(p-\gamma p_s)^\alpha))$, and if $w_o - s > h\eta/\lambda$ then $I^* = I^o \Leftrightarrow w_\eta \in (s, s + H(p-\gamma p_s)^\alpha)$.

The remaining scenarios of w_η are attributed to when the reseller selects $I^* = 0$. Instantiating the regions corresponding to each of the reseller's inventory policy into $\Pi(w_\eta)$ is sufficient to verify that Lemma 2 holds. ▫

Proof of Proposition 3. Given the profit function in Lemma 2, we solve the problem of optimizing $\Pi(w_\eta)$ over $w_\eta \in (s, w_o]$. For those discount prices w_η which are close enough to w_o and elicit $I^* = 0$, the profit function $\Pi(w_\eta) = \lambda(w_o - c_o)$ remains a constant. It is sufficient to set $w_\eta = w_o$ to generate this profit. Consider $H \geqslant (w_o - s)/[2(p-\gamma p_s)^\alpha]$. By Lemma 2, if $w_\eta < s + (w_o - s)^2/(4H(p-\gamma p_s)^\alpha)$, then $I^* > 0$. Under the assumption that $s < c_\eta$, $\Pi(w_\eta)$ is increasing over $w_\eta \in (s, s + (w_o - s)^2/(4H(p-\gamma p_s)^\alpha)]$, and is maximized at $w_\eta = s + (w_o - s)^2/(4H(p-\gamma p_s)^\alpha)$ generating profit $\lambda[(w_o - s)/2 - 2H(p-\gamma p_s)^\alpha(c_\eta - s)/(w_o - s)]$ for the supplier. Consider $H < (w_o - s)/[2(p-\gamma p_s)^\alpha]$. If $w_\eta \leqslant w_o - H(p-\gamma p_s)^\alpha$, then $I^* > 0$. Again since $s < c_\eta$, it is easy to see that $\Pi(w_\eta)$ is continuous at $w_\eta = s + H(p-\gamma p_s)^\alpha$ and increasing over $w_\eta \in (s, w_o - H(p-\gamma p_s)^\alpha]$, and is therefore maximized at $w_\eta = w_o - H(p-\gamma p_s)^\alpha$ generating profit $\lambda(w_o - H(p-\gamma p_s)^\alpha - c_\eta)$ for the supplier. We can now compare supplier's profits given the reseller's best response of $I^* > 0$ or $I^* = 0$ depending on whether $(w_o - s)/2 > H(p-\gamma p_s)^\alpha$. When $H(p-\gamma p_s)^\alpha \geqslant (w_o - s)/2$, setting $w_\eta = w_o$ generates greater profit for the supplier than $w_\eta = s + (w_o - s)^2/(4H(p-\gamma p_s)^\alpha)$ if and only if $H(p-\gamma p_s)^\alpha > (2c_o - w_o - s)(w_o - s)/[4(c_\eta - s)]$. When $H(p-\gamma p_s)^\alpha < (w_o - s)/2$, setting $w_\eta = w_o$ generates greater profit for the supplier than $w_\eta = w_o - H(p-\gamma p_s)^\alpha$ if and only if $H(p-\gamma p_s)^\alpha > c_o - c_\eta$. ▫

Proof of Corollary 1. By Proposition 3, if $H(p-\gamma p_s)^\alpha < \min\{(w_o - s)/2, c_o - c_\eta\}$, then $w_\eta^* = w_o - H(p-\gamma p_s)^\alpha$ and the reseller's best response is $I^* = \eta$. The quantity discount $H(p-\gamma p_s)^\alpha$ per unit off w_o is just to offset the increased holding cost in induced strategy $I^* = \eta$ compared to the inventory strategy $I^* = 0$ when no quantity discount is offered. Again by Proposition 3, if $(w_o - s)/2 \leqslant H(p-\gamma p_s)^\alpha \leqslant (2c_o - w_o - s)(w_o - s)/[4(c_\eta - s)]$, then $w_\eta^* = s + (w_o - s)^2/(4H(p-\gamma p_s)^\alpha)$ and the best response is $I^* = I^o = \sqrt{(w_\eta^* - s)/H(p-\gamma p_s)^\alpha}\eta = (w_o - s)\eta/(2H(p-\gamma p_s)^\alpha)$. For the reseller, the increased holding cost per cycle with length I^o/λ is $(I^o/2)(I^o/\lambda)h = (w_o - s)^2\eta/(4H(p-\gamma p_s)^\alpha)$ as compared to the inventory strategy $I^* = 0$; the loss in gray market diversion within the same cycle is $(w_\eta^* - s)(\eta - I^o) = (w_o - s)^2\eta[1 - (w_o - s)/(2H(p-\gamma p_s)^\alpha)]/(4H(p-\gamma p_s)^\alpha)$. The reseller's gain from the quantity discount for the same cycle is $(w_o - w_\eta^*)I^o = [w_o - s - (w_o - s)^2/(4H(p-\gamma p_s)^\alpha)](w_o - s)\eta/(2H(p-\gamma p_s)^\alpha)$, which is exactly equal to the sum of the increased holding cost and loss in gray market diversion for the same cycle. Therefore, we can conclude that the supplier's optimal all-unit quantity discount leaves the reseller with zero profits. ▫

Proof of Proposition 4. The reseller selects his resale price p and inventory policy I to maximize the expected profit per unit of time. We write the reseller's profit in terms of the endogenously determined demand rate λ: $\pi(I, \lambda) = \sqrt{m\lambda} + \gamma p_s \lambda - g(I, \lambda)$, where $\sqrt{m\lambda} + \gamma p_s \lambda$ is the reseller's revenue per unit of time and $g(I, \lambda)$ is the total costs per unit of time among all zero-inventory policies characterized by the initial cycle inventory level I. By Proposition 2, for any given resale price p and its corresponding demand rate λ, the reseller will choose the unique inventory policy that minimizes $g(I, \lambda)$. Using the reseller's optimal inventory response to an arrival rate, we derive the minimum inventory cost function $g^*(\lambda) = g(I^*(\lambda))$ as follows: if $s < w_\eta < (w_o + s)/2$, $g^*(\lambda) = w_o\lambda$ when $0 < \lambda < [4(w_\eta - s)mH]/[(w_o - s)^2]$, $g^*(\lambda) = \sqrt{4(w_\eta - s)mH\lambda} + s\lambda$ when $[4(w_\eta - s)mH]/[(w_o - s)^2] \leqslant$

$\lambda < [mH]/[(w_\eta - s)]$, $g^*(\lambda) = w_\eta\lambda + mH$ otherwise; if $(w_o + s)/2 \leqslant w_\eta \leqslant w_o$, $g^*(\lambda) = w_o\lambda$ when $0 < \lambda < \limsup_{c\to w_\eta -} mH/[(w_o - c)]$, and $g^*(\lambda) = w_\eta\lambda + mH$ otherwise.

The above cost function includes the cost of ordering, diversion and holding inventory. When the order size is η and there is no gray market diversion, the reseller enjoys the low unit cost w_η but suffers an average holding cost of mH per unit. In the case where the reseller diverts to the gray market, the reseller optimizes the diversion quantity $\eta - I$ by comparing the holding cost $hI/2$ with the diversion cost $w_\eta - s$. The reseller profit function with demand function $\lambda(p) = m/(p - \gamma p_s)^2$ is $\pi(\lambda) = \pi_0(\lambda) \equiv \sqrt{m\lambda} + (\gamma p_s - w_o)\lambda$ if $0 < \lambda \leqslant [4(w_\eta - s)mH]/[(w_o - s)^2]$, $\pi(\lambda) = \pi^o(\lambda) \equiv \sqrt{m\lambda} + (\gamma p_s - s)\lambda - \sqrt{4(w_\eta - s)mH\lambda}$ if $[4(w_\eta - s)mH]/[(w_o - s)^2] < \lambda < mH/[(w_\eta - s)]$ and $\pi(\lambda) = \pi_\eta(\lambda) \equiv \sqrt{m\lambda} + (\gamma p_s - w_\eta)\lambda - mH$ otherwise, where π_0, π^o and π_η correspond to when the reseller adopts the inventory policy $I^* = 0$, $I^* = I^o$ and $I^* = \eta$ respectively. Note that $\pi_0(\lambda)$ and $\pi_\eta(\lambda)$ are concave since $m > 0$ and $\pi^o(\lambda)$ is concave if $\sqrt{m} - \sqrt{4(w_\eta - s)mH} > 0$. We take the derivative of $\pi_0(\lambda)$, $\pi^o(\lambda)$ and $\pi_\eta(\lambda)$ with respect to λ as $\partial\pi_0(\lambda)/\partial\lambda = \sqrt{m}/(2\sqrt{\lambda}) + \gamma p_s - w_o$, $\partial\pi^o(\lambda)/\partial\lambda = \sqrt{m}/(2\sqrt{\lambda}) + (\gamma p_s - s) - \sqrt{(w_\eta - s)mH/(2\lambda)}$ and $\partial\pi_\eta(\lambda)/\partial\lambda = \sqrt{m}/(2\sqrt{\lambda}) + \gamma p_s - w_\eta$. The local optima satisfying the first-order conditions are $\lambda_1^* = m/(4(w_o - \gamma p_s)^2)$, $\lambda_2^* = (\sqrt{m} - \sqrt{4(w_\eta - s)mH})^2/(4(s - \gamma p_s)^2)$ and $\lambda_3^* = m/(4(w_\eta - \gamma p_s)^2)$ respectively, and the corresponding profits are $\pi_0(\lambda_1^*) = m/(4(w_o - \gamma p_s))$, $\pi^o(\lambda_2^*) = (\sqrt{m} - \sqrt{4(w_\eta - s)mH})^2/(4(s - \gamma p_s)$, $\pi_\eta(\lambda_3^*) = m/(4(w_\eta - \gamma p_s)) - mH$. The continuity of $\pi(\lambda)$ is easily verified by checking at the two breakpoints $\lambda_A = 4(w_\eta - s)mH/[(w_o - s)^2]$ and $\lambda_B = mH/[(w_\eta - s)]$. Since $\lim_{\lambda\to\lambda_A -} \partial\pi_0(\lambda)/\partial\lambda < \lim_{\lambda\to\lambda_A +} \partial\pi^o(\lambda)/\partial\lambda$, we eliminate the breakpoint λ_A as a global optimum. Since $\lim_{\lambda\to\lambda_B -} \partial\pi^o(\lambda)/\partial\lambda = \lim_{\lambda\to\lambda_B +} \partial\pi_\eta(\lambda)/\partial\lambda$, the global optimum $\lambda^* = \lambda_B$ only if $\lambda_2^* = \lambda_3^* = \lambda_B$. Hence, we conclude that the global optimum λ^* must be one of the local optima λ_1^*, λ_2^* and λ_3^*.

It remains to check under what conditions each local optimum dominates. First, note that $\lim_{\lambda\to\lambda_B -} \partial\pi^o(\lambda)/\partial\lambda = \lim_{\lambda\to\lambda_B +} \partial\pi_\eta(\lambda)/\partial\lambda \leqslant 0$ if and only if $H \geqslant (w_\eta - s)/[4(w_\eta - \gamma p_s)^2]$. Hence, a necessary condition for λ_2^* to be a global optimum is $H \geqslant (w_\eta - s)/[4(w_\eta - \gamma p_s)^2]$ and a necessary condition for λ_3^* to be a global optimum is $H \leqslant (w_\eta - s)/[4(w_\eta - \gamma p_s)^2]$. Second, we compare the profit of each of the batch order policies $I^* = I^o$ or $I^* = \eta$ to the profit of the order-as-you-go policy $I^* = 0$: $\pi_0(\lambda_1^*) > \pi^o(\lambda_2^*) \Leftrightarrow H > (1 - \sqrt{(s - \gamma p_s)/(w_o - \gamma p_s)})^2/[4(w_\eta - s)]$ and $\pi_0(\lambda_1^*) > \pi_\eta(\lambda_3^*) \Leftrightarrow H > 1/[4(w_\eta - \gamma p_s)] - 1/[4(w_o - \gamma p_s)]$. Lastly, conditioned on whether $w_\eta - \gamma p_s < \sqrt{(w_o - \gamma p_s)(s - \gamma p_s)}$, the break points on H can be ordered as follows: if $w_\eta - \gamma p_s < \sqrt{(w_o - \gamma p_s)(s - \gamma p_s)}$, $(1 - \sqrt{(s - \gamma p_s)/(w_o - \gamma p_s)})^2/[4(w_\eta - s)] > (w_\eta - s)/[4(w_\eta - \gamma p_s)^2]$ and if $w_\eta - \gamma p_s \geqslant \sqrt{(w_o - \gamma p_s)(s - \gamma p_s)}$, $1/[4(w_\eta - \gamma p_s)] - 1/[4(w_o - \gamma p_s)] \geqslant (w_\eta - s)/[4(w_\eta - \gamma p_s)^2]$. Therefore, it is not hard to conclude that when $w_\eta - \gamma p_s \geqslant \sqrt{(w_o - \gamma p_s)(s - \gamma p_s)}$, the optimal demand rate is $\lambda^* = \lambda_1^*$ if $H > (1 - \sqrt{(s - \gamma p_s)/(w_o - \gamma p_s)})^2/[4(w_\eta - s)]$, $\lambda^* = \lambda_2^*$ if $(w_\eta - s)/[4(w_\eta - \gamma p_s)^2] < H \leqslant (1 - \sqrt{(s - \gamma p_s)/(w_o - \gamma p_s)})^2/[4(w_\eta - s)]$ and $\lambda^* = \lambda_3^*$ otherwise; when $w_\eta \geqslant \sqrt{w_o s}$ the optimal demand rate is $\lambda^* = \lambda_1^*$ if $H > 1/[4(w_\eta - \gamma p_s)] - 1/[4(w_o - \gamma p_s)]$ and $\lambda^* = \lambda_3^*$ otherwise. By the relationship between price and demand rate $p(\lambda) = m/\sqrt{\lambda} + \gamma p_s$ and Proposition 2, the corresponding reseller's optimal pricing and inventory policy follows immediately. □

Proof of Proposition 5. It is readily apparent that when the reseller's best response is to order in batches without any gray market diversion, the supplier enjoys economies of scale from batch processing at the same rate as the demand rate $\lambda(p^* = 2w_\eta - \gamma p_s) = m/[4(w_\eta - \gamma p_s)^2]$ in the authorized channel. As a result, the supplier's profit per unit of time is $\Pi_\eta(w_\eta) \equiv (w_\eta - c_\eta)m/[4(w_\eta - \gamma p_s)^2]$. When the reseller's best response is to order on demand and not to hold inventory at all, the supplier delivers the product at the list price and the same rate as the demand rate $\lambda(p^* = 2w_o - \gamma p_s) = m/[4(w_o - \gamma p_s)^2]$ in the authorized channel, and hence the supplier's profit per unit of time is $\Pi_0 \equiv (w_o - c_o)m/[4(w_o - \gamma p_s)^2]$ which is independent of the size of the all-unit discount. Finally, when the reseller's best response is to order in batches with part of the order diverted to the gray market, the supplier enjoys economies of scale from orders of size η

every $I^*/\lambda(p^*)$ time units, and hence the supplier's profit per unit of time can be shown to be $\Pi^o(w_\eta) \equiv m(w_\eta - c_\eta)(\sqrt{H/[w_\eta - s]} - 2H)/[2(s-\gamma p_s)]$.

As a precursor to establishing this proposition we derive the supplier's profit function. By Proposition 4, the supplier's profit function given that the reseller employs the optimal pricing and inventory decisions can be described as $\Pi(w_\eta) = \Pi_\eta(w_\eta)$ if $w_\eta \in R_\eta^L \bigcup R_\eta^H$, $\Pi(w_\eta) = \Pi^0(w_\eta)$ if $w_\eta \in R^o$ and $\Pi(w_\eta) = \Pi_0$ if $w_\eta \in R_0^L \bigcup R_0^H$. We let $\tilde{w} \equiv 1/(4H + 1/(w_o - \gamma p_s)) + \gamma p_s$, $\hat{w} \equiv s + (1 - \sqrt{(s-\gamma p_s)/(w_o - \gamma p_s)})^2/4H$, $\underline{w} \equiv (1 - \sqrt{1 - 16(s-\gamma p_s)H})/(8H) + \gamma p_s$, $\overline{w} \equiv (1 + \sqrt{1 - 16(s-\gamma p_s)H})/(8H) + \gamma p_s$, where $\underline{w}$ and $\overline{w}$ are the two real roots, if they exist, of the quadratic equation $f(w) \equiv 4H(w-\gamma p_s)^2 - w + s = 0$. Then, the regions of quantity discount that induces different reseller pricing and inventory decisions are $R_0^H \equiv \{\sqrt{(w_o - \gamma p_s)(s-\gamma p_s)} + \gamma p_s \leqslant w_\eta \leqslant w_o \mid w_\eta > \tilde{w}\}$, $R_\eta^H \equiv \{\sqrt{(w_o - \gamma p_s)(s-\gamma p_s)} + \gamma p_s \leqslant w_o \mid w_\eta \leqslant \tilde{w}\}$, $R_\eta^L \equiv \{s < w_\eta < \sqrt{w_o s} \mid 4H(w_\eta - \gamma p_s)^2 - w_\eta + s \leqslant 0\}$, $R^o \equiv \{s < w_\eta < \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s \mid w_\eta \leqslant \hat{w}, 4H(w_\eta - \gamma p_s)^2 - w_\eta + s > 0\}$ and $R_0^L \equiv \{s < w_\eta < \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s \mid w_\eta > \hat{w}\}$. Regions R_0^H and R_η^H are mutually exclusive, with one of them possibly being an empty set. Regions R_0^L, R^o and R_0^H are mutually exclusive, with no more than two of them possibly being an empty set.

To simplify the profit function, we condition on the magnitude of H according to the following three cases.

case (i). Consider $4H < 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o - \gamma p_s)$, which is equivalent to $\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s < \tilde{w}$, hence $R_\eta^H = [\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s, \tilde{w}]$ and $R_0^H = (\tilde{w}, w_o]$. It can be easily verified that $4H < 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o - \gamma p_s)$ is also equivalent to $\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s < \hat{w}$, hence $R_0^L = \emptyset$. Note that $f(w=s) = 4H(w-\gamma p_s)^2 \geqslant 0$ and $f(w = \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s) < 0$ when $4H < 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s)$, hence the smaller root $\underline{w}$ must be real-valued and exist between s and $\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s$, i.e., $\underline{w} \in [s, \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s)$. Hence $R_\eta^L = [\underline{w}, \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s)$ and $R^o = (s, \underline{w})$ (R^o degenerates to $\emptyset$ if $\underline{w} = s$ which is equivalent to $s4H = 0$). In summary, $R_0^L \bigcup R_0^H = (\tilde{w}, w_o]$, $R_\eta^L \bigcup R_\eta^H = [\underline{w}, \tilde{w}]$ and $R^o = (s, \underline{w})$.

case (ii). Consider $1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s) \leqslant 4H \leqslant 1/[4(s-\gamma p_s)]$. Such an interval of H indeed exists since $1/[4(s-\gamma p_s)] + 1/(w_o - \gamma p_s) \geqslant 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}$ with equality holding if and only if $w_o - \gamma p_s = 4(s-\gamma p_s)$. Note that $4H \geqslant 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s)$ is equivalent to $\tilde{w} \leqslant \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s$, hence $R_\eta^H = \emptyset$ and $R_0^H = [\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s, w_o]$. Also note that $4H \geqslant 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s)$ is equivalent to $\hat{w} \leqslant \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s$, hence $R_0^L = (\hat{w}, \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s)$. Furthermore, when $4H \leqslant 1/[4(s-\gamma p_s)]$, the discriminant of the quadratic equation $f(w) = 0$ is non-negative, hence the roots $\underline{w}$ and $\overline{w}$ of equation $f(w) = 0$ must be real-valued. It is easy to check that $4H \geqslant 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s)$ is equivalent to $\overline{w} \leqslant \hat{w}$. Since $f(w=s) \geqslant 0$, $f(w=\hat{w}) \geqslant 0$ and $f(w = \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s) \geqslant 0$ with the last two inequalities ensured by $4H \geqslant 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s)$, the roots $\underline{w}$ and $\overline{w}$ must exist between s and $\hat{w}$, i.e., $[\underline{w}, \overline{w}] \subseteq [s, \hat{w}]$. Hence $R_\eta^L = [\underline{w}, \overline{w}]$ and $R^o = (s, \underline{w}) \bigcup (\overline{w}, \hat{w}]$. In summary, $R_0^H \bigcup R_0^L = (\hat{w}, w_o]$, $R^o = (s, \underline{w}) \bigcup (\overline{w}, \hat{w}]$ and $R_\eta^L \bigcup R_\eta^H = [\underline{w}, \overline{w}]$.

case (iii). Consider $4H \leqslant 1/[4(s-\gamma p_s)]$. According to the case (ii), we know that $4H \leqslant 1/[4(s-\gamma p_s)] \geqslant 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s)$; $4H > 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s)$ leads to that $R_\eta^H = \emptyset$, $R_0^H = [\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s, w_o]$ and $R_0^L = (\hat{w}, \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} + \gamma p_s)$. Moreover, $4H > 1/[4(s-\gamma p_s)]$ guarantees that $f(w) > 0$ for any w, hence $R_\eta^L = \emptyset$ and $R^o = (s, \hat{w}]$. In summary, $R_0^H \bigcup R_0^L = (\hat{w}, w_o]$, $R^o = (s, \hat{w}]$ and $R_\eta^L \bigcup R_\eta^H = \emptyset$.

Thus, the profit function can be expressed as follows completing the derivation: case (i). if $4H < 1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s)$, $\Pi(w_\eta) = \Pi^o(w_\eta)$ when $s < w_\eta < \underline{w}$ ($I^* = I^o$), $\Pi(w_\eta) = \Pi_\eta(w_\eta)$ when $\underline{w} \leqslant w_\eta \leqslant \tilde{w}$, ($I^* = \eta$), $\Pi(w_\eta) = \Pi_0$ when $\tilde{w} < w_\eta \leqslant w_o$ ($I^* = 0$); case (ii). if

$1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)} - 1/(w_o-\gamma p_s) \leqslant 4H \leqslant 1/[4(s-\gamma p_s)]$, $\Pi(w_\eta)=\Pi_\eta(w_\eta)$ when $\underline{w} \leqslant w_\eta \leqslant \overline{w}, (I^*=\eta)$, $\Pi(w_\eta)=\Pi^o(w_\eta)$ when $s<w_\eta<\underline{w}$ and $\overline{w}<w_\eta \leqslant \hat{w}$, $(I^*=I^o)$, $\Pi(w_\eta)=\Pi_0$ when $\hat{w}<w_\eta \leqslant w_o$ $(I^*=0)$; case (iii). if $4H \leqslant 1/[4(s-\gamma p_s)]$, $\Pi(w_\eta)=\Pi^o(w_\eta)$ when $s<w_\eta \leqslant \hat{w}$ $(I^*=I^o)$, $\Pi(w_\eta)=\Pi_0$ if $\hat{w}<w_\eta \leqslant w_o$ $(I^*=0)$.

Given that the reseller employs the optimal pricing and inventory policy in response to a discount wholesale price w^*_η, the supplier earns the following profit per unit of time:

Taking the first-order derivative of $\Pi_\eta(w_\eta)$ with respect to w_η, we have $\partial \Pi_\eta(w_\eta)/\partial w_\eta = m[1-2(w_\eta-c_\eta)/(w_\eta-\gamma p_s)]/[4(w_\eta-\gamma p_s)^2]$, hence the function $\Pi_\eta(w_\eta)$ is increasing on $(0, 2c_\eta-\gamma p_s]$ and decreasing on $[2c_\eta-\gamma p_s, \infty)$. Note that under the assumption that $(w_o-\gamma p_s)/4<(s-\gamma p_s)<(c_\eta-\gamma p_s)$, we have $\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}<2c_\eta-\gamma p_s$. Taking the first-order derivative of $\Pi^o(w_\eta)$ with respect to w_η, we have

$$\frac{\partial \Pi^o(w_\eta)}{\partial w_\eta} = \frac{m\sqrt{H}}{4(s-\gamma p_s)(w_\eta-s)^{\frac{3}{2}}}\left[-4\sqrt{H}(w_\eta-s)^{\frac{3}{2}}+w_\eta+c_\eta-2s\right].$$

Taking the second-order derivative of $\Pi^o(w_\eta)$ with respect to w_η, we have

$$\frac{\partial^2 \Pi^o(w_\eta)}{\partial {w_\eta}^2} = \frac{m\sqrt{H}(4s-3c_\eta-w_\eta)}{8(s-\gamma p_s)(w_\eta-s)^{\frac{5}{2}}}<0.$$

Under the assumption that $s<c_\eta$, then $\partial^2\Pi^o(w_\eta)/(\partial {w_\eta}^2)<0$ for $w_\eta>s$, namely, $\Pi^o(w_\eta)$ is strictly concave on (s,∞). Furthermore, since $s<c_\eta$, $\lim_{w_\eta \to s^+} \partial\Pi^o(w_\eta)/\partial w_\eta=\infty$. Under the assumption that $w_o-\gamma p_s)/4<(s-\gamma p_s)<(c_\eta-\gamma p_s)$,

$$\left.\frac{\partial \Pi^o(w_\eta)}{\partial w_\eta}\right|_{w_\eta=\hat{w}} = \frac{m\sqrt{H}}{4(s-\gamma p_s)(\hat{w}-s)^{\frac{3}{2}}}\left[\frac{(1-\sqrt{(s-\gamma p_s)/w_o-\gamma p_s)})^2}{4H}(2\sqrt{\frac{s-\gamma p_s}{w_o-\gamma p_s}}-1)+c_\eta-s\right]>0.$$

Therefore, $\Pi^o(w_\eta)$ is strictly increasing on $(s,\hat{w}]$.

cases (i) and (ii). In both cases, $4H<1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}-1/(w_o-\gamma p_s)$ with the form of the supplier's profit function corresponding to case (i) of the supplier profit function. If $4H<1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}-1/(w_o-\gamma p_s)$, by the derivation of the supplier profit function, $\underline{w}<\hat{w}$. Hence, $\Pi^o(w_\eta)$ is increasing on $(s,\underline{w}]$. Note that $\Pi_\eta(w_\eta)$ and $\Pi^o(w_\eta)$ are continuous at $\underline{w}$. Furthermore, by the derivation of the supplier profit function, $\underline{w}<\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}$ and we know under the assumption $(w_o-\gamma p_s)/4<(s-\gamma p_s)<(c_\eta-\gamma p_s)$, $\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}<2c_\eta-\gamma p_s$, hence $\underline{w}<2c_\eta-\gamma p_s$. Recall that $\Pi_\eta(w_\eta)$ is increasing on $(0,2c_\eta-\gamma p_s]$ and decreasing on $[2c_\eta-\gamma p_s,\infty)$. Therefore, if $2c_\eta-\gamma p_s<\tilde{w}$, $\Pi_\eta(w_\eta)$ is increasing on $[\underline{w},2c_\eta-\gamma p_s]$ and decreasing on $[2c_\eta-\gamma p_s,\tilde{w}]$; otherwise, $\Pi_\eta(w_\eta)$ is increasing on $[\underline{w},\tilde{w}]$. Finally, it remains to compare the supplier's profit at the list price w_o and the one at the discount price $\min\{2c_\eta-\gamma p_s,\tilde{w}\}$ that maximizes the profit when a discount is offered.

case (iii). We consider $4H<1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}-1/(w_o-\gamma p_s)$. First, we consider $1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}-1/(w_o-\gamma p_s) \leqslant 4H \leqslant 1/(4(s-\gamma p_s))$ with the form of the supplier's profit function corresponding to case (ii) of the supplier profit function. Note that if $4H \leqslant 1/(4(s-\gamma p_s))$, $\underline{w} \leqslant \overline{w} \leqslant \sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}<2c_\eta-\gamma p_s$, where the second inequality is shown in the derivation of the supplier profit function and the last is due to the assumption $(w_o-\gamma p_s)/4<(s-\gamma p_s)<(c_\eta-\gamma p_s)$. Hence $\Pi_\eta(w_\eta)$ is increasing on $[\underline{w},\overline{w}]$ and maximized at $\overline{w}$. We also know that $\Pi^o(w_\eta)$ is increasing on $(s,\underline{w}]$ and $[\overline{w},\hat{w}]$, and obtains its maximum at $\hat{w}$. Note that $\Pi_\eta(w_\eta)$ and $\Pi^o(w_\eta)$ are continuous at $\underline{w}$ and $\overline{w}$. Therefore, if $1/\sqrt{(w_o-\gamma p_s)(s-\gamma p_s)}-1/(w_o-\gamma p_s)) \leqslant 4H \leqslant 1/(4(s-\gamma p_s))$, the discount price $\hat{w}$ maximizes the supplier's profit when a discount is offered. Second, we consider $4H>1/(4(s-\gamma p_s))$ with the form of the supplier's profit function corresponding to case

(iii) of the supplier profit function. Recall that $\Pi^o(w_\eta)$ is strictly increasing on $(s, \hat{w}]$. Therefore, if $4H > 1/(4(s-\gamma p_s))$, the discount price $\hat{w}$ maximizes the supplier's profit when a discount is offered. Finally, combining the two subcases, it remains to compare the supplier's profit at the list price w_o and the one at the optimal discount price $\hat{w}$. □

Proof of Corollary 2. First we check case by case how in Proposition 5 the benefits from economies of scale are allocated between the supplier and reseller. When it is optimal for the supplier to offer a discount price $w_\eta^* < w_o$, the supplier enjoys economies of scale. By Proposition 4, the reseller's best response $(p^*(w_\eta^*), I^*(w_\eta^*))$ of pricing and inventory is either to take the $I^* = I^o$ strategy, namely, $(p^*(w_\eta^*), I^*(w_\eta^*)) = \big(2(s-\gamma p_s)/[1-2\sqrt{(w_\eta - s)H}] + \gamma p_s,\ [\sqrt{[w_\eta - s]/(4H)} - (w_\eta - s)]\eta/(s-\gamma p_s)\big)$ or the $I^* = \eta$ strategy, namely, $(p^*(w_\eta^*), I^*(w_\eta^*)) = (2w_\eta^* - \gamma p_s, \eta)$.

If the best response of the reseller $(p^*(w_\eta^*), I^*(w_\eta^*)) = (2w_\eta^* - \gamma p_s, \eta)$, then the reseller's profit per unit of time is $\pi(\lambda(p^*(w_\eta^*))) = m/(4(w_\eta - \gamma p_s)) - mH$. When the supplier sets $w_\eta = w_o$ and does not enjoy economies of scale, the reseller's profit per unit of time in the best response is $\pi(\lambda(p^*(w_o))) = m/(4(w_o - \gamma p_s)^2)$. It is easy to see that if $w_\eta^* \leqslant$ (resp. $<$)$\tilde{w}$, the reseller is (resp. strictly) better off when the supplier offers a discount price $w_\eta = w_\eta^*$ as compared to when the supplier does not, i.e., $\pi(\lambda(p^*(w_\eta^*))) \geqslant$ (resp. $>$)$\pi(\lambda(p^*(w_o)))$. By the proof of Proposition 5, we can verify that to elicit $I^* = \eta$, in cases (i) and (ii) of Proposition 5, it is optimal for the supplier to set $w_\eta^* = \min\{2c_\eta - \gamma p_s, \tilde{w}\} \leqslant \tilde{w}$. Hence, in case (i), $w_\eta^* = 2c_\eta - \gamma p_s < \tilde{w}$ and the reseller shares part of the benefits from economies of scale; in case (ii), $w_\eta^* = \tilde{w}$ and the reseller shares no benefits.

By Proposition 5, the other scenario is that the supplier sets $w_\eta^* = \hat{w}$ to induce the reseller to take the $I^* = I^o$ strategy. By Proposition 4, the corresponding best response of the reseller is $(p^*(\hat{w}), I^*(\hat{w})) = \big(2\sqrt{(w_o - \gamma p_s)(s-\gamma p_s)} - \gamma p_s, [\sqrt{(s-\gamma p_s)/(w_o - \gamma p_s)} - (s-\gamma p_s)/(w_o - \gamma p_s)]\eta/(s-\gamma p_s)4H)\big)$ and the reseller's profit per unit of time is $\pi(\lambda(p^*(\hat{w}))) = (\sqrt{m} - \sqrt{4(\hat{w} - s)mH})^2/(4(s-\gamma p_s), = m/(4(w_o - \gamma p_s)) = \pi(\lambda(p^*(w_o)))$. Hence in this scenario, the reseller earns the same profit as without the quantity discount and shares no benefits from economies of scale.

From the analysis of all three cases, we can see that the resale price in equilibrium when the supplier enjoys economies of scale is always strictly smaller than the resale price $2w_o - \gamma p_s$ that the reseller will charge if the supplier does not enjoy economies of scale and offers the wholesale price at w_o. Hence, consumers will always enjoy a lower resale price with the supplier's economies of scale than without. □

Proof of Proposition 6. Fix an arbitrary $w_\eta \leqslant w_o$. Since $\pi_d(p, I)$ is a continuous function defined over a compact set, there exists a unique largest optimizer $I^*(h) = \max\{I' \in [0,\eta] \mid \pi_d(p', I'; h) = \max_{(p,I)\in\mathbb{R}\times[0,\eta]} \pi_d(p, I; h)\}$. It is easy to see that $\pi_d(p, I; h)$ has decreasing differences in $((p, I); h)$, hence the cycle inventory level $I^*(h)$ in the optimal diversion strategy is a decreasing function of h. Furthermore, in one extreme case when $h = 0$, intuitively, $I^*(h=0) = \eta$. In the other extreme case when $h = \infty$, intuitively, $I^*(h=\infty) = 0$. Therefore, there must exist an interval $\hat{\mathcal{H}} \subseteq (0, \infty)$ of holding cost parameters such that the optimal diversion strategy resorts to the gray market (i.e., $I^*(h) < \eta$) if and only if $h \in \hat{\mathcal{H}}$. This is the condition on the holding cost parameter h to ensure the optimal diversion strategy is not degenerate into a batch strategy. Now let us investigate when the optimal diversion strategy is more profitable than the non-diversion strategies: pay-as-you-go strategy and batch strategy. First, if the reseller adopts the pay-as-you-go strategy, then the associated long-run average profit rate is $\pi_p \equiv \max_p (p - w_o)\lambda(p, p_s(0))$, which is independent of the holding cost parameter h. Since $\pi_d(p, I; h)$ is strictly decreasing in h, $\pi_d^*(h) \equiv \max_{(p,I)\in\mathbb{R}\times[0,\eta]} \pi_d(p, I; h)$ is strictly decreasing in h. Moreover, note that $\pi_d^*(0) = \max_{p\in\mathbb{R}} (p - w_\eta)\lambda(p, p_s(0)) \geqslant \pi_p$. Hence there must exist a finite upper bound $\tilde{h}$ on the holding cost parameters such that the profit rate under the optimal diversion strategy is weakly higher than that of the pay-as-you-go strategy if and only if $h \in \tilde{\mathcal{H}} \equiv [0, \tilde{h}]$, where $\tilde{h}$ is the solution to the equation $\pi_d^*(\tilde{h}) = \pi_p$. Second, the associated long-run average profit rate of the batch strategy is a special case of $\pi_d(p, I)$ with $I = \eta$. Hence the profit rate under the optimal diversion strategy is higher than that of the batch strategy if and only if $I^*(h) < \eta$, which is guaranteed as long as $h \in \hat{\mathcal{H}}$. Combining $\hat{\mathcal{H}}$ and $\tilde{\mathcal{H}}$, we have the desired result. A necessary and sufficient condition to guarantee that $\mathcal{H} = \hat{\mathcal{H}} \bigcap \tilde{\mathcal{H}}$ is non-empty is that $\partial(\max_{p\in\mathbb{R}} \pi_d(p, I; h = \tilde{h}))/\partial I \mid_{I=\eta} < 0$. ◻

Proof of Proposition 7. Fix any h such that $h \in \mathcal{H}(w'_\eta) \neq \emptyset$ for some $w'_\eta \in [s(\eta), w_o]$. By implicit function theorem, $g(p, I)$ is decreasing in $I \in [0, \eta]$, because by taking derivative with respect to p at both sides of $g = (\eta - I)/(I/\lambda(p, p_s(g)))$ and re-arranging terms, we have $\partial g/\partial p = [\frac{\eta - I}{I} \frac{\partial \lambda(p, p_s)}{\partial p}]/[1 - \frac{\eta - I}{I} \frac{\partial \lambda(p, p_s)}{\partial p_s} \frac{\partial p_s(g)}{\partial g}] < 0$. Under Assumption (CL), $\pi_d(p, I; w_\eta)$ has increasing differences in $(I; w_\eta)$. Moreover, by $\partial g/\partial p < 0$, we have $I/\lambda(p, p_s(g(p, I))) = (\eta - I)/g(p, I)$ is increasing in p. Hence, $\pi_d(p, I; w_\eta)$ has increasing differences in $(p; w_\eta)$. Therefore, the optimal diversion strategy $I^*(w_\eta) = \max\{I' \in [0, \eta] \mid \pi_d(p', I'; w_\eta) = \max_{(p,I)\in\mathbb{R}\times[0,\eta]} \pi_d(p, I; w_\eta)\}$ and $p^*(w_\eta)$ are increasing in w_η. In one extreme when $w_\eta = s(\eta)$, the reseller enjoys the discount price by ordering in batches but diverting the most fraction of the order without any diversion loss, i.e., $I^*(w_\eta = s(\eta)) = 0$. In the other extreme when $w_\eta = \infty$, intuitively, the unit diversion loss is tremendous and the reseller does not have any incentive for diversion, i.e., $I^*(w_\eta = \infty) = \eta$. Then there exists an interval $\mathcal{W}(h) \subseteq [s(\eta), w_o]$ of discount wholesale prices such that the optimal diversion strategy indeed resorts to the gray market (i.e., $I^*(w_\eta) < \eta$) if and only if $w_\eta \in \mathcal{W}(h)$. The set $\mathcal{W}(h)$ is non-empty since w'_η must be an element in it.

When the reseller strategy is a diversion strategy, the supplier's profit rate is $\Pi_d(w_\eta; c_\eta; p, I) \equiv (w_\eta - c_\eta)\eta/[I/\lambda(p, p_s(g(p, I)))]$. Since $I/\lambda(p, p_s(g(p, I)))$ is increasing in p by deviation and increasing in I by Assumption (CL), and $(p^*(w_\eta), I^*(w_\eta))$ is increasing in w_η, then $I/\lambda(p, p_s(g(p, I))) \mid_{(p,I)=(p^*(w_\eta), I^*(w_\eta))}$ is increasing in w_η. Hence, $\Pi_d^*(w_\eta; c_\eta) \equiv \Pi_d\big(w_\eta; c_\eta; (p, I) = (p^*(w_\eta), I^*(w_\eta))\big)$ has increasing differences in $(w_\eta; c_\eta)$, and therefore $w_\eta^*(c_\eta) = \arg\max_{w_\eta \in [s(\eta), w_o]} \Pi_d^*(w_\eta; c_\eta)$ is increasing in c_η. Then there exists an interval $\hat{\mathcal{C}}$ of batch supply costs such that $w_\eta^*(c_\eta) \in \mathcal{W}(h)$ if and only if $c_\eta \in \hat{\mathcal{C}}$. Moreover, the supplier's profit rate under the reseller's pay-as-you-go strategy is $\Pi_p \equiv (w_o - c_o)\lambda(p^*(I = \eta; w_o), p_s(0))$, where $p^*(I = \eta; w_o) = \arg\max_{p\in\mathbb{R}}(p - w_o)\lambda(p, p_s(0))$. Since $\Pi_d^*(w_\eta; c_\eta)$ is strictly decreasing in c_η, then $\Pi_d^*(w_\eta^*(c_\eta); c_\eta)$ is strictly decreasing in c_η. Moreover, under the additional Assumption (P), $p^*(I) = \arg\max_{p\in\mathbb{R}} \pi_d(p, I)$ is increasing in I. Hence $\Pi_d^*(w_\eta^*(0); 0) \geqslant \Pi_d^*(w_o; 0) = w_o\eta/[I/\lambda(p, p_s(g(p, I)))] \mid_{(p,I)=(p^*(w_o), I^*(w_o))} \geqslant w_o\eta/[I/\lambda(p, p_s(g(p, I)))] \mid_{(p,I)=(p^*(I=\eta; w_o), \eta)} \geqslant \Pi_p$, where the second-to-last inequality is due to that $I/\lambda(p, p_s(g(p, I)))$ is increasing in (I, p) and that $(p^*(w_o), I^*(w_o)) = (p^*(I = I^*(w_o); w_o), I^*(w_o)) \leqslant (p^*(I = \eta; w_o), \eta)$ ensured by that $p^*(I)$ is increasing in I. Therefore there exists a finite upper bound $\tilde{c}$ on the batch supply cost such that the supplier's profit rate under the induced optimal diversion strategy with a discount wholesale price is strictly higher than under the induced pay-as-you-go strategy without offering a discount wholesale price if and only if $c_\eta \in \tilde{\mathcal{C}} \equiv (0, \tilde{c})$, where $\tilde{c}$ is the solution to $\Pi_d^*(w_\eta^*(\tilde{c}); \tilde{c}) = \Pi_p$. Finally, the supplier's profit rate under the reseller's batch strategy is $(w_\eta - c_\eta)\lambda(p^*(I = \eta; w_\eta), p_s(0))$. If $I^*(w_\eta) < \eta$, then $p^*(I = I^*(w_\eta); w_\eta) \leqslant p^*(I = \eta; w_\eta)$ and $\Pi_d(w_\eta; c_\eta; p^*(w_\eta), I^*(w_\eta)) \geqslant \Pi_d(w_\eta; c_\eta; p^*(I = \eta; w_\eta), \eta)$, since $\Pi_d(w_\eta; c_\eta; p, I)$ is decreasing in (p, I). In other words, the supplier's profit rate under the reseller's optimal diversion strategy is higher than under the reseller's batch strategy if and only if $I^*(w_\eta^*(c_\eta)) < \eta$ if and only if $c_\eta \in \hat{\mathcal{C}}$. Let $\mathcal{C} \equiv \hat{\mathcal{C}} \bigcap \tilde{\mathcal{C}}$ and we have the desired result. A necessary and sufficient condition to guarantee that $\mathcal{C}$ is non-empty is that $\partial(\max_{p\in\mathbb{R}} \pi_d(p, I; w_\eta^*(\tilde{c}), \tilde{c}))/\partial I \mid_{I=\eta} < 0$. ◻