Conditional and Lenient Drug Approvals

Chenxi Xu, Giuseppe Lopomo, David Ridley, Peng Sun
Fuqua School of Business, Duke University, Durham, North Carolina 27708,
chenxi.xu@duke.edu, glopomo@duke.edu, david.ridley@duke.edu, psun@duke.edu

Regulators often allow firms to sell new drugs based on preliminary efficacy evidence, with final approval contingent on confirmatory testing. We characterize optimal approval policies when firms have private information about testing costs and the regulator's payoff depends on expected efficacy. The optimal policy may include partial conditional approval for drugs with low expected efficacy, and leniency in granting final approval below conventional standards. Our calibration suggests these tools can generate hundreds of millions of dollars in annual social value.

Key words: mechanism design, adverse selection, accelerated approval, incentives, health, pharmaceutical
History: The current version was updated on September 13, 2025.

1. Introduction

In 1988, HIV-AIDS activists protesting outside the Food and Drug Administration (FDA) chanted, "Hey, hey, FDA, how many people have you killed today?" With tens of thousands dying annually from HIV-AIDS, activists demanded access to promising antiviral drugs still under development (United Press International 1988). In response, the FDA introduced a conditional approval mechanism called "accelerated approval." In 2024, the FDA granted conditional approval to seven drugs, six of which target rare diseases (Food and Drug Administration 2025). Conditional approval has also been granted in many regulatory jurisdictions, including the U.S., Europe, and Japan.

Conditional approval allows a new drug to reach the market after phase II testing, with full approval contingent on confirmatory phase III tests. Phase II tests are relatively short and inexpensive, but provide inconclusive evidence. In contrast, phase III tests are costly and time-consuming, but provide statistically and clinically significant evidence of efficacy. Without conditional approval, firms might forego phase III testing for slow-progressing diseases, which require long and large tests (Budish et al. 2015).[1]

Conditional approval benefits patients through faster access and encourages investment by increasing revenue. It allows firms to generate sales while funding costly phase III testing, making the development of marginally profitable drugs viable. Revenue rises early from immediate sales and later through market diffusion.

[1] More generally, regulatory restrictions and testing costs can stifle innovation and reduce social welfare (Peltzman 1973, Philipson and Sun 2008, Grennan and Town 2020).

We use the term "leniency" to describe the regulator committing to a final approval threshold below what it would choose later. Leniency may involve approving drugs without persuasive confirmatory trials or relying on surrogate endpoints (e.g., tumor shrinkage) instead of clinical outcomes (e.g., survival). Indeed, 9 of 46 cancer indications granted final approval after conditional approval lacked statistically significant evidence of benefit in the medical literature (Liu et al. 2024).

The regulator can commit to leniency before phase III. For example, the FDA's "breakthrough" designation gives the firm access to timely advice and interaction with the regulator to facilitate drug development. Through this pathway, the regulator can guide trial design and specify what (possibly lowered) evidence is required.

Leniency is always costly to society, as it raises the risk of approving ineffective drugs. Conditional approval, by contrast, may or may not be costly to society, depending on the drug's expected effectiveness and the extent to which early access boosts future adoption.

Regulators set conditional approval and leniency with limited information about testing costs for three reasons. First, many drugs lack predicates for benchmarking; five of the seven conditionally approved drugs in 2024 were first in their therapeutic class. Second, recruitment and enrollment costs are substantial, especially for rare diseases with few patients. Third, transparency is limited: firms anticipate ex post profits, so they conceal costs to avoid criticism about excess profit and calls for price regulation. The regulator's belief distribution depends on the degree of informational asymmetry with the firm.[2]

Private information about phase III testing cost motivates our mechanism design approach, which yields optimal conditional approval and leniency thresholds, based on the regulator's belief about the testing cost and the drug's efficacy potential. In essence, the regulator's problem is a procurement problem, where the firm supplying the new drug can be paid in two ways: conditional approval and leniency.

In our framework, however, three features distinguish the regulator's problem from standard procurement problems. First, the final approval decision is based on a testing result that does not arise in a procurement setting. Second, the two types of payments are bounded by constraints analogous to budgets: there is a finite amount of feasible leniency and the scale of conditional approval is bounded by the market size. Third, the regulator's payoff function is nonlinear in leniency. This nonlinearity precludes the use of standard mechanism design techniques to solve the regulator's problem.

[2] We focus on private information about testing costs, though firms may also have private information about efficacy. Both high costs and low expected efficacy can deter development, motivate government incentives, and give firms information rents. The distinction between the two types of private information matters most when the regulator provides funding up front and ex post profit is not positive, as with foundation grants for tropical disease research. In such cases, under private information about expected efficacy, a moral hazard problem arises. The firm has little incentive to exert effort because it already has the money and little to gain from success (Ridley et al. 2025).

If a drug is "low-cost," i.e., the regulator's belief about the firm's testing cost is sufficiently optimistic, the optimal policy involves no costly incentives. Conditional approval is granted only if the drug is ex-ante promising (not simply to induce testing), and there is no leniency.

If, instead, the regulator believes that testing costs could be high enough to deter investment, the regulator can implement a take-it-or-leave-it offer by setting a scale of conditional approval and a threshold for final approval. The firm will find it profitable to accept the offer if and only if its testing cost falls below a threshold. The optimal levels of conditional approval and leniency balance the goal of inducing the firm to run confirmatory testing against the risk of paying for ultimately ineffective drugs.

We begin with a two-period version of the model, where the regulator's problem can be framed as an (infinite-dimensional) linear program. The two-period model is appropriate for rare-disease drugs, which are exempt from price regulation and face limited generic competition due to their small market size. Duality theory implies that a take-it-or-leave-it offer is indeed optimal, and yields its full characterization for all parameter values in our framework.

We then analyze a three-period version of the model, in which generic competition or price regulation drives the price down to marginal cost in the third period. In this version, we derive a linear relaxation of the regulator's problem, which provides an upper bound for the regulator's payoff. Based on its solution, we construct a feasible policy that yields a lower bound for the regulator's payoff. In our calibrated numerical analysis, these bounds are remarkably close.

Based on our analysis, lawmakers should grant regulators the following powers: First, regulators should be permitted to conditionally approve drugs with low expected efficacy. Second, they should have the flexibility to grant partial conditional approval, such as limiting access to specific patient subgroups. Insurance coverage could also be restricted. Third, regulators should be able to use leniency in setting the final approval threshold below the myopic standard.

1.1. Literature Review

Myerson (1981) laid the foundation for optimal mechanism design. Baron and Myerson (1982) extend the framework to regulation economics with adverse selection concerns. In that paper, the agent (a firm) privately knows the unit production cost, and the principal (a regulator) maximizes a linear combination of consumer welfare and the firm's profit. The regulator sets the unit price the firm can charge, which determines the resulting sales quantity and may also subsidize or tax the firm. The basic Myersonian framework assumes at least two instruments in the mechanism: allocation and monetary transfer. Our principal uses entry, as well as the conditional and final approvals, as instruments. This also implies that we have more "budgetary" constraints in the optimization model, which complicates the analysis and results. Furthermore, we need to consider

testing with uncertain outcomes, an important feature that Baron and Myerson (1982) does not have.

Some diseases lack commercial prospects because they are concentrated in lower-income countries, where profit margins are near zero (Ridley et al. 2025). In such cases, conditional approval may help patients but offers little incentive to firms, because speeding an unprofitable drug to market does not improve returns. For tropical diseases, push and pull mechanisms are more effective than conditional approval. Push mechanisms subsidize early-stage testing, often through agencies such as the National Institutes of Health (NIH) or the Gates Foundation. Pull mechanisms raise revenues, for example through advance market commitments that guarantee purchases (Berndt et al. 2007, Kremer et al. 2022), or transferable vouchers such as U.S. priority review vouchers (Gans and Ridley 2013) and exclusivity-extension vouchers (Dubois et al. 2022).

For diseases that are profitable ex post but not ex ante due to long testing timelines (Budish et al. 2015), conditional approval offers two advantages over push and pull funding. First, it grants early access and accelerates diffusion, generating benefits both immediately and over time. Second, it is often more politically feasible than direct subsidies to drug makers, which remain unpopular. Indeed, NIH funding is directed largely to universities for early-stage testing rather than to firms for late-stage trials (Zhou et al. 2023). Push, pull, and conditional approval are not mutually exclusive, however. When calibrating the model, push funding could be incorporated as reducing testing costs, and pull funding could increase revenue, complementing conditional approval.

This paper also contributes to the literature on optimal regulatory approval. Carpenter and Ting (2007) analyze a regulator-firm interaction where approval decisions rely on imperfect information. They examine the trade-offs between false positives and negatives in regulatory decisions. Conditional approval often depends on surrogate endpoints, which must be carefully validated to ensure reliability for regulators and payers (Bognar et al. 2017). Beyond conditional approval, regulators employ pathways like the FDA's "breakthrough" designation to facilitate market entry (Chandra et al. 2024).

Permissive conditional approval policies may lead to withdrawals of drug indications if further testing reveals insufficient efficacy. Orlov et al. (2020) show that firms might delay reporting negative information to postpone withdrawals. Xu et al. (2021) demonstrate that firms balance regulatory costs by choosing between extended and shorter testing periods. To address delays, regulators like the FDA now require firms to initiate confirmatory testing before granting conditional approval. The dynamics differ for most medical devices, especially those similar to previous devices. Testing is less rigorous, markets see frequent entry, and competition is dynamic (Grennan and Town 2020, Collard-Wexler et al. 2024).

Some papers study continuous-time models in which approval decisions evolve with accumulating evidence. Henry and Ottaviani (2019) extends Wald's sequential hypothesis testing model to a game between a firm and a regulator. The firm conducts costly testing, generating a publicly observed Brownian motion with drift based on unobservable efficacy. The firm benefits from approval; the regulator's payoff depends on true efficacy. The firm chooses when to stop and seek approval. Commitment by the regulator leads to Pareto improvements over outcomes without commitment. Henry et al. (2022) allow for approval revocation when negative evidence emerges. The results highlight the importance of flexible regulatory policies.

The continuous-time models are ideal for settings with adaptable protocols, such as early-stage drug testing in animals or review of drugs after final approval. However, discrete time is sufficient in settings such as late-stage human testing which involves fixed endpoints, with outcomes revealed simultaneously. Furthermore, our model captures the diffusion effect triggered by conditional approval, which is not present in Henry et al. (2022).

McClellan (2022) presents another continuous-time model in which the regulator decides when to stop experimentation and approve a drug. The setup generalizes optimal stopping problems by accounting for the firm's incentive to exit prematurely. Even with symmetric information, the regulator lowers the approval threshold as the firm nears exit, terminating only if the drug's prospects fall too low. Under asymmetric information, if the firm claims the drug is less promising, the structure resembles the symmetric case. If the firm reports a more promising drug, the regulator offers a "fast track" with a lower approval threshold and a floor on efficacy. If the drug hits the floor, the regulator reverts to a higher approval threshold. Likewise, in our model, the regulator can be more lenient to allow a more marginal drug that might otherwise exit. However, our model differs by including conditional approval, allowing the firm to earn revenue and begin diffusion during confirmatory testing.

2. Model

A regulator and a pharmaceutical firm interact in a three-period game. At the beginning, the firm has already conducted preliminary clinical tests for a new drug. Based on the outcome of these tests, both parties believe that the drug's efficacy η is a random variable with support $[0,1]$, cumulative distribution function G, and strictly positive density g. To ascertain the true efficacy of the drug $\eta \in [0,1]$, the firm must conduct a final round of confirmatory tests which cost c. The firm knows the value of c privately, while the regulator believes that c is drawn from a distribution with a cumulative distribution function F and positive density f on its support $[c_L, c_H]$, with $0 \leq c_L \leq c_H$. We maintain the following regularity assumption throughout the paper.

Assumption 1. *The inverse hazard ratio $F(c)/f(c)$ is non-decreasing.*

Assumption 1 is routinely invoked in the mechanism design literature, following Myerson's pioneering work (Myerson 1981). It implies that the "virtual cost" function $c + F(c)/f(c)$ is increasing, and holds for several commonly used distributions, including uniform, exponential, and Pareto distributions. In our setting, this assumption rules out stochastic regulatory policies.

The game begins with the regulator committing publicly to a regulatory policy. Any feasible regulatory policy specifies, for each cost realization $c \in [c_L, c_H]$, a probability $\tau(c) \in [0,1]$ of the firm conducting confirmatory testing, and a scale of conditional approval $a_1(c) \in [0,1]$ contingent on the firm testing; and then for each pair $(c, \eta) \in [c_L, c_H] \times [0,1]$, a probability of granting final approval $a_2(c,\eta) \in [0,1]$.[3] If the firm chooses not to test, the game ends immediately with both parties earning zero payoffs. Otherwise, the firm enters the market in period one at scale a_1 while testing is taking place.

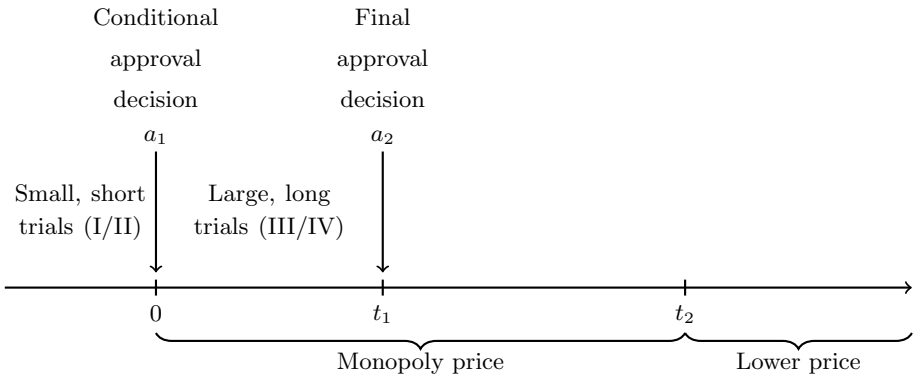

Figure 1 Timeline for testing, regulatory review, and falling prices. After preliminary testing, the regulator may grant conditional approval for a portion of the population (a_1). Later, the regulator can grant full approval (a_2). After conditional or final approval, the firm will charge a monopoly price until t_2 when the price falls.

Figure 1 illustrates the timeline. At time 0, the regulator decides whether to induce testing ($\tau(c)$) and grants conditional approval ($a_1(c)$) according to the agent's reported cost c. From time 0 to t_1, the firm sells a_1 fraction of the market at a given price. At time t_1, the test result (η) is revealed, and the regulator decides whether to grant final approval ($a_2(c,\eta)$). If the drug receives final approval, the firm continues selling at the monopoly profit margin until time t_2, after which generic drugs are allowed to enter the market and the price falls to marginal cost. In the rest of the paper, we refer to time intervals $[0, t_1)$, $[t_1, t_2)$, and $[t_2, \infty)$ as periods one, two, and three, respectively.

[3] Any regulatory policy can be interpreted as a direct revelation mechanism, as implied by the revelation principle (Myerson 1981).

Without loss of generality, we normalize both the firm's monopolistic profit per unit and the demand in period one, when full conditional approval is granted, to 1. If the firm is granted conditional approval at scale a_1, its payoff in period one is a_1.

The firm's revenue in period two, after the drug's final approval, is affected by its revenue in period one. Drug sales tend to rise over time as doctors and patients learn about the drug through experience and word of mouth.[4] Hence, entering the market in period one at a larger scale increases later sales and profit in period two. We assume that the total discounted demand in period two, if final approval is granted, is $\lambda_0 + \lambda_1 \cdot a_1$, where λ_0 and λ_1 are fixed parameters. When the monopolistic period ends, generic drugs are allowed to enter the market, and the firm's net revenue becomes 0.

The firm's total expected profit is given by the testing probability multiplied by its total revenue across both periods minus the testing cost. Therefore, if the firm has a true cost c and chooses the option designed for cost c', its expected total profit is

$$\hat{\Pi}(c, c') \equiv \tau(c') \cdot \left[a_1(c') - c + [\lambda_0 + \lambda_1 \cdot a_1(c')] \cdot \int_0^1 a_2(c', \eta) \, \mathrm{d}G(\eta) \right]. \tag{1}$$

To be feasible, any regulatory policy must satisfy incentive compatibility (IC), meaning that the firm should not benefit from choosing any option other than the one designed for its true cost c, i.e.,

$$\forall c, c' \in [c_L, c_H]: \qquad \hat{\Pi}(c, c) \geq \hat{\Pi}(c, c'). \tag{IC}$$

Also, because the firm can earn zero profit by declining the request to conduct the final round of tests, the following individual rationality (IR) constraint must hold:

$$\forall c \in [c_L, c_H]: \qquad \hat{\Pi}(c, c) \geq 0. \tag{IR}$$

Finally, the probability of testing τ, the scale of conditional approval a_1, and the probability of final approval a_2, must be between 0 and 1:

$$\forall c \in [c_L, c_H]: \qquad 0 \leq \tau(c) \leq 1, \tag{2}$$

$$\forall c \in [c_L, c_H]: \qquad 0 \leq a_1(c) \leq 1, \text{ and} \tag{3}$$

$$\forall (c, \eta) \in [c_L, c_H] \times [0, 1]: \qquad 0 \leq a_2(c, \eta) \leq 1. \tag{4}$$

The regulator's payoff per unit of the drug $v(\eta)$ depends on the efficacy level. The function $v : [0, 1] \to \mathbb{R}$ reflects the drug's value net of its monopoly price, which the firm charges in periods one and two. We assume that v is strictly increasing and satisfies

$$v(0) \leq -1 \text{ and } 0 < v(1), \tag{5}$$

[4] A drug's revenue often peaks in year six (Robey and David 2016).

where the price difference between the monopolistic and competitive periods equals the firm's net profit per unit, which is normalized to 1. That is, $v(0)$ equals zeros minus the drug's production cost, minus the firm's monopoly profit (which is normalized to 1), and minus the cost of toxicity.[5] Therefore, if the test outcome is sufficiently poor, i.e., η is close to 0, then the regulator's net payoff remains negative even after the firm loses its monopoly power.

If market access is granted in period one at scale a_1, the regulator's expected payoff is

$$v_E \cdot \int_{c_L}^{c_H} \tau(c) \cdot a_1(c) \, \mathrm{d}F(c), \tag{6}$$

where v_E is the expected value of $v(\eta)$ given by

$$v_E \equiv \int_0^1 v(\eta) \, \mathrm{d}G(\eta).$$

Given the firm's reported cost c and the revealed efficacy level η, the drug is finally approved with probability $a_2(c, \eta)$. Therefore, the regulator's expected payoff in period two is

$$\int_{c_L}^{c_H} \tau(c) \cdot [\lambda_0 + \lambda_1 \cdot a_1(c)] \cdot \int_0^1 v(\eta) \cdot a_2(c, \eta) \, \mathrm{d}G(\eta) \, \mathrm{d}F(c). \tag{7}$$

At the end of period two, the firm loses its monopoly power, due to generic entry or the initiation of government price controls. We assume that the price drops to the marginal cost. Therefore the firm's profit vanishes. The regulator's unit payoff increases to $v(\eta) + 1$.[6] Letting Δ denote the total demand after period two, the regulator's payoff in period three is given by

$$\Delta \cdot \int_{c_L}^{c_H} \int_0^1 \tau(c) \cdot [v(\eta) + 1] \cdot a_2(c, \eta) \; \mathrm{d}G(\eta) \, \mathrm{d}F(c). \tag{8}$$

The regulator maximizes its expected payoff, given by the sum of (6), (7), and (8):

$$U_R(\tau, a_1, a_2; \Delta) \equiv \int_{c_L}^{c_H} \tau(c) \left[v_E \, a_1(c) + \int_0^1 \Phi(a_1(c), \eta; \Delta) \; a_2(c, \eta) \, \mathrm{d}G(\eta) \right] \mathrm{d}F(c), \tag{9}$$

where the coefficient of a_2 in the inner integral is defined as

$$\Phi(a_1, \eta; \Delta) \equiv (\lambda_0 + \lambda_1 \, a_1) \; v(\eta) + \Delta \; [v(\eta) + 1] \, . \tag{10}$$

The monotonicity of $v(\eta)$ and assumption (5) imply the following "single-crossing" property of the function Φ.

[5] Many conditionally approved drugs show toxicity in phase I and II trials. For example, many conditionally approved drugs target cancer, such as chemotherapy, which is inherently toxic but potentially effective. The key questions are whether the drug is sufficiently efficacious relative to its toxicity, and whether this justifies spending on the drug.

[6] Recall that the profit margin before period three is normalized to 1.

LEMMA 1. *For any* $a_1 \in [0,1]$ *and* $\Delta \geq 0$*, the function* $\Phi(a_1, \cdot; \Delta)$ *is strictly increasing, and satisfies*
$\Phi(a_1, 0; \Delta) < 0 < \Phi(a_1, 1; \Delta)$.

In light of Lemma 1, we can define, for each level of conditional approval $a_1 \in [0,1]$, the "no-leniency" threshold for final approval as the unique solution to the equation $\Phi\big(a_1, \cdot\big) = 0$, i.e.,

$$\eta_{\text{NL}}(a_1) \equiv v^{-1}\left(-\frac{\Delta}{\lambda_0 + \lambda_1\, a_1 + \Delta}\right). \tag{11}$$

If $\Delta = 0$, the no-leniency threshold becomes

$$\bar{\eta} \equiv v^{-1}(0). \tag{12}$$

A regulator that could renege on its initial commitment after seeing the realization of η would always set the final approval threshold at the no-leniency level, because this threshold maximizes its continuation payoff after any amount of conditional approval granted in period one.

Under the assumption that any commitment made at the beginning of the first period is irrevocable, the regulator's problem can be stated as

$$\begin{aligned} U_R^*(\Delta) \equiv \max_{\tau, a_1, a_2} \; & U_R(\tau, a_1, a_2; \Delta), \\ & \text{subject to (IC), (IR), (2), (3), and (4).} \end{aligned} \tag{13}$$

The optimization problem in (13) is infinite-dimensional and non-convex, hence well beyond the reach of standard solvers. Our approach consists in partitioning the parameter space into three regions, and using a different approach for each region.

In Section 3, we show that, if the drug is "low-cost," i.e., the regulator's belief about the firm's testing cost is concentrated on relatively low values, all (IC) and (IR) constraints in (13) can be ignored. In this case, the highest cost is not too high to be excluded. The problem is solved by policies that provide no *costly* incentive to the firm, i.e., no leniency, and conditional approval only if the belief in the drug's efficacy is sufficiently optimistic.

In Section 4, we focus on the version of the model where the testing cost could be high, i.e., the regulator's belief is not as optimistic as the previous case, and there is no third period, i.e., $\Delta = 0$. We reformulate the problem in (13) as a linear optimization problem, and provide analytical expressions of optimal policies for all parameter values in this second region. The optimal solution can be implemented by a "take-it-or-leave-it" policy which specifies a conditional approval scale and a final approval threshold, in exchange for testing.

In Section 5, we let $\Delta > 0$ and provide a linear relaxation of the regulator's problem, which yields an upper bound for the regulator's expected payoff. Based on the solution to the upper-bound

problem, we construct a take-it-or-leave-it policy that is feasible for the original problem (13). This policy provides a lower bound on the regulator's expected payoff. Under the calibrated model parameters, the lower bound is within 97.5% of the upper bound. Moreover, when $\Delta = 0$, the upper and lower bounds coincide with the optimal objective value.

Optimal conditional approval and leniency may depend on the length of exclusivity. Longer exclusivity raises firm profits but delays savings for the regulator. As a result, longer exclusivity can strengthen development incentives and reduce the need for conditional approval and leniency, all else equal. We report how optimal strategies vary with exclusivity length in Appendix EC.3.

Before closing this section, we provide an in-depth discussion of the main assumptions in the model.

2.1. Discussion of the Model

In this section, we discuss features of the model that correspond to the institutional detail in the Appendix EC.1. The interaction between the regulator and the firm testing a new drug unfolds over two periods. Later, we add a third period in which the price falls. Period one begins at time 0, and period two begins at time t_1 (Figure 1).

At time 0, the firm has already conducted preliminary tests, which are informative but insufficient to yield statistically significant results. Consequently, the regulator's decision to approve the drug is complicated by residual uncertainty about the efficacy of the drug (Fact 2). Throughout the game, the firm's belief about the drug's efficacy (η) is shared with the regulator.

A key question is whether efficacy is high enough to justify any toxicity. Some toxicity is acceptable for serious diseases with limited treatment options, e.g., chemotherapy is inherently toxic but potentially effective. Toxicity is generally known after phase I and II testing, and is especially well known for new indications of an already approved drug, such as a cancer drug tested for a different type of cancer.

While efficacy is publicly observed, confirmatory testing costs may be private for two reasons. First, joint development makes it hard to attribute costs to a single drug. Second, rare-disease drugs are often first in class and lack a predicate (Fact 5).

At time 0, the regulator can grant conditional approval in exchange for confirmatory testing (Fact 3), which allows the drug to be sold to a fraction of the market (Fact 4).

If there is no confirmatory testing, either because the regulator does not grant conditional approval or the firm decides to exit, the game ends with both parties earning zero payoffs. If instead, the regulator requests testing and the firm accepts, the firm can start selling the drug in period one, up to the market share conditionally approved. Testing is the only source of information about efficacy; they do not learn more about the drug's efficacy from sales (Fact 6).

The firm has monopoly power until t_2 (Fact 7) after which the price falls to marginal cost due to generic entry or price controls (Figure 1). In Appendix EC.3, we discuss how changes in the monopoly-pricing period affect the optimal policy.

Revenue rises gradually in the years following approval as doctors and patients gradually become aware of the drug (Fact 8), as shown in Figure 2(a). Because of rising revenue, the firm benefits from conditional approval not just because it can begin selling earlier, but also because its later sales are higher.

For drugs with high testing costs, profit can be positive only if the drug is conditionally approved at a large scale (Fact 9), as illustrated in Figure 2(b). The cumulative profit in the case of high testing costs without conditional approval remains negative for the duration of the game.

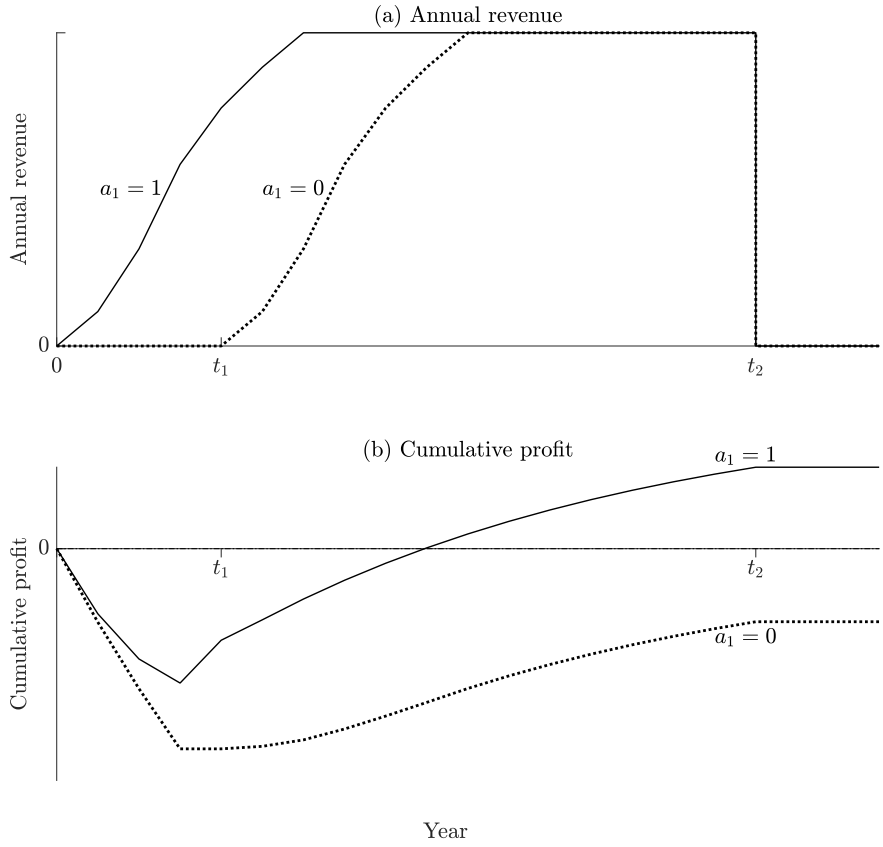

Figure 2 Annual revenue (top) and cumulative profit (bottom) depend on whether the regulator conditionally approves ($a_1 > 0$) or not ($a_1 = 0$) (assuming final approval). Source: Authors' analysis.

The regulator makes three decisions: (i) whether to induce the firm to conduct confirmatory testing, (ii) at what scale the firm can enter the market before final approval, and (iii) where to

set the efficacy threshold for final approval. In the next three sections, we characterize optimal regulatory policies for various configurations of parameter values.

3. Low Testing Cost

In this section, we focus on cases in which the regulator's belief about the firm's testing cost is relatively optimistic. For these "low testing cost" parameter configurations, the firm finds it profitable to participate without any leniency, even at the highest possible testing cost c_H.

DEFINITION 1. Let $\check{U}_R(a_1, \eta_m)$ denote the regulator's expected payoff conditional on testing, generated by conditional approval level a_1 and final approval threshold η_m,[7] i.e.,

$$\check{U}_R(a_1, \eta_m) \;\equiv\; v_E\, a_1 + \int_{\eta_m}^{1} \Big[(\lambda_0 + \lambda_1\, a_1)\, v(\eta) + \Delta\, [v(\eta) + 1]\Big]\, \mathrm{d}G(\eta), \quad \forall (a_1, \eta_m) \in [0,1]^2. \tag{14}$$

We say that a drug is

- "promising", if

$$\check{U}_R(0, \eta_{\mathrm{NL}}(0)) \;\leq\; \check{U}_R(1, \eta_{\mathrm{NL}}(1)); \tag{15}$$

- "not promising", if

$$\check{U}_R(0, \eta_{\mathrm{NL}}(0)) \;\geq\; \check{U}_R(1, \eta_{\mathrm{NL}}(1)); \tag{16}$$

- "low-cost", if either (15) holds and

$$c_H \leq 1 + (\lambda_0 + \lambda_1)\, [1 - G(\eta_{\mathrm{NL}}(1))]\,, \tag{17}$$

  or (16) holds and

$$c_H \leq \lambda_0\, [1 - G(\eta_{\mathrm{NL}}(0))]\,. \tag{18}$$

□

The expression in (14) is obtained from the expression within the square brackets in (9), after substituting $a_1(c) = a_1$ and $a_2(c, \eta) = \mathbb{1}_{\eta \geq \eta_m}$. It represents the regulator's expected payoff when the firm runs confirmatory testing, is granted conditional approval at scale a_1, and is promised final approval only if the revealed efficacy η exceeds the threshold η_m. The definition of "promising," as specified in (15), implies that the regulator achieves a higher expected payoff by granting full conditional approval ($a_1 = 1$) than not ($a_1 = 0$).

The next proposition establishes that it is optimal to provide no leniency if the drug is low-cost.

[7] If we replace any function a_2 with the step function $a_2(c, \eta) = \mathbb{1}_{\eta \geq \eta_m(c)}$, where the threshold η_m is given by

$$1 - G(\eta_m(c)) = \int_0^1 a_2(c, \eta)\, \mathrm{d}G(\eta), \quad \forall c \in [c_L, c_H],$$

the firm's expected profit remains unchanged, and the regulator's expected payoff does not decrease, because the unit payoff function v is increasing. Therefore we can restrict attention without loss of generality to threshold final approval policies.

PROPOSITION 1. *(1) If the drug is promising and low-cost, i.e.,* (15) *and* (17) *hold, it is optimal for the regulator to grant full conditional approval and no leniency, which induces all cost types to test, i.e.,*

$$\forall (c,\eta) \in [c_L, c_H] \times [0,1]: \qquad \tau(c) = 1, \quad a_1(c) = 1, \quad a_2(c,\eta) = \mathbb{1}_{\eta \geq \eta_{\mathrm{NL}}(1)}. \tag{19}$$

(2) If the drug is not promising and low-cost, i.e., (16) *and* (18) *hold, on the other hand, it is optimal for the regulator to grant no conditional approval and no leniency, which induces all cost types to test, i.e.,*

$$\forall (c,\eta) \in [c_L, c_H] \times [0,1]: \qquad \tau(c) = 1, \quad a_1(c) = 0, \quad a_2(c,\eta) = \mathbb{1}_{\eta \geq \eta_{\mathrm{NL}}(0)}. \tag{20}$$

When the drug is low-cost, it is optimal to provide no *costly* incentives to the firm: always no leniency, full conditional approval if the drug is promising, and no conditional approval if the drug is not promising. Therefore the optimal policies in (19) and (20) also solve both "first-best" problem, where the incentive constraints (IC) are ignored, and the "dictatorial" problem where both (IC) and (IR) are ignored. We record this result in the next corollary.

COROLLARY 1. *The "dictatorial" problem, defined as* (13) *without* (IC) *and* (IR)*, and the "first-best" problem, defined as* (13) *without* (IC)*, are solved by the policy in* (19) *under conditions* (15) *and* (17)*, and by the policy in* (20) *under conditions* (16) *and* (18)*.*

If the inequalities in (17) or (18) hold strictly, the optimal policies described in Proposition 1 leave the firm with a strictly positive expected profit, even when its production cost is c_H. This never happens in standard procurement problems, where the (IR) constraint for the worst type always binds.

4. High Testing Cost in the Two-Period Game ($\Delta = 0$)

Starting from this section, we allow the testing cost to violate the low-cost conditions of Definition 1. In other words, the regulator believes the firm's testing cost could be high relative to the drug's potential. Furthermore, we also assume prices never fall from generic entry or price controls. This implies $\Delta = 0$. In the next section, we will relax this assumption about prices falling.

Why assume prices do not fall? Generic competition may never emerge if the drug is displaced by newer technology before patent expiry. Moreover, rare-disease drugs, the majority of conditionally approved treatments (see fact 7), often maintain high prices for two reasons. First, markets are typically too small to support multiple firms. Second, Medicare price controls do not currently apply to rare-disease drugs.

4.1. Optimal Policies

When $\Delta = 0$, we can reformulate the regulator's problem (13) as a linear optimization problem, by introducing the following variables:

$$\forall c \in [c_L, c_H]: \quad \alpha_1(c) \equiv \tau(c)\, a_1(c), \text{ and} \tag{21}$$

$$\forall (c,\eta) \in [c_L, c_H] \times [0,1]: \quad \alpha_2(c,\eta) \equiv \tau(c)\, [\lambda_0 + \lambda_1\, a_1(c)]\, a_2(c,\eta). \tag{22}$$

The feasibility conditions (2)-(4) can be written as

$$\forall c \in [c_L, c_H]: \quad 0 \le \alpha_1(c) \le \tau(c), \tag{23}$$

$$\forall (c,\eta) \in [c_L, c_H] \times [0,1]: \quad 0 \le \alpha_2(c,\eta) \le \lambda_0\, \tau(c) + \lambda_1\, \alpha_1(c). \tag{24}$$

The firm's total expected profit function in (1) becomes

$$\hat{\Pi}(c,c') = \alpha_1(c') - c\,\tau(c') + \int_0^1 \alpha_2(c',\eta)\, \mathrm{d}G(\eta), \tag{25}$$

and the regulator's expected payoff in (9) becomes

$$U_R(\tau,\alpha_1,\alpha_2;\Delta) = \int_{c_L}^{c_H} \left[v_E\, \alpha_1(c) + \int_0^1 v(\eta)\, \alpha_2(c,\eta)\, \mathrm{d}G(\eta) \right] \mathrm{d}F(c). \tag{26}$$

The next proposition establishes that, when $\Delta = 0$, we can transform the original regulator's problem defined in (13) into an equivalent linear optimization formulation.

PROPOSITION 2. *Recall $U_R^*(\Delta)$ defined in* (13)*. We have*

$$U_R^*(0) = \max_{\tau,\alpha_1,\alpha_2} \; U_R(\tau,\alpha_1,\alpha_2;\Delta), \tag{27}$$
$$\text{subject to} \;\; (\mathrm{IC}), (\mathrm{IR}), (2), (23), \text{ and } (24).$$

Furthermore, given an optimal solution $\{\tau^*(c), \alpha_1^*(c), \alpha_2^*(c,\eta)\}$ to (27)*, define*

$$a_1^*(c) \equiv \frac{\alpha_1^*(c)}{\tau^*(c)} \cdot \mathbb{1}_{\tau^*(c)>0} \quad \text{and} \quad a_2^*(c,\eta) \equiv \frac{\alpha_2^*(c,\eta)}{\lambda_0\, \tau^*(c) + \lambda_1\, \alpha_1^*(c)} \cdot \mathbb{1}_{\tau^*(c)>0}. \tag{28}$$

Then, $\{\tau^*(c), a_1^*(c), a_2^*(c,\eta)\}$ is an optimal solution to (13)*.*

The expressions for a_1^* and a_2^* in (28) follow directly from the change of variables defined in (21) and (22), given the optimal solution to (27). Proposition 2 provides a translation between the optimal incentive mechanism that solves (13) and an optimal solution to the linear optimization problem (27).

In the rest of this section, we focus on presenting the optimal solution to (27). A rigorous proof, based on linear programming and duality, is presented in Appendix EC.4.3. Here, we present an intuitive (albeit non-rigorous) line of argument to motivate the result.

We show that the optimal incentive mechanism when $\Delta = 0$ can be implemented as a take-it-or-leave-it offer. That is, the firm is granted a single conditional approval scale a_1 and is promised a single final approval threshold η, in exchange for testing, if and only if its cost is below a cutoff type c_0.

In the previous section, we examined the low-cost cases. Here, we turn to the cases where the testing cost could be high, defined as the complement of the "low-cost" condition in Definition 1, where the regulator may need to screen out a high cost firm. The incentive constraints and optimality conditions imply that a firm of cutoff type c_0 is indifferent between participating and not. That is, its expected revenue across both periods is equal to c_0, i.e.,

$$c_0 = a_1 + (\lambda_0 + \lambda_1\, a_1) \cdot [1 - G(\eta)]\,. \tag{29}$$

Therefore a firm with testing cost below c_0 has no incentive to decline the offer. The regulator's problem can be rewritten as:

$$\begin{aligned} \max_{a_1,\eta,c_0}\ & \hat{U}(a_1,\eta,c_0) \equiv F(c_0) \cdot \left[v_E\, a_1 \,+\, (\lambda_0 + \lambda_1\, a_1) \int_\eta^1 v(y)\, \mathrm{d}G(y) \right] \\ \text{subject to}\ & \ (29) \text{ and } (a_1,\eta,c_0) \in [0,1] \times [0,\bar{\eta}] \times [c_L, c_H], \end{aligned} \tag{30}$$

where $F(c_0)$ denotes the probability that a firm's testing cost is below the cut-off type c_0. Constraint (29) guarantees that such a firm is willing to participate. The term that multiplies $F(c_0)$ represents the regulator's expected payoff from committing to a conditional approval scale a_1 and final approval threshold η, conditional on the firm's participation. It is without loss of generality to restrict the search for the optimal final approval threshold from within the interval $[0, \bar{\eta}]$, following the definition (12) and the fact that v is increasing.

The optimization problem in (30) can be interpreted as a procurement problem in which the regulator wants to buy a new drug with the goal of maximizing total consumer surplus, and has two instruments to incentivize the supplier: i) conditional approval a_1, which generates revenue in period one and stimulates demand after period one, and ii) "leniency" $\eta < \bar{\eta}$, which increases the firm's expected revenue after period one, by lowering the final approval threshold η below its no-leniency level $\bar{\eta}$, thus increasing the probability of final approval.

Now we consider first-order conditions (FOCs) to maximize $\hat{U}$. We first eliminate a_1 using (29), and substitute

$$\tilde{a}_1(\eta, c_0) \equiv \frac{c_0 - \lambda_0\, [1 - G(\eta)]}{1 + \lambda_1\, [1 - G(\eta)]}. \tag{31}$$

in place of a_1 in the regulator's expected payoff $\hat{U}$. The partial derivative with respect to η is

$$\frac{\partial \hat{U}(\tilde{a}_1(\eta,c_0),\eta,c_0)}{\partial \eta} = \frac{(\lambda_0 + \lambda_1\, c_0)\; F(c_0)\, g(\eta)}{\left[1 + \lambda_1\, [1 - G(\eta)]\right]^2} \cdot h(\eta), \tag{32}$$

in which

$$h(\eta) \equiv v_E - v(\eta) + \lambda_1 \cdot \int_{\eta}^{1} [v(y) - v(\eta)] \, \mathrm{d}G(y), \quad \forall \eta \in [0, \bar{\eta}]. \tag{33}$$

It is clear that the sign of h determines the sign of $\hat{U}$'s partial derivative with respect to η. The function h is decreasing and satisfies $h(0) > 0$. Therefore, for any given c_0, the objective function $\hat{U}$ admits a unique maximizer $\eta^*(c_0) \in (0, \bar{\eta}]$. If the drug is promising, i.e., $h(\bar{\eta}) \geq 0$, then the derivative is nonnegative over the entire interval, and the maximizer is attained at the upper bound, i.e., $\eta^*(c_0) = \bar{\eta}$. Conversely, if the drug is not promising, i.e., $h(\bar{\eta}) < 0$, then the maximizer $\eta^*(c_0)$ corresponds to the unique root of h in the interval $(0, \bar{\eta})$.

The function h plays a central role in determining the testing and final approval thresholds under the optimal regulatory policy. Before presenting the details, it is convenient to define the following two functions, from setting a_1 in the right-hand side of (29) to be 0 and 1, respectively:

$$\underline{c}(\eta) \equiv \lambda_0 \left[1 - G(\eta)\right], \quad \forall \eta \in [0, 1], \tag{34}$$

and

$$\bar{c}(\eta) \equiv 1 + (\lambda_0 + \lambda_1) \left[1 - G(\eta)\right], \quad \forall \eta \in [0, 1]. \tag{35}$$

The following lemma defines the critical testing threshold on cost c and the final approval threshold on η, to be used in the optimal mechanism.

LEMMA 2. *The function h defined in (33) is strictly decreasing with $h(0) > 0$, leading to the following results:*[8]

(1) If $h(\bar{\eta}) \leq 0$, there exists a unique $\check{\eta} \in [0, \bar{\eta}]$ such that $h(\check{\eta}) = 0$. Define

$$\check{c} \equiv \sup \left\{ c \in [c_L, c_H] : \frac{\mathrm{d}\hat{U}(\tilde{a}_1(\check{\eta}, c), \check{\eta}, c)}{\mathrm{d}c} \geq 0 \right\}. \tag{36}$$

(2) If $h(\bar{\eta}) \leq 0$ and $\check{c} < \underline{c}(\check{\eta})$, there exists a unique $\tilde{\eta} \in [\check{\eta}, \bar{\eta}]$ such that

$$\left. \frac{\mathrm{d}\hat{U}\big(0, \eta, \underline{c}(\eta)\big)}{\mathrm{d}\eta} \right|_{\eta = \tilde{\eta}} = 0, \tag{37}$$

and we have $h(\tilde{\eta}) \leq 0$.

(3) If $h(\bar{\eta}) > 0$ or $h(\bar{\eta}) \leq 0$ and $\check{c} > \bar{c}(\check{\eta})$, define

$$\underline{\eta} \equiv \inf \left\{ \eta \in [0, \check{\eta}] : \frac{\mathrm{d}\hat{U}(1, \eta, \bar{c}(\eta))}{\mathrm{d}\eta} \leq 0 \right\}. \tag{38}$$

We have $h(\underline{\eta}) \geq 0$.

[8] We extend the definition of $F(c)/f(c)$ for $c \notin [c_L, c_H]$ as follows: $F(c)/f(c) = 0$ for $c < c_L$ and $F(c)/f(c) = 1/f(c_H)$ for $c > c_H$.

Intuitively, the pair of testing threshold $\check{c}$ and final approval threshold $\check{\eta}$ maximizes the principal's expected payoff $\hat{U}$ if we ignore feasibility constraints. The final approval thresholds $\tilde{\eta}$ and $\underline{\eta}$ correspond to the maximizers of $\hat{U}$ when the conditional approval scales are fixed at 0 and 1, respectively, and the cutoff types c_0 are set to $\underline{c}(\tilde{\eta})$ and $\bar{c}(\underline{\eta})$, respectively.

The next proposition characterizes the optimal incentive mechanism for all parameter values.

PROPOSITION 3. *Define three values, a conditional approval scale α^*, a testing threshold c^*, and a final approval threshold η^*, that depend on model parameters in the following cases.*

(1) If $h(\bar{\eta}) \leq 0$ and $\underline{c}(\check{\eta}) \leq \check{c} \leq \bar{c}(\check{\eta})$, we have $c^* = \check{c}$ and $\eta^* = \check{\eta}$, and the optimal conditional approval scale below the threshold is $\alpha^* = \tilde{a}_1(\eta^*, c^*)$, where $\tilde{a}_1$ is defined by (31).

(2) If $h(\bar{\eta}) \leq 0$ and $\check{c} < \underline{c}(\check{\eta})$, the regulator grants no conditional approval, i.e., $\alpha^* = 0$, and the thresholds c^* and η^* are determined as follows:

(a) if $\underline{c}(\tilde{\eta}) < c_L$, we have $c^* = c_L$ and $\eta^* = 1$;

(b) if $c_L \leq \underline{c}(\tilde{\eta}) < c_H$, we have $c^* = \underline{c}(\tilde{\eta})$ and $\eta^* = \tilde{\eta}$; and

(c) if $\underline{c}(\bar{\eta}) < c_H \leq \underline{c}(\tilde{\eta})$, we have $c^* = c_H$ and $\eta^* = \sup\{\eta \in [0,1] : \underline{c}(\eta) \geq c_H\}$.

(3) If $h(\bar{\eta}) > 0$ or $h(\bar{\eta}) \leq 0$ and $\check{c} > \bar{c}(\check{\eta})$, the regulator grants full conditional approval to firms whose types are below the threshold, i.e., $\alpha^* = 1$. The thresholds c^* and η^* are determined as follows:

(a) if $\bar{c}(\underline{\eta}) < c_L$, we have $c^* = c_L$ and $\eta^* = 1$;

(b) if $c_L \leq \bar{c}(\underline{\eta}) < c_H$, we have $c^* = \bar{c}(\underline{\eta})$ and $\eta^* = \underline{\eta}$; and

(c) if $\bar{c}(\bar{\eta}) < c_H \leq \bar{c}(\underline{\eta})$, we have $c^* = c_H$ and $\eta^* = \sup\{\eta \in [0,1] : \bar{c}(\eta) \geq c_H\}$.

The following solution is optimal to (27):

$$\tau^*(c) = \mathbb{1}_{c \in [c_L, c^*)}, \quad \alpha_1^*(c) = \alpha^* \cdot \mathbb{1}_{c \in [c_L, c^*)}, \quad \alpha_2^*(c, \eta) = (\lambda_0 + \lambda_1 \alpha^*) \cdot \mathbb{1}_{(c,\eta) \in [c_L, c^*) \times [\eta^*, 1]}. \tag{39}$$

Cases (2) and (3) in Proposition 3 do not consider conditions $c_H \leq \underline{c}(\bar{\eta})$ or $c_H \leq \bar{c}(\bar{\eta})$, respectively, because these are special cases of Proposition 1 for $\Delta = 0$. It can be verified that when $\Delta = 0$, conditions (15)-(18) become $h(\bar{\eta}) \geq 0$, $h(\bar{\eta}) \leq 0$, $c_H \leq \bar{c}(\bar{\eta})$, and $c_H \leq \underline{c}(\bar{\eta})$, respectively.

Consider case (2) of Proposition 3. When $h(\bar{\eta}) \leq 0$ and $c_H \leq \underline{c}(\bar{\eta})$, we have

$$\check{c} \leq c_H \leq \underline{c}(\bar{\eta}) \leq \underline{c}(\check{\eta}).$$

Therefore, the condition $\check{c} < \underline{c}(\check{\eta})$ is already implied, and the optimal mechanism corresponds to case (2) in Proposition 1.

Next, consider case (3) of Proposition 3. When $c_H \leq \bar{c}(\bar{\eta})$, we have

$$\check{c} \leq c_H \leq \bar{c}(\bar{\eta}) \leq \bar{c}(\check{\eta}),$$

implying that the condition $\check{c} > \bar{c}(\check{\eta})$ is irrelevant. The optimal mechanism corresponds to case (1) in Proposition 1.

4.2. Calibration

Here, we use the calibrated model parameters with $\Delta = 0$. Figure 3 shows how the optimal policy changes as the drug's expected effectiveness increases. In this and later figures, except Figure 6, we use the calibrated values of the model parameters from Table EC.3 in Appendix EC.2, a uniform distribution of private information c, the cdf $G(\eta) = \eta^k$ for various values of k to obtain the corresponding ex ante expected values v_E, and a linear regulator's utility function v. We plot c^*, α^*, and η^* against v_E.

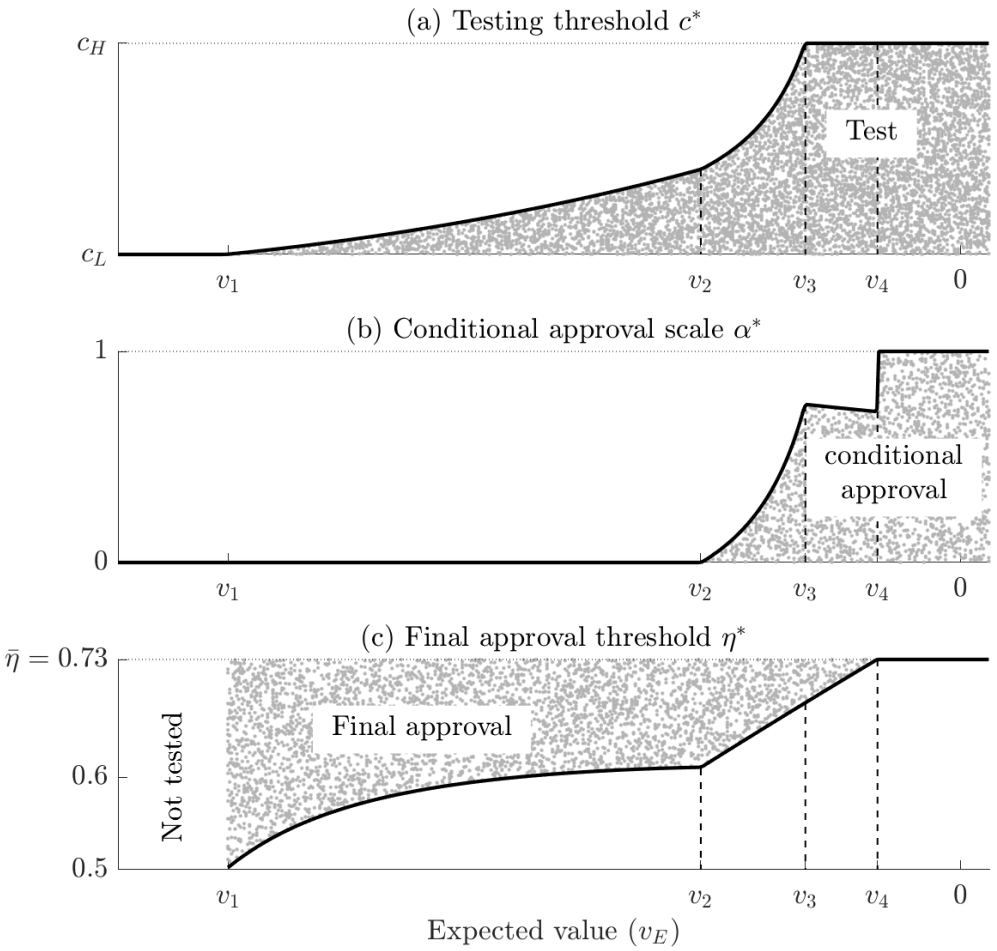

Figure 3 Drugs with high expected value and low testing cost are tested (top panel), receive conditional approval (middle panel), and have a high final approval threshold, meaning less leniency (bottom panel).

We partition the v_E space into five intervals. In the first interval ($v_E < v_1$), the drug's expected effectiveness is so low that it is not worth inducing testing. Therefore, in this first interval, the drug will never be tested ($c^* = c_L$), and hence the regulator grants no conditional approval ($\alpha^* = 0$). In

the second interval, $v_E \in [v_1, v_2)$, the optimal policy grants no conditional approval but induces the firm to test if the testing cost is relatively low by providing some leniency, i.e., $\eta^* < \bar{\eta}$. In the third interval $[v_2, v_3)$, as the expected value v_E increases, the regulator progressively grants greater scales of conditional approval while becoming more stringent on the final approval threshold. Because the scale of conditional approval increases, the cutoff type c^* also increases. In the fourth interval $v_E \in [v_3, v_4)$, the regulator reduces the scale of conditional approval and continues raising the final approval threshold. The firm is now willing to test regardless of the cost ($c^* = c_H$). Here, the regulator lowers the conditional approval scale α^* because the higher likelihood of final approval already provides sufficient incentive for the firm to test. Finally, in the fifth interval $v_E \in [v_4, v(1)]$, the assumptions of Proposition 1 hold, hence, the optimal policy coincides with the "first-best" policy. The regulator grants full conditional approval and no leniency, and the firm tests regardless of the cost.

Note that the scale of conditional approval, α^*, exhibits a discontinuous jump at $v_E = v_4$. The intuition behind this behavior is as follows. When v_E approaches v_4 from the left, the final approval threshold converges to $\bar{\eta}$. Correspondingly, the coefficient of a_1 in the regulator's payoff function (9) is given by

$$v_E + \lambda_1 \int_{\bar{\eta}}^{1} v(\eta)\, \mathrm{d}G(\eta),$$

which crosses 0 at v_4 from below. Consequently, it becomes optimal to grant full conditional approval, i.e., $\alpha^* = 1$, when v_E increases beyond v_4.

The regulator's final approval threshold (η^*) is lower for drugs with lower expected value (v_E) under the parameters in our calibration exercise (Figure 3(c)). Under a higher expected value, the regulator can set a higher final approval threshold without discouraging the firm from testing. The particularly interesting observation in Figure 3(c) is that the threshold η^* may be strictly less than the myopic threshold $\bar{\eta}$. Recall that if the efficacy is at $\bar{\eta}$, the drug's societal value $v(\bar{\eta}) = 0$. It may be counterintuitive that the regulator should grant final approval to a drug whose societal value is negative. The reason is that if the ex ante effectiveness v_E is fairly negative and the regulator still holds the final approval threshold at $\bar{\eta}$, the chance of success is too low for the firm to be willing to test. This may forfeit the possibility of developing a drug that could turn out to be efficacious. In order to ensure sufficient profit for the firm to invest in testing, the regulator can either use conditional approval or commit to a final approval threshold that is lower than the "no-leniency" threshold $\bar{\eta}$. That is, committing to the final approval of a drug with societal value $v(\eta) < 0$ can be interpreted as the society sharing part of the testing cost with the firm, in the hope that the drug may ultimately prove more efficacious.

In practice, missing the threshold might mean that the trials were not conclusive in demonstrating a benefit for the average patient, but the regulator asks for no more. The regulator can commit to

a lower threshold during discussions with the firm before confirmatory testing. Indeed, the FDA grants what it calls "breakthrough status" to certain drugs, providing the developer with timely advice on clinical trial design (Fact 11).

4.3. Benchmark: First-Best Optimal Policies

How do the optimal policies under private information compare to the first-best policies, where the regulator observes the firm's testing cost? The answer depends on whether the regulator believes testing costs are low or high. The previous section showed that the first-best and second-best solutions are the same under low-cost beliefs (see Corollary 1). We now examine the case where testing costs may be high. More precisely, these first-best policies solve the regulator's problem (13) with $\Delta = 0$, without the incentive compatibility constraints (IC). That is, we define the first-best optimization problem as

$$U_{FB} \equiv \max_{\tau, a_1, a_2} U_R(\tau, a_1, a_2; 0), \quad \text{subject to (IR), (2), (3), and (4).} \tag{40}$$

Because the firm's testing cost is potentially high, any first-best policy generates zero expected profit for the firm, thus solving the optimization problem (30) with $c_0 = c$. Among the multiple combinations of leniency and conditional approval that yield zero expected profit, it is optimal to use the combination that maximizes the regulator's expected payoff.

The next proposition provides a formal characterization of the first-best policy when $\Delta = 0$.

PROPOSITION 4. *Suppose that either* (15) *and* $c > 1 + (\lambda_0 + \lambda_1)\,[1 - G(\eta_{\mathrm{NL}}(1))]$ *hold together, or* (16) *and* $c > \lambda_0\,[1 - G(\eta_{\mathrm{NL}}(0))]$ *hold together. The combinations of conditional approval* a_1^* *and final approval threshold* η^**, as defined below, solve the optimization problem on the right-hand side of* (40)*, when restricting to policies that induce testing.*

(1) If the drug is promising, i.e., $h(\bar{\eta}) \geq 0$*, then:*
- *(a) if* $\bar{c}(\bar{\eta}) < c \leq \bar{c}(0)$*, we have* $a_1^* = 1$ *and* $\eta^* = G^{-1}\left(1 - \dfrac{c-1}{\lambda_0 + \lambda_1}\right)$*; and*
- *(b) if* $c > \bar{c}(0)$*, the regulator cannot induce the firm to conduct testing.*

(2) If the drug is not promising, i.e., $h(\bar{\eta}) \leq 0$*, then:*
- *(a) if* $\underline{c}(\bar{\eta}) < c < \underline{c}(\check{\eta})$*, we have* $a_1^* = 0$ *and* $\eta^* = G^{-1}\left(1 - \dfrac{c}{\lambda_0}\right)$*;*
- *(b) if* $\underline{c}(\check{\eta}) \leq c \leq \bar{c}(\check{\eta})$*, we have* $a_1^* = \tilde{a}_1(\check{\eta}, c)$ *and* $\eta^* = \check{\eta}$*;*
- *(c) if* $\bar{c}(\check{\eta}) < c \leq \bar{c}(0)$*, we have* $a_1^* = 1$ *and* $\eta^* = G^{-1}\left(1 - \dfrac{c-1}{\lambda_0 + \lambda_1}\right)$*; and*
- *(d) if* $c > \bar{c}(0)$*, the regulator cannot induce the firm to conduct testing.*

It is optimal to induce testing (i.e., $\tau = 1$*) if and only if the regulator's resulting expected payoff is nonnegative.*

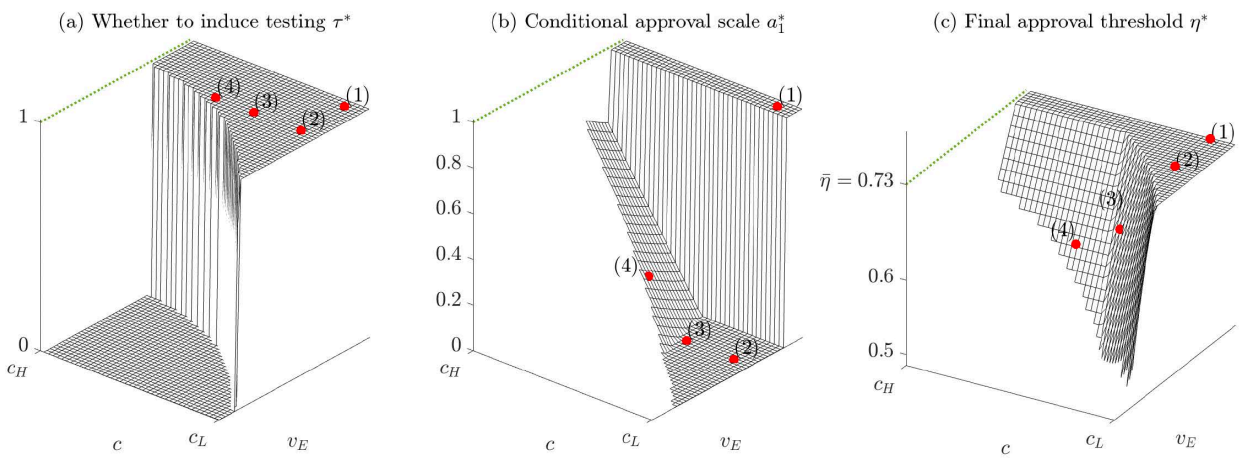

Figure 4 **First-best policies for various values of the regulator's expected unit payoff v_E and the testing cost c.**

Figure 4 illustrates first-best optimal policies for four representative parameter configurations. At point (1), the first-best policy grants full conditional approval and no leniency. At point (2), the drug is less ex-ante efficacious (while the testing cost remains unchanged), so the first-best policy grants neither conditional approval nor leniency. These two points correspond to the two cases considered in Proposition 1 and Corollary 1. At point (3), the testing cost is higher (while efficacy remains the same as at point 2), so it is optimal to grant no conditional approval, and some leniency is required to induce testing. This corresponds to case (2)(a) in Proposition 4. At point (4), the even higher testing cost c requires a combination of leniency and a positive scale of conditional approval to induce testing. This corresponds to case (2)(b) in Proposition 4.

Next, we illustrate the value of information, i.e., the gap between the regulator's expected payoffs in the first-best and second-best scenarios (Figure 5). When the belief about efficacy is pessimistic (e.g., $v_E < -40$), the value of information is small. This indicates that learning the firm's true testing cost provides little benefit to the regulator, because only limited information rent is paid in this case. When the belief is optimistic (e.g., $v_E > -8$), the value of information is zero because the first-best and second-best policies coincide, as explained in Corollary 1. When the belief is intermediate, the value of information first increases and then decreases as the drug becomes more efficacious (e.g., v_E increases from -40 to -8), reflecting the regulator's willingness to pay to learn the true testing cost.

5. High Testing Cost in the Three-Period Game ($\Delta > 0$)

We now consider the case of potentially high testing costs and a third period in which the drug price falls due to generic competition or price controls. A lower price in the third period implies an even higher net benefit to the regulator. For $\Delta > 0$, the regulator's problem (13) cannot be formulated as an equivalent linear optimization as in Section 4.1. Therefore, we present a linear

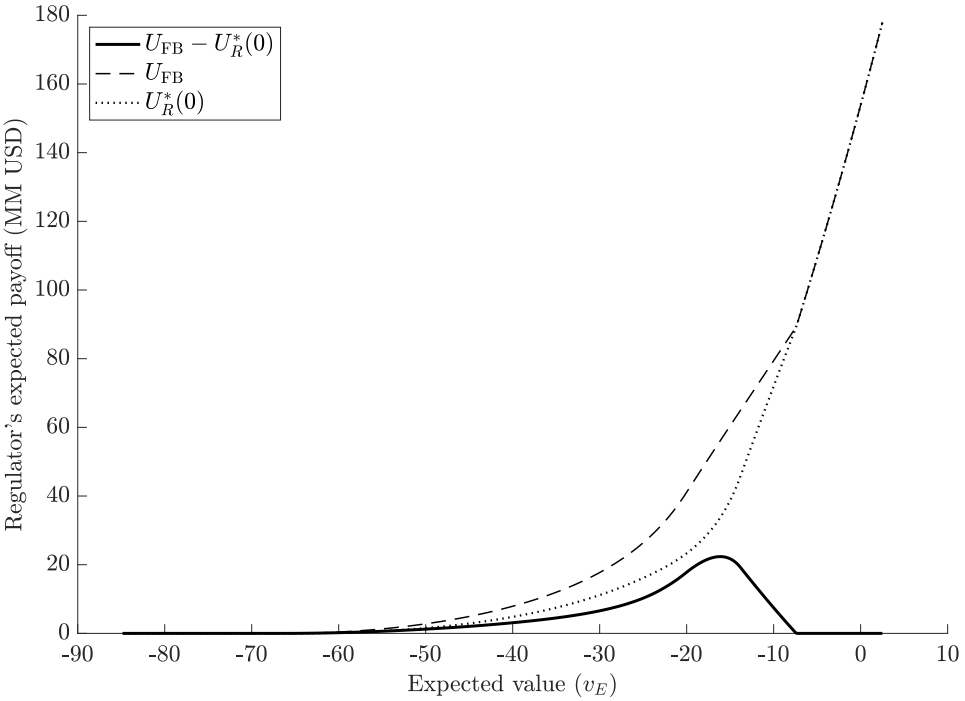

Figure 5 **The regulator's expected payoff in the first-best (U_{FB}) and second-best ($U_R^*(0)$), and the value of information ($U_{FB} - U_R^*(0)$) as beliefs about efficacy η change.**

optimization relaxation, which not only provides an upper bound for $U_R^*(\Delta)$ but also allows us to construct a feasible mechanism.

5.1. Upper-Bound Linear Optimization

In addition to α_1 defined in (21), we define the following new decision variables for any $(c, \eta) \in [c_L, c_H] \times [0, 1]$:

$$\hat{\alpha}_2(c, \eta) \equiv \tau(c)\, a_2(c, \eta), \text{ and} \tag{41}$$

$$\hat{\alpha}_3(c, \eta) \equiv a_1(c)\, \hat{\alpha}_2(c, \eta) = \alpha_1(c)\, a_2(c, \eta). \tag{42}$$

Using these notations, the regulator's payoff function (9) becomes

$$\hat{U}_R(\tau, \alpha_1, \hat{\alpha}_2, \hat{\alpha}_3; \Delta) \equiv \int_{c_L}^{c_H} \left[v_E\, \alpha_1(c) + \int_0^1 \left[(\lambda_0 + \Delta)\, v(\eta) + \Delta \right] \hat{\alpha}_2(c, \eta)\, \mathrm{d}G(\eta) \right. \\ \left. + \lambda_1 \int_0^1 v(\eta)\, \hat{\alpha}_3(c, \eta)\, \mathrm{d}G(\eta) \right] \mathrm{d}F(c), \tag{43}$$

and the firm's payoff function (1) becomes

$$\hat{\Pi}(c, c') = \alpha_1(c') - c\, \tau(c') + \int_0^1 \left[\lambda_0\, \hat{\alpha}_2(c', \eta) + \lambda_1\, \hat{\alpha}_3(c', \eta) \right] \mathrm{d}G(\eta). \tag{44}$$

We further impose the following constraints on the decision variables $\hat{\alpha}_2$ and $\hat{\alpha}_3$: for any $(c,\eta) \in [c_L, c_H] \times [0,1]$, we require

$$0 \le \hat{\alpha}_2(c,\eta) \le \tau(c), \tag{45}$$

$$0 \le \hat{\alpha}_3(c,\eta) \le \hat{\alpha}_2(c,\eta), \text{ and} \tag{46}$$

$$\hat{\alpha}_3(c,\eta) \le \alpha_1(c). \tag{47}$$

The regulator's upper-bound problem can be formulated as the following linear optimization problem:

$$\begin{aligned} \bar{U}_R^*(\Delta) \equiv \max_{\tau,\alpha_1,\hat{\alpha}_2,\hat{\alpha}_3} \ & \hat{U}_R(\tau,\alpha_1,\hat{\alpha}_2,\hat{\alpha}_3;\Delta), \\ \text{subject to } \ & \text{(IC)}, \text{(IR)}, (2), (23), (45), (46), \text{ and } (47), \end{aligned} \tag{48}$$

where $\hat{U}_R$ and $\hat{\Pi}$ follow (43) and (44), respectively.

The next proposition formalizes that the objective function value of the optimization problem (48) is an upper bound for the original problem (13).

PROPOSITION 5. *$\bar{U}_R^*(\Delta) \ge U_R^*(\Delta)$ for any $\Delta \ge 0$.*

Under Assumption 1, the optimal solution to the regulator's upper-bound problem (48) is fully characterized by four constants: $\hat{c}$, $\hat{\alpha}$, $\hat{\eta}_1$, and $\hat{\eta}_2$. The following proposition summarizes this result.

PROPOSITION 6. *Consider the optimization problem (48) under Assumption 1. There exists an optimal solution $\{\hat{\tau}(c), \hat{\alpha}_1(c), \hat{\alpha}_2(c,\eta), \hat{\alpha}_3(c,\eta)\}$ that can be summarized by four values: a cost threshold $\hat{c} \in [c_L, c_H]$, a scale of conditional approval $\hat{\alpha} \in [0,1]$, and two efficacy thresholds $\hat{\eta}_1$ and $\hat{\eta}_2$ with $0 \le \hat{\eta}_1 \le \hat{\eta}_2 \le \bar{\eta}$, as follows: for any $(c,\eta) \in [c_L, c_H] \times [0,1]$,*

$$\hat{\tau}(c) = \mathbb{1}_{c<\hat{c}}, \quad \hat{\alpha}_1(c) = \hat{\alpha} \cdot \mathbb{1}_{c<\hat{c}}.$$

Furthermore, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ demonstrate one of the following two cases:

Case 1:

$$\hat{\alpha}_2(c,\eta) = \mathbb{1}_{(c,\eta)\in[c_L,\hat{c})\times[\hat{\eta}_1,1]}, \quad \hat{\alpha}_3(c,\eta) = \hat{\alpha} \cdot \mathbb{1}_{(c,\eta)\in[c_L,\hat{c})\times[\hat{\eta}_2,1]}; \tag{49}$$

Case 2:

$$\hat{\alpha}_2(c,\eta) = \begin{cases} 1, \ \forall (c,\eta) \in [c_L,\hat{c}) \times [\hat{\eta}_2, 1], \\ \hat{\alpha}, \ \forall (c,\eta) \in [c_L,\hat{c}) \times [\hat{\eta}_1, \hat{\eta}_2), \\ 0, \ \textit{otherwise}, \end{cases} \quad \hat{\alpha}_3(c,\eta) = \hat{\alpha} \cdot \mathbb{1}_{(c,\eta)\in[c_L,\hat{c})\times[\hat{\eta}_1,1]}. \tag{50}$$

The proof of the proposition, presented in Appendix EC.4.4, provides closed-form expressions of $\hat{c}$, $\hat{\alpha}$, $\hat{\eta}_1$, and $\hat{\eta}_2$, which vary depending on the model parameters. For simplicity of exposition, we leave the detailed expressions to Proposition EC.1 in the appendix.

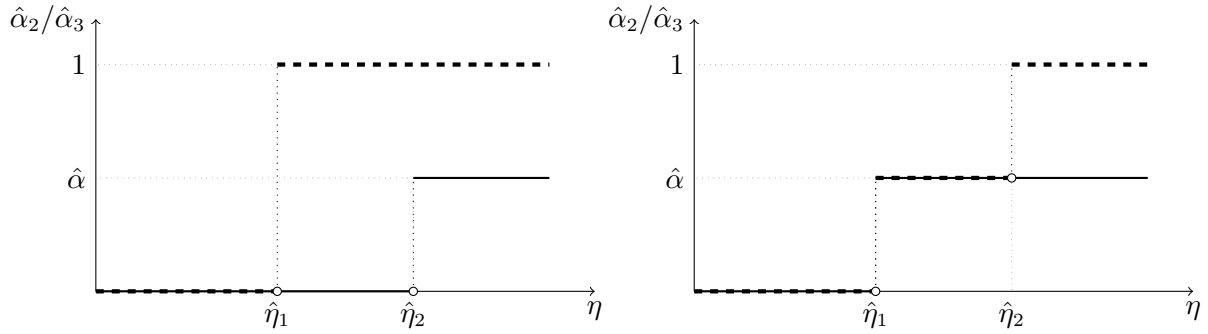

Figure 6 Illustration of $\hat{\alpha}_2$ and $\hat{\alpha}_3$ from Proposition 6. The left and right panels correspond to cases 1 and 2, respectively, with the dashed and solid lines representing $\hat{\alpha}_2$ and $\hat{\alpha}_3$, respectively, for a given $c \in [c_L, \hat{c})$.

It is worth mentioning that the solution described in Proposition 6 may not be feasible to the original problem (13). To explain this, we observe Figure 6, which illustrates the two cases of the optimal $\hat{\alpha}_2$ and $\hat{\alpha}_3$ for a given $c \in [c_L, \hat{c})$ as described in Proposition 6. As shown in the figure, the thresholds $\hat{\eta}_1$ and $\hat{\eta}_2$ may differ. As a result, we cannot identify a final approval policy based on $\hat{\alpha}_2$ and $\hat{\alpha}_3$ that achieves the optimal value of the objective function in (48).

As a special case, if the optimal $\hat{\alpha} = 0$, then $\hat{\eta}_2$ and $\hat{\eta}_1$ are irrelevant in Cases 1 and 2, respectively. This implies that the upper-bound optimal solution is feasible to the original problem (13). In particular, the principal offers no conditional approval while including a firm with cost less than $\hat{c}$ to test. If Case 1 (resp. Case 2) of Proposition 6 holds, then the final approval decision follows the threshold $\hat{\eta}_1$ (resp. $\hat{\eta}_2$). This result is summarized in the following corollary.

COROLLARY 2. *If the optimal solution described in Proposition 6 satisfies $\hat{\alpha} = 0$, we have $\bar{U}_R^*(\Delta) = U_R^*(\Delta)$.*

If $\Delta = 0$, the two thresholds $\hat{\eta}_1$ and $\hat{\eta}_2$ always coincide,[9] ensuring that the optimal solution to the upper-bound problem remains feasible for the original one. Hence, the upper bound $\bar{U}_R^*$ is tight when $\Delta = 0$. This property is documented in the following corollary.

COROLLARY 3. *We have $\bar{U}_R^*(0) = U_R^*(0)$.*

5.2. Feasible Mechanism and Lower Bound

When the drug is low-cost (Definition 1), the policies described in Proposition 1 are optimal. For drugs with potentially high testing costs, the following proposition states that, given the optimal solution to the upper-bound problem (48), we can always construct a feasible solution to the original problem (13), with its objective function value serving as a lower bound.

[9] See Lemmas EC.3, EC.4, and EC.6 for details.

PROPOSITION 7. *Let* $\{\hat{\tau}(c), \hat{\alpha}_1(c), \hat{\alpha}_2(c,\eta), \hat{\alpha}_3(c,\eta)\}$ *denote the optimal solution to the upper-bound problem* (48) *described in Proposition 6 (i.e.,* $\hat{\tau}(c) = \mathbb{1}_{c<\hat{c}}$ *and* $\hat{\alpha}_1(c) = \hat{\alpha} \cdot \mathbb{1}_{c<\hat{c}}$*). For any* $c \in [c_L, \hat{c})$*, define a threshold* $\eta_1 \in [0,1]$ *as the unique solution to*

$$\lambda_0 \int_0^1 \hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta) + \lambda_1 \int_0^1 \hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta) = (\lambda_0 + \lambda_1\,\hat{\alpha})\,[1 - G(\eta_1)]\,. \tag{51}$$

Based on these, define

$$a_1'(c) \equiv \hat{\alpha} \cdot \mathbb{1}_{c<\hat{c}} \quad \textit{and} \quad a_2'(c,\eta) \equiv \mathbb{1}_{(c,\eta)\in[c_L,\hat{c})\times[\eta_1,1]}; \tag{52}$$

and

$$\begin{aligned} &c_2 \equiv \hat{\alpha} + (\lambda_0 + \lambda_1\,\hat{\alpha})\,[1 - G(\eta_{\mathrm{NL}}(\hat{\alpha}))]\,, \quad \tau''(c) \equiv \mathbb{1}_{c<c_2}\,, \\ &a_1''(c) \equiv \hat{\alpha} \cdot \mathbb{1}_{c<c_2}, \quad \textit{and} \quad a_2''(c,\eta) \equiv \mathbb{1}_{(c,\eta)\in[c_L,c_2)\times[\eta_{\mathrm{NL}}(\hat{\alpha}),1]}. \end{aligned} \tag{53}$$

Both $\{\hat{\tau}(c), a_1'(c), a_2'(c,\eta)\}$ *and* $\{\tau''(c), a_1''(c), a_2''(c,\eta)\}$ *are feasible solutions to* (13)*. Therefore, we have*

$$\tilde{U}_R^*(\Delta) \equiv \max\{U_R(\hat{\tau}, a_1', a_2'; \Delta),\ U_R(\tau'', a_1'', a_2''; \Delta)\} \le U_R^*(\Delta). \tag{54}$$

Finally, if $\eta_1 > \eta_{\mathrm{NL}}(\hat{\alpha})$*, we have* $\tilde{U}_R^*(\Delta) = U_R(\tau'', a_1'', a_2''; \Delta)$*.*

As in Section 4, the feasible solutions to the optimization problem (13) constructed in Proposition 7 can be implemented as follows. The regulator chooses between two take-it-or-leave-it offers. Expression (52) (along with the function $\hat{\tau}$) indicates that the drug is tested if its cost c is below the threshold $\hat{c}$, and receives a conditional approval at scale $\hat{\alpha}$ and final approval if the efficacy result η is above η_1. Similarly, (53) specifies the testing threshold, conditional approval scale, and final approval threshold as c_2, $\hat{\alpha}$, and $\eta_{\mathrm{NL}}(\hat{\alpha})$, respectively. Expression (54) then implies that the regulator chooses the better one between the two aforementioned mechanisms, which provides a lower bound on the optimal value $U_R^*(\Delta)$.

In the first construction (52), we take both $\hat{c}$ and $\hat{\alpha}$ directly from the upper-bound optimal solution to construct a feasible mechanism. In the second construction (53), by contrast, we use only $\hat{\alpha}$ as the conditional approval scale, and set the final approval threshold to $\eta_{\mathrm{NL}}(\hat{\alpha})$.

It is instructive to explain how we obtain the other thresholds accordingly. The final approval threshold η_1 is determined by (51), which guarantees that the firm receives the same expected payoff as in the upper bound optimal solution (and hence all incentive constraints (IC) and (IR) are automatically satisfied). In particular, the left- and right-hand sides of (51) represent the firm's expected payoff in period two under the optimal solution to the upper-bound problem (48) and our candidate solution, respectively.

In the second case, we set the final approval threshold to $\eta_{\mathrm{NL}}(\hat{\alpha})$ because it is never optimal to choose a higher value. Given this threshold, we then determine the testing threshold c_2 from (53)

such that the (IR) constraint for the threshold type binds. This choice of c_2 also ensures that the incentive compatibility constraint (IC) is satisfied.

Finally, when $\eta_{\mathrm{NL}}(\hat{\alpha}) < \eta_1$, the coefficient of a_2 remains nonnegative for all $\eta \in [\eta_{\mathrm{NL}}(\hat{\alpha}), \eta_1]$, implying that lowering the final approval threshold from η_1 to $\eta_{\mathrm{NL}}(\hat{\alpha})$ increases the regulator's expected payoff from any participating firm. Moreover, since $c_2 \geq \hat{c}$, the second policy induces greater participation. Therefore, it strictly outperforms the first policy. This explains the last sentence of Proposition 7.

5.3. Calibration

We again use the calibrated model parameters following Appendix EC.2, and obtain the optimal upper-bound solution according to Proposition EC.1 for various G, similar to Section 4.2. Based on these, we construct feasible solutions to the original problem following Proposition 7.

Figure 7(a) illustrates the performance ratio (PR) of the candidate solution compared with the upper bound, i.e.,

$$\mathrm{PR} = \frac{\tilde{U}_R^*}{\bar{U}_R^*}.$$

Under the given model parameters, the candidate solution consistently achieves a performance ratio exceeding 0.975, implying that it is very close to optimal. In particular, the performance ratio equals 1 when the beliefs about efficacy are either very pessimistic ($v_E < v_2$) or sufficiently optimistic ($v_E > v_5$). In the pessimistic case ($v_E < v_2$), it is optimal to grant no conditional approval, and the upper and lower bound objective values coincide. This corresponds to Corollary 2. In the optimistic case ($v_E > v_5$), the "no-leniency" policy described in Proposition 1 is optimal, yielding the exact optimal solution to the original problem.

Our candidate mechanism shares the same characteristics as the optimal one when $\Delta = 0$: the regulator induces testing if the cost c is low (below the threshold $\hat{c}$) and grants final approval if the efficacy η is high (above the threshold $\hat{\eta}$). Against the corresponding expected values v_E, we plot the testing threshold $\hat{c}$ (Figure 7(b)), the scale of conditional approval $\hat{\alpha}$ (Figure 7(c)), and the threshold of final approval $\hat{\eta}$ (Figure 7(d)). Compared to Figure 3, the testing threshold $\hat{c}$ is weakly higher due to the regulator's positive payoff after period two, as shown in panel (b).

The discontinuous jump of $\hat{\alpha}$ in panel (c) at $v_E = v_4$ arises for the same reason as discussed in Section 4.2. When $v_E > v_4$, as the drug appears more promising, the coefficient of a_1 in the regulator's payoff function (9) remains nonnegative, implying that it is optimal to grant full conditional approval, i.e., $\hat{\alpha} = 1$.

In panel (d), we use the notation $\bar{\eta}_2$ to represent the highest final approval threshold. Its value coincides with η_{NL} (defined in (11)) when $a_1 = 1$, i.e., $\bar{\eta}_2 \equiv \eta_{\mathrm{NL}}(1)$. Given $\hat{\alpha} = 1$, the coefficient of

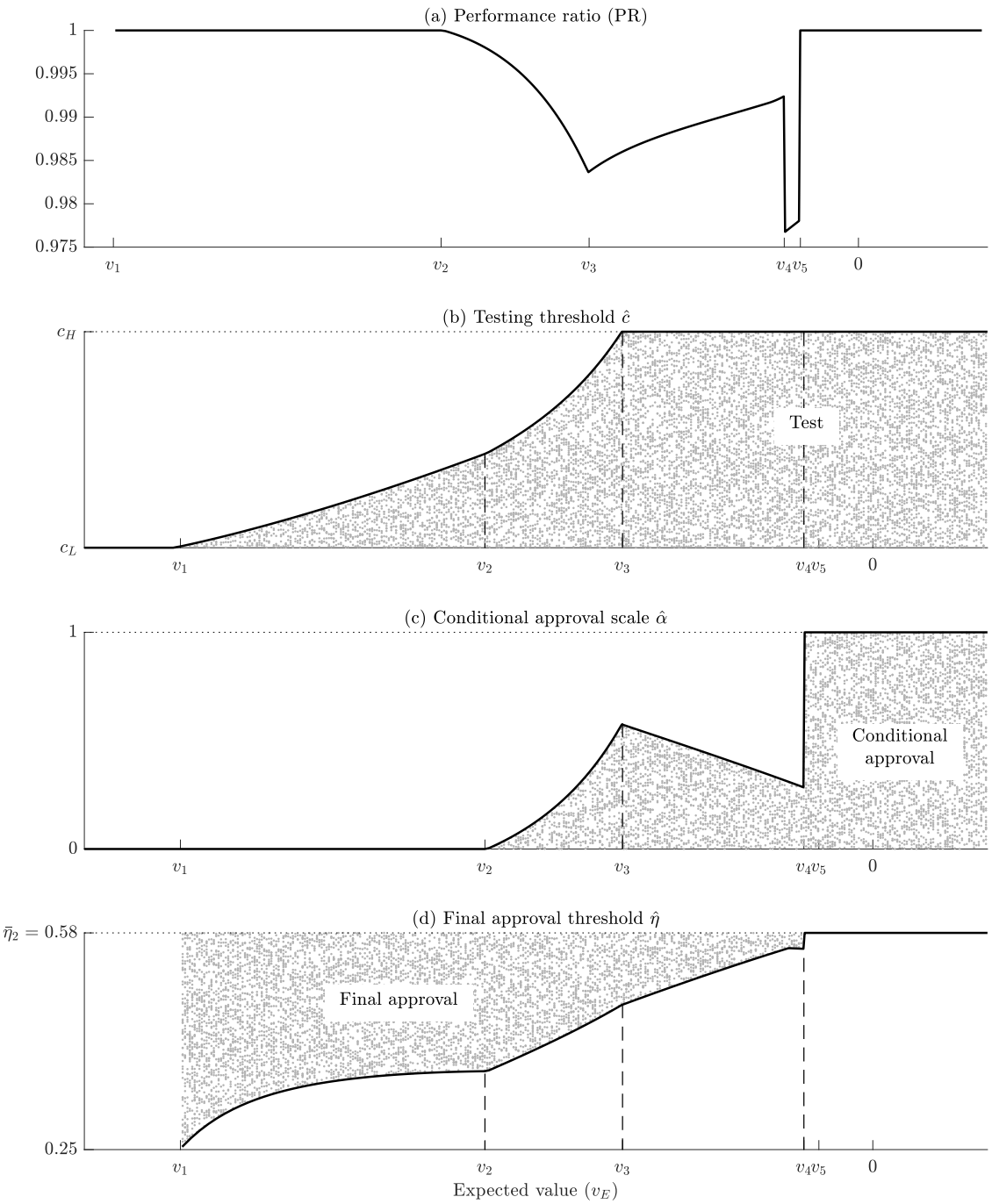

Figure 7 The lower-bound mechanism derived from the upper-bound problem performs well (top panel). Drugs with high expected value and low testing cost are tested (second panel), receive conditional approval (third panel), and have a high final approval threshold, meaning less leniency (bottom panel).

a_2 in the integrand of the objective function is nonnegative for all $\eta \in [\bar{\eta}_2, 1]$. Therefore, it is never optimal to set the final approval threshold above this value.

We compare policy performance with and without the tools of conditional approval and leniency (Figure 8). First, consider the scenario where the regulator uses conditional approval but not

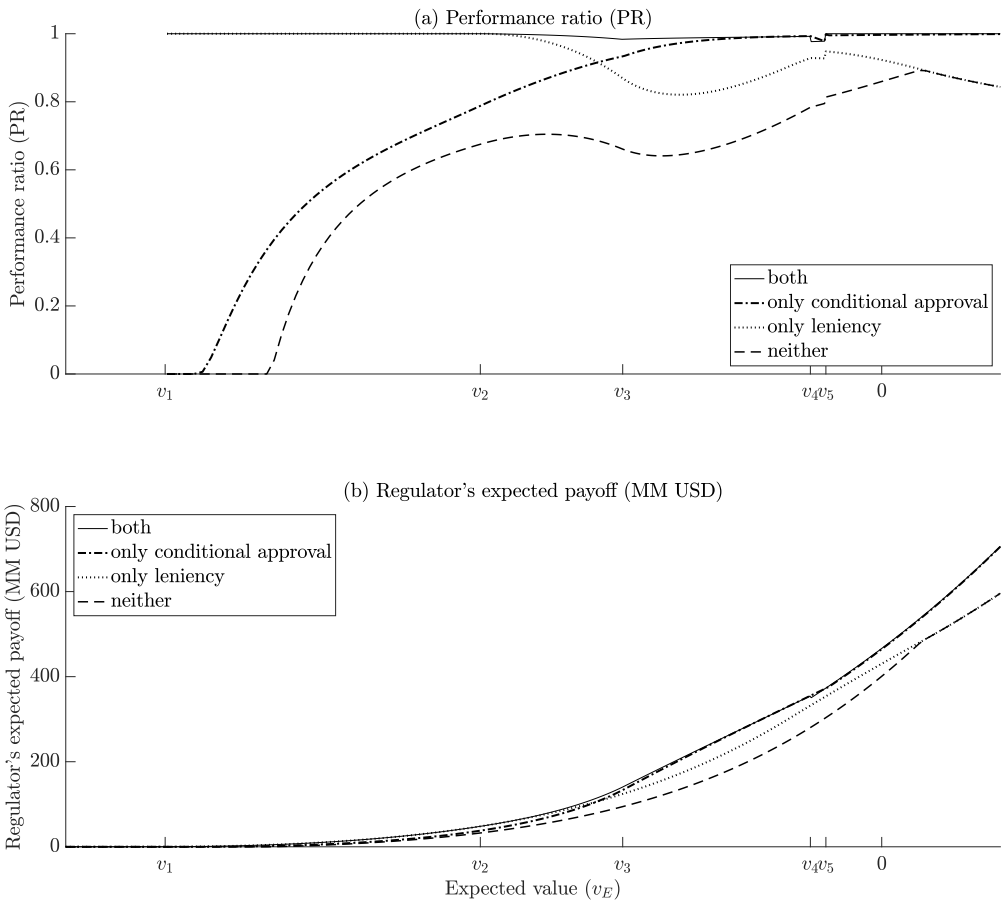

Figure 8 Performance ratios and the regulator's expected payoff (in million USD) of our candidate policy, the conditional approval-only policy, the leniency-only policy, and the policy without both tools.

leniency. Here, it fixes a naive final approval threshold, $\bar{\eta}'$, defined as the no-leniency threshold when a_1=0:

$$\bar{\eta}' \equiv \eta_{\mathrm{NL}}(0) = \inf\left\{\eta \in [0,1] : (\lambda_0 + \Delta)\, v(\eta) + \Delta \geq 0\right\}. \tag{55}$$

When the regulator uses only a leniency threshold and not conditional approval, we have $a_1(c) = 0$ for all $c \in [c_L, c_H]$. With neither tool, the final approval threshold is $\bar{\eta}'$, and only firms with low testing costs proceed with trials. For completeness, we present the optimal policies in closed form in Appendix EC.4.5.

In Figure 8(a), the solid line matches the curve in Figure 7(a). We observe that when either tool is unavailable, performance drops significantly under certain model parameters, highlighting the importance of both tools. It is worth noting that when the expected value v_E is close to v_4, the policy without leniency performs best among the four. This is because our candidate policy, which utilizes both tools, is constructed from an infeasible solution and is not necessarily optimal, making it possible for another policy to outperform it.

We plot the regulator's expected payoffs (in millions of USD) for the four scenarios in Figure 8(b). For instance, when $G(\eta) = \eta^{3/2}$, we have $v_E = -15.2$, which falls between v_3 and v_4. At this point, the regulator's payoffs are \$187, \$231, \$270, and \$270 million under the policies with neither tool, only leniency, only conditional approval, and both tools available, respectively. That is, considering both conditional approval and leniency yields a value as high as \$83 million.

Hence, conditional approval and leniency can generate more than \$6 billion in social value during a decade, compared to using neither, assuming a value of about \$83 million per drug and 79 drugs per decade. Additional value goes to shareholders, but we do not estimate it here. Nearly all of the social value comes from conditional approval rather than leniency, although leniency is relatively important for drugs with lower expected value (Figure 8(b)).

6. Conclusion

In 1992, the FDA granted its first conditional approval to an HIV-AIDS drug. The original aim was to accelerate patient access. We show that conditional approval offers a second benefit: encouraging investment in drugs that might otherwise be undeveloped.

We estimate that optimal use of conditional approval and leniency, compared to using neither, can generate hundreds of millions of dollars in social value each year, in addition to commercial value. Nearly all of the social value comes from conditional approval, rather than leniency, although leniency is relatively important for drugs with lower expected value.

Our analysis demonstrates the high value of conditional approval, even under our conservative framework that omits two additional benefits. First, conditional approval can accelerate price reductions if the end of pricing power depends on the approval date.[10] Second, conditional approval creates opportunities to learn from real-world use. Observational data may reduce the required sample size in clinical trials and lower testing costs. In this way, conditional approval can raise profits not only through earlier sales (as modeled), but also through reduced testing costs. This benefit, however, depends on the regulator's willingness to use non-randomized evidence, which is subject to selection bias.[11]

Our results yield three policy recommendations for lawmakers regarding regulatory authority. First, regulators should be allowed to conditionally approve drugs with low expected efficacy, accepting that some will later be withdrawn. The FDA's actual withdrawal rate was 22% for drugs conditionally approved between 2013 and 2017 (Figure EC.1). Our back-of-the-envelope calibration implies that this is consistent with the optimal withdrawal rate for drugs with zero expected value

[10] Earlier approval can lead to earlier price declines if the patent expires before exclusivity ends, or if the drug is subject to Medicare price regulation. See Appendix EC.3 for details on the duration of market power.

[11] For example, if only high-income patients use the drug, their outcomes may reflect socioeconomic advantages rather than drug efficacy.

(Appendix EC.2.4). This indicates that, in the past, the agency approved a mix of drugs with positive and negative expected values. However, parts of the agency have become more conservative, as the head of gene therapy appointed in 2025 has called for stricter conditional approval standards and less leniency (The Economist 2025).

Second, regulators should have the flexibility to grant partial conditional approval, such as restricting access to specific patient subgroups. Limiting insurance coverage could achieve a similar effect. Partial access should not be motivated solely by evidentiary uncertainty, though, but by whether limited access suffices to motivate continued development.

Third, regulators should be able to commit to a final approval threshold below the myopic threshold to encourage testing. The FDA appears to follow this practice, though the policy is not formally codified.

Acknowledgments

The authors appreciate feedback from Angelique Acquatella, Eric Budish, Pierre Dubois, Jennifer Kao, Margaret Kyle, Erzo Luttmer, Kathleen Miller, Paul Novosad, Marco Ottaviani, Jonathan Skinner, Christopher Snyder, Ariel Stern, Joe Stevens, and seminar participants at the American Society of Health Economists, the City University of Hong Kong, Congressional Budget Office, Dartmouth College, Duke University, the Shanghai University of Finance and Economics, the Toulouse School of Economics, Tsinghua University, the University of Chicago, and conference participants at the CEPR-TSE Health Economics Conference and the INFORMS annual meeting. The authors received no external funding for this research.

References

Armstrong A (2024) Preventing "3 billion visits" to FDA a year with CBER's Peter Marks. *Fierce Biotech* URL `https://www.fiercebiotech.com/biotech/preventing-3-billion-visits-fda-year-cbers-peter-marks`, accessed 2024-07-02.

Baron DP, Myerson RB (1982) Regulating a monopolist with unknown costs. *Econometrica* 911–930.

Berndt ER, Glennerster R, Kremer MR, Lee J, Levine R, Weizsäcker G, Williams H (2007) Advance market commitments for vaccines against neglected diseases: Estimating costs and effectiveness. *Health Economics* 16(5):491–511.

Bognar K, Romley JA, Bae JP, Murray J, Chou JW, Lakdawalla DN (2017) The role of imperfect surrogate endpoint information in drug approval and reimbursement decisions. *Journal of Health Economics* 51:1–12, ISSN 0167-6296, URL `http://dx.doi.org/https://doi.org/10.1016/j.jhealeco.2016.12.001`.

Budish E, Roin BN, Williams H (2015) Do firms underinvest in long-term research? Evidence from cancer clinical trials. *American Economic Review* 105(7):2044–2085.

Carpenter D, Ting MM (2007) Regulatory errors with endogenous agendas. *American Journal of Political Science* 51(4):835–852.

Centers for Medicare & Medicaid Services (2022) CMS finalizes Medicare coverage policy for monoclonal antibodies directed against amyloid for the treatment of Alzheimer's disease. URL https://www.cms.gov/newsroom/press-releases/cms-finalizes-medicare-coverage-policy-monoclonal-antibodies-directed-against-amyloid-treatment.

Centers for Medicare & Medicaid Services (2024) Medicare & Medicaid spending by drug. URL https://data.cms.gov/summary-statistics-on-use-and-payments/medicare-medicaid-spending-by-drug.

Chandra A, Kao J, Miller KL, Stern AD (2024) Regulatory incentives for innovation: The FDA's breakthrough therapy designation. *The Review of Economics and Statistics* 1–46, ISSN 0034-6535, URL http://dx.doi.org/10.1162/rest_a_01434.

Collard-Wexler A, Grennan M, Steck A (2024) Regulating new product testing: The FDA vs. the invisible hand. *Working Paper* 1–32.

DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics* 47:20–33, ISSN 01676296.

Dubois P, Moisson PH, Tirole J (2022) The economics of transferable patent extensions. *Toulouse School of Economics Working Paper* 1377:1–55.

Food and Drug Administration (2025) Advancing health through innovation: New drug therapy approvals 2024. URL https://www.fda.gov/drugs/novel-drug-approvals-fda/novel-drug-approvals-2024.

Frank RG, Shahzad M, Emanuel EJ (2022) Accelerated approval of cancer drugs: No economic reward for drug makers that conduct confirmatory trials. *Health Affairs* 41(9):1273–1280.

Gans JS, Ridley DB (2013) Innovation incentives under transferable fast-track regulatory review. *Journal of Industrial Economics* 61(3):789–816, ISSN 00221821.

Grennan M, Town RJ (2020) Regulating innovation with uncertain quality: Information, risk, and access in medical devices. *American Economic Review* 110(1):120–161.

Henry E, Loseto M, Ottaviani M (2022) Regulation with experimentation: Ex ante approval, ex post withdrawal, and liability. *Management Science* 68(7):5330–5347.

Henry E, Ottaviani M (2019) Research and the approval process: The organization of persuasion. *American Economic Review* 109(3):911–955, ISSN 19447981, URL http://dx.doi.org/10.1257/aer.20171919.

Kong E, Zhao O (2024) Market exclusivity and innovation: Evidence from antibiotics. https://papers.ssrn.com/abstract=4676384, SSRN Working Paper.

Kremer M, Levin J, Snyder CM (2022) Designing advance market commitments for new vaccines. *Management Science* 68(7):4786–4814, ISSN 15265501, URL http://dx.doi.org/10.1287/mnsc.2021.4163.

Light DW, Warburton RN (2005) Extraordinary claims require extraordinary evidence. *Journal of Health Economics* 24(5):1030–1033, ISSN 01676296, URL http://dx.doi.org/10.1016/j.jhealeco.2005.07.001.

Liu ITT, Kesselheim AS, Cliff ERS (2024) Clinical benefit and regulatory outcomes of cancer drugs receiving accelerated approval. *Journal of the American Medical Association* 331(17):1471–1479, URL http://dx.doi.org/10.1001/jama.2024.2396.

McClellan A (2022) Experimentation and approval mechanisms. *Econometrica* 90(5):2215–2247, ISSN 0012-9682, URL http://dx.doi.org/10.3982/ecta17021.

Myerson RB (1981) Optimal auction design. *Mathematics of Operations Research* 6(1):58–73.

Orlov D, Skrzypacz A, Zryumov P (2020) Persuading the principal to wait. *Journal of Political Economy* 128(7):2542–2578.

Peltzman S (1973) An evaluation of consumer protection legislation: The 1962 drug amendments. *Journal of Political Economy* 81(5):1049–1091.

Philipson TJ, Sun E (2008) Is the Food and Drug Administration safe and effective? *Journal of Economic Perspectives* 22(1):85–102, ISSN 08953309, URL http://dx.doi.org/10.1257/jep.22.1.85.

Ridley DB, Lee CY (2020) Does Medicare reimbursement drive up drug launch prices? *The Review of Economics and Statistics* 102(5):980–993.

Ridley DB, Sun P, Xu C (2025) Optimal push, pull, and failure funding for global health. Technical report, Duke University, URL http://dx.doi.org/10.2139/ssrn.5266214, posted: 10 June 2025.

Robey S, David FS (2016) Drug launch curves in the modern era. *Nature Reviews Drug Discovery* 16(1):13–14, ISSN 14741784.

Snyder CM, Hoyt K, Gouglas D (2023) An optimal mechanism to fund the development of vaccines against emerging epidemics. *Journal of Health Economics* 91:102795, ISSN 18791646, URL http://dx.doi.org/10.1016/j.jhealeco.2023.102795.

The Economist (2025) America's drug regulator is in turmoil. https://www.economist.com/united-states/2025/08/10/americas-drug-regulator-is-in-turmoil, accessed August 11, 2025.

United Press International (1988) Police arrest aids protesters blocking access to fda offices. *Los Angeles Times* URL https://www.latimes.com/archives/la-xpm-1988-10-11-mn-3909-story.html.

US Congress (2023) Consolidated appropriations act, 2023. URL https://www.congress.gov/117/bills/hr2617/BILLS-117hr2617enr.pdf, pub. L. No. 117-328, div. FF, title III, §3210(d)(D), 136 Stat. 4459.

Wouters OJ, McKee M, Luyten J (2020) Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *Journal of the American Medical Association* 323(9):844–853.

Xu L, Zhao H, Petruzzi NC (2021) Inducing compliance with postmarket studies for drugs under FDA's accelerated approval pathway. *Manufacturing & Service Operations Management* 23(1):170–190.

Zhou EW, Jackson MJ, Ledley FD (2023) Spending on phased clinical development of approved drugs by the us national institutes of health compared with industry. *JAMA Health Forum* 4(7):e231921–e231921.

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Online Appendix

This online appendix includes facts, model calibration, and proofs.

EC.1. Institutional background

Details about drug regulation and competition provide a foundation for the model.

Fact 1: The regulator does not typically directly finance late-stage testing. About 3% of NIH's total research budget was devoted to financing the clinical testing of drugs ultimately approved between 2010 and 2019 (Zhou et al. 2023). Rather than supporting late-stage trials for firms, NIH funding is typically directed toward lower-cost, pre-clinical research conducted at universities. A notable exception was the financing of late-stage testing for COVID-19 vaccines through Operation Warp Speed (Snyder et al. 2023).

Fact 2: Phase II clinical testing provides limited information about a drug's efficacy. Sample sizes are typically smaller in phase II than in phase III and are insufficient for definitive inference. Phase I and II trials tend to be short and small to minimize patient risk and financial cost, given that most drugs entering phase II do not progress to phase III (DiMasi et al. 2016). We treat phase I and II testing costs as sunk, but the model could be reframed to account for earlier decision-making by the firm.

Fact 3: The regulator can require initiation of confirmatory testing as a condition for granting conditional approval. Beginning in fiscal year 2023, the U.S. Congress granted the FDA the authority to mandate that trials begin before a drug receives conditional approval (US Congress 2023). Previously, some firms selling conditionally-approved drugs cited difficulty enrolling enough patients in trials for rare diseases. This rationale was often a convenient excuse, as confirmatory testing could reveal that the drug should be withdrawn (Xu et al. 2021, Frank et al. 2022). Among 46 cancer drugs granted conditional approval between 2013 and 2017, 15% had not completed testing after a median of six years (Liu et al. 2024) (Figure EC.1). However, we expect that under the new policy, the share of drugs failing to do timely testing will drop.

Fact 4: The regulator can grant conditional approval for only a share of patients. For example, approval may be limited to those with the most severe conditions, or coverage may be restricted so that only some patients are reimbursed while others pay out-of-pocket. In the United States, the FDA determines approval scope, while Medicare sets coverage rules, both under the Secretary of Health and Human Services. After conditional approval of an Alzheimer's drug, for instance, Medicare covered it only for patients who agreed to participate in a study of its efficacy (Centers for Medicare & Medicaid Services 2022). For tractability, we assume conditional approval applies to a share of uniform patients.

Fact 5: Testing costs are privately known by the firm. According to Light and Warburton (2005), "this science-based industry refuses to allow independent parties to check the validity of their cost

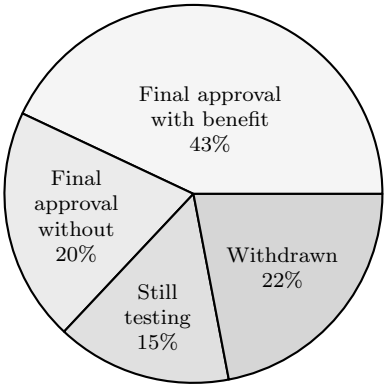

Figure EC.1 For 46 cancer indications conditionally approved in the U.S. between 2013 and 2017, 43% received final approval after demonstrating statistically significant benefits in confirmatory trials, 20% received final approval despite Liu et al. (2024) finding no evidence of statistically significant benefits, 22% were withdrawn, and 15% remained in ongoing testing. Source: Authors' figure based on data from Liu et al. (2024).

.

data and analyze it so that policy can be based on solid, objective, reproducible evidence." Firms do not disclose testing costs, because it is hard to allocate joint costs. Also, firm executives fear being held up. Disclosing testing costs could lead the government to reimburse only the testing costs, without compensating for the failures of other drugs or the opportunity cost of capital. For external parties, it is especially hard to benchmark costs for conditionally-approved drugs because they tend to be novel with few comparable examples.

Fact 6: Information from use of the drug outside randomized trials is not very informative, as it is noisy and subject to selection bias. For example, if healthier, high-income patients are more likely to purchase a drug, observed outcomes may overstate efficacy. Accordingly, our model assumes that usage after conditional approval does not affect phase III testing. If "real-world evidence" were informative, it could reduce firms' testing costs by allowing smaller or shorter trials. Hence, we might underestimate the value of conditional approval.

Fact 7: Drugs given conditional approval are typically for rare diseases and do not have competition. From 2015 to 2024, more than 80% of FDA conditional approvals were for rare diseases (see Figure EC.2). At launch, drugs for rare diseases typically have no competition, due to orphan drug exclusivity of seven years in the U.S. and ten years in Europe. Even years later, markets for rare diseases are often too small to attract multiple firms.

Fact 8: Conditional approval increases the firm's revenue during conditional approval and several years afterward. Drug revenue typically rises during the first five years after approval (Robey and David 2016), as physicians and patients become more familiar with the drug. Demand is higher for drugs with which physicians and patients have more experience (Ridley and Lee 2020).

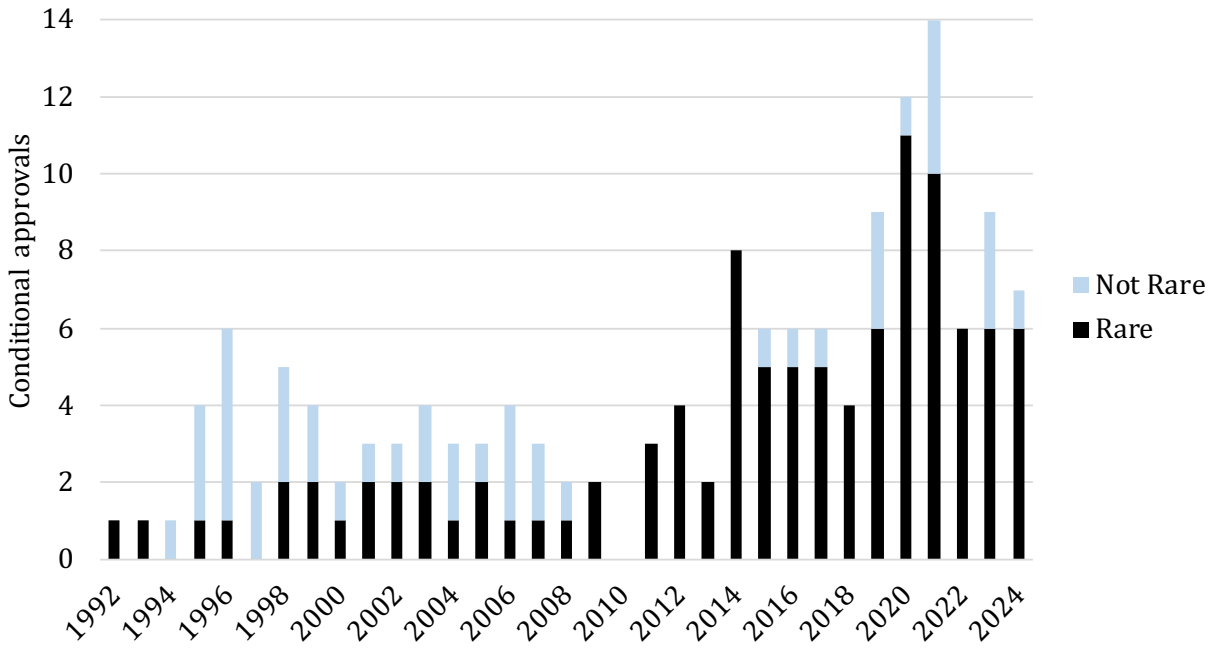

Figure EC.2 The FDA granted conditional approval to 79 drugs between 2015 and 2024. Most of the conditionally-approved drugs treat rare diseases. The official FDA term for conditional approval is "accelerated approval" and for rare disease is "orphan disease." Source: Authors' analysis of data on approvals of new molecules by the Center for Drug Evaluation and Research.

Fact 9: Conditional approval provides earlier revenue for some firms that would otherwise not continue testing. According to the head of the FDA center responsible for gene therapies and vaccines, "The wherewithal to do a three-year study or a four-year study without having a revenue stream, it's just beyond many companies that are startups. So having the accelerated approval process is a way to get there" (Armstrong 2024).

Fact 10: A discrete-time model is appropriate in our setting because human clinical trials typically involve pre-specified endpoints and results that are announced all at once. Typically, phase III trials have pre-specified endpoints, followed by a comprehensive analysis by the firm before regulatory submission. However, there are exceptions, such as early-stage animal testing and the rolling review of the COVID-19 vaccine, which have a more flexible and exploratory set of protocols.

Fact 11: The regulator can signal to the firm that it will grant final approval to a drug that narrowly misses the efficacy threshold. The FDA has granted final approval to several drugs that did not prove efficacy (Figure EC.1). Furthermore, the FDA will advise some firms before confirmatory testing about how to design their trials and what evidence will merit approval. Under a 2012 law, FDA staff support the developer of a potential breakthrough drug by "providing timely advice to, and interactive communication with, the sponsor regarding the development of the drug to ensure

that the development program to gather the nonclinical and clinical data necessary for approval is as efficient as practicable."

EC.2. Calibration

We calibrate the model to illustrate its application and to inform policy recommendations. Our approach proceeds in five steps: First, we select a representative drug. Second, we use the literature on drug development to determine values for the probability of technical success and the discount rate. Third, we estimate the testing cost for a representative drug. Fourth, we estimate drug doses and revenue for a representative drug. Fifth, we apply these parameter values to calculate the cumulative net present value for a representative drug.

We select a representative drug that received conditional approval: avelumab (brand name Bavencio). In 2017, the FDA granted conditional approval for avelumab to treat skin cancer based on tumor shrinkage. In 2023, full approval was granted based on clinical benefits. Following its initial approval for skin cancer, avelumab was also approved for bladder and kidney cancers.

The discount rate (δ) is set at 10.5% (DiMasi et al. 2016), and for simplicity, we assume that both the regulator and the firm use the same discount rate.

EC.2.1. Estimated testing costs

We assume that phase III clinical testing costs are evenly distributed over the first three years, with the regulator making the final approval decision at the start of the fifth year. The firm sells at monopoly prices for thirteen years, after which we assume the net revenue following generic entry is negligible.

To establish the lower and upper bounds for avelumab's testing costs, we use estimates from similar drugs. A prior study analyzed financial filings from companies with few drugs and limited joint costs, which made it easier to attribute expenses to a specific drug. Among 355 drugs approved by the FDA between 2009 and 2018, Wouters et al. (2020) identified sufficient testing cost data for 63 drugs. Of these, 13 drugs had phase III testing cost estimates in the same therapeutic class (antineoplastic and immunomodulating agents) as avelumab. For the low estimate, we used brigatinib's phase III testing cost of \$98.1 million (in 2018 dollars). For the high estimate, we used \$630.7 million for sarilumab. However, these estimates are imprecise. For example, the authors expressed low confidence in the phase III cost estimate for sarilumab.

EC.2.2. Estimated doses and sales

Next, we estimate the total doses and revenue for the representative drug. We have only publicly available data on Medicare and Medicaid patients so we extrapolate to the total market. Also, our analysis uses data on total sales of avelumab across all indications, meaning the reported sales

	Avelumab
Medicaid+Medicare ($MM)[a]	121
Total sales ($MM)[b]	290
Net revenue ($MM)[c]	217
Price per dose ($)[d]	86
Production cost per dose ($)[e]	21.5
Total doses (MM)[f]	3.37
Willingness to pay per dose ($)[g]	118

[a] Medicaid and Medicare Part B spending (Source: (Centers for Medicare & Medicaid Services 2024))

[b] "Medicaid+Medicare" multiplied by 2.4 because Medicaid and Medicare accounted for 42% of total U.S. prescription drug spending at the time (Source: National Health Expenditures)

[c] "Total sales" multiplied by 0.75 to reflect cost of goods sold is 25% of total sales (Source: Harmand, `https://hardmanandco.com/2021-pharma-statistics-long-term-cost-underlying-ebit-analysis/`)

[d] We use Medicare prices because they are negotiated by commercial insurers and are similar to commercial prices. We use Part B spending for avelumab, because it is mainly administered by providers. (Source: Centers for Medicare and Medicaid Services)

[e] "Price per dose" multiplied by 0.25 to reflect production cost of goods sold is 25% of price (Source: Harmand, `https://hardmanandco.com/2021-pharma-statistics-long-term-cost-underlying-ebit-analysis/`)

[f] "Total sales"/"Price per dose"

[g] We assume that the buyer's surplus is half of the seller's. Hence, the willingness to pay is 1.375 times the unit price.

Table EC.1 Annual doses and prices for avelumab in 2022.

figures are larger than those specific to the skin cancer indication. Table EC.1 summarizes the annual sales data for avelumab in 2022.

We denote the firm's net profit per dose by π and the number of doses sold during the conditional approval period (under full conditional approval) by q. Both values are normalized to 1 in the theoretical model (see Section 2).

The price per dose is $86, the production cost per dose is $21.5, and the patient's willingness to pay $118 (Table EC.1), and we have

$$\pi = \$86 - \$21.5 = \$64.5, \quad v(0) = \$86, \quad v(1) = \$118 - \$86 = \$32. \tag{EC.1}$$

Table EC.2 summarizes the (discounted) annual doses of avelumab. When full conditional approval is granted, the firm's total sales in the first 4 years are 4.36 million doses,[12] and from years 5 to 17, they are 15.43 million doses.[13] When no conditional approval is granted, if the drug

[12] Summation of Years 1 to 4 of "Discounted annual sales" for "Full conditional approval" in Table EC.2.

[13] Summation of Years 5 to 17 of "Discounted annual sales" for "Full conditional approval" in Table EC.2.

receives final approval, the firm's total sales are 11.34 million.[14] By their definitions in Table EC.3, the values of q, λ_0, and λ_1 are determined by

$$\begin{cases} q = 4.36 \\ \lambda_0 = 11.34 \\ \lambda_0 + \lambda_1 = 15.43 \end{cases} \Rightarrow \begin{cases} q = 4.36 \\ \lambda_0 = 11.34 \\ \lambda_1 = 4.09 \end{cases}. \tag{EC.2}$$

The discounted quantities sold from year 18 onward are given by[15]

$$\Delta = \sum_{i=18}^{\infty} \frac{3.37}{(1+10.5\%)^i} = 5.88. \tag{EC.3}$$

Under fixed duration, if conditional approval is granted, the firm is only able to sell at the monopolistic price ($86) until year 13. In this case, the total discounted doses in the final approval period decrease by 2.89 million.[16] Conversely, the total discounted doses during the competitive period increase by the same amount, as Years 14 to 17 are now included in the competitive period.

The estimated model parameters are summarized in Table EC.3.

EC.2.3. Estimated cumulative net present value

Next, we apply the parameter values to estimate the cumulative net present value for avelumab when the testing cost is at its maximum value, $c_H = \$630.7$ million, is evenly distributed over the first three years, and full conditional approval is granted. Table EC.4 displays the firm's cumulative net present value.

Figure EC.3 shows the firm's cumulative net present value under various scales of conditional approval. The upper and lower curves in each panel represent the firm's cumulative net profit when testing costs are at their lower bound (c_L) and upper (c_H) bound. For testing costs that fall within $[c_L, c_H]$, the corresponding curve lies in the shaded area. In the absence of conditional approval ($\alpha_1 = 0$), firms with high testing costs will forgo testing, causing a potential societal loss.

EC.2.4. Calibrated optimal policy

Next, we estimate the optimal withdrawal rate, i.e., the optimal share of conditionally approved drugs ultimately denied final approval (Figure EC.4). If the regulator grants conditional approval, given the final approval threshold $\hat{\eta}$, the probability of withdrawal is $G(\hat{\eta})$.

In Figure EC.4, there are two discontinuities. First, the regulator does not grant conditional approval to drugs with very low expected value ($v_E < v_2$). Second, there is a discontinuity at v_4 because the final approval threshold jumps to $\bar{\eta}_2$ at this point.

[14] Summation of Years 5 to 17 of "Discounted annual sales" for "No conditional approval" in Table EC.2.

[15] The values 3.37 and 10.5% come from the "Total doses" line in Table EC.1 and the discount rate (δ), respectively.

[16] Summation of Years 14 to 17 of "Discounted annual sales" for "Full conditional approval" in Table EC.2.

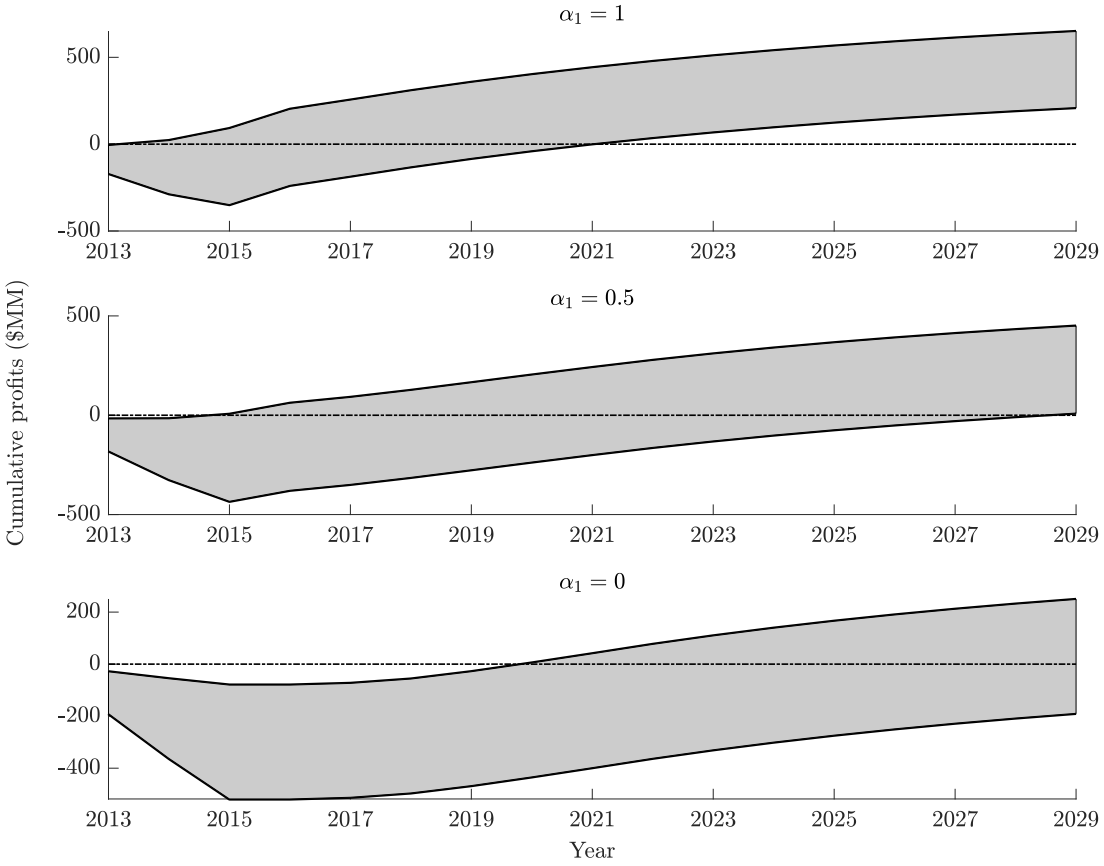

Figure EC.3 Firm's cumulative net profit under different scales of conditional approval and testing costs. Source: Authors' analysis using a representative drug avelumab.

The optimal withdrawal rate for drugs with $v_E = 0$ is 23%. How does this compare to the FDA's actual rate? A study of cancer drugs granted conditional approval between 2013 and 2017 found that 22% were later withdrawn (Figure EC.1). If the withdrawal rate for drugs with $v_E = 0$ is equal to the FDA's actual rate (Figure EC.1), then the agency was conditionally approving a mix of drugs with positive and negative expected values.

These results suggest the FDA was historically close to the optimum and, if anything, slightly too conservative. However, the leader of the FDA gene therapy unit appointed in 2025 has adopted a more conservative stance on conditional approval (The Economist 2025). Our calibrated results suggest that this change will reduce social welfare.

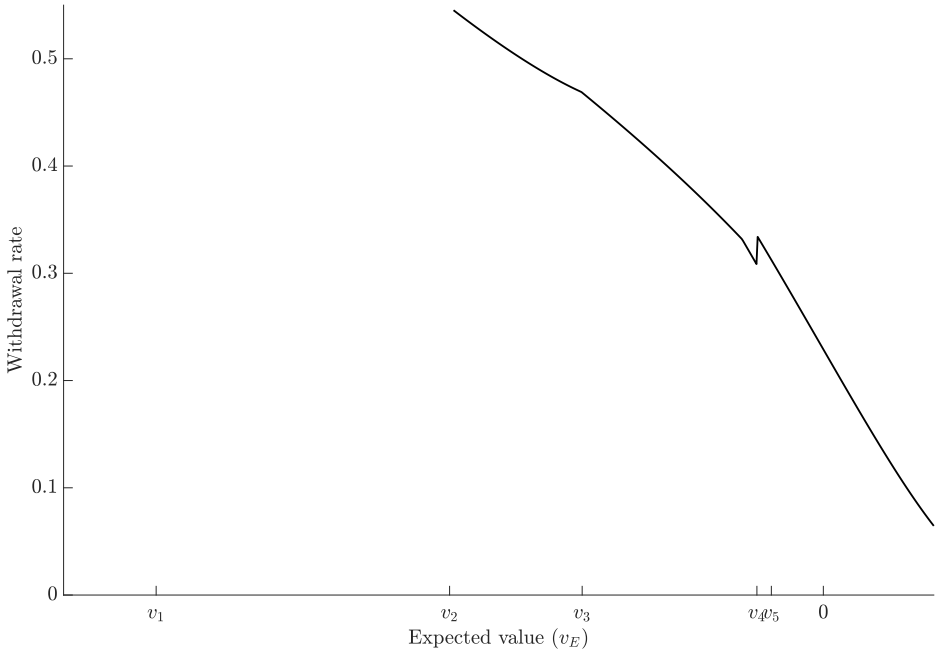

Figure EC.4 Proportion of drugs withdrawn after receiving conditional approval. Source: Authors' analysis based on a representative drug.

	Full conditional approval			No conditional approval		
Year	% of peak sales[a]	Annual doses (MM)[b]	Discounted annual doses (MM)[c]	% of peak sales[a]	Annual doses (MM)[b]	Discounted annual doses (MM)[c]
1	11%	0.37	0.34			
2	31%	1.05	0.86			
3	58%	1.96	1.45			
4	76%	2.56	1.72			
5	89%	3.00	1.82	11%	0.37	0.23
6	100%	3.37	1.85	31%	1.05	0.57
7	100%	3.37	1.68	58%	1.96	0.97
8	100%	3.37	1.52	76%	2.56	1.15
9	100%	3.37	1.37	89%	3.00	1.22
10	100%	3.37	1.24	100%	3.37	1.24
11	100%	3.37	1.12	100%	3.37	1.12
12	100%	3.37	1.02	100%	3.37	1.02
13	100%	3.37	0.92	100%	3.37	0.92
14	100%	3.37	0.83	100%	3.37	0.83
15	100%	3.37	0.75	100%	3.37	0.75
16	100%	3.37	0.68	100%	3.37	0.68
17	100%	3.37	0.62	100%	3.37	0.62

[a] Source: Robey and David (2016)

[b] Annual doses=Peak annual doses $\times$ % of peak sales

[c] Discounted annual doses in year i=Annual doses in year $i/(1+\delta)^i$, where $\delta = 10.5\%$ is the annual discount rate.

Table EC.2: The firm's annual (discounted) doses when full/no conditional approval is granted to avelumab.

Notation	Description	Value	Source
δ	discount rate	10.5%	DiMasi et al. (2016)
c_L	lower bound of the testing cost	98.1($MM)	The estimated phase III testing cost of a similar drug (brigatinib) as estimated by Wouters et al. (2020).
c_H	upper bound of the testing cost	630.7($MM)	The estimated phase III testing cost of a similar drug (sarilumab) as estimated by Wouters et al. (2020).
$v(0)$	regulator's loss from administering an ineffective dose	$86	See equation (EC.1).
$v(1)$	regulator's payoff from administering an effective dose	$32	See equation (EC.1).
Γ	price drop per dose after the monopolistic period	$64.5	See equation (EC.1).
q	total discounted doses with full conditional approval before final approval	4.36(MM)	See equation (EC.2).
λ_0	total discounted doses from final approval when $\alpha_1 = 0$	11.34(MM)	See equation (EC.2).
λ_1	total boosted discounted doses in the monopolistic period after final approval when $\alpha_1 = 1$	4.09(MM)	See equation (EC.2).
Δ	total discounted doses after the monopolistic period	5.88(MM)	See equation (EC.3).

Table EC.3 Estimated model parameters for avelumab.

Year	Discounted net revenue (if in market) ($MM)[a]	Discounted expected net revenue ($MM)[b]	Annual testing cost ($MM)[c]	Discounted annual testing cost ($MM)	Discounted annual net profit ($MM)[d]	Cumulative net present value ($MM)[e]
1	22	22	210	190	-169	-169
2	55	55	210	172	-117	-286
3	93	93	210	156	-63	-348
4	111	111	0	0	111	-238
5	117	53	0	0	53	-185
6	119	54	0	0	54	-131
7	108	49	0	0	49	-83
8	98	44	0	0	44	-39
9	88	40	0	0	40	1
10	80	36	0	0	36	37
11	72	33	0	0	33	70
12	65	29	0	0	29	99
13	59	27	0	0	27	126
14	54	24	0	0	24	150
15	49	22	0	0	22	172
16	44	20	0	0	20	191
17	40	18	0	0	18	209

[a] "Discounted annual doses" column from Table EC.2 times "Price per dose" from Table EC.1

[b] Discounted expected net revenue = Discounted net revenue if the drug is in mraket × Pr(the drug is in market), and the probabilistic term is 1 and 0.45 in years 1 to 4 and 5 to 17, respectively.

[c] The testing cost $ 630.7 million is evenly distributed over the first three years.

[d] Discounted annual net profit = Discounted expected net revenue (column 3) - Discounted annual testing cost (column 5).

[e] Cumulative NPV in year i = sum of the discounted annual net profit from years 1 to i.

Table EC.4: The firm's cumulative net present value when its testing cost is the maximum value.

EC.3. Flexible exclusivity

Optimal conditional approval and leniency may depend on the length of exclusivity. Some drug classes receive extensions. For example, antibiotics gain five additional years under the GAIN Act (Kong and Zhao 2024). Longer exclusivity increases monopoly sales, boosting firm profits but delaying savings for payers. Exclusivity extensions can therefore strengthen development incentives and reduce the need for conditional approval and leniency, all else equal.

Recall that λ_0 denotes the number of doses sold before the competitive period under no conditional approval. Likewise, $\lambda_0 + \lambda_1$ denotes the number of doses sold under full conditional approval. In Figure EC.5, the curve starting at time 0 shows annual sales with full conditional approval; the parallel curve beginning at t_1 represents sales without conditional approval. Because the competitive period begins at time t_2 (as illustrated in Figure 1), λ_0 and λ_1 correspond to the gray areas in the middle and on the left, respectively. The number of doses sold during the competitive period, denoted by Δ, is represented by the rectangle on the right. Because annual sales peak in six years (Robey and David 2016), extending the firm's monopolistic period beyond 13 years does not affect λ_1 and only increases λ_0. Given that $\lambda_0 + \Delta$ represents the total number of doses sold after final approval, corresponding to the area under the curve starting at t_1, any increase in λ_0 leads to an equivalent decrease in Δ. This adjustment corresponds to a rightward shift of time t_2 in Figure EC.5.

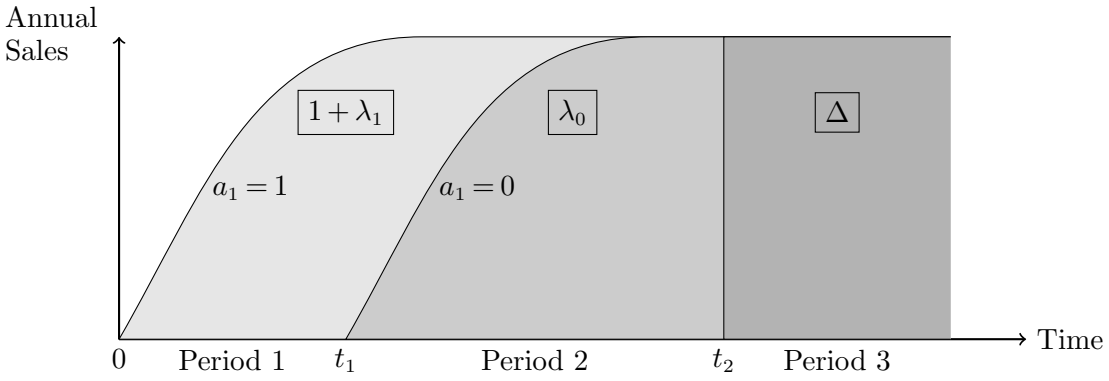

Figure EC.5 Annual sales of the drug under no conditional approval and full conditional approval.

Let $\hat{\lambda}_0$ and Δ denote the calibrated values of λ_0 and Δ, respectively.[17] In Figure EC.6, we plot the testing threshold, the scale of conditional approval, the final approval threshold, and the expected payoffs of both the regulator and the firm (with testing costs \$200 and \$500 million) as functions of λ_0, varying from $\hat{\lambda}_0$ to $\hat{\lambda}_0 + \Delta$. The value $\hat{\lambda}_0$ corresponds to a scenario in which the firm's monopoly power ends at the end of year 13, while the upper bound $\hat{\lambda}_0 + \Delta$ reflects a case where the regulator

[17] See Appendix EC.2 for more details.

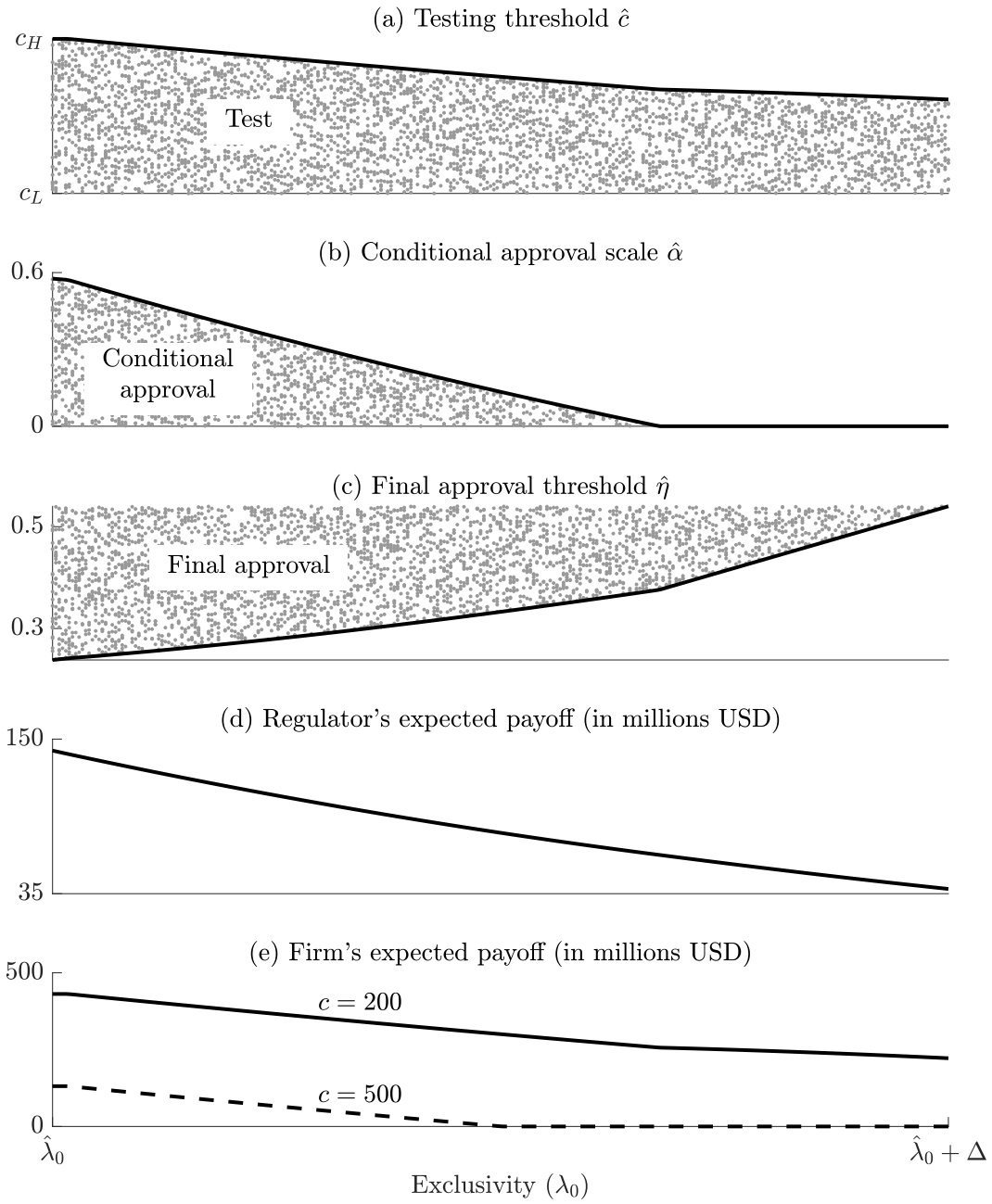

Figure EC.6 The testing threshold, scale of conditional approval, threshold for final approval, and the expected payoffs of the regulator and the firm (with different testing costs) under calibrated model parameters. The x-axis ranges from calibrated $\hat{\lambda}_0$ to $\hat{\lambda}_0 + \Delta$.

extends monopoly power indefinitely. Fixing λ_0, the number of doses sold in the competitive period is $\hat{\lambda}_0 + \Delta - \lambda_0$.

As shown in panels (b) and (c), under the calibrated model parameters, when λ_0 increases, the scale of conditional approval decreases, and the final approval threshold increases. The testing threshold, illustrated in panel (a), also declines, indicating that the regulator induces less firm participation as the length of exclusivity increases. Finally, both the regulator (panel (d)) and the firm's expected payoffs (panel (e)) decrease as λ_0 increases, suggesting that it is optimal to maintain the current exclusivity timeline rather than granting additional exclusivity. Given the firm's testing cost, the monotonicity of its expected payoff in λ_0 follows from the monotonicity of the testing threshold $\hat{c}$. Specifically, since the expected payoff of a firm with production cost $\hat{c}$ is zero, the expected payoff for a firm with cost c is given by $(c-\hat{c})^+$. This observation also highlights that, although firms generally favor extended exclusivity, they could be worse off if the regulator reoptimizes its policy in response to the change.

EC.4. Proofs

EC.4.1. Proofs for Section 2

Proof of Lemma 1. By definition of Φ, we have

$$\frac{\partial \Phi(a_1, \eta; \Delta)}{\partial \eta} = (\lambda_0 + \lambda_1 a_1 + \Delta)\, v'(\eta) > 0.$$

Furthermore, by assumption (5), we have

$$\Phi(a_1, 0; \Delta) = (\lambda_0 + \lambda_1 a_1) \underbrace{v(0)}_{<0} + \Delta \underbrace{[v(0) + 1]}_{\leq 0} < 0$$

and

$$\Phi(a_1, 1; \Delta) = (\lambda_0 + \lambda_1 a_1)\, v(1) + \Delta\, [v(1) + 1] > 0.$$

This completes the proof. □

EC.4.2. Proofs for Section 3

Proof of Proposition 1. For any c, $a_1(c)$, and $a_2(c, \eta)$, we have

$$\begin{aligned}
& v_E\, a_1(c) + \int_0^1 \Big[[\lambda_0 + \lambda_1\, a_1(c)]\; v(\eta) + \Delta\; [v(\eta) + 1] \Big]\; a_2(c, \eta)\, \mathrm{d}G(\eta) \\
\leq\; & v_E\, a_1(c) + \int_{\eta_{\mathrm{NL}}(a_1(c))}^1 \Big[[\lambda_0 + \lambda_1\, a_1(c)]\; v(\eta) + \Delta\; [v(\eta) + 1] \Big]\, \mathrm{d}G(\eta) \\
=\; & \check{U}_R(a_1(c), \eta_{\mathrm{NL}}(a_1(c))),
\end{aligned}$$

because the no-leniency threshold η_{NL} maximizes the regulator's continuation payoff after period one.

The function $\check{U}_R$ in the previous expression is the restriction of the function defined in (14), and it is convex in a_1 because

$$\frac{\mathrm{d}}{\mathrm{d}a_1} \check{U}_R(a_1, \eta_{\mathrm{NL}}(a_1)) = v_E + \lambda_1 \int_{\eta_{\mathrm{NL}}(a_1)}^1 v(\eta)\, \mathrm{d}G(\eta)$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}a_1^2} \check{U}_R(a_1, \eta_{\mathrm{NL}}(a_1)) = -\lambda_1\, v\,(\eta_{\mathrm{NL}}(a_1)) \cdot g\,(\eta_{\mathrm{NL}}(a_1)) \geq 0.$$

Therefore, the maximum is attained at either $a_1(c) = 0$ or $a_1(c) = 1$.

If the drug is promising and low-cost, i.e., (15) and (17) hold, we have

$$U_R^*(\Delta) \leq \max_{(a_1, \eta) \in [0,1]^2} \check{U}_R(a_1, \eta) = \check{U}_R(1, \eta_{\mathrm{NL}}(1)),$$

because $\tau(c) \in [0, 1]$ and the upper bound is attained by the candidate solution defined in (19).

If the drug is not promising and low-cost, i.e., (16) and (18) hold, we have

$$U_R^*(\Delta) \leq \max_{(a_1, \eta) \in [0,1]^2} \check{U}_R(a_1, \eta) = \check{U}_R(0, \eta_{\mathrm{NL}}(0)),$$

because $\tau(c) \in [0, 1]$ and the upper bound is attained by the candidate solution defined in (20).

This completes the proof. □

Proof of Corollary 1. The proof of Proposition 1 ignores both the (IC) and (IR) constraints. Hence, the same argument applies to the first-best problem and the dictatorial problem.

This completes the proof. □

EC.4.3. Proofs for Section 4

Proof of Proposition 2. When $\Delta = 0$, any feasible solution $\{\tau(c), a_1(c), a_2(c,\eta)\}$ to the optimization problem (13) yields a corresponding solution $\{\tau(c), \alpha_1(c), \alpha_2(c,\eta)\}$, with α_1 and α_2 defined in (21) and (22), respectively—that is feasible for the optimization problem (27) and attains the same objective value.

Conversely, any feasible solution $\{\tau(c), \alpha_1(c), \alpha_2(c,\eta)\}$ to the optimization problem (27) yields a corresponding solution $\{\tau(c), a_1(c), a_2(c,\eta)\}$, with a_1 and a_2 defined in (28), that is feasible for the optimization problem (13) and likewise attains the same objective value.

This completes the proof. □

The following lemma is commonly used to simplify mechanism design problems involving direct payments between buyers and sellers. For completeness, we provide a proof here.

LEMMA EC.1. *The incentive constraints* (IC) *and* (IR) *can be replaced by*

$$\tau \text{ non-increasing}, \tag{EC.4}$$

$$\alpha_1(c) = \Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y + c\,\tau(c) - \int_0^1 \alpha_2(c,\eta)\,\mathrm{d}G(\eta), \tag{EC.5}$$

and

$$\Pi(c_H) \geq 0, \tag{EC.6}$$

where $\Pi(c_H) \equiv \hat{\Pi}(c_H, c_H)$.

Proof of Lemma EC.1. The incentive-compatibility constraint can be rewritten as

$$\Pi(c) \geq \Pi(c') + \tau(c') \cdot (c' - c),$$

where $\Pi(c) \equiv \hat{\Pi}(c,c)$. Flipping the role of c and c' yields

$$\Pi(c') \geq \Pi(c) + \tau(c) \cdot (c - c').$$

Combining the two inequalities above yields the monotonicity of τ.

By envelope theorem, we have

$$\Pi'(c) = -\tau(c) \leq 0,$$

which implies that

$$\Pi(c) = \Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y.$$

Therefore, it is sufficient to impose the individual rationality constraint on type c_H and

$$\alpha_1(c) = \Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y + c\,\tau(c) - \int_0^1 \alpha_2(c,\eta)\,\mathrm{d}G(\eta).$$

To prove the other direction, for any $c < c'$, we have

$$\begin{aligned}\Pi(c) &= \Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y \\ &= \Pi(c_H) + \int_{c'}^{c_H} \tau(y)\,\mathrm{d}y + \int_c^{c'} \tau(y)\,\mathrm{d}y \\ &= \Pi(c') + \int_c^{c'} \tau(y)\,\mathrm{d}y \\ &\geq \Pi(c') + \tau(c')\cdot(c'-c),\end{aligned}$$

where the inequality follows from the monotonicity of τ. Constraint (IR) is implied by

$$\Pi(c) \geq \Pi(c_H) \geq 0.$$

This completes the proof. □

Lemma EC.2. *The optimization problem* (27) *is equivalent to*

$$\begin{aligned}\max_{\{\tau,\alpha_2,\Pi(c_H)\}\in\Omega'_C}\ & v_E\,\Pi(c_H) + \int_{c_L}^{c_H}\int_0^1 \left[v(\eta) - v_E\right]\,\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c) \\ & + v_E \int_{c_L}^{c_H} \left[c + \frac{F(c)}{f(c)}\right]\,\tau(c)\,f(c)\,\mathrm{d}c,\end{aligned} \tag{EC.7}$$

where the feasible set Ω'_C *is determined by* (2), (EC.4), (EC.6), *and for any* $(c,\eta)\in[c_L,c_H]\times[0,1]$,

$$0 \leq \Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y + c\,\tau(c) - \int_0^1 \alpha_2(c,\eta)\,\mathrm{d}G(\eta) \leq \tau(c), \tag{EC.8}$$

and

$$0 \leq \alpha_2(c,\eta) \leq \lambda_0\,\tau(c) + \lambda_1\left[\Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y + c\,\tau(c) - \int_0^1 \alpha_2(c,z)\,\mathrm{d}G(z)\right]. \tag{EC.9}$$

Proof of Lemma EC.2. Substituting the equation (EC.5) in Lemma EC.1 into the objective function yields the desired result. □

Proof of Lemma 2. We first reformulate the definitions of $\check{c}$, $\tilde{\eta}$, and $\underline{\eta}$ given in (36)–(38) in a more explicit form:

$$\begin{aligned}&\check{c} = \sup\left\{c\in[c_L,c_H] : v(\check{\eta})\,\left[c + \frac{F(c)}{f(c)}\right] + \lambda_0\int_{\check{\eta}}^1 \left[v(y) - v(\check{\eta})\right]\mathrm{d}G(y) \geq 0\right\}, \\ &v(\tilde{\eta})\,\frac{F(\underline{c}(\tilde{\eta}))}{f(\underline{c}(\tilde{\eta}))} + \lambda_0\int_{\tilde{\eta}}^1 v(y)\,\mathrm{d}G(y) = 0, \text{ and} \\ &\underline{\eta} = \inf\left\{\eta\in[0,\check{\eta}] : v(\eta)\,\frac{F(\bar{c}(\eta))}{f(\bar{c}(\eta))} + v_E + (\lambda_0+\lambda_1)\int_\eta^1 v(y)\,\mathrm{d}G(y) \geq 0\right\}.\end{aligned}$$

We start with the first part of the lemma. Because $\int_\eta^1 [v(y)-v(\eta)]\,\mathrm{d}G(y)$ and $v(\eta)$ are decreasing and increasing in η, respectively, we know that function h is decreasing. Therefore, the unique existence of $\check{\eta}\in[0,\bar{\eta}]$ such that $h(\check{\eta})=0$ is implied by $h(\bar{\eta})\le 0\le h(0)$, or equivalently,

$$\lambda_1\int_{\bar{\eta}}^1 v(y)\,\mathrm{d}G(y)+v_E\le 0\le \lambda_1\left(v_E-v(0)\right)-v(0)+v_E.$$

The second part of the inequality holds by $v_E-v(0)>0$.

To prove the second part, we first verify that $v_E\le 0$, which follows from

$$\check{\eta}\text{ uniquely exists}\quad\Rightarrow\quad\lambda_1\underbrace{\int_{\bar{\eta}}^1 v(y)\,\mathrm{d}G(y)}_{>0}+v_E\le 0\quad\Rightarrow\quad v_E<0.$$

Define

$$t_1(\eta)\equiv v(\eta)\,\frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}+\lambda_0\int_\eta^1 v(y)\,\mathrm{d}G(y),\quad\forall\eta\in[0,\bar{\eta}].$$

Taking derivative yields

$$\begin{aligned}t_1'(\eta)&=v'(\eta)\,\frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}+v(\eta)\,\frac{\mathrm{d}}{\mathrm{d}\eta}\left[\frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\right]-\lambda_0\,v(\eta)\,g(\eta)\\&=v'(\eta)\,\frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}-\lambda_0\,v(\eta)\,g(\eta)\;\frac{\mathrm{d}}{\mathrm{d}c}\left[c+\frac{F(c)}{f(c)}\right]\Big|_{c=\underline{c}(\eta)},\end{aligned}$$

which is positive for any $\eta\in[0,\bar{\eta}]$. Therefore, to prove the unique existence of $\tilde{\eta}$, it suffices to verify that $t_1(\check{\eta})\le 0\le t_1(\bar{\eta})$, where the second inequality follows from

$$t_1(\bar{\eta})=\lambda_0\int_{\bar{\eta}}^1 v(y)\,\mathrm{d}G(y)>0.$$

Because $\check{c}<\underline{c}(\check{\eta})$, by definition of $\check{c}$, we have

$$v(\check{\eta})\;\left[\underline{c}(\check{\eta})+\frac{F(\underline{c}(\check{\eta}))}{f(\underline{c}(\check{\eta}))}\right]+\lambda_0\int_{\check{\eta}}^1[v(y)-v(\check{\eta})]\,\mathrm{d}G(y)<0,$$

which implies that

$$\begin{aligned}t_1(\check{\eta})&=v(\check{\eta})\,\frac{F(\underline{c}(\check{\eta}))}{f(\underline{c}(\check{\eta}))}+\lambda_0\int_{\check{\eta}}^1 v(y)\,\mathrm{d}G(y)\\&<-v(\check{\eta})\,\underline{c}(\check{\eta})-\lambda_0\int_{\check{\eta}}^1[v(y)-v(\check{\eta})]\,\mathrm{d}G(y)+\lambda_0\int_{\check{\eta}}^1 v(y)\,\mathrm{d}G(y)\\&=-v(\check{\eta})\,\underline{c}(\check{\eta})+\lambda_0\,v(\check{\eta})\,[1-G(\check{\eta})]\\&=0.\end{aligned}$$

Therefore, we know the unique existence of $\tilde{\eta}\in[\check{\eta},\bar{\eta}]$, and the monotonicity property implies the desired result, i.e.,

$$h(\tilde{\eta})\le h(\check{\eta})=0.$$

It remains to prove the last part of the lemma. Define

$$t_2(\eta) \equiv v(\eta)\,\frac{F(\bar{c}(\eta))}{f(\bar{c}(\eta))} + \left[v_E + (\lambda_0+\lambda_1)\int_\eta^1 v(y)\,\mathrm{d}G(y)\right], \quad \forall \eta \in [0,\bar{\eta}].$$

Taking derivative yields

$$\begin{aligned} t_2'(\eta) &= v'(\eta)\,\frac{F(\bar{c}(\eta))}{f(\bar{c}(\eta))} + v(\eta)\,\frac{\mathrm{d}}{\mathrm{d}\eta}\left[\frac{F(\bar{c}(\eta))}{f(\bar{c}(\eta))}\right] - (\lambda_0+\lambda_1)\,v(\eta)\,g(\eta) \\ &= v'(\eta)\,\frac{F(\bar{c}(\eta))}{f(\bar{c}(\eta))} - (\lambda_0+\lambda_1)\,v(\eta)\,g(\eta)\;\frac{\mathrm{d}}{\mathrm{d}c}\left[c+\frac{F(c)}{f(c)}\right]\Big|_{c=\bar{c}(\eta)}, \end{aligned}$$

which is positive for any $\eta \in [0,\bar{\eta}]$. When $h(\bar{\eta}) > 0$, we have

$$t_2(\bar{\eta}) = v_E + (\lambda_0+\lambda_1)\int_{\bar{\eta}}^1 v(y)\,\mathrm{d}G(y) > \lambda_0 \int_{\bar{\eta}}^1 v(y)\,\mathrm{d}G(y) > 0.$$

When $h(\bar{\eta}) \le 0$ and $\check{c} > \bar{c}(\check{\eta})$, by definition of $\check{c}$, we have

$$v(\check{\eta})\,\left[\bar{c}(\check{\eta}) + \frac{F(\bar{c}(\check{\eta}))}{f(\bar{c}(\check{\eta}))}\right] + \lambda_0 \int_{\check{\eta}}^1 [v(y) - v(\check{\eta})]\,\mathrm{d}G(y) > 0,$$

which can be rewritten as

$$v(\check{\eta})\,\frac{F(\bar{c}(\check{\eta}))}{f(\bar{c}(\check{\eta}))} + \lambda_0 \int_{\check{\eta}}^1 v(y)\,\mathrm{d}G(y) > -v(\check{\eta})\,[1+\lambda_1\,[1-G(\check{\eta})]]\,.$$

Because of the monotonicity of t_2, to prove the desired result, it suffices to verify that $t_2(\check{\eta}) \ge 0$, which follows from

$$\begin{aligned} t_2(\check{\eta}) &= v(\check{\eta})\,\frac{F(\bar{c}(\check{\eta}))}{f(\bar{c}(\check{\eta}))} + \left[v_E + (\lambda_0+\lambda_1)\int_{\check{\eta}}^1 v(y)\,\mathrm{d}G(y)\right] \\ &> -v(\check{\eta})\,[1+\lambda_1\,[1-G(\check{\eta})]] + v_E + \lambda_1 \int_{\check{\eta}}^1 v(y)\,\mathrm{d}G(y) \\ &= \underbrace{v_E - v(\check{\eta})}_{=-\lambda_1\int_{\check{\eta}}^1 [v(y)-v(\check{\eta})]\,\mathrm{d}G(y)} + \lambda_1\left[\int_{\check{\eta}}^1 v(y)\,\mathrm{d}G(y) - v(\check{\eta})\,[1-G(\check{\eta})]\right] \\ &= 0. \end{aligned}$$

Therefore, the threshold $\underline{\eta}$ is defined as

$$\underline{\eta} = \begin{cases} \mathring{\eta}, \text{ if } t_2(0) \le 0 \\ 0, \text{ if } t_2(0) > 0, \end{cases}$$

where $\mathring{\eta} \in [0,\check{\eta}]$ is the unique zero of t_2.

To prove $h(\underline{\eta}) \ge 0$, note that when $h(\bar{\eta}) > 0$, we have

$$h(\underline{\eta}) \ge h(\bar{\eta}) > 0.$$

When $\check{c} > \bar{c}(\check{\eta})$, the monotonicity of h implies that

$$h(\underline{\eta}) \ge h(\check{\eta}) = 0.$$

This completes the proof. □

Proof of Proposition 3. It is straightforward to verify that our candidate solutions are feasible in all cases.

To prove the optimality, we start with the first case. Define

$$\kappa = -v(\check{\eta}),$$

$$\nu(c,\eta) = [v(\check{\eta}) - v(\eta)]\ f(c)\, g(\eta), \quad \forall (c,\eta) \in [c_L, c_H] \times [0, \check{\eta}],$$

$$\mu(c,\eta) = [v(\eta) - v(\check{\eta})]\ f(c)\, g(\eta), \quad \forall (c,\eta) \in [c_L, c_H] \times [\check{\eta}, 1],$$

$$\sigma(c) = v_E\,[c\, f(c) + F(c)] + (\lambda_0 + \lambda_1\, c) \int_{\check{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta + \lambda_1 \int_{c_L}^{c}\int_{\check{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y, \qquad \forall c \in \mathcal{C}_1, \text{ and}$$

$$\beta(c) = -v_E\,[c\, f(c) + F(c)] - (\lambda_0 + \lambda_1\, c) \int_{\check{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \lambda_1 \int_{c_L}^{c}\int_{\check{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y, \quad \forall c \in \mathcal{C}_2,$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are subsets of $[c_L, c_H]$ satisfying: (i) $\mathcal{C}_1 \cup \mathcal{C}_2 = [c_L, c_H]$; and (ii) the measure of $\mathcal{C}_1 \cap \mathcal{C}_2$ is 0. These subsets will be specified later.

The nonnegativity of κ, ν, and μ follows from the definition of $\check{\eta}$ and the monotonicity of v directly.

The expressions of σ and β can be rewritten as

$$\sigma(c) = \underbrace{\left[v(\check{\eta})\ \left[c + \frac{F(c)}{f(c)}\right] + \lambda_0 \int_{\check{\eta}}^{1} [v(\eta) - v(\check{\eta})]\ g(\eta)\,\mathrm{d}\eta\right]}_{\text{decreasing in } c} f(c), \quad \forall c \in \mathcal{C}_1,$$

and

$$\beta(c) = -\left[v(\check{\eta})\ \left[c + \frac{F(c)}{f(c)}\right] + \lambda_0 \int_{\check{\eta}}^{1} [v(\eta) - v(\check{\eta})]\ g(\eta)\,\mathrm{d}\eta\right] f(c), \quad \forall c \in \mathcal{C}_2.$$

In the two possible cases, we have:

1. if $c_L \le \check{c} < c_H$, we know that σ and β are nonnegative on $[c_L, \check{c}]$ and $[\check{c}, c_H]$, respectively. In this case, we have $\mathcal{C}_1 = [c_L, \check{c}]$ and $\mathcal{C}_2 = [\check{c}, c_H]$; and
2. if $\check{c} = c_H$, we know that σ is nonnegative on $[c_L, c_H]$. In this case, we have $\mathcal{C}_1 = [c_L, c_H]$ and $\mathcal{C}_2 = \emptyset$.

Because

$$-\kappa - \lambda_1 \int_{c_L}^{c_H}\int_{\check{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c = v_E,$$

$\forall (c,\eta) \in [c_L, c_H] \times [0, \check{\eta}]$:

$$-\nu(c,\eta) + \lambda_1\, g(\eta) \int_{\check{\eta}}^{1} \mu(c,z)\,\mathrm{d}z = [v(\eta) - v_E]\ g(\eta)\, f(c),$$

$\forall (c,\eta) \in [c_L, c_H] \times [\check{\eta}, 1]$:

$$\mu(c,\eta) + \lambda_1\, g(\eta) \int_{\check{\eta}}^{1} \mu(c,z)\,\mathrm{d}z = [v(\eta) - v_E]\ g(\eta)\, f(c),$$

$\forall c \in \mathcal{C}_1$:

$$\sigma(c) - (\lambda_0 + \lambda_1\, c) \int_{\check{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \lambda_1 \int_{c_L}^{c}\int_{\check{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y = v_E\,[c\,f(c) + F(c)],$$

$\forall c \in \mathcal{C}_2$:

$$-\beta(c) - (\lambda_0 + \lambda_1\, c) \int_{\check{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \lambda_1 \int_{c_L}^{c}\int_{\check{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y = v_E\,[c\,f(c) + F(c)],$$

we have

$$\begin{aligned}
&\kappa\,[-\Pi(c_H)] + \int_{\mathcal{C}_1} \sigma(c)\,\tau(c)\,\mathrm{d}c + \int_{\mathcal{C}_2} \beta(c)\,[-\tau(c)]\,\mathrm{d}c + \int_{c_L}^{c_H}\int_{0}^{\check{\eta}} \nu(c,\eta)\,[-\alpha_2(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c \\
&\quad + \int_{c_L}^{c_H}\int_{\check{\eta}}^{1} \mu(c,\eta)\left[\alpha_2(c,\eta) - \lambda_0\,\tau(c) - \lambda_1\left[\Pi(c_H) + \int_{c}^{c_H} \tau(y)\,\mathrm{d}y + c\,\tau(c) - \int_{0}^{1} \alpha_2(c,z)\,\mathrm{d}G(z)\right]\right]\mathrm{d}\eta\,\mathrm{d}c \\
=\ &\left[-\kappa - \lambda_1 \int_{c_L}^{c_H}\int_{\check{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\right]\Pi(c_H) \\
&\quad + \int_{\mathcal{C}_1}\left[\sigma(c) - (\lambda_0 + \lambda_1\, c)\int_{\check{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \lambda_1 \int_{c_L}^{c}\int_{\check{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c \\
&\quad + \int_{\mathcal{C}_2}\left[-\beta(c) - (\lambda_0 + \lambda_1\, c)\int_{\check{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \lambda_1 \int_{c_L}^{c}\int_{\check{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c \\
&\quad + \int_{c_L}^{c_H}\int_{0}^{\check{\eta}}\left[-\nu(c,\eta) + \lambda_1\, g(\eta)\int_{\check{\eta}}^{1} \mu(c,z)\,\mathrm{d}z\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&\quad + \int_{c_L}^{c_H}\int_{\check{\eta}}^{1}\left[\mu(c,\eta) + \lambda_1\, g(\eta)\int_{\check{\eta}}^{1} \mu(c,z)\,\mathrm{d}z\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
=\ &v_E\,\Pi(c_H) + \int_{c_L}^{c_H}\int_{0}^{1} [v(\eta) - v_E]\,\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c) + v_E \int_{c_L}^{c_H}\left[c + \frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\,\mathrm{d}c.
\end{aligned}$$

Therefore, the objective function is bounded by

$$\int_{\mathcal{C}_1} \sigma(c)\,\mathrm{d}c.$$

Substituting our candidate solution into the objective function achieves the upper bound in all the subcases, which follows from:

$$\begin{aligned}
&\int_{c_L}^{\check{c}} \sigma(c)\,\mathrm{d}c - \left[v_E\,\alpha^* + (\lambda_0 + \lambda_1\,\alpha^*)\int_{\check{\eta}}^{1} v(\eta)\,\mathrm{d}G(\eta)\right]F(\check{c}) \\
=\ &\int_{c_L}^{\check{c}}\left[v(\check{\eta})\left[c + \frac{F(c)}{f(c)}\right] + \lambda_0 \int_{\check{\eta}}^{1} [v(\eta) - v(\check{\eta})]\,g(\eta)\,\mathrm{d}\eta\right]f(c)\,\mathrm{d}c \\
&\quad - \left[v_E\,\alpha^* + (\lambda_0 + \lambda_1\,\alpha^*)\int_{\check{\eta}}^{1} v(\eta)\,\mathrm{d}G(\eta)\right]F(\check{c}) \\
=\ &v(\check{\eta})\int_{c_L}^{\check{c}} [(c - \check{c})\,f(c) + F(c)]\,\mathrm{d}c \\
=\ &0,
\end{aligned}$$

where $\alpha^* = \frac{\check{c}-\lambda_0\,[1-G(\tilde{\eta})]}{1+\lambda_1\,[1-G(\tilde{\eta})]}$.

This establishes the optimality in the first case.

To prove the optimality in the second case, we first define

$$\kappa = -v(\tilde{\eta}),$$

$$\nu(c,\eta) = [v(\tilde{\eta}) - v(\eta)]\; f(c)\, g(\eta), \qquad \forall (c,\eta) \in [c_L, c_H] \times [0, \tilde{\eta}],$$

$$\mu(c,\eta) = [v(\eta) - v(\tilde{\eta})]\; f(c)\, g(\eta), \qquad \forall (c,\eta) \in [c_L, c_H] \times [\tilde{\eta}, 1],$$

$$\phi(c) = -\lambda_1 \int_{\tilde{\eta}}^{1} \mu(c,y)\,\mathrm{d}y + [v(\tilde{\eta}) - v_E]\; f(c), \qquad \forall c \in [c_L, c_H],$$

$$\begin{aligned}\sigma(c) = {} & v_E\,[c\,f(c) + F(c)] + (\lambda_0 + \lambda_1\,c) \int_{\tilde{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta \\ & + \lambda_1 \int_{c_L}^{c} \int_{\tilde{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y + \int_{c_L}^{c} \phi(y)\,\mathrm{d}y + c\,\phi(c), \qquad \forall c \in \mathcal{C}_1,\end{aligned}$$

$$\begin{aligned}\beta(c) = {} & -v_E\,[c\,f(c) + F(c)] - (\lambda_0 + \lambda_1\,c) \int_{\tilde{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta \\ & - \lambda_1 \int_{c_L}^{c} \int_{\tilde{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y - \int_{c_L}^{c} \phi(y)\,\mathrm{d}y - c\,\phi(c), \qquad \forall c \in \mathcal{C}_2,\end{aligned}$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are subsets of $[c_L, c_H]$ satisfying: (i) $\mathcal{C}_1 \cup \mathcal{C}_2 = [c_L, c_H]$; and (ii) the measure of $\mathcal{C}_1 \cap \mathcal{C}_2$ is 0. These subsets will be specified later.

The nonnegativity of κ, ν, and μ follows from the definition of $\tilde{\eta}$ and the monotonicity of v directly.

The nonnegativity of ϕ follows from the second part of Lemma 2.

The expressions for σ and β can be simplified as

$$\sigma(c) = v(\tilde{\eta})\; [(c - \underline{c}(\tilde{\eta}))\, f(c) + F(c)] + \lambda_0\, f(c) \int_{\tilde{\eta}}^{1} v(\eta)\,\mathrm{d}G(\eta), \quad \forall c \in \mathcal{C}_1,$$

and

$$\beta(c) = -v(\tilde{\eta})\; [(c - \underline{c}(\tilde{\eta}))\, f(c) + F(c)] - \lambda_0\, f(c) \int_{\tilde{\eta}}^{1} v(\eta)\,\mathrm{d}G(\eta), \quad \forall c \in \mathcal{C}_2.$$

To prove the nonnegativity of σ, taking derivative yields

$$\sigma'(c) = v(\tilde{\eta})\; [(c - \underline{c}(\tilde{\eta}))\, f'(c) + 2\, f(c)] + f'(c) \underbrace{\lambda_0 \int_{\tilde{\eta}}^{1} v(\eta)\,\mathrm{d}G(\eta)}_{= -v(\tilde{\eta})\,\frac{F(\underline{c}(\tilde{\eta}))}{f(\underline{c}(\tilde{\eta}))} \geq 0} \leq 0, \quad \forall c \leq \underline{c}(\tilde{\eta}).$$

Meanwhile, we have

$$\sigma(\underline{c}(\tilde{\eta})) = \left[v(\tilde{\eta})\, \frac{F(\underline{c}(\tilde{\eta}))}{f(\underline{c}(\tilde{\eta}))} + \lambda_0 \int_{\tilde{\eta}}^{1} v(\eta)\,\mathrm{d}G(\eta) \right] f(\underline{c}(\tilde{\eta})) = 0.$$

In the first two possible cases, we have:

1. if $\underline{c}(\tilde{\eta}) < c_L$, we know that β is nonnegative on $[c_L, c_H]$. In this case, we have $\mathcal{C}_1 = \emptyset$ and $\mathcal{C}_2 = [c_L, c_H]$; and
2. if $c_L \le \underline{c}(\tilde{\eta}) \le c_H$, we know that σ and β are nonnegative on $[c_L, \underline{c}(\tilde{\eta})]$ and $[\underline{c}(\tilde{\eta}), c_H]$, respectively. In this case, we have $\mathcal{C}_1 = [c_L, \underline{c}(\tilde{\eta})]$ and $\mathcal{C}_2 = [\underline{c}(\tilde{\eta}), c_H]$.

Therefore, the nonnegativity of σ and β holds automatically.

Because

$$-\kappa - \int_{c_L}^{c_H} \phi(c)\,\mathrm{d}c - \lambda_1 \int_{c_L}^{c_H} \int_{\tilde{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c = v_E,$$

$\forall (c,\eta) \in [c_L, c_H] \times [0, \tilde{\eta}]$:

$$\phi(c)\,g(\eta) - \nu(c,\eta) + \lambda_1\,g(\eta) \int_{\tilde{\eta}}^{1} \mu(c,z)\,\mathrm{d}z = [v(\eta) - v_E]\;g(\eta)\,f(c),$$

$\forall (c,\eta) \in [c_L, c_H] \times [\tilde{\eta}, 1]$:

$$\phi(c)\,g(\eta) + \mu(c,\eta) + \lambda_1\,g(\eta) \int_{\tilde{\eta}}^{1} \mu(c,z)\,\mathrm{d}z = [v(\eta) - v_E]\;g(\eta)\,f(c),$$

$\forall c \in \mathcal{C}_1$:

$$-\int_{c_L}^{c} \phi(y)\,\mathrm{d}y - c\,\phi(c) + \sigma(c) - (\lambda_0 + \lambda_1\,c) \int_{\tilde{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \lambda_1 \int_{c_L}^{c} \int_{\tilde{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y = v_E\,[c\,f(c) + F(c)],$$

$\forall c \in \mathcal{C}_2$:

$$-\int_{c_L}^{c} \phi(y)\,\mathrm{d}y - c\,\phi(c) - \beta(c) - (\lambda_0 + \lambda_1\,c) \int_{\tilde{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \lambda_1 \int_{c_L}^{c} \int_{\tilde{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y = v_E\,[c\,f(c) + F(c)],$$

we have

$$\begin{aligned}
&\kappa\,[-\Pi(c_H)] + \int_{\mathcal{C}_1} \sigma(c)\,\tau(c)\,\mathrm{d}c + \int_{\mathcal{C}_2} \beta(c)\,[-\tau(c)]\,\mathrm{d}c \\
&\quad + \int_{c_L}^{c_H} \phi(c) \left[-\Pi(c_H) - \int_{c}^{c_H} \tau(y)\,\mathrm{d}y - c\,\tau(c) + \int_{0}^{1} \alpha_2(c,\eta)\,\mathrm{d}G(\eta) \right] \mathrm{d}c \\
&\quad + \int_{c_L}^{c_H} \int_{0}^{\tilde{\eta}} \nu(c,\eta)\,[-\alpha_2(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c \\
&\quad + \int_{c_L}^{c_H} \int_{\tilde{\eta}}^{1} \mu(c,\eta) \left[\alpha_2(c,\eta) - \lambda_0\,\tau(c) - \lambda_1 \left[\Pi(c_H) + \int_{c}^{c_H} \tau(y)\,\mathrm{d}y + c\,\tau(c) - \int_{0}^{1} \alpha_2(c,z)\,\mathrm{d}G(z) \right] \right] \mathrm{d}\eta\,\mathrm{d}c \\
= &\left[-\kappa - \int_{c_L}^{c_H} \phi(c)\,\mathrm{d}c - \lambda_1 \int_{c_L}^{c_H} \int_{\tilde{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \right] \Pi(c_H) \\
&\quad + \int_{\mathcal{C}_1} \left[\sigma(c) - \int_{c_L}^{c} \phi(y)\,\mathrm{d}y - c\,\phi(c) - (\lambda_0 + \lambda_1\,c) \int_{\tilde{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \lambda_1 \int_{c_L}^{c} \int_{\tilde{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y \right] \tau(c)\,\mathrm{d}c \\
&\quad + \int_{\mathcal{C}_2} \left[-\beta(c) - \int_{c_L}^{c} \phi(y)\,\mathrm{d}y - c\,\phi(c) - (\lambda_0 + \lambda_1\,c) \int_{\tilde{\eta}}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \lambda_1 \int_{c_L}^{c} \int_{\tilde{\eta}}^{1} \mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y \right] \tau(c)\,\mathrm{d}c \\
&\quad + \int_{c_L}^{c_H} \int_{0}^{\tilde{\eta}} \left[\phi(c)\,g(\eta) - \nu(c,\eta) + \lambda_1\,g(\eta) \int_{\tilde{\eta}}^{1} \mu(c,z)\,\mathrm{d}z \right] \alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c
\end{aligned}$$

$$+\int_{c_L}^{c_H}\int_{\tilde{\eta}}^{1}\left[\phi(c)\,g(\eta)+\mu(c,\eta)+\lambda_1\,g(\eta)\int_{\tilde{\eta}}^{1}\mu(c,z)\,\mathrm{d}z\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c$$
$$=v_E\,\Pi(c_H)+\int_{c_L}^{c_H}\int_{0}^{1}\left[v(\eta)-v_E\right]\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c)+v_E\int_{c_L}^{c_H}\left[c+\frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\;\mathrm{d}c.$$

Therefore, the objective function is bounded by

$$\int_{\mathcal{C}_1}\sigma(c)\,\mathrm{d}c.$$

Substituting our candidate solution into the objective function achieves the upper bound in the first two cases, which follows from:

1. if $\underline{c}(\tilde{\eta})\le c_L$, we have $c^*=c_L$ and the measure of $\mathcal{C}_1$ is 0. The regulator's expected payoff generated by our candidate policy and its upper bound are both 0;
2. if $c_L<\underline{c}(\tilde{\eta})\le c_H$, we have

$$\begin{aligned}
&\int_{c_L}^{\underline{c}(\tilde{\eta})}\sigma(c)\,\mathrm{d}c-\lambda_0\,F(\underline{c}(\tilde{\eta}))\int_{\tilde{\eta}}^{1}v(\eta)\,\mathrm{d}G(\eta)\\
&=\int_{c_L}^{\underline{c}(\tilde{\eta})}\left[v(\tilde{\eta})\,\left[[c-\underline{c}(\tilde{\eta})]\,f(c)+F(c)\right]+\lambda_0\,f(c)\int_{\tilde{\eta}}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]\mathrm{d}c-\lambda_0\,F(\underline{c}(\tilde{\eta}))\int_{\tilde{\eta}}^{1}v(\eta)\,\mathrm{d}G(\eta)\\
&=v(\tilde{\eta})\int_{c_L}^{\underline{c}(\tilde{\eta})}\left[[c-\underline{c}(\tilde{\eta})]\,f(c)+F(c)\right]\mathrm{d}c\\
&=0.
\end{aligned}$$

To verify the optimality in case (2)(c), define

$$\begin{aligned}
&\kappa=-v(\eta^*),\\
&\nu(c,\eta)=\left[v(\eta^*)-v(\eta)\right]\,f(c)\,g(\eta), && \forall(c,\eta)\in[c_L,c_H]\times[0,\eta^*],\\
&\mu(c,\eta)=\left[v(\eta)-v(\eta^*)\right]\,f(c)\,g(\eta), && \forall(c,\eta)\in[c_L,c_H]\times[\eta^*,1],\\
&\phi(c)=-\lambda_1\int_{\eta^*}^{1}\mu(c,y)\,\mathrm{d}y+\left[v(\eta^*)-v_E\right]\,f(c), && \forall c\in[c_L,c_H],\\
&\sigma(c)=v_E\left[c\,f(c)+F(c)\right]+(\lambda_0+\lambda_1\,c)\int_{\eta^*}^{1}\mu(c,\eta)\,\mathrm{d}\eta\\
&\qquad\quad+\lambda_1\int_{c_L}^{c}\int_{\eta^*}^{1}\mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y+\int_{c_L}^{c}\phi(y)\,\mathrm{d}y+c\,\phi(c), && \forall c\in[c_L,c_H],
\end{aligned}$$

where $\eta^*\equiv\sup\{\eta\in[0,1]:\underline{c}(\eta)\ge c_H\}$. Because $\underline{c}(\tilde{\eta})>c_H>\underline{c}(\bar{\eta})$, we know that $\tilde{\eta}<\eta^*<\bar{\eta}$. The nonnegativity of κ, ν, and μ follows from the corresponding definitions directly. Because

$$\phi(c)\ge-\lambda_1\int_{\tilde{\eta}}^{1}\left[v(y)-v(\tilde{\eta})\right]\,f(c)\,g(y)\,\mathrm{d}y+\left[v(\tilde{\eta})-v_E\right]\,f(c)\ge 0,\quad\forall c\in[c_L,c_H],$$

the desired nonnegativity condition holds automatically. Finally, we have

$$\sigma(c) = v(\eta^*)\,[(c-c_H)\,f(c)+F(c)] + \lambda_0\,f(c)\int_{\eta^*}^1 v(\eta)\,\mathrm{d}G(\eta), \quad \forall c\in[c_L,c_H].$$

The monotonicity of σ still holds, and we have

$$\begin{aligned}\sigma(c_H) &= \left[v(\eta^*)\,\frac{F(c_H)}{f(c_H)} + \lambda_0\int_{\eta^*}^1 v(\eta)\,\mathrm{d}G(\eta)\right] f(c_H)\\ &\geq \left[v(\tilde\eta)\,\frac{F(\underline{c}(\tilde\eta))}{f(\underline{c}(\tilde\eta))} + \lambda_0\int_{\tilde\eta}^1 v(\eta)\,\mathrm{d}G(\eta)\right] f(c_H)\\ &= 0,\end{aligned}$$

where the inequality follows from the monotonicity property of v and the fact that $\eta^*>\tilde\eta$.

Because

$$\begin{aligned}&\kappa\,[-\Pi(c_H)] + \int_{c_L}^{c_H}\sigma(c)\,\tau(c)\,\mathrm{d}c\\ &+\int_{c_L}^{c_H}\phi(c)\left[-\Pi(c_H) - \int_c^{c_H}\tau(y)\,\mathrm{d}y - c\,\tau(c) + \int_0^1\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\right]\mathrm{d}c\\ &+\int_{c_L}^{c_H}\int_0^{\eta^*}\nu(c,\eta)\,[-\alpha_2(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c\\ &+\int_{c_L}^{c_H}\int_{\eta^*}^1\mu(c,\eta)\left[\alpha_2(c,\eta)-\lambda_0\,\tau(c)-\lambda_1\left[\Pi(c_H)+\int_c^{c_H}\tau(y)\,\mathrm{d}y + c\,\tau(c) - \int_0^1\alpha_2(c,z)\,\mathrm{d}G(z)\right]\right]\mathrm{d}\eta\,\mathrm{d}c\\ =\ &\left[-\kappa - \int_{c_L}^{c_H}\phi(c)\,\mathrm{d}c - \lambda_1\int_{c_L}^{c_H}\int_{\eta^*}^1\mu(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\right]\Pi(c_H)\\ &+\int_{c_L}^{c_H}\left[\sigma(c) - \int_{c_L}^{c}\phi(y)\,\mathrm{d}y - c\,\phi(c) - (\lambda_0+\lambda_1\,c)\int_{\eta^*}^1\mu(c,\eta)\,\mathrm{d}\eta - \lambda_1\int_{c_L}^{c}\int_{\eta^*}^1\mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\ &+\int_{c_L}^{c_H}\int_0^{\eta^*}\left[\phi(c)\,g(\eta) - \nu(c,\eta) + \lambda_1\,g(\eta)\int_{\eta^*}^1\mu(c,z)\,\mathrm{d}z\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\ &+\int_{c_L}^{c_H}\int_{\eta^*}^1\left[\phi(c)\,g(\eta) + \mu(c,\eta) + \lambda_1\,g(\eta)\int_{\eta^*}^1\mu(c,z)\,\mathrm{d}z\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\ =\ &v_E\,\Pi(c_H) + \int_{c_L}^{c_H}\int_0^1\left[v(\eta)-v_E\right]\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c) + v_E\int_{c_L}^{c_H}\left[c+\frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\,\mathrm{d}c,\end{aligned}$$

the objective function is bounded by

$$\int_{c_L}^{c_H}\sigma(c)\,\mathrm{d}c.$$

Substituting our candidate solution into the objective function achieves the upper bound, which follows from

$$\begin{aligned}&\int_{c_L}^{c_H}\sigma(c)\,\mathrm{d}c - \lambda_0\int_{\eta^*}^1 v(\eta)\,\mathrm{d}G(\eta)\\ =\ &\int_{c_L}^{c_H}\left[v(\eta^*)\,[(c-c_H)\,f(c)+F(c)] + f(c)\,\lambda_0\int_{\eta^*}^1 v(\eta)\,\mathrm{d}G(\eta)\right]\mathrm{d}c - \lambda_0\int_{\eta^*}^1 v(\eta)\,\mathrm{d}G(\eta)\end{aligned}$$

$$= v(\eta^*) \int_{c_L}^{c_H} [(c - c_H) f(c) + F(c)] \, \mathrm{d}c$$
$$= 0.$$

It remains to prove the last part of the proposition, where the optimal mechanism is to grant full conditional approval to firms whose types are below the threshold c^*. In the first two subcases, we have $\bar{c}(\underline{\eta}) < c_H \leq \bar{c}(0)$, implying that $\underline{\eta} = \eta^*$.

Define

$$\kappa = -v(\underline{\eta}),$$

$$\nu(c,\eta) = \left[v(\underline{\eta}) - v(\eta)\right] f(c)\, g(\eta), \qquad \forall (c,\eta) \in [c_L, c_H] \times [0, \underline{\eta}],$$

$$\mu(c,\eta) = \left[v(\eta) - v(\underline{\eta})\right] f(c)\, g(\eta), \qquad \forall (c,\eta) \in [c_L, c_H] \times [\underline{\eta}, 1],$$

$$\psi(c) = \lambda_1 \int_{\underline{\eta}}^{1} \mu(c,y) \, \mathrm{d}y - \left[v(\underline{\eta}) - v_E\right] f(c), \qquad \forall c \in [c_L, c_H],$$

$$\sigma(c) = v_E \left[c\, f(c) + F(c)\right] + (\lambda_0 + \lambda_1 c) \int_{\underline{\eta}}^{1} \mu(c,\eta) \, \mathrm{d}\eta$$
$$+ \lambda_1 \int_{c_L}^{c} \int_{\underline{\eta}}^{1} \mu(y,\eta) \, \mathrm{d}\eta \, \mathrm{d}y - \int_{c_L}^{c} \psi(y) \, \mathrm{d}y - (c-1)\, \psi(c), \qquad \forall c \in \mathcal{C}_1,$$

$$\beta(c) = -v_E \left[c\, f(c) + F(c)\right] - (\lambda_0 + \lambda_1 c) \int_{\underline{\eta}}^{1} \mu(c,\eta) \, \mathrm{d}\eta$$
$$- \lambda_1 \int_{c_L}^{c} \int_{\underline{\eta}}^{1} \mu(y,\eta) \, \mathrm{d}\eta \, \mathrm{d}y + \int_{c_L}^{c} \psi(y) \, \mathrm{d}y + (c-1)\, \psi(c), \qquad \forall c \in \mathcal{C}_2,$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are subsets of $[c_L, c_H]$ satisfying: (i) $\mathcal{C}_1 \cup \mathcal{C}_2 = [c_L, c_H]$; and (ii) the measure of $\mathcal{C}_1 \cap \mathcal{C}_2$ is 0. These subsets will be specified later.

The nonnegativity of κ, ν, and μ follows from the definition of $\underline{\eta}$ and the monotonicity of v directly.

The nonnegativity of ψ follows from the last part of Lemma 2.

The expressions of σ and β can be rewritten as

$$\sigma(c) = v(\underline{\eta}) \left[(c - \bar{c}(\underline{\eta}))\, f(c) + F(c)\right] + f(c) \left[v_E + (\lambda_0 + \lambda_1) \int_{\underline{\eta}}^{1} v(y) \, \mathrm{d}G(y)\right]$$

and

$$\beta(c) = -v(\underline{\eta}) \left[(c - \bar{c}(\underline{\eta}))\, f(c) + F(c)\right] - f(c) \left[v_E + (\lambda_0 + \lambda_1) \int_{\underline{\eta}}^{1} v(y) \, \mathrm{d}G(y)\right].$$

Differentiating σ yields

$$\sigma'(c) = v(\underline{\eta}) \left[(c - \bar{c}(\underline{\eta}))\, f'(c) + 2\, f(c)\right] + f'(c) \underbrace{\left[v_E + (\lambda_0 + \lambda_1) \int_{\underline{\eta}}^{1} v(y) \, \mathrm{d}G(y)\right]}_{= -v(\underline{\eta}) \frac{F(\bar{c}(\underline{\eta}))}{f(\bar{c}(\underline{\eta}))}} \leq 0, \quad \forall c \leq \bar{c}(\underline{\eta}).$$

Meanwhile, we have

$$\sigma(\bar{c}(\underline{\eta})) = \left[v(\underline{\eta}) \frac{F(\bar{c}(\underline{\eta}))}{f(\bar{c}(\underline{\eta}))} + \left[v_E + (\lambda_0 + \lambda_1) \int_{\underline{\eta}}^{1} v(y) \, \mathrm{d}G(y) \right] \right] f(\bar{c}(\underline{\eta})) = 0.$$

In the first two possible cases, we have:

1. if $\bar{c}(\underline{\eta}) < c_L$, we know that β is nonnegative on $[c_L, c_H]$. In this case, we have $\mathcal{C}_1 = \emptyset$ and $\mathcal{C}_2 = [c_L, c_H]$;
2. if $c_L \le \bar{c}(\underline{\eta}) \le c_H$, we know that σ and β are nonnegative on $[c_L, \bar{c}(\underline{\eta})]$ and $[\bar{c}(\underline{\eta}), c_H]$, respectively. In this case, we have $\mathcal{C}_1 = [c_L, \bar{c}(\underline{\eta})]$ and $\mathcal{C}_2 = [\bar{c}(\underline{\eta}), c_H]$.

Therefore, the nonnegativity of σ and β holds automatically.

Because

$$-\kappa + \int_{c_L}^{c_H} \psi(c) \, \mathrm{d}c - \lambda_1 \int_{c_L}^{c_H} \int_{\underline{\eta}}^{1} \mu(c, \eta) \, \mathrm{d}\eta \, \mathrm{d}c = v_E,$$

$\forall (c, \eta) \in [c_L, c_H] \times [0, \underline{\eta}]$:

$$-\psi(c) \, g(\eta) - \nu(c, \eta) + \lambda_1 \, g(\eta) \int_{\underline{\eta}}^{1} \mu(c, y) \, \mathrm{d}y = [v(\eta) - v_E] \, g(\eta) \, f(c),$$

$\forall (c, \eta) \in [c_L, c_H] \times [\underline{\eta}, 1]$:

$$-\psi(c) \, g(\eta) + \mu(c, \eta) + \lambda_1 \, g(\eta) \int_{\underline{\eta}}^{1} \mu(c, y) \, \mathrm{d}y = [v(\eta) - v_E] \, g(\eta) \, f(c),$$

$\forall c \in \mathcal{C}_1$:

$$\int_{c_L}^{c} \psi(y) \, \mathrm{d}y + (c-1) \, \psi(c) + \sigma(c) - (\lambda_0 + \lambda_1 \, c) \int_{\underline{\eta}}^{1} \mu(c, \eta) \, \mathrm{d}\eta - \lambda_1 \int_{c_L}^{c} \int_{\underline{\eta}}^{1} \mu(y, \eta) \, \mathrm{d}\eta \, \mathrm{d}y = v_E \, [c \, f(c) + F(c)],$$

$\forall c \in \mathcal{C}_2$:

$$\int_{c_L}^{c} \psi(y) \, \mathrm{d}y + (c-1) \, \psi(c) - \beta(c) - (\lambda_0 + \lambda_1 \, c) \int_{\underline{\eta}}^{1} \mu(c, \eta) \, \mathrm{d}\eta - \lambda_1 \int_{c_L}^{c} \int_{\underline{\eta}}^{1} \mu(y, \eta) \, \mathrm{d}\eta \, \mathrm{d}y = v_E \, [c \, f(c) + F(c)],$$

we have

$$\begin{aligned}
&\kappa \, [-\Pi(c_H)] + \int_{\mathcal{C}_1} \sigma(c) \, \tau(c) \, \mathrm{d}c + \int_{\mathcal{C}_2} \beta(c) \, [-\tau(c)] \, \mathrm{d}c \\
&+ \int_{c_L}^{c_H} \psi(c) \left[\Pi(c_H) + \int_{c}^{c_H} \tau(y) \, \mathrm{d}y + c \, \tau(c) - \int_{0}^{1} \alpha_2(c, \eta) \, \mathrm{d}G(\eta) - \tau(c) \right] \mathrm{d}c \\
&+ \int_{c_L}^{c_H} \int_{0}^{\underline{\eta}} \nu(c, \eta) \, [-\alpha_2(c, \eta)] \, \mathrm{d}\eta \, \mathrm{d}c \\
&+ \int_{c_L}^{c_H} \int_{\underline{\eta}}^{1} \mu(c, \eta) \left[\alpha_2(c, \eta) - \lambda_0 \, \tau(c) - \lambda_1 \left[\Pi(c_H) + \int_{c}^{c_H} \tau(y) \, \mathrm{d}y + c \, \tau(c) - \int_{0}^{1} \alpha_2(c, z) \, \mathrm{d}G(z) \right] \right] \mathrm{d}\eta \, \mathrm{d}c \\
= &\left[-\kappa + \int_{c_L}^{c_H} \psi(c) \, \mathrm{d}c - \lambda_1 \int_{c_L}^{c_H} \int_{\underline{\eta}}^{1} \mu(c, \eta) \, \mathrm{d}\eta \, \mathrm{d}c \right] \Pi(c_H)
\end{aligned}$$

$$
\begin{aligned}
&+\int_{\mathcal{C}_1}\left[\sigma(c)+\int_{c_L}^{c}\psi(y)\,\mathrm{d}y+(c-1)\,\psi(c)-(\lambda_0+\lambda_1\,c)\int_{\underline{\eta}}^{1}\mu(c,\eta)\,\mathrm{d}\eta-\lambda_1\int_{c_L}^{c}\int_{\underline{\eta}}^{1}\mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
&+\int_{\mathcal{C}_2}\left[-\beta(c)+\int_{c_L}^{c}\psi(y)\,\mathrm{d}y+(c-1)\,\psi(c)-(\lambda_0+\lambda_1\,c)\int_{\underline{\eta}}^{1}\mu(c,\eta)\,\mathrm{d}\eta-\lambda_1\int_{c_L}^{c}\int_{\underline{\eta}}^{1}\mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{0}^{\underline{\eta}}\left[-\psi(c)\,g(\eta)-\nu(c,\eta)+\lambda_1\,g(\eta)\int_{0}^{1}\mu(c,z)\,\mathrm{d}z\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\underline{\eta}}^{1}\left[-\psi(c)\,g(\eta)+\mu(c,\eta)+\lambda_1\,g(\eta)\int_{0}^{1}\mu(c,z)\,\mathrm{d}z\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
=\;& v_E\,\Pi(c_H)+\int_{c_L}^{c_H}\int_{0}^{1}\left[v(\eta)-v_E\right]\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c)+v_E\int_{c_L}^{c_H}\left[c+\frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\;\mathrm{d}c.
\end{aligned}
$$

Therefore, the objective function is bounded by

$$\int_{\mathcal{C}_1}\sigma(c)\,\mathrm{d}c.$$

Substituting our candidate solution into the objective function achieves the upper bound in both cases, which follows from

1. if $\bar{c}(\underline{\eta})<c_L$, we have $c^*=c_L$ and the measure of $\mathcal{C}_1$ is 0. The regulator's expected payoff generated by our candidate policy and its upper bound are both 0;
2. if $c_L\le\bar{c}(\underline{\eta})\le c_H$, we have

$$
\begin{aligned}
&\int_{c_L}^{\bar{c}(\underline{\eta})}\sigma(c)\,\mathrm{d}c-\left[v_E+(\lambda_0+\lambda_1)\int_{\underline{\eta}}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]F(\bar{c}(\underline{\eta}))\\
=\;&\int_{c_L}^{\bar{c}(\underline{\eta})}\left[v(\underline{\eta})\,\left[(c-\bar{c}(\underline{\eta}))\,f(c)+F(c)\right]+f(c)\,\left[v_E+(\lambda_0+\lambda_1)\int_{\underline{\eta}}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]\right]\mathrm{d}c\\
&-\left[v_E+(\lambda_0+\lambda_1)\int_{\underline{\eta}}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]F(\bar{c}(\underline{\eta}))\\
=\;&v(\underline{\eta})\int_{c_L}^{\bar{c}(\underline{\eta})}\left[(c-\bar{c}(\underline{\eta}))\,f(c)+F(c)\right]\mathrm{d}c\\
=\;&0.
\end{aligned}
$$

To prove the optimality in case (3)(c), define

$$
\begin{aligned}
&\kappa=-v(\eta^*),\\
&\nu(c,\eta)=\left[v(\eta^*)-v(\eta)\right]\,f(c)\,g(\eta), && \forall(c,\eta)\in[c_L,c_H]\times[0,\eta^*],\\
&\mu(c,\eta)=\left[v(\eta)-v(\eta^*)\right]\,f(c)\,g(\eta), && \forall(c,\eta)\in[c_L,c_H]\times[\eta^*,1],\\
&\psi(c)=\lambda_1\int_{\eta^*}^{1}\mu(c,y)\,\mathrm{d}y-\left[v(\eta^*)-v_E\right]\,f(c), && \forall c\in[c_L,c_H],
\end{aligned}
$$

$$\sigma(c) = v_E\left[c\,f(c) + F(c)\right] + (\lambda_0 + \lambda_1\,c)\int_{\eta^*}^{1}\mu(c,\eta)\,\mathrm{d}\eta + \lambda_1\int_{c_L}^{c}\int_{\eta^*}^{1}\mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y - \int_{c_L}^{c}\psi(y)\,\mathrm{d}y - (c-1)\,\psi(c), \qquad \forall c\in[c_L,c_H].$$

The monotonicity of $\bar{c}$ implies that $\underline{\eta}\le\eta^*<\bar{\eta}$. The nonnegativity of κ, ν, and μ follows from their definitions directly.

When $h(\bar{\eta})>0$, we have

$$\psi(c) = h(\eta^*)\,f(c) \ge h(\bar{\eta})\,f(c) > 0.$$

When $\check{c}>\bar{c}(\check{\eta})$, by Lemma 2, we know that ψ is nonnegative if and only if $\eta^*\le\check{\eta}$, which follows from

$$\bar{c}(\check{\eta})<\check{c}\le c_H \quad\Rightarrow\quad \check{\eta}>\eta^*.$$

Similar to the previous case, the expression of σ can be simplified as

$$\sigma(c) = v(\eta^*)\,\left[(c-c_H)\,f(c)+F(c)\right] + f(c)\,\left[v_E + (\lambda_0+\lambda_1)\int_{\eta^*}^{1}v(y)\,\mathrm{d}G(y)\right], \quad \forall c\in[c_L,c_H].$$

The monotonicity of σ still holds, and we have

$$\begin{aligned}\sigma(c_H) &= v(\eta^*) + f(c_H)\,\left[v_E + (\lambda_0+\lambda_1)\int_{\eta^*}^{1}v(y)\,\mathrm{d}G(y)\right]\\ &\ge \left[v(\underline{\eta})\,\frac{F(\bar{c}(\underline{\eta}))}{f(\bar{c}(\underline{\eta}))} + \left[v_E + (\lambda_0+\lambda_1)\int_{\underline{\eta}}^{1}v(y)\,\mathrm{d}G(y)\right]\right]\,f(c_H)\\ &\ge 0,\end{aligned}$$

implying the nonnegativity of σ. Because

$$\begin{aligned}&\kappa\,[-\Pi(c_H)] + \int_{c_L}^{c_H}\sigma(c)\,\tau(c)\,\mathrm{d}c\\ &+\int_{c_L}^{c_H}\psi(c)\,\left[\Pi(c_H) + \int_{c}^{c_H}\tau(y)\,\mathrm{d}y + c\,\tau(c) - \int_{0}^{1}\alpha_2(c,\eta)\,\mathrm{d}G(\eta) - \tau(c)\right]\mathrm{d}c\\ &+\int_{c_L}^{c_H}\int_{0}^{\eta^*}\nu(c,\eta)\,[-\alpha_2(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c\\ &+\int_{c_L}^{c_H}\int_{\eta^*}^{1}\mu(c,\eta)\,\left[\alpha_2(c,\eta) - \lambda_0\,\tau(c) - \lambda_1\,\left[\Pi(c_H) + \int_{c}^{c_H}\tau(y)\,\mathrm{d}y + c\,\tau(c) - \int_{0}^{1}\alpha_2(c,z)\,\mathrm{d}G(z)\right]\right]\mathrm{d}\eta\,\mathrm{d}c\\ =\;&\left[-\kappa + \int_{c_L}^{c_H}\psi(c)\,\mathrm{d}c - \lambda_1\int_{c_L}^{c_H}\int_{\eta^*}^{1}\mu(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\right]\,\Pi(c_H)\\ &+\int_{c_L}^{c_H}\left[\sigma(c) + \int_{c_L}^{c}\psi(y)\,\mathrm{d}y + (c-1)\,\psi(c) - (\lambda_0+\lambda_1\,c)\int_{\eta^*}^{1}\mu(c,\eta)\,\mathrm{d}\eta - \lambda_1\int_{c_L}^{c}\int_{\eta^*}^{1}\mu(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\,\tau(c)\,\mathrm{d}c\\ &+\int_{c_L}^{c_H}\int_{0}^{\eta^*}\left[-\psi(c)\,g(\eta) - \nu(c,\eta) + \lambda_1\,g(\eta)\int_{\eta^*}^{1}\mu(c,z)\,\mathrm{d}z\right]\,\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\end{aligned}$$

$$+\int_{c_L}^{c_H}\int_{\eta^*}^{1}\left[-\psi(c)\,g(\eta)+\mu(c,\eta)+\lambda_1\,g(\eta)\int_{\eta^*}^{1}\mu(c,z)\,\mathrm{d}z\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c$$
$$=v_E\,\Pi(c_H)+\int_{c_L}^{c_H}\int_{0}^{1}\left[v(\eta)-v_E\right]\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c)+v_E\int_{c_L}^{c_H}\left[c+\frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\,\mathrm{d}c,$$

the objective function is bounded by

$$\int_{c_L}^{c_H}\sigma(c)\,\mathrm{d}c.$$

Substituting our candidate solution into the objective function achieves the upper bound, which follows from

$$\begin{aligned}
&\int_{c_L}^{c_H}\sigma(c)\,\mathrm{d}c-\left[v_E+(\lambda_0+\lambda_1)\int_{\eta^*}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]\\
&=\int_{c_L}^{c_H}\left[v(\eta^*)\left[(c-c_H)\,f(c)+F(c)\right]+f(c)\left[v_E+(\lambda_0+\lambda_1)\int_{\eta^*}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]\right]\mathrm{d}c\\
&\quad-\left[v_E+(\lambda_0+\lambda_1)\int_{\eta^*}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]\\
&=v(\eta^*)\int_{c_L}^{c_H}\left[(c-c_H)\,f(c)+F(c)\right]\mathrm{d}c\\
&=0.
\end{aligned}$$

This completes the proof. □

Proof of Proposition 4. When the testing cost is high and the regulator induces testing, the firm's individual rationality constraint (IR) must be binding. Otherwise, the regulator could strictly increase its expected payoff by reducing the costly incentives offered to the firm. That is, the regulator's first-best optimization problem can be written as follows:

$$\begin{aligned}
&\max_{\eta}\ v_E\,\tilde{a}_1(\eta,c)+(\lambda_0+\lambda_1\,\tilde{a}_1(\eta,c))\int_{\eta}^{1}v(y)\,\mathrm{d}G(y),\\
&\text{subject to}\ \ 0\le\tilde{a}_1(\eta,c)\le 1,
\end{aligned}$$

where $\tilde{a}_1$ is defined by (31). We refer to the above optimization problem without the feasibility constraint on a_1 as the "relaxed problem".

Since the firm's revenue across both periods is bounded above by $\bar{c}(0)=1+\lambda_0+\lambda_1$, the regulator cannot induce testing when the testing cost exceeds this value. In the remainder of the proof, we focus on the cases where $c\le\bar{c}(0)$.

As established in Section 4.1, the derivative of the objective function with respect to η has the same sign as the function h defined in (33). If the drug is promising ($h(\bar{\eta})\ge 0$) and $c>1+(\lambda_0+\lambda_1)\,[1-G(\bar{\eta})]$, since $\tilde{a}_1(\bar{\eta},c)>1$, the optimal solution to the relaxed problem is infeasible. Because η is increasing in a_1, it is optimal to set $a_1=1$, and the corresponding value of η is determined such that the individual rationality constraint (IR) binds.

Next, we consider the case in which the drug is not promising ($h(\bar{\eta}) \leq 0$) and $c > \lambda_0 [1 - G(\bar{\eta})]$. Recall that $\check{\eta}$ is the unique root of the function h on $(0, \bar{\eta}]$. If $\tilde{a}_1(\check{\eta}, c) \in [0, 1]$, then the optimal solution to the relaxed problem is feasible, and hence also optimal for the original problem. Instead, if $\tilde{a}_1(\check{\eta}, c) < 0$, then any feasible value of η must lie above $\check{\eta}$. Since η is increasing in a_1 and the objective function is decreasing in η for all $\eta \in [\check{\eta}, \bar{\eta}]$, it is optimal to set $a_1 = 0$, which uniquely determines the final approval threshold $\eta^* = G^{-1}\left(1 - \frac{c}{\lambda_0}\right)$. Finally, if $\tilde{a}_1(\check{\eta}, c) > 1$, then any feasible value of η must lie below $\check{\eta}$. Since η is increasing in a_1 and the objective function is increasing in η for all $\eta \in [0, \check{\eta}]$, it is optimal to set $a_1 = 1$, and the corresponding final approval threshold is given by $\eta^* = G^{-1}\left(1 - \frac{c-1}{\lambda_0 + \lambda_1}\right)$.

When the regulator's expected payoff from inducing testing ($\tau = 1$) is negative, it is optimal to forgo testing by setting $\tau = 0$.

This completes the proof. □

EC.4.4. Proofs for Section 5

Proof of Proposition 5. To prove the desired result, it suffices to show that for any feasible solution to (13), we can construct a feasible solution to (48) that yields the same objective function value.

Let $\mathcal{M} \equiv \{\tau, a_1, a_2\}$ denote a feasible solution to the optimization problem (13). By construction, it is straightforward to verify that $\bar{\mathcal{M}} \equiv \{\tau, \alpha_1, \alpha_2, \alpha_3\}$, where

$$\begin{aligned}
&\alpha_1(c) = \tau(c)\, a_1(c), && \forall c \in [c_L, c_H],\\
&\alpha_2(c, \eta) = \tau(c)\, a_2(c, \eta), && \forall (c, \eta) \in [c_L, c_H] \times [0, 1],\\
&\alpha_3(c, \eta) = \tau(c)\, a_1(c)\, a_2(c, \eta), && \forall (c, \eta) \in [c_L, c_H] \times [0, 1],
\end{aligned}$$

is feasible to the upper-bound problem (48) and yields the same objective function value.

This completes the proof. □

Similar to Lemma EC.1, after eliminating α_1 using the envelope theorem, the regulator's upper-bound problem can be formulated as follows:

$$\begin{aligned}
\max_{\{\tau, \hat{\alpha}_2, \hat{\alpha}_3, \Pi(c_H)\} \in \Omega'_R} \; & v_E\, \Pi(c_H) + v_E \int_{c_L}^{c_H} \left[c + \frac{F(c)}{f(c)}\right] \tau(c)\, f(c)\, \mathrm{d}c \\
& + \int_{c_L}^{c_H} \int_0^1 \left[\lambda_0 \left[v(\eta) - v_E\right] + \Delta \left[v(\eta) + 1\right]\right] \hat{\alpha}_2(c, \eta)\, \mathrm{d}G(\eta)\, \mathrm{d}F(c) \qquad \text{(EC.10)} \\
& + \lambda_1 \int_{c_L}^{c_H} \int_0^1 \left[v(\eta) - v_E\right] \hat{\alpha}_3(c, \eta)\, \mathrm{d}G(\eta)\, \mathrm{d}F(c),
\end{aligned}$$

where the feasible set Ω'_R is determined by (2), (45), (47), (EC.4), (EC.6), and for any $(c, \eta) \in [c_L, c_H] \times [0, 1]$,

$$0 \leq \Pi(c_H) + \int_c^{c_H} \tau(y)\, \mathrm{d}y + c\, \tau(c) - \lambda_0 \int_0^1 \hat{\alpha}_2(c, \eta)\, \mathrm{d}G(\eta) - \lambda_1 \int_0^1 \hat{\alpha}_3(c, \eta)\, \mathrm{d}G(\eta) \leq \tau(c), \quad \text{(EC.11)}$$

and

$$\hat{\alpha}_3(c,\eta) \le \Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y + c\,\tau(c) - \lambda_0 \int_0^1 \hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta) - \lambda_1 \int_0^1 \hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta). \quad \text{(EC.12)}$$

With a slight abuse of notation, in this subsection, let

$$\bar{c}(\eta_1,\eta_2) \equiv 1 + \lambda_0\,[1-G(\eta_1)] + \lambda_1\,[1-G(\eta_2)], \quad \forall \eta_1,\eta_2 \in [0,1]. \quad \text{(EC.13)}$$

By Lemma 2(1), we know that there exists a unique $\check{\eta}_2 \in [0,\bar{\eta}]$ such that $h(\check{\eta}_2)=0$ if $h(\bar{\eta}) \le 0$. Let

$$\Psi_1(\eta_1,\eta_2) \equiv (\lambda_0+\Delta)\,v(\eta_1) + \Delta - \lambda_0\,v(\eta_2) \quad \text{(EC.14)}$$

and

$$\Psi_2(\eta_1,\eta_2) \equiv (\lambda_0+\lambda_1+\Delta)\,[v(\eta_2)-v(\eta_1)] + \frac{\lambda_1}{\lambda_0}\,\Delta\,[v(\eta_2)+1]. \quad \text{(EC.15)}$$

In order to prove Proposition 6, we need the following technical results.

Lemma EC.3. *When $h(\bar{\eta}) \le 0$, we have the following results.*

(1) If $v(\check{\eta}_2)+1 \ge 0$, there exists a unique $\check{\eta}_1 \in [0,\check{\eta}_2]$ such that $\Psi_1(\check{\eta}_1,\check{\eta}_2)=0$.

(2) If $v(\check{\eta}_2)+1<0$, there uniquely exist $(\mathring{\eta}_1,\mathring{\eta}_2)$ such that $0 \le \mathring{\eta}_1 \le \mathring{\eta}_2 \le \tilde{\eta}$, $\Psi_2(\mathring{\eta}_1,\mathring{\eta}_2)=0$, and $\iota(\mathring{\eta}_1,\mathring{\eta}_2)=0$, where

$$\eta^\dagger \equiv \inf\{\eta \in [0,1] : v(\eta)+1 \ge 0\} \quad \text{(EC.16)}$$

and

$$\begin{aligned}\iota(\eta_1,\eta_2) \equiv\ & (\lambda_0+\lambda_1+\Delta)\int_{\eta_1}^{\eta_2} [v(\eta_2)-v(z)]\,g(z)\,\mathrm{d}z - \lambda_1 \int_{\eta_2}^1 [v(z)-v(\eta_2)]\,g(z)\,\mathrm{d}z \\ & + \frac{\Delta}{\lambda_0}\,[v(\eta_2)+1]\,[1+\lambda_1\,[1-G(\eta_1)]] + v(\eta_2) - v_E. \end{aligned} \quad \text{(EC.17)}$$

Proof of Lemma EC.3. When $\check{\eta}_2$ exists and $v(\check{\eta}_2)+1 \ge 0$, we have

$$\Psi_1(0,\check{\eta}_2) = \lambda_0\,[v(0)-v(\check{\eta}_2)] + \Delta\,[v(0)+1] \le 0,$$

and

$$\Psi_1(\check{\eta}_2,\check{\eta}_2) = \Delta\,[v(\check{\eta}_2)+1] \ge 0.$$

The unique existence of $\check{\eta}_1$ follows directly from the monotonicity of Ψ_1 in η_1.

If $v(\check{\eta}_2)+1<0$, define

$$\begin{aligned}\Xi(\eta_2) \equiv\ & -\lambda_0\,\iota(\eta_1(\eta_2),\eta_2) \\ =\ & \lambda_0\left[-(\lambda_0+\lambda_1+\Delta)\int_{\eta_1(\eta_2)}^{\eta_2} [v(\eta_2)-v(z)]\,g(z)\,\mathrm{d}z + \lambda_1\int_{\eta_2}^1 [v(z)-v(\eta_2)]\,g(z)\,\mathrm{d}z\right] \\ & - \Delta\,[v(\eta_2)+1]\,[1+\lambda_1\,[1-G(\eta_1(\eta_2))]] - \lambda_0\,[v(\eta_2)-v_E], \end{aligned} \quad \text{(EC.18)}$$

where $\eta_1(\eta_2)$ is determined by

$$\Psi_2(\eta_1(\eta_2),\eta_2)=0 \quad \Leftrightarrow \quad v(\eta_1(\eta_2))=v(\eta_2)+\frac{\lambda_1\,\Delta\,[v(\eta_2)+1]}{\lambda_0\,(\lambda_0+\lambda_1+\Delta)}.$$

Note that $\eta_1(\eta_2)$ is increasing.

The monotonicity of v implies that $\check{\eta}_2<\eta^\dagger$. We have $\eta_1(\eta^\dagger)=\eta^\dagger$, implying that

$$\begin{aligned}\Xi(\eta^\dagger) &\equiv \lambda_0\left[\lambda_1\int_{\eta^\dagger}^1\left[v(z)-v(\eta^\dagger)\right]\,g(z)\,\mathrm{d}z-v(\eta^\dagger)+v_E\right]\\ &\le \lambda_0\left[\lambda_1\int_{\check{\eta}_2}^1\left[v(z)-v(\check{\eta}_2)\right]\,g(z)\,\mathrm{d}z-v(\check{\eta}_2)+v_E\right]\\ &= 0,\end{aligned}$$

where the inequality follows from the monotonicity of the function h. To prove the desired result, it suffices to show that $\Xi(\check{\eta}_2)\ge 0$ and Ξ is non-increasing.

Taking derivative yields

$$\begin{aligned}\Xi'(\eta_2)= &-\lambda_0\,(\lambda_0+\lambda_1+\Delta)\int_{\eta_1(\eta_2)}^{\eta_2}v'(\eta_2)\,g(z)\,\mathrm{d}z+\lambda_0\,(\lambda_0+\lambda_1+\Delta)\,[v(\eta_2)-v(\eta_1(\eta_2))]\,g(\eta_1(\eta_2))\,\eta_1'(\eta_2)\\ &-\lambda_0\,\lambda_1\int_{\eta_2}^1 v'(\eta_2)\,g(z)\,\mathrm{d}z-\Delta\,v'(\eta_2)\,[1+\lambda_1\,(1-G(\eta_1(\eta_2)))]\\ &+\Delta\,[v(\eta_2)+1]\,\lambda_1\,g(\eta_1(\eta_2))\,\eta_1'(\eta_2)-\lambda_0\,v'(\eta_2)\\ =&\left\{-\lambda_0\,(\lambda_0+\lambda_1+\Delta)\,[G(\eta_2)-G(\eta_1(\eta_2))]-\lambda_0\,\lambda_1\,[1-G(\eta_2)]\right.\\ &\left.-\lambda_0-\Delta\,[1+\lambda_1\,(1-G(\eta_1(\eta_2)))]\right\}v'(\eta_2)\\ \le\ &0,\end{aligned}$$

where the second equality follows from the definition of $\eta_1(\eta_2)$, i.e., $\Psi_2(\eta_1(\eta_2),\eta_2)=0$.

Since $h(\check{\eta}_2)=0$, the desired condition $\Xi(\check{\eta}_2)\ge 0$ reduces to $\tilde{\Xi}(\check{\eta}_2)\ge 0$, where

$$\tilde{\Xi}(\eta)\equiv-\lambda_0\,(\lambda_0+\lambda_1+\Delta)\int_{\eta_1(\eta)}^{\eta}[v(\eta)-v(z)]\,g(z)\,\mathrm{d}z-\Delta\,[v(\eta)+1]\,[1+\lambda_1\,[1-G(\eta_1(\eta))]].$$

Taking derivative yields

$$\begin{aligned}\tilde{\Xi}'(\eta)= &-\lambda_0\,(\lambda_0+\lambda_1+\Delta)\left[\int_{\eta_1(\eta)}^{\eta}v'(\eta)\,g(z)\,\mathrm{d}z-[v(\eta)-v(\eta_1(\eta))]\,g(\eta_1(\eta))\,\eta_1'(\eta)\right]\\ &-\Delta\,v'(\eta)\,[1+\lambda_1\,[1-G(\eta_1(\eta))]]+\Delta\,[v(\eta)+1]\,\lambda_1\,g(\eta_1(\eta))\,\eta_1'(\eta)\\ =&-\{\lambda_0\,(\lambda_0+\lambda_1+\Delta)\,[G(\eta)-G(\eta_1(\eta))]+\Delta\,[1+\lambda_1\,[1-G(\eta_1(\eta))]]\}\,v'(\eta)\\ \le&\ 0.\end{aligned}$$

Therefore, the desired result is implied by

$$\tilde{\Xi}(\check{\eta}_2)\ge\tilde{\Xi}(\eta^\dagger)=0.$$

This completes the proof. □

Let

$$\check{c}_1 \equiv \sup\left\{c \in [c_L, c_H] : (\lambda_0+\Delta)\int_{\check{\eta}_1}^1 [v(\eta)-v(\check{\eta}_1)]\, g(\eta)\,\mathrm{d}\eta + v(\check{\eta}_2)\left[c+\frac{F(c)}{f(c)}\right] \geq 0\right\} \tag{EC.19}$$

and

$$\check{c}_2 \equiv \sup\left\{c \in [c_L, c_H] : (\lambda_0+\Delta)\int_{\mathring{\eta}_2}^1 [v(\eta)-v(\mathring{\eta}_2)]\, g(\eta)\,\mathrm{d}\eta + \left[c+\frac{F(c)}{f(c)}\right]\left[v(\mathring{\eta}_2)+\frac{\Delta}{\lambda_0}\,[v(\mathring{\eta}_2)+1]\right] \geq 0\right\}. \tag{EC.20}$$

By Lemma EC.3, we know that $\check{c}_1$ and $\check{c}_2$ are well-defined when $v(\check{\eta}_2)+1 \geq 0$ and $v(\check{\eta}_2)+1<0$, respectively, with the existence of $\check{\eta}_2$ requiring $h(\bar{\eta}) \leq 0$.

Lemma EC.4. *When $h(\bar{\eta}) \leq 0$, we have the following results.*

(1) If $\check{c}_2 < \underline{c}(\mathring{\eta}_2)$, there exists a unique $\tilde{\eta}_2 \in [\mathring{\eta}_2, \bar{\eta}']$ such that

$$(\lambda_0+\Delta)\int_{\tilde{\eta}_2}^1 [v(\eta)-v(\tilde{\eta}_2)]\, g(\eta)\,\mathrm{d}\eta + \left[\underline{c}(\tilde{\eta}_2)+\frac{F(\underline{c}(\tilde{\eta}_2))}{f(\underline{c}(\tilde{\eta}_2))}\right]\left[v(\tilde{\eta}_2)+\frac{\Delta}{\lambda_0}\,[v(\tilde{\eta}_2)+1]\right] = 0, \tag{EC.21}$$

where $\bar{\eta}'$ is defined by (55). If $v(\tilde{\eta}_2)+1 \leq 0$, there exists $\tilde{\eta}_1 \leq \tilde{\eta}_2$ such that $\Psi_2(\tilde{\eta}_1, \tilde{\eta}_2) = 0$. In this case, we have $\iota(\tilde{\eta}_1, \tilde{\eta}_2) \geq 0$, where ι is defined by (EC.17).

(2) If (i) $v(\tilde{\eta}_2)+1>0$, there exists a unique $\acute{\eta}_1 \in [\mathring{\eta}_2, \bar{\eta}']$, or if (ii) $\check{c}_1 < \underline{c}(\check{\eta}_1)$, there exists a unique $\acute{\eta}_1 \in [\check{\eta}_1, \bar{\eta}']$, such that

$$v(\acute{\eta}_2)\left[\underline{c}(\acute{\eta}_1)+\frac{F(\underline{c}(\acute{\eta}_1))}{f(\underline{c}(\acute{\eta}_1))}\right] + (\lambda_0+\Delta)\int_{\acute{\eta}_1}^1 [v(\eta)-v(\acute{\eta}_1)]\, g(\eta)\,\mathrm{d}\eta = 0 \tag{EC.22}$$

and $\Psi_1(\acute{\eta}_1, \acute{\eta}_2) = 0$, respectively. Moreover, we have $h(\acute{\eta}_2) \leq 0$.

Proof of Lemma EC.4. To prove the first part, define

$$h_2(\eta) \equiv (\lambda_0+\Delta)\int_{\eta}^1 [v(z)-v(\eta)]\, g(z)\,\mathrm{d}z + \left[\underline{c}(\eta)+\frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\right]\left[v(\eta)+\frac{\Delta}{\lambda_0}\,[v(\eta)+1]\right]. \tag{EC.23}$$

Taking derivative yields

$$\begin{aligned} h_2'(\eta) = & -(\lambda_0+\Delta)\,v'(\eta)\,[1-G(\eta)] + \frac{\mathrm{d}}{\mathrm{d}\eta}\left[\underline{c}(\eta)+\frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\right]\left[v(\eta)+\frac{\Delta}{\lambda_0}\,[v(\eta)+1]\right] \\ & + \left[\underline{c}(\eta)+\frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\right]\left(\frac{\Delta}{\lambda_0}+1\right)v'(\eta) \\ = & \underbrace{\frac{\mathrm{d}}{\mathrm{d}\eta}\left[\underline{c}(\eta)+\frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\right]}_{\leq 0}\left[v(\eta)+\frac{\Delta}{\lambda_0}\,[v(\eta)+1]\right] + \left(\frac{\Delta}{\lambda_0}+1\right)\frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\,v'(\eta), \end{aligned}$$

which is nonnegative if $\eta \leq \bar{\eta}'$. We have

$$h_2(\bar{\eta}') = (\lambda_0+\Delta)\int_{\bar{\eta}'}^1 [v(z)-v(\bar{\eta}')]\, g(z)\,\mathrm{d}z \geq 0$$

and

$$h_2(\mathring{\eta}_2) = (\lambda_0 + \Delta) \int_{\mathring{\eta}_2}^{1} [v(z) - v(\mathring{\eta}_2)]\, g(z)\, \mathrm{d}z + \left[\underline{c}(\mathring{\eta}_2) + \frac{F(\underline{c}(\mathring{\eta}_2))}{f(\underline{c}(\mathring{\eta}_2))}\right] \left[v(\mathring{\eta}_2) + \frac{\Delta}{\lambda_0}\,[v(\mathring{\eta}_2) + 1]\right] < 0,$$

where the inequality follows from $\check{c}_2 < \underline{c}(\mathring{\eta}_2)$, implying the unique existence of $\tilde{\eta}_2 \in [\mathring{\eta}_2, \bar{\eta}']$. The existence of $\tilde{\eta}_1 \in [0, \tilde{\eta}_2]$ is ensured by $v(\tilde{\eta}_2) + 1 \le 0$. It remains to verify that $\iota(\tilde{\eta}_1, \tilde{\eta}_2) \ge 0$, which follows from

$$\iota(\tilde{\eta}_1, \tilde{\eta}_2) = -\frac{\Xi(\tilde{\eta}_2)}{\lambda_0} \ge -\frac{\Xi(\mathring{\eta}_2)}{\lambda_0} = 0.$$

To prove the second part, we first verify that $v_E \le 0$, which follows from

$$h(\bar{\eta}) \le 0 \quad \Rightarrow \quad \lambda_1 \underbrace{\int_{\bar{\eta}}^{1} v(y)\, \mathrm{d}G(y)}_{>0} + v_E \le 0 \quad \Rightarrow \quad v_E < 0.$$

Define

$$h_1(\eta) \equiv \frac{1}{\lambda_0}\,[(\lambda_0 + \Delta)\, v(\eta) + \Delta] \left[\underline{c}(\eta) + \frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\right] + (\lambda_0 + \Delta) \int_{\eta}^{1} [v(z) - v(\eta)]\, g(z)\, \mathrm{d}z. \tag{EC.24}$$

Taking derivative yields

$$\begin{aligned} h_1'(\eta) = {} & \frac{1}{\lambda_0}\,(\lambda_0 + \Delta)\, v'(\eta) \left[\underline{c}(\eta) + \frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\right] + \frac{1}{\lambda_0}\,[(\lambda_0 + \Delta)\, v(\eta) + \Delta] \underbrace{\frac{\mathrm{d}}{\mathrm{d}\eta} \left[\underline{c}(\eta) + \frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\right]}_{\le 0} \\ & - (\lambda_0 + \Delta)\, v'(\eta)\, [1 - G(\eta)] \\ = {} & \frac{1}{\lambda_0} \left\{(\lambda_0 + \Delta)\, v'(\eta)\, \frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))} + [(\lambda_0 + \Delta)\, v(\eta) + \Delta]\, \frac{\mathrm{d}}{\mathrm{d}\eta} \left[\underline{c}(\eta) + \frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))}\right]\right\}. \end{aligned}$$

We know that function h_1 is increasing for any $\eta \in [0, \bar{\eta}']$.

To prove the unique existence of $\acute{\eta}_1$, it suffices to show that, in cases (i) and (ii), we have $h_1(\mathring{\eta}_2) \le 0 \le h_1(\bar{\eta}')$ and $h_1(\check{\eta}_1) \le 0 \le h_1(\bar{\eta}')$, respectively. The nonnegative condition follows from

$$h_1(\bar{\eta}') = (\lambda_0 + \Delta) \int_{\bar{\eta}'}^{1} [v(z) - v(\bar{\eta}')]\, g(z)\, \mathrm{d}z \ge 0.$$

It remains to prove the other desired inequality condition.

- When $v(\tilde{\eta}_2) + 1 > 0$, we have

$$\begin{aligned} h_1(\mathring{\eta}_2) = {} & \left[v(\mathring{\eta}_2) + \frac{\Delta}{\lambda_0}\,[v(\mathring{\eta}_2) + 1]\right] \left[\underline{c}(\mathring{\eta}_2) + \frac{F(\underline{c}(\mathring{\eta}_2))}{f(\underline{c}(\mathring{\eta}_2))}\right] + (\lambda_0 + \Delta) \int_{\mathring{\eta}_2}^{1} [v(z) - v(\mathring{\eta}_2)]\, g(z)\, \mathrm{d}z \\ < {} & 0, \end{aligned}$$

  where the inequality follows from $\check{c}_2 < \underline{c}(\mathring{\eta}_2)$.

- When $v(\check{\eta}_2)+1\geq 0$ and $\check{c}_1<\underline{c}(\check{\eta}_1)$, we have

$$\begin{aligned} h_1(\check{\eta}_1) &= \left[v(\check{\eta}_1)+\frac{\Delta}{\lambda_0}\left[v(\check{\eta}_1)+1\right]\right]\left[\underline{c}(\check{\eta}_1)+\frac{F(\underline{c}(\check{\eta}_1))}{f(\underline{c}(\check{\eta}_1))}\right]+(\lambda_0+\Delta)\int_{\check{\eta}_1}^1\left[v(z)-v(\check{\eta}_1)\right]g(z)\,\mathrm{d}z \\ &< \left[v(\check{\eta}_1)+\frac{\Delta}{\lambda_0}\left[v(\check{\eta}_1)+1\right]\right]\left[\underline{c}(\check{\eta}_1)+\frac{F(\underline{c}(\check{\eta}_1))}{f(\underline{c}(\check{\eta}_1))}\right]-v(\check{\eta}_2)\left[\underline{c}(\check{\eta}_1)+\frac{F(\underline{c}(\check{\eta}_1))}{f(\underline{c}(\check{\eta}_1))}\right] \\ &= 0, \end{aligned}$$

  where the inequality follows from $\check{c}_1<\underline{c}(\check{\eta}_1)$.

Therefore, the given conditions guarantee the unique existence of $\acute{\eta}_1$.

It remains to verify that $h(\acute{\eta}_2)\leq 0$, which follows from

$$\begin{aligned} &(i)\ \acute{\eta}_2\geq\acute{\eta}_1\geq\mathring{\eta}_2\geq\check{\eta}_2 \quad\Rightarrow\quad h(\acute{\eta}_2)\leq 0,\ \text{and} \\ &(ii)\ \acute{\eta}_1\geq\check{\eta}_1 \quad\Rightarrow\quad \acute{\eta}_2\geq\check{\eta}_2 \quad\Rightarrow\quad h(\acute{\eta}_2)\leq 0, \end{aligned}$$

where $\mathring{\eta}_2\geq\check{\eta}_2$ follows from the proof of Lemma EC.3.

This completes the proof. □

Lemma EC.5. *If $\check{c}_2>\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)$, we have $t_1(\eta^\dagger)\geq 0$, where $\eta^\dagger$ is defined by* (EC.16),

$$\begin{aligned} t_1(\eta_2)\equiv\ & v(\eta_2)\left[\bar{c}(\eta_1(\eta_2),\eta_2)+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_2))}{f(\bar{c}(\eta_1(\eta_2),\eta_2))}\right]+(\lambda_0+\Delta)\int_{\eta_1(\eta_2)}^1\left[v(z)-v(\eta_1(\eta_2))\right]g(z)\,\mathrm{d}z \\ &+\lambda_1\int_{\eta_2}^1\left[v(z)-v(\eta_2)\right]g(z)\,\mathrm{d}z-v(\eta_2)+v_E \end{aligned} \tag{EC.25}$$

and function $\eta_1(\eta_2)$ is determined by $\Psi_1(\eta_1(\eta_2),\eta_2)=0$.

Proof of Lemma EC.5. The given condition $\check{c}_2>\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)$ can be rewritten as

$$(\lambda_0+\Delta)\int_{\mathring{\eta}_2}^1\left[v(\eta)-v(\mathring{\eta}_2)\right]g(\eta)\,\mathrm{d}\eta+\left[\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)+\frac{F(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}{f(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}\right]\left[v(\mathring{\eta}_2)+\frac{\Delta}{\lambda_0}\left[v(\mathring{\eta}_2)+1\right]\right]>0.$$

Recall that the two thresholds $\mathring{\eta}_1$ and $\mathring{\eta}_2$ satisfy $\iota(\mathring{\eta}_1,\mathring{\eta}_2)=0$, we have

$$\begin{aligned} &(\lambda_0+\Delta)\int_{\mathring{\eta}_2}^1\left[v(\eta)-v(\mathring{\eta}_2)\right]g(\eta)\,\mathrm{d}\eta+\left[\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)+\frac{F(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}{f(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}\right]\left[v(\mathring{\eta}_2)+\frac{\Delta}{\lambda_0}\left[v(\mathring{\eta}_2)+1\right]\right] \\ =\ &(\lambda_0+\Delta)\int_{\mathring{\eta}_2}^1\left[v(\eta)-v(\mathring{\eta}_2)\right]g(\eta)\,\mathrm{d}\eta+\left[\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)+\frac{F(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}{f(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}\right]v(\mathring{\eta}_2) \\ &+\frac{\Delta}{\lambda_0}\left[\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)+\frac{F(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}{f(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}\right]\left[v(\mathring{\eta}_2)+1\right] \\ \leq\ &(\lambda_0+\Delta)\int_{\mathring{\eta}_2}^1\left[v(\eta)-v(\mathring{\eta}_2)\right]g(\eta)\,\mathrm{d}\eta+\left[\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)+\frac{F(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}{f(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}\right]v(\mathring{\eta}_2) \\ &+\frac{\Delta}{\lambda_0}\left[1+\lambda_1\left[1-G(\mathring{\eta}_1)\right]\right]\left[v(\mathring{\eta}_2)+1\right]+\Delta\left[1-G(\mathring{\eta}_1)\right]\left[v(\mathring{\eta}_2)+1\right] \\ =\ &(\lambda_0+\lambda_1+\Delta)\int_{\mathring{\eta}_1}^1\left[v(\eta)-v(\mathring{\eta}_2)\right]g(\eta)\,\mathrm{d}\eta+\left[\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)+\frac{F(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}{f(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}\right]v(\mathring{\eta}_2)-v(\mathring{\eta}_2)+v_E \end{aligned}$$

$$
\begin{aligned}
&+\Delta\left[1-G(\mathring{\eta}_1)\right]\left[v(\mathring{\eta}_2)+1\right]\\
\le\ &(\lambda_0+\lambda_1+\Delta)\int_{\mathring{\eta}_2}^{1}\left[v(\eta)-v(\mathring{\eta}_2)\right]g(\eta)\,\mathrm{d}\eta+\left[\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)+\frac{F(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}{f(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}\right]v(\mathring{\eta}_2)-v(\mathring{\eta}_2)+v_E\\
&+\Delta\left[1-G(\mathring{\eta}_2)\right]\left[v(\mathring{\eta}_2)+1\right],
\end{aligned}
$$

where the first inequality and the second equality follow from $v(\mathring{\eta}_2)+1\le v(\eta^\dagger)+1=0$ and $\iota(\mathring{\eta}_1,\mathring{\eta}_2)=0$, respectively.

Define

$$
\begin{aligned}
\Upsilon(\eta_2)\equiv\ &(\lambda_0+\lambda_1+\Delta)\int_{\eta_2}^{1}\left[v(\eta)-v(\eta_2)\right]g(\eta)\,\mathrm{d}\eta+\left[\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2))+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]v(\eta_2)\\
&-v(\eta_2)+v_E+\Delta\left[1-G(\eta_2)\right]\left[v(\eta_2)+1\right],\quad\forall\eta_2\in[0,\eta^\dagger],
\end{aligned}
$$

where $\eta_1(\eta_2)$ is determined by $\Psi_2(\eta_1(\eta_2),\eta_2)=0$. Taking derivative yields

$$
\begin{aligned}
\Upsilon'(\eta_2)=\ &-(\lambda_0+\lambda_1+\Delta)\,v'(\eta_2)\left[1-G(\eta_2)\right]+\left[\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2))+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]v'(\eta_2)\\
&+\frac{\mathrm{d}}{\mathrm{d}\eta_2}\left[\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2))+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]v(\eta_2)-v'(\eta_2)\\
&+\Delta\left[-\left[v(\eta_2)+1\right]g(\eta_2)+\left[1-G(\eta_2)\right]v'(\eta_2)\right]\\
\ge\ &\frac{\mathrm{d}}{\mathrm{d}\eta_2}\left[\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2))+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]v(\eta_2)-\Delta\left[v(\eta_2)+1\right]g(\eta_2)+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}v'(\eta_2)\\
\ge\ &0,
\end{aligned}
$$

where the first inequality follows from $\eta_1(\eta_2)\le\eta_2$ for any $\eta_2\in[0,\eta^\dagger]$, as established in Lemma EC.3(2). Therefore, we have

$$
t_1(\eta^\dagger)=\Upsilon(\eta^\dagger)\ge\Upsilon(\mathring{\eta}_2)>0,
$$

which completes the proof. □

Lemma EC.6. *When one of the following conditions holds, (i) $h(\bar{\eta})\le 0$, $\check{c}_2>\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)$, (ii) $h(\bar{\eta})\le 0$, $\check{c}_1>\bar{c}(\check{\eta}_1,\check{\eta}_2)$, or (iii) $h(\bar{\eta})>0$, we have the following results.*

(1) If $t_1(\eta^\dagger)\le 0$, there uniquely exist $\eta^\dagger\le\dot{\eta}_1\le\dot{\eta}_2\le\bar{\eta}$ such that $t_1(\dot{\eta}_2)=0$, where function t_1 is defined by (EC.25)*, $\dot{\eta}_1=\eta_1(\dot{\eta}_2)$, and function $\eta_1(\eta_2)$ is determined by $\Psi_1(\eta_1(\eta_2),\eta_2)=0$. Moreover, we have $h(\dot{\eta}_2)\ge 0$.*

(2) If $t_2(\breve{\eta})\le 0\le t_1(\eta^\dagger)=t_2(\eta^\dagger)$, there uniquely exist $\breve{\eta}\le\grave{\eta}_1\le\grave{\eta}_2\le\eta^\dagger$ such that $t_2(\grave{\eta}_2)=0$, where $\grave{\eta}_1=\eta_1(\grave{\eta}_2)$, $\breve{\eta}$ is determined by $\eta_1(\breve{\eta})=0$,

$$
\begin{aligned}
t_2(\eta_2)\equiv\ &\left[v(\eta_2)+\frac{\Delta}{\lambda_0}\left[v(\eta_2)+1\right]\right]\left[\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2))+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]\\
&+(\lambda_0+\Delta)\int_{\eta_2}^{1}\left[v(\eta)-v(\eta_2)\right]g(\eta)\,\mathrm{d}\eta-\iota(\eta_1(\eta_2),\eta_2),
\end{aligned}
\tag{EC.26}
$$

and function $\eta_1(\eta_2)$ is determined by $\Psi_2(\eta_1(\eta_2),\eta_2)=0$. Moreover, we have $\iota(\grave{\eta}_1,\grave{\eta}_2)\le 0$, where ι is defined by (EC.17)*.*

Proof of Lemma EC.6. To prove the result in case (1), taking derivative yields

$$\begin{aligned}
t_1'(\eta_2) = {} & v'(\eta_2)\left[\bar{c}(\eta_1(\eta_2),\eta_2)+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_2))}{f(\bar{c}(\eta_1(\eta_2),\eta_2))}\right]+v(\eta_2)\frac{\mathrm{d}}{\mathrm{d}\eta_2}\left(\bar{c}(\eta_1(\eta_2),\eta_2)+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_2))}{f(\bar{c}(\eta_1(\eta_2),\eta_2))}\right)\\
& -(\lambda_0+\Delta)\,v'(\eta_1(\eta_2))\,[1-G(\eta_1(\eta_2))]\,\eta_1'(\eta_2)-\lambda_1\,v'(\eta_2)\,[1-G(\eta_2)]-v'(\eta_2)\\
= {} & v'(\eta_2)\frac{F(\bar{c}(\eta_1(\eta_2),\eta_2))}{f(\bar{c}(\eta_1(\eta_2),\eta_2))}+v(\eta_2)\frac{\mathrm{d}}{\mathrm{d}\eta_2}\left(\bar{c}(\eta_1(\eta_2),\eta_2)+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_2))}{f(\bar{c}(\eta_1(\eta_2),\eta_2))}\right)\\
& +[\lambda_0\,v'(\eta_2)-(\lambda_0+\Delta)\,v'(\eta_1(\eta_2))\,\eta_1'(\eta_2)]\,[1-G(\eta_1(\eta_2))]\\
= {} & v'(\eta_2)\frac{F(\bar{c}(\eta_1(\eta_2),\eta_2))}{f(\bar{c}(\eta_1(\eta_2),\eta_2))}+v(\eta_2)\frac{\mathrm{d}}{\mathrm{d}\eta_2}\left(\bar{c}(\eta_1(\eta_2),\eta_2)+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_2))}{f(\bar{c}(\eta_1(\eta_2),\eta_2))}\right),
\end{aligned}$$

which is positive for any $\eta_2\in[0,\bar{\eta}]$. The unique existence of $\dot{\eta}_2$ is ensured by the monotonicity of t_1, $t_1(\eta^\dagger)\le 0$, and the following:

- In Lemma EC.5, we have proved that given $t_1(\eta^\dagger)\le 0$, it is impossible to have $\check{c}_2>\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)$.
- If $h(\bar{\eta})\le 0$ and $\check{c}_1>\bar{c}(\check{\eta}_1,\check{\eta}_2)$, we have

$$\begin{aligned}
t_1(\check{\eta}_2) = {} & v(\check{\eta}_2)\left[\bar{c}(\check{\eta}_1,\check{\eta}_2)+\frac{F(\bar{c}(\check{\eta}_1,\check{\eta}_2))}{f(\bar{c}(\check{\eta}_1,\check{\eta}_2))}\right]+(\lambda_0+\Delta)\int_{\check{\eta}_1}^1[v(z)-v(\check{\eta}_1)]\,g(z)\,\mathrm{d}z\\
& +\lambda_1\int_{\check{\eta}_2}^1[v(z)-v(\check{\eta}_2)]\,g(z)\,\mathrm{d}z-v(\check{\eta}_2)+v_E\\
= {} & v(\check{\eta}_2)\left[\bar{c}(\check{\eta}_1,\check{\eta}_2)+\frac{F(\bar{c}(\check{\eta}_1,\check{\eta}_2))}{f(\bar{c}(\check{\eta}_1,\check{\eta}_2))}\right]+(\lambda_0+\Delta)\int_{\check{\eta}_1}^1[v(z)-v(\check{\eta}_1)]\,g(z)\,\mathrm{d}z\\
\ge {} & 0,
\end{aligned}$$

  where the inequality follows from $\check{c}_1>\check{c}(\check{\eta}_1,\check{\eta}_2)$.

- If $h(\bar{\eta})>0$, or equivalently,

$$\lambda_1\int_{\bar{\eta}}^1 v(z)\,g(z)\,\mathrm{d}z+v_E>0,$$

  we have

$$t_1(\bar{\eta})=(\lambda_0+\Delta)\int_{\eta_1(\bar{\eta})}^1[v(z)-v(\eta_1(\bar{\eta}))]\,g(z)\,\mathrm{d}z+\underbrace{\lambda_1\int_{\bar{\eta}}^1 v(z)\,g(z)\,\mathrm{d}z+v_E}_{>0}>0.$$

The remaining condition $h(\dot{\eta}_2)\ge 0$ follows from either $\dot{\eta}_2\le\check{\eta}_2$ or $h(\bar{\eta})>0$.

To prove the second part, differentiating t_2 yields

$$\begin{aligned}
t_2'(\eta_2) = {} & \left(\frac{\Delta}{\lambda_0}+1\right)v'(\eta_2)\left[\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2))+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]\\
& +\left[v(\eta_2)+\frac{\Delta}{\lambda_0}\,[v(\eta_2)+1]\right]\frac{\mathrm{d}}{\mathrm{d}\eta_2}\left[\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2))+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]\\
& -(\lambda_0+\Delta)\,v'(\eta_2)\,[1-G(\eta_2)]+\frac{\Xi'(\eta_2)}{\lambda_0}\\
= {} & \left(\frac{\Delta}{\lambda_0}+1\right)v'(\eta_2)\left[1+(\lambda_0+\lambda_1)\,[1-G(\eta_1(\eta_2))]-\lambda_0\,[1-G(\eta_2)]+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]
\end{aligned}$$

$$
\begin{aligned}
&+\left[v(\eta_2)+\frac{\Delta}{\lambda_0}\left[v(\eta_2)+1\right]\right]\frac{\mathrm{d}}{\mathrm{d}\eta_2}\left[\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2))+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]\\
&-\left\{(\lambda_0+\lambda_1+\Delta)\left[G(\eta_2)-G(\eta_1(\eta_2))\right]+\lambda_1\left[1-G(\eta_2)\right]+\frac{\Delta}{\lambda_0}\left[1+\lambda_1\left[1-G(\eta_1(\eta_2))\right]\right]+1\right\}v'(\eta_2)\\
=&\left[v(\eta_2)+\frac{\Delta}{\lambda_0}\left[v(\eta_2)+1\right]\right]\frac{\mathrm{d}}{\mathrm{d}\eta_2}\left[\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2))+\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}\right]\\
&+\left(\frac{\Delta}{\lambda_0}+1\right)v'(\eta_2)\frac{F(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}{f(\bar{c}(\eta_1(\eta_2),\eta_1(\eta_2)))}.
\end{aligned}
$$

For any $\eta_2\in[0,\bar{\eta}']$, we have $(\lambda_0+\Delta)\,v(\eta_2)+\Delta\le 0$, implying that $t_2'(\eta_2)\ge 0$. The unique existence of $\grave{\eta}_2\in[\breve{\eta},\eta^\dagger]$ is ensured by $t_2(\breve{\eta})\le 0\le t_2(\eta^\dagger)$ and the monotonicity of t_2.

By definition of Ξ in Lemma EC.3, to prove the non-positivity of $\iota(\grave{\eta}_1,\grave{\eta}_2)$, it is equivalent to showing that $\Xi(\grave{\eta}_2)\ge 0$. We prove the desired result in the following cases:

- If $h(\bar{\eta})\le 0$ and $\breve{c}_1>\bar{c}(\breve{\eta}_1,\breve{\eta}_2)$, the desired result follows from

$$
\iota(\grave{\eta}_1,\grave{\eta}_2)=-\frac{\Xi(\grave{\eta}_2)}{\lambda_0}\le-\frac{\Xi(\eta^\dagger)}{\lambda_0}\le-\frac{\Xi(\breve{\eta}_2)}{\lambda_0}\le 0,
$$

  where the three inequalities follow from $\grave{\eta}_2\le\eta^\dagger$, $\eta^\dagger\le\breve{\eta}_2$, and $\Xi(\breve{\eta}_2)\ge 0$, respectively;

- If $h(\bar{\eta})\le 0$ and $\breve{c}_2>\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)$, since $\Xi(\mathring{\eta}_2)=0$, it suffices to verfiy that $\grave{\eta}_2\le\mathring{\eta}_2$, or equivalently, $t_2(\mathring{\eta}_2)\ge 0$. The desired result follows from

$$
\begin{aligned}
t_2(\mathring{\eta}_2)=&\left[v(\mathring{\eta}_2)+\frac{\Delta}{\lambda_0}\left[v(\mathring{\eta}_2)+1\right]\right]\left[\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)+\frac{F(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}{f(\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1))}\right]\\
&+(\lambda_0+\Delta)\int_{\mathring{\eta}_2}^1\left[v(\eta)-v(\mathring{\eta}_2)\right]g(\eta)\,\mathrm{d}\eta-\underbrace{\iota(\mathring{\eta}_1,\mathring{\eta}_2)}_{=0}\\
\ge&\ 0,
\end{aligned}
$$

  where the inequality is implied by $\breve{c}_2>\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)$.

- If $h(\bar{\eta})>0$, the desired result follows from

$$
\iota(\grave{\eta}_1,\grave{\eta}_2)=-\frac{\Xi(\grave{\eta}_2)}{\lambda_0}\le-\frac{\Xi(\eta^\dagger)}{\lambda_0}\le 0,
$$

  where the second inequality is implied by $\Xi(\eta^\dagger)=\lambda_0\,h(\eta^\dagger)\ge\lambda_0\,h(\bar{\eta})\ge 0$.

This completes the proof. □

The following proposition summarizes the values of $\hat{c}$, $\hat{\alpha}$, $\hat{\eta}_1$, and $\hat{\eta}_2$ in all possible cases. We assume $t_2(\breve{\eta})\le 0$ to exclude scenarios where the optimal $\hat{\eta}_1=0$, meaning final approval is granted regardless of the revealed efficacy level. Such cases imply that testing is not utilized, which is unrealistic in practice.

PROPOSITION EC.1. *The values of $\hat{c}$, $\hat{\alpha}$, $\hat{\eta}_1$, and $\hat{\eta}_2$, depend on model parameters in the following cases.*

(1) If $\underline{c}(\check{\eta}_1) \le \check{c}_1 \le \bar{c}(\check{\eta}_1, \check{\eta}_2)$*, the optimal policy follows case 1, as defined in* (49)*. We have* $\hat{c} = \check{c}_1$*,* $\hat{\alpha} = \frac{\hat{c} - \lambda_0 [1 - G(\hat{\eta}_1)]}{1 + \lambda_1 [1 - G(\hat{\eta}_2)]}$*,* $\hat{\eta}_1 = \check{\eta}_1$*, and* $\hat{\eta}_2 = \check{\eta}_2$*.*

(2) If $\underline{c}(\mathring{\eta}_2) \le \check{c}_2 \le \bar{c}(\mathring{\eta}_1, \mathring{\eta}_1)$*, the optimal policy follows case 2, as defined in* (50)*. We have* $\hat{c} = \check{c}_2$*,* $\hat{\alpha} = \frac{\hat{c} - \lambda_0 [1 - G(\hat{\eta}_2)]}{1 + \lambda_0 [G(\hat{\eta}_2) - G(\hat{\eta}_1)] + \lambda_1 [1 - G(\hat{\eta}_1)]}$*,* $\hat{\eta}_1 = \mathring{\eta}_1$*, and* $\hat{\eta}_2 = \mathring{\eta}_2$*.*

(3) If $\check{c}_1 < \underline{c}(\check{\eta}_1)$ *or* $\check{c}_2 < \underline{c}(\mathring{\eta}_2)$*, the regulator never grants conditional approval, i.e.,* $\hat{\alpha} = 0$*, implying that* $\hat{\alpha}_3(c, \eta) = 0$ *for any* $(c, \eta) \in [c_L, c_H] \times [0, 1]$*. The thresholds* $\hat{c}$*,* $\hat{\eta}_1$*, and* $\hat{\eta}_2$ *are determined as follows.*

(a) When $\check{c}_2 < \underline{c}(\mathring{\eta}_2)$ *and* $v(\tilde{\eta}_2) + 1 \le 0$*, the optimal policy follows case 2, as defined in* (50)*,*

(i) if $\underline{c}(\tilde{\eta}_2) < c_L$*, we have* $\hat{c} = c_L$ *and* $\hat{\eta}_2 = 1$*;*

(ii) if $c_L \le \underline{c}(\tilde{\eta}_2) < c_H$*, we have* $\hat{c} = \underline{c}(\tilde{\eta}_2)$ *and* $\hat{\eta}_2 = \tilde{\eta}_2$*;*

(iii) if $\underline{c}(\bar{\eta}') < c_H \le \underline{c}(\tilde{\eta}_2)$*, we have* $\hat{c} = c_H$ *and* $\hat{\eta}_2 = \sup\{\eta \in [0, 1] : \underline{c}(\eta) \ge c_H\}$*; and*

(iv) if $c_H \le \underline{c}(\bar{\eta}')$*, we have* $\hat{c} = c_H$ *and* $\hat{\eta}_2 = \bar{\eta}'$*.*

The other threshold $\hat{\eta}_1$ *can take any values between* 0 *and* $\hat{\eta}_2$*.*

(b) When (i) $\check{c}_2 < \underline{c}(\mathring{\eta}_2)$ *and* $v(\tilde{\eta}_2) + 1 > 0$*, or (ii)* $\check{c}_1 < \underline{c}(\check{\eta}_1)$*, the optimal policy follows case 1, as defined in* (49)*,*

(i) if $\underline{c}(\acute{\eta}_1) < c_L$*, we have* $\hat{c} = c_L$ *and* $\hat{\eta}_1 = 1$*;*

(ii) if $c_L \le \underline{c}(\acute{\eta}_1) < c_H$*, we have* $\hat{c} = \underline{c}(\acute{\eta}_1)$ *and* $\hat{\eta}_1 = \acute{\eta}_1$*;*

(iii) if $\underline{c}(\bar{\eta}') < c_H \le \underline{c}(\acute{\eta}_1)$*, we have* $\hat{c} = c_H$ *and* $\hat{\eta}_1 = \sup\{\eta \in [0, 1] : \underline{c}(\eta) \ge c_H\}$*; and*

(iv) if $c_H \le \underline{c}(\bar{\eta}')$*, we have* $\hat{c} = c_H$ *and* $\hat{\eta}_1 = \bar{\eta}'$*.*

The other threshold $\hat{\eta}_2$ *can take any values between* $\hat{\eta}_1$ *and* 1*.*

(4) If $h(\bar{\eta}) > 0$*, or* $\check{c}_1 > \bar{c}(\check{\eta}_1, \check{\eta}_2)$*, or* $\check{c}_2 > \bar{c}(\mathring{\eta}_1, \mathring{\eta}_1)$*, the regulator grants full conditional approval to firms whose types are below the threshold, i.e.,* $\hat{\alpha} = 1$*. The thresholds* $\hat{c}$*,* $\hat{\eta}_1$*, and* $\hat{\eta}_2$ *are determined as follows.*

(a) When $t_1(\eta^\dagger) \le 0$*, the optimal policy follows case 1, as defined in* (49)*,*

(i) if $\bar{c}(\dot{\eta}_1, \dot{\eta}_2) < c_L$*, we have* $\hat{c} = c_L$ *and* $\hat{\eta}_1 = \hat{\eta}_2 = 1$*;*

(ii) if $c_L \le \bar{c}(\dot{\eta}_1, \dot{\eta}_2) < c_H$*, we have* $\hat{c} = \bar{c}(\dot{\eta}_1, \dot{\eta}_2)$*,* $\hat{\eta}_1 = \dot{\eta}_1$*, and* $\hat{\eta}_2 = \dot{\eta}_2$*;*

(iii) if $\bar{c}(\bar{\eta}', \bar{\eta}) < c_H \le \bar{c}(\dot{\eta}_1, \dot{\eta}_2)$*, we have* $\hat{c} = c_H$ *and* $(\hat{\eta}_1, \hat{\eta}_2)$ *such that* $\bar{c}(\hat{\eta}_1, \hat{\eta}_2) = c_H$ *and* $\Psi_1(\hat{\eta}_1, \hat{\eta}_2) = 0$*; and*

(iv) if $c_H \le \bar{c}(\bar{\eta}', \bar{\eta})$*, we have* $\hat{c} = c_H$*,* $\hat{\eta}_1 = \bar{\eta}'$*, and* $\hat{\eta}_2 = \bar{\eta}$*.*

(b) When $0 \le t_1(\eta^\dagger) = t_2(\eta^\dagger)$*, for cases (i)–(iii), the optimal policy follows case 2, as defined in* (50)*; and for cases (iv) and (v), the optimal policy follows case 1, as defined in* (49)*,*

(i) if $\bar{c}(\grave{\eta}_1, \grave{\eta}_1) < c_L$*, we have* $\hat{c} = c_L$ *and* $\hat{\eta}_1 = 1$*;*

(ii) if $c_L \le \bar{c}(\grave{\eta}_1, \grave{\eta}_1) < c_H$*, we have* $\hat{c} = \bar{c}(\grave{\eta}_1, \grave{\eta}_1)$ *and* $\hat{\eta}_1 = \grave{\eta}_1$*;*

(iii) if $\bar{c}(\eta^\dagger, \eta^\dagger) \le c_H \le \bar{c}(\grave{\eta}_1, \grave{\eta}_1)$*, we have* $\hat{c} = c_H$ *and* $\hat{\eta}_1 = \sup\{\eta \in [0, 1] : \bar{c}(\eta, \eta) \ge c_H\}$*;*

(iv) if $\bar{c}(\bar{\eta}', \bar{\eta}) < c_H \le \bar{c}(\eta^\dagger, \eta^\dagger)$*, we have* $\hat{c} = c_H$ *and* $(\hat{\eta}_1, \hat{\eta}_2)$ *such that* $\bar{c}(\hat{\eta}_1, \hat{\eta}_2) = c_H$ *and* $\Psi_1(\hat{\eta}_1, \hat{\eta}_2) = 0$*; and*

(v) if $c_H \le \bar{c}(\bar{\eta}', \bar{\eta})$*, we have* $\hat{c} = c_H$*,* $\hat{\eta}_1 = \bar{\eta}'$*, and* $\hat{\eta}_2 = \bar{\eta}$*.*

For cases (i)–(iii), the other threshold $\hat{\eta}_2$ *can take any values between* $\hat{\eta}_1$ *and* 1*.*

Proof of Proposition EC.1. The mechanism is a "take-it-or-leave-it" offer, and a firm of threshold type $\hat{c} < c_H$'s expected payoff is always 0. When $\hat{c} = c_H$, the mechanism is pooling and clearly feasible. Moreover, by Lemmas EC.3–EC.6, we have $\hat{\eta}_1 \le \hat{\eta}_2$, ensuring the feasibility of our candidate solutions in all cases.

We start with proving the optimality in case (1). Define

$$
\begin{aligned}
&\kappa \equiv -v(\check{\eta}_2),\\
&\chi(c,\eta) \equiv (\lambda_0 + \Delta)\,[v(\check{\eta}_1) - v(\eta)]\,g(\eta)\,f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [0, \check{\eta}_1],\\
&\omega(c,\eta) \equiv \begin{cases} \{(\lambda_0 + \lambda_1 + \Delta)\,[v(\check{\eta}_1) - v(\eta)] + \lambda_1\,[v(\check{\eta}_2) - v(\check{\eta}_1)]\}\,g(\eta)\,f(c), & \forall (c,\eta) \in [c_L, c_H] \times [0, \check{\eta}_1],\\ \lambda_1\,[v(\check{\eta}_2) - v(\eta)]\,g(\eta)\,f(c), & \forall (c,\eta) \in [c_L, c_H] \times [\check{\eta}_1, \check{\eta}_2], \end{cases}\\
&\mu(c,\eta) \equiv (\lambda_0 + \Delta)\,[v(\eta) - v(\check{\eta}_1)]\,g(\eta)\,f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [\check{\eta}_1, 1],\\
&\xi(c,\eta) \equiv \lambda_1\,[v(\eta) - v(\check{\eta}_2)]\,g(\eta)\,f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [\check{\eta}_2, 1],\\
&\sigma(c) \equiv \left\{ (\lambda_0 + \Delta) \int_{\check{\eta}_1}^{1} [v(\eta) - v(\check{\eta}_1)]\,g(\eta)\,\mathrm{d}\eta + v(\check{\eta}_2) \left[c + \frac{F(c)}{f(c)} \right] \right\} f(c), \quad \forall c \in [c_L, \hat{c}], \text{ and}\\
&\beta(c) \equiv -\sigma(c), \quad \forall c \in [\hat{c}, c_H].
\end{aligned}
$$

The nonnegativity of κ, χ, the second part of ω, μ, and ξ follows directly from their respective definitions, while the nonnegativity of the first part of ω follows from its monotonicity in η. The nonnegativity of σ and β is ensured by the monotonicity of $\sigma(c)/f(c)$ and $\beta(c)/f(c)$ (Assumption 1), along with the definition of the testing threshold $\hat{c} = \check{c}_1$. Because

$$
-\kappa - \int_{c_L}^{c_H} \int_{\check{\eta}_2}^{1} \xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c = v_E,
$$

$\forall (c,\eta) \in [c_L, c_H] \times [0, \check{\eta}_1]$:

$$
-\chi(c,\eta) + \lambda_0\,g(\eta) \int_{\check{\eta}_2}^{1} \xi(c,z)\,\mathrm{d}z = [\lambda_0\,[v(\eta) - v_E] + \Delta\,[v(\eta) + 1]]\,g(\eta)\,f(c),
$$

$\forall (c,\eta) \in [c_L, c_H] \times [\check{\eta}_1, 1]$:

$$
\mu(c,\eta) + \lambda_0\,g(\eta) \int_{\check{\eta}_2}^{1} \xi(c,z)\,\mathrm{d}z = [\lambda_0\,[v(\eta) - v_E] + \Delta\,[v(\eta) + 1]]\,g(\eta)\,f(c),
$$

$\forall (c,\eta) \in [c_L, c_H] \times [0, \check{\eta}_1]$:

$$
-\omega(c,\eta) + \chi(c,\eta) + \lambda_1\,g(\eta) \int_{\check{\eta}_2}^{1} \xi(c,z)\,\mathrm{d}z = \lambda_1\,[v(\eta) - v_E]\,g(\eta)\,f(c),
$$

$\forall(c,\eta)\in[c_L,c_H]\times[\check{\eta}_1,\check{\eta}_2]$:

$$-\omega(c,\eta)+\lambda_1\,g(\eta)\int_{\check{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,[v(\eta)-v_E]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\check{\eta}_2,1]$:

$$\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\check{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,[v(\eta)-v_E]\,g(\eta)\,f(c),$$

$\forall c\in[c_L,\hat{c}]$:

$$\sigma(c)-\int_{\check{\eta}_1}^{1}\mu(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\check{\eta}_2}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y-c\int_{\check{\eta}_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta=v_E\,[c\,f(c)+F(c)],$$

$\forall c\in[\hat{c},c_H]$:

$$-\beta(c)-\int_{\check{\eta}_1}^{1}\mu(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\check{\eta}_2}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y-c\int_{\check{\eta}_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta=v_E\,[c\,f(c)+F(c)],$$

we have

$$\begin{aligned}
&\kappa\,[-\Pi(c_H)]+\int_{c_L}^{\hat{c}}\sigma(c)\,\tau(c)\,\mathrm{d}c+\int_{\hat{c}}^{c_H}\beta(c)\,[-\tau(c)]\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{\check{\eta}_1}^{1}\mu(c,\eta)\,[\hat{\alpha}_2(c,\eta)-\tau(c)]\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{0}^{\check{\eta}_2}\omega(c,\eta)\,[-\hat{\alpha}_3(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{0}^{\check{\eta}_1}\chi(c,\eta)\,[\hat{\alpha}_3(c,\eta)-\hat{\alpha}_2(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\check{\eta}_2}^{1}\xi(c,\eta)\left[\hat{\alpha}_3(c,\eta)-\Pi(c_H)-\int_{c}^{c_H}\tau(y)\,\mathrm{d}y-c\,\tau(c)+\int_{0}^{1}[\lambda_0\,\hat{\alpha}_2(c,z)+\lambda_1\,\hat{\alpha}_3(c,z)]\,\mathrm{d}G(z)\right]\mathrm{d}\eta\,\mathrm{d}c\\
=\;&\left[-\kappa-\int_{c_L}^{c_H}\int_{\check{\eta}_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\right]\Pi(c_H)\\
&+\int_{c_L}^{c_H}\int_{0}^{\check{\eta}_1}\left[-\chi(c,\eta)+\lambda_0\,g(\eta)\int_{\check{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\check{\eta}_1}^{1}\left[\mu(c,\eta)+\lambda_0\,g(\eta)\int_{\check{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{0}^{\check{\eta}_1}\left[-\omega(c,\eta)+\chi(c,\eta)+\lambda_1\,g(\eta)\int_{\check{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\check{\eta}_1}^{\check{\eta}_2}\left[-\omega(c,\eta)+\lambda_1\,g(\eta)\int_{\check{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\check{\eta}_2}^{1}\left[\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\check{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\left[\sigma(c)-\int_{\check{\eta}_1}^{1}\mu(c,\eta)\,\mathrm{d}\eta-c\int_{\check{\eta}_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\check{\eta}_2}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\left[-\beta(c)-\int_{\check{\eta}_1}^{1}\mu(c,\eta)\,\mathrm{d}\eta-c\int_{\check{\eta}_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\check{\eta}_2}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
=\;&v_E\,\Pi(c_H)+v_E\int_{c_L}^{c_H}\left[c+\frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{0}^{1}[\lambda_0\,[v(\eta)-v_E]+\Delta\,[v(\eta)+1]]\,\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c)\\
&+\lambda_1\int_{c_L}^{c_H}\int_{0}^{1}[v(\eta)-v_E]\,\hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c).
\end{aligned}$$

Therefore, the objective function is bounded by $\int_{c_L}^{\hat{c}} \sigma(c)\,\mathrm{d}c$, which is attained by our candidate primal solution, i.e.,

$$
\begin{aligned}
&\int_{c_L}^{\hat{c}} \sigma(c)\,\mathrm{d}c - \left[v_E\,\hat{\alpha} + \lambda_0 \int_{\check{\eta}_1}^{1} v(\eta)\,\mathrm{d}G(\eta) + \lambda_1\,\hat{\alpha} \int_{\check{\eta}_2}^{1} v(\eta)\,\mathrm{d}G(\eta) + \Delta \int_{\check{\eta}_1}^{1} [v(\eta)+1]\,\mathrm{d}G(\eta)\right] F(\hat{c}) \\
=\ & (\lambda_0+\Delta)\,F(\hat{c}) \int_{\check{\eta}_1}^{1} v(\eta)\,g(\eta)\,\mathrm{d}\eta - (\lambda_0+\Delta)\,F(\hat{c})\,v(\check{\eta}_1)\,[1-G(\check{\eta}_1)] + \hat{c}\,F(\hat{c})\,v(\check{\eta}_2) \\
&- \left[v_E\,\hat{\alpha} + \lambda_0 \int_{\check{\eta}_1}^{1} v(\eta)\,\mathrm{d}G(\eta) + \lambda_1\,\hat{\alpha} \int_{\check{\eta}_2}^{1} v(\eta)\,\mathrm{d}G(\eta) + \Delta \int_{\check{\eta}_1}^{1} [v(\eta)+1]\,\mathrm{d}G(\eta)\right] F(\hat{c}) \\
=\ & -(\lambda_0+\Delta)\,F(\hat{c})\,v(\check{\eta}_1)\,[1-G(\check{\eta}_1)] + \hat{c}\,F(\hat{c})\,v(\check{\eta}_2) \\
&- \left[v_E\,\hat{\alpha} + \underbrace{\lambda_1\,\hat{\alpha} \int_{\check{\eta}_2}^{1} v(\eta)\,\mathrm{d}G(\eta)}_{=\hat{\alpha}\,[v(\check{\eta}_2)-v_E+\lambda_1\,v(\check{\eta}_2)\,[1-G(\check{\eta}_2)]]} + \Delta\,[1-G(\check{\eta}_1)]\right] F(\hat{c}) \\
=\ & -(\lambda_0+\Delta)\,F(\hat{c})\,v(\check{\eta}_1)\,[1-G(\check{\eta}_1)] + \hat{c}\,F(\hat{c})\,v(\check{\eta}_2) - [[\hat{c}-\lambda_0\,[1-G(\check{\eta}_1)]]\,v(\check{\eta}_2) + \Delta\,[1-G(\check{\eta}_1)]]\,F(\hat{c}) \\
=\ & [-(\lambda_0+\Delta)\,v(\check{\eta}_1) + \lambda_0\,v(\check{\eta}_2) - \Delta]\,[1-G(\check{\eta}_1)]\,F(\hat{c}) \\
=\ & 0.
\end{aligned}
$$

To prove the optimality in case (2), define

$$
\begin{aligned}
&\kappa \equiv -\frac{1}{\lambda_0}\,[(\lambda_0+\Delta)\,v(\mathring{\eta}_2)+\Delta], \\
&\chi(c,\eta) \equiv (\lambda_0+\Delta)\,[v(\mathring{\eta}_2)-v(\eta)]\,g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[0,\mathring{\eta}_2], \\
&\mu(c,\eta) \equiv (\lambda_0+\Delta)\,[v(\eta)-v(\mathring{\eta}_2)]\,g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[\mathring{\eta}_2,1], \\
&\omega(c,\eta) \equiv \left\{(\lambda_0+\lambda_1+\Delta)\,[v(\mathring{\eta}_2)-v(\eta)] + \frac{\lambda_1}{\lambda_0}\,\Delta\,[v(\mathring{\eta}_2)+1]\right\} g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[0,\mathring{\eta}_1], \\
&\xi(c,\eta) \equiv \begin{cases} -\left\{(\lambda_0+\lambda_1+\Delta)\,[v(\mathring{\eta}_2)-v(\eta)] + \frac{\lambda_1}{\lambda_0}\,\Delta\,[v(\mathring{\eta}_2)+1]\right\} g(\eta)\,f(c), & \forall (c,\eta)\in[c_L,c_H]\times[\mathring{\eta}_1,\mathring{\eta}_2], \\ \left\{\lambda_1\,[v(\eta)-v(\mathring{\eta}_2)] - \frac{\lambda_1}{\lambda_0}\,\Delta\,[v(\mathring{\eta}_2)+1]\right\} g(\eta)\,f(c), & \forall (c,\eta)\in[c_L,c_H]\times[\mathring{\eta}_2,1], \end{cases} \\
&\sigma(c) \equiv \left\{(\lambda_0+\Delta)\int_{\mathring{\eta}_2}^{1}[v(\eta)-v(\mathring{\eta}_2)]\,g(\eta)\,\mathrm{d}\eta + \left[c+\frac{F(c)}{f(c)}\right]\left[v(\mathring{\eta}_2)+\frac{\Delta}{\lambda_0}\,[v(\mathring{\eta}_2)+1]\right]\right\} f(c), \quad \forall c\in[c_L,\hat{c}], \\
&\beta(c) \equiv -\sigma(c), \quad \forall c\in[\hat{c},c_H].
\end{aligned}
$$

The nonnegativity of κ follows from $\mathring{\eta}_2 \le \eta^\dagger$, as established in Lemma EC.3(2). The nonnegativity of χ and μ follows from their definitions directly. The nonnegativity of ω and ξ follows from the relationship between $\mathring{\eta}_1$ and $\mathring{\eta}_2$ (i.e., $\Psi_2(\mathring{\eta}_1,\mathring{\eta}_2)=0$, implying $\omega(c,\mathring{\eta}_1)=\xi(c,\mathring{\eta}_1)=0$ for any $c\in[c_L,c_H]$) and their monotonicity in η. The nonnegativity of σ and β is ensured by the monotonicity of $\sigma(c)/f(c)$ and $\beta(c)/f(c)$, along with the definition of the testing threshold $\hat{c}=\check{c}_2$. Because

$$
-\kappa - \int_{c_L}^{c_H}\int_{\mathring{\eta}_1}^{1} \xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c = v_E,
$$

$\forall(c,\eta)\in[c_L,c_H]\times[0,\mathring{\eta}_2]$:

$$-\chi(c,\eta)+\lambda_0\,g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\left[\lambda_0\,\left[v(\eta)-v_E\right]+\Delta\,\left[v(\eta)+1\right]\right]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\mathring{\eta}_2,1]$:

$$\mu(c,\eta)+\lambda_0\,g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\left[\lambda_0\,\left[v(\eta)-v_E\right]+\Delta\,\left[v(\eta)+1\right]\right]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[0,\mathring{\eta}_1]$:

$$-\omega(c,\eta)+\chi(c,\eta)+\lambda_1\,g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,\left[v(\eta)-v_E\right]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\mathring{\eta}_1,\mathring{\eta}_2]$:

$$\chi(c,\eta)+\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,\left[v(\eta)-v_E\right]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\mathring{\eta}_2,1]$:

$$\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,\left[v(\eta)-v_E\right]\,g(\eta)\,f(c),$$

$\forall c\in[c_L,\hat{c}]$:

$$\sigma(c)-\int_{\mathring{\eta}_2}^{1}\mu(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\mathring{\eta}_1}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y-c\int_{\mathring{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta=v_E\left[c\,f(c)+F(c)\right],$$

$\forall c\in[\hat{c},c_H]$:

$$-\beta(c)-\int_{\mathring{\eta}_2}^{1}\mu(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\mathring{\eta}_1}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y-c\int_{\mathring{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta=v_E\left[c\,f(c)+F(c)\right],$$

we have

$$\begin{aligned}
&\kappa\,\left[-\Pi(c_H)\right]+\int_{c_L}^{\hat{c}}\sigma(c)\,\tau(c)\,\mathrm{d}c+\int_{\hat{c}}^{c_H}\beta(c)\,\left[-\tau(c)\right]\mathrm{d}c+\int_{c_L}^{c_H}\int_{\mathring{\eta}_2}^{1}\mu(c,\eta)\,\left[\hat{\alpha}_2(c,\eta)-\tau(c)\right]\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{0}^{\mathring{\eta}_1}\omega(c,\eta)\,\left[-\hat{\alpha}_3(c,\eta)\right]\mathrm{d}\eta\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{0}^{\mathring{\eta}_2}\chi(c,\eta)\,\left[\hat{\alpha}_3(c,\eta)-\hat{\alpha}_2(c,\eta)\right]\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\mathring{\eta}_1}^{1}\xi(c,\eta)\,\left[\hat{\alpha}_3(c,\eta)-\Pi(c_H)-\int_{c}^{c_H}\tau(y)\,\mathrm{d}y-c\,\tau(c)+\int_{0}^{1}\left[\lambda_0\,\hat{\alpha}_2(c,z)+\lambda_1\,\hat{\alpha}_3(c,z)\right]\mathrm{d}G(z)\right]\mathrm{d}\eta\,\mathrm{d}c\\
=\;&\left[-\kappa-\int_{c_L}^{c_H}\int_{\mathring{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\right]\,\Pi(c_H)\\
&+\int_{c_L}^{c_H}\int_{0}^{\mathring{\eta}_2}\left[-\chi(c,\eta)+\lambda_0\,g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\,\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\mathring{\eta}_2}^{1}\left[\mu(c,\eta)+\lambda_0\,g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\,\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{0}^{\mathring{\eta}_1}\left[-\omega(c,\eta)+\chi(c,\eta)+\lambda_1\,g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\,\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c
\end{aligned}$$

$$
\begin{aligned}
&+\int_{c_L}^{c_H}\int_{\mathring{\eta}_1}^{\mathring{\eta}_2}\left[\chi(c,\eta)+\xi(c,\eta)+\lambda_1\, g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\mathring{\eta}_2}^{1}\left[\xi(c,\eta)+\lambda_1\, g(\eta)\int_{\mathring{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\left[\sigma(c)-\int_{\mathring{\eta}_2}^{1}\mu(c,\eta)\,\mathrm{d}\eta-c\int_{\mathring{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\mathring{\eta}_1}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\left[-\beta(c)-\int_{\mathring{\eta}_2}^{1}\mu(c,\eta)\,\mathrm{d}\eta-c\int_{\mathring{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\mathring{\eta}_1}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
=\;& v_E\,\Pi(c_H)+v_E\int_{c_L}^{c_H}\left[c+\frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{0}^{1}\left[\lambda_0\,[v(\eta)-v_E]+\Delta\,[v(\eta)+1]\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c)\\
&+\lambda_1\int_{c_L}^{c_H}\int_{0}^{1}[v(\eta)-v_E]\,\hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c).
\end{aligned}
$$

Therefore, the objective function is bounded by $\int_{c_L}^{\hat{c}}\sigma(c)\,\mathrm{d}c$, which is attained by our candidate primal solution, i.e.,

$$
\begin{aligned}
&\int_{c_L}^{\hat{c}}\sigma(c)\,\mathrm{d}c-\left[v_E\,\hat{\alpha}+\hat{\alpha}\int_{\mathring{\eta}_1}^{\mathring{\eta}_2}\left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)\right.\\
&\qquad\left.+\int_{\mathring{\eta}_2}^{1}\left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)+\lambda_1\,\hat{\alpha}\int_{\mathring{\eta}_1}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]F(\hat{c})\\
=\;&(\lambda_0+\Delta)\,F(\hat{c})\int_{\mathring{\eta}_2}^{1}[v(\eta)-v(\mathring{\eta}_2)]\,g(\eta)\,\mathrm{d}\eta+\hat{c}\,F(\hat{c})\left[v(\mathring{\eta}_2)+\frac{\Delta}{\lambda_0}\,[v(\mathring{\eta}_2)+1]\right]\\
&-\left[v_E\,\hat{\alpha}+\hat{\alpha}\int_{\mathring{\eta}_1}^{\mathring{\eta}_2}\left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)\right.\\
&\qquad\left.+\int_{\mathring{\eta}_2}^{1}\left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)+\lambda_1\,\hat{\alpha}\int_{\mathring{\eta}_1}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]F(\hat{c})\\
=\;&\left\{-(\lambda_0+\Delta)\,v(\mathring{\eta}_2)\,[1-G(\mathring{\eta}_2)]+\hat{c}\left[v(\mathring{\eta}_2)+\frac{\Delta}{\lambda_0}\,[v(\mathring{\eta}_2)+1]\right]\right.\\
&\left.-\left[v_E\,\hat{\alpha}+\hat{\alpha}\int_{\mathring{\eta}_1}^{\mathring{\eta}_2}\left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)+\Delta\,[1-G(\mathring{\eta}_2)]+\lambda_1\,\hat{\alpha}\int_{\mathring{\eta}_1}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]\right\}F(\hat{c}),
\end{aligned}
$$

where $\hat{\alpha}=\frac{\hat{c}-\lambda_0\,[1-G(\hat{\eta}_2)]}{1+\lambda_0\,[G(\hat{\eta}_2)-G(\hat{\eta}_1)]+\lambda_1\,[1-G(\hat{\eta}_1)]}$ represents the scale of conditional approval. Because

$$
\begin{aligned}
\lambda_1\int_{\mathring{\eta}_1}^{1}v(\eta)\,\mathrm{d}G(\eta)&=\lambda_1\left[\int_{\mathring{\eta}_1}^{\mathring{\eta}_2}v(\eta)\,\mathrm{d}G(\eta)+\int_{\mathring{\eta}_2}^{1}v(\eta)\,\mathrm{d}G(\eta)\right]\\
&=\lambda_1\int_{\mathring{\eta}_1}^{\mathring{\eta}_2}v(\eta)\,\mathrm{d}G(\eta)+[v(\mathring{\eta}_2)-v_E]+\frac{\Delta}{\lambda_0}\,[v(\mathring{\eta}_2)+1]\,[1+\lambda_1\,[1-G(\mathring{\eta}_1)]]\\
&\quad+(\lambda_0+\lambda_1+\Delta)\int_{\mathring{\eta}_1}^{\mathring{\eta}_2}[v(\mathring{\eta}_2)-v(z)]\,g(z)\,\mathrm{d}z+\lambda_1\,v(\mathring{\eta}_2)\,[1-G(\mathring{\eta}_2)]\,,
\end{aligned}
$$

where the second equality follows from $\iota(\mathring{\eta}_1,\mathring{\eta}_2)=0$, and

$$
\hat{c}=\hat{\alpha}\,[\lambda_0\,[G(\mathring{\eta}_2)-G(\mathring{\eta}_1)]+\lambda_1\,[1-G(\mathring{\eta}_1)]+1]+\lambda_0\,[1-G(\mathring{\eta}_2)]\,,
$$

the duality gap boils down to

$$
-(\lambda_0+\Delta)\,v(\mathring{\eta}_2)\,[1-G(\mathring{\eta}_2)]-\Delta\,[1-G(\mathring{\eta}_2)+\hat{\alpha}\,[G(\mathring{\eta}_2)-G(\mathring{\eta}_1)]]
$$

$$
\begin{aligned}
&+ \left[\hat{\alpha}\left[\lambda_0\left[G(\mathring{\eta}_2) - G(\mathring{\eta}_1)\right] + \lambda_1\left[1 - G(\mathring{\eta}_1)\right]\right] + \lambda_0\left[1 - G(\mathring{\eta}_2)\right]\right] v(\mathring{\eta}_2) \\
&+ \left[\hat{\alpha}\,\lambda_0\left[G(\mathring{\eta}_2) - G(\mathring{\eta}_1)\right] + \lambda_0\left[1 - G(\mathring{\eta}_2)\right]\right] \frac{\Delta}{\lambda_0}\left[v(\mathring{\eta}_2) + 1\right] \\
&- \hat{\alpha}\left[(\lambda_0 + \lambda_1 + \Delta)\, v(\mathring{\eta}_2)\left[G(\mathring{\eta}_2) - G(\mathring{\eta}_1)\right] + \lambda_1\, v(\mathring{\eta}_2)\left[1 - G(\mathring{\eta}_2)\right]\right] \\
&= 0.
\end{aligned}
$$

To prove the optimality of cases (3)(a)(i) and (3)(a)(ii), define

$$
\begin{aligned}
&\kappa \equiv -\frac{1}{\lambda_0}\left[(\lambda_0 + \Delta)\, v(\tilde{\eta}_2) + \Delta\right], \\
&\chi(c,\eta) \equiv (\lambda_0 + \Delta)\left[v(\tilde{\eta}_2) - v(\eta)\right] g(\eta)\, f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [0, \tilde{\eta}_2], \\
&\mu(c,\eta) \equiv (\lambda_0 + \Delta)\left[v(\eta) - v(\tilde{\eta}_2)\right] g(\eta)\, f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [\tilde{\eta}_2, 1], \\
&\omega(c,\eta) \equiv \left\{(\lambda_0 + \lambda_1 + \Delta)\left[v(\tilde{\eta}_2) - v(\eta)\right] + \frac{\lambda_1}{\lambda_0}\Delta\left[v(\tilde{\eta}_2) + 1\right]\right\} g(\eta)\, f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [0, \tilde{\eta}_1], \\
&\xi(c,\eta) \equiv \begin{cases} -\left\{(\lambda_0 + \lambda_1 + \Delta)\left[v(\tilde{\eta}_2) - v(\eta)\right] + \frac{\lambda_1}{\lambda_0}\Delta\left[v(\tilde{\eta}_2) + 1\right]\right\} g(\eta)\, f(c), & \forall (c,\eta) \in [c_L, c_H] \times [\tilde{\eta}_1, \tilde{\eta}_2], \\ \left\{\lambda_1\left[v(\eta) - v(\tilde{\eta}_2)\right] - \frac{\lambda_1}{\lambda_0}\Delta\left[v(\tilde{\eta}_2) + 1\right]\right\} g(\eta)\, f(c), & \forall (c,\eta) \in [c_L, c_H] \times [\tilde{\eta}_2, 1], \end{cases} \\
&\phi(c) \equiv \iota(\tilde{\eta}_1, \tilde{\eta}_2)\, f(c), \quad \forall c \in [c_L, c_H], \\
&\sigma(c) \equiv \left\{(\lambda_0 + \Delta)\int_{\tilde{\eta}_2}^{1}\left[v(\eta) - v(\tilde{\eta}_2)\right] g(\eta)\,\mathrm{d}\eta + \left[c + \frac{F(c)}{f(c)}\right]\left[v(\tilde{\eta}_2) + \frac{\Delta}{\lambda_0}\left[v(\tilde{\eta}_2) + 1\right]\right]\right\} f(c), \quad \forall c \in \mathcal{C}_1, \\
&\beta(c) \equiv -\sigma(c), \quad \forall c \in \mathcal{C}_2,
\end{aligned}
$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are subsets of $[c_L, c_H]$ satisfying: (i) $\mathcal{C}_1 \cup \mathcal{C}_2 = [c_L, c_H]$; and (ii) the measure of $\mathcal{C}_1 \cap \mathcal{C}_2$ is 0. These subsets will be specified later.

The nonnegativity of κ is implied by $\tilde{\eta}_2 \le \bar{\eta}'$, as established in Lemma EC.4(1). The nonnegativity of χ and μ follows from their definitions directly. The nonnegativity of ω and ξ follows from the relationship between $\tilde{\eta}_1$ and $\tilde{\eta}_2$ (i.e., $\Psi_2(\tilde{\eta}_1, \tilde{\eta}_2) = 0$, implying $\omega(c, \tilde{\eta}_1) = \xi(c, \tilde{\eta}_1) = 0$ for any $c \in [c_L, c_H]$) and their monotonicity in η. The nonnegativity of ϕ follows from the fact that $\iota(\tilde{\eta}_1, \tilde{\eta}_2) \ge 0$, as established in Lemma EC.4(1).

In the first two possible cases, we have:

1. if $\underline{c}(\tilde{\eta}_2) < c_L$, we know that β is nonnegative on $[c_L, c_H]$. In this case, we have $\mathcal{C}_1 = \emptyset$ and $\mathcal{C}_2 = [c_L, c_H]$; and
2. if $c_L \le \underline{c}(\tilde{\eta}_2) \le c_H$, we know that σ and β are nonnegative on $[c_L, \underline{c}(\tilde{\eta}_2)]$ and $[\underline{c}(\tilde{\eta}_2), c_H]$, respectively. In this case, we have $\mathcal{C}_1 = [c_L, \underline{c}(\tilde{\eta}_2)]$ and $\mathcal{C}_2 = [\underline{c}(\tilde{\eta}_2), c_H]$.

Therefore, the nonnegativity of σ and β holds automatically.

Because

$$
-\kappa - \int_{c_L}^{c_H} \phi(c)\,\mathrm{d}c - \int_{c_L}^{c_H}\int_{\tilde{\eta}_1}^{1} \xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c = v_E,
$$

$\forall(c,\eta)\in[c_L,c_H]\times[0,\tilde{\eta}_2]$:

$$\lambda_0\,\phi(c)\,g(\eta)-\chi(c,\eta)+\lambda_0\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\left[\lambda_0\,\left[v(\eta)-v_E\right]+\Delta\,\left[v(\eta)+1\right]\right]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\tilde{\eta}_2,1]$:

$$\lambda_0\,\phi(c)\,g(\eta)+\mu(c,\eta)+\lambda_0\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\left[\lambda_0\,\left[v(\eta)-v_E\right]+\Delta\,\left[v(\eta)+1\right]\right]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[0,\tilde{\eta}_1]$:

$$\lambda_1\,\phi(c)\,g(\eta)-\omega(c,\eta)+\chi(c,\eta)+\lambda_1\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,\left[v(\eta)-v_E\right]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\tilde{\eta}_1,\tilde{\eta}_2]$:

$$\lambda_1\,\phi(c)\,g(\eta)+\chi(c,\eta)+\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,\left[v(\eta)-v_E\right]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\tilde{\eta}_2,1]$:

$$\lambda_1\,\phi(c)\,g(\eta)+\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,\left[v(\eta)-v_E\right]\,g(\eta)\,f(c),$$

$\forall c\in\mathcal{C}_1$:

$$\begin{aligned}&-\int_{c_L}^{c}\phi(y)\,\mathrm{d}y-c\,\phi(c)+\sigma(c)-c\int_{\tilde{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta\\&\qquad-\int_{c_L}^{c}\int_{\tilde{\eta}_1}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y-\int_{\tilde{\eta}_2}^{1}\mu(c,\eta)\,\mathrm{d}\eta=v_E\,[c\,f(c)+F(c)],\end{aligned}$$

$\forall c\in\mathcal{C}_2$:

$$\begin{aligned}&-\int_{c_L}^{c}\phi(y)\,\mathrm{d}y-c\,\phi(c)-\beta(c)-c\int_{\tilde{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta\\&\qquad-\int_{c_L}^{c}\int_{\tilde{\eta}_1}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y-\int_{\tilde{\eta}_2}^{1}\mu(c,\eta)\,\mathrm{d}\eta=v_E\,[c\,f(c)+F(c)],\end{aligned}$$

we have

$$\begin{aligned}&\kappa\,[-\Pi(c_H)]+\int_{\mathcal{C}_1}\sigma(c)\,\tau(c)\,\mathrm{d}c+\int_{\mathcal{C}_2}\beta(c)\,[-\tau(c)]\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{\tilde{\eta}_2}^{1}\mu(c,\eta)\,[\hat{\alpha}_2(c,\eta)-\tau(c)]\,\mathrm{d}\eta\,\mathrm{d}c\\&+\int_{c_L}^{c_H}\int_{0}^{\tilde{\eta}_1}\omega(c,\eta)\,[-\hat{\alpha}_3(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{0}^{\tilde{\eta}_2}\chi(c,\eta)\,[\hat{\alpha}_3(c,\eta)-\hat{\alpha}_2(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c\\&+\int_{c_L}^{c_H}\int_{\tilde{\eta}_1}^{1}\xi(c,\eta)\,\left[\hat{\alpha}_3(c,\eta)-\Pi(c_H)-\int_{c}^{c_H}\tau(y)\,\mathrm{d}y-c\,\tau(c)+\int_{0}^{1}\left[\lambda_0\,\hat{\alpha}_2(c,z)+\lambda_1\,\hat{\alpha}_3(c,z)\right]\mathrm{d}G(z)\right]\mathrm{d}\eta\,\mathrm{d}c\\&+\int_{c_L}^{c_H}\phi(c)\,\left[-\Pi(c_H)-\int_{c}^{c_H}\tau(y)\,\mathrm{d}y-c\,\tau(c)+\lambda_0\int_{0}^{1}\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta)+\lambda_1\int_{0}^{1}\hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta)\right]\mathrm{d}c\\=&\,\left[-\kappa-\int_{c_L}^{c_H}\phi(c)\,\mathrm{d}c-\int_{c_L}^{c_H}\int_{\tilde{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\right]\,\Pi(c_H)\end{aligned}$$

$$
\begin{aligned}
&+\int_{c_L}^{c_H}\int_{0}^{\tilde{\eta}_2}\left[\lambda_0\,\phi(c)\,g(\eta)-\chi(c,\eta)+\lambda_0\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\tilde{\eta}_2}^{1}\left[\lambda_0\,\phi(c)\,g(\eta)+\mu(c,\eta)+\lambda_0\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{0}^{\tilde{\eta}_1}\left[\lambda_1\,\phi(c)\,g(\eta)-\omega(c,\eta)+\chi(c,\eta)+\lambda_1\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\tilde{\eta}_1}^{\tilde{\eta}_2}\left[\lambda_1\,\phi(c)\,g(\eta)+\chi(c,\eta)+\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\tilde{\eta}_2}^{1}\left[\lambda_1\,\phi(c)\,g(\eta)+\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\tilde{\eta}_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{\mathcal{C}_1}\left[-\int_{c_L}^{c}\phi(y)\,\mathrm{d}y-c\,\phi(c)+\sigma(c)-\int_{\tilde{\eta}_2}^{1}\mu(c,\eta)\,\mathrm{d}\eta-c\int_{\tilde{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\tilde{\eta}_1}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
&+\int_{\mathcal{C}_2}\left[-\int_{c_L}^{c}\phi(y)\,\mathrm{d}y-c\,\phi(c)-\beta(c)-\int_{\tilde{\eta}_2}^{1}\mu(c,\eta)\,\mathrm{d}\eta-c\int_{\tilde{\eta}_1}^{1}\xi(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\tilde{\eta}_1}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
=\;& v_E\,\Pi(c_H)+v_E\int_{c_L}^{c_H}\left[c+\frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{0}^{1}\left[\lambda_0\,\left[v(\eta)-v_E\right]+\Delta\,\left[v(\eta)+1\right]\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c)\\
&+\lambda_1\int_{c_L}^{c_H}\int_{0}^{1}\left[v(\eta)-v_E\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c).
\end{aligned}
$$

Therefore, the objective function is bounded by $\int_{\mathcal{C}_1}\sigma(c)\,\mathrm{d}c$. Substituting our candidate solution into the objective function achieves the upper bound in the first two cases:

1. If $\underline{c}(\tilde{\eta}_2)\le c_L$, we have $\hat{c}=c_L$ and the measure of $\mathcal{C}_1$ is 0. The regulator's expected payoff from our candidate policy and its upper bound are both 0.
2. If $c_L<\underline{c}(\tilde{\eta}_2)\le c_H$, we have

$$
\begin{aligned}
&\int_{c_L}^{\underline{c}(\tilde{\eta}_2)}\sigma(c)\,\mathrm{d}c-F(\underline{c}(\tilde{\eta}_2))\int_{\tilde{\eta}_2}^{1}\left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)\\
=\;&\int_{c_L}^{\underline{c}(\tilde{\eta}_2)}\left\{(\lambda_0+\Delta)\int_{\tilde{\eta}_2}^{1}\left[v(\eta)-v(\tilde{\eta}_2)\right]\,g(\eta)\,\mathrm{d}\eta+\left[c+\frac{F(c)}{f(c)}\right]\,\left[v(\tilde{\eta}_2)+\frac{\Delta}{\lambda_0}\,\left[v(\tilde{\eta}_2)+1\right]\right]\right\}\,f(c)\,\mathrm{d}c\\
&-F(\underline{c}(\tilde{\eta}_2))\int_{\tilde{\eta}_2}^{1}\left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)\\
=\;&\left\{-(\lambda_0+\Delta)\,v(\tilde{\eta}_2)\,\left[1-G(\tilde{\eta}_2)\right]+\underline{c}(\tilde{\eta}_2)\,\left[v(\tilde{\eta}_2)+\frac{\Delta}{\lambda_0}\,\left[v(\tilde{\eta}_2)+1\right]\right]-\Delta\,\left[1-G(\tilde{\eta}_2)\right]\right\}\,F(\underline{c}(\tilde{\eta}_2))\\
=\;&0,
\end{aligned}
$$

establishing the optimality.

In case (3)(a)(iii), we set $\mathcal{C}_1=[c_L,c_H]$ and $\mathcal{C}_2=\emptyset$, and replace the thresholds $\tilde{\eta}_1$ and $\tilde{\eta}_2$ with $\eta_1(\breve{\eta}_2)$ and $\breve{\eta}_2$, respectively, where $\breve{\eta}_2$ is determined by $\underline{c}(\breve{\eta}_2)=c_H$, and $\eta_1(\eta_2)$ is obtained by solving $\Psi_2(\eta_1(\eta_2),\eta_2)=0$. In this case, the nonnegativity of κ, χ, μ, ω, and ξ follows from their definitions directly. The desired nonnegativity of ϕ follows from

$$
\iota(\breve{\eta}_1,\breve{\eta}_2)=-\frac{\Xi(\breve{\eta}_2)}{\lambda_0}\ge-\frac{\Xi(\tilde{\eta}_2)}{\lambda_0}=\iota(\tilde{\eta}_1,\tilde{\eta}_2)\ge 0,
$$

where function Ξ is defined by (EC.18). The first inequality follows from the monotonicity of Ξ and the fact that $\breve{\eta}_2 \geq \tilde{\eta}_2$. Due to the monotonicity of σ, its desired nonnegativity condition boils down to $\sigma(c_H) \geq 0$, which follows from

$$\sigma(c_H) = h_2(\breve{\eta}_2)\, f(c_H) \geq h_2(\tilde{\eta}_2)\, f(c_H) = 0,$$

where the function h_2 is defined in (EC.23), and the inequality follows from the monotonicity of h_2, as established in Lemma EC.4. Because all the multipliers remain nonnegative, the objective function is bounded by $\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c$, which is attained by our candidate solution, i.e.,

$$\begin{aligned}
&\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c - \int_{\breve{\eta}_2}^{1} [(\lambda_0+\Delta)\, v(\eta)+\Delta]\,\mathrm{d}G(\eta) \\
=\ &\int_{c_L}^{c_H} \left\{ (\lambda_0+\Delta) \int_{\breve{\eta}_2}^{1} [v(\eta)-v(\breve{\eta}_2)]\, g(\eta)\,\mathrm{d}\eta + \left[c+\frac{F(c)}{f(c)}\right] \left[v(\breve{\eta}_2)+\frac{\Delta}{\lambda_0}\,[v(\breve{\eta}_2)+1]\right] \right\} f(c)\,\mathrm{d}c \\
&- \int_{\breve{\eta}_2}^{1} [(\lambda_0+\Delta)\, v(\eta)+\Delta]\,\mathrm{d}G(\eta) \\
=\ &-(\lambda_0+\Delta)\, v(\breve{\eta}_2)\,[1-G(\breve{\eta}_2)] + c_H \left[v(\breve{\eta}_2)+\frac{\Delta}{\lambda_0}\,[v(\breve{\eta}_2)+1]\right] - \Delta\,[1-G(\breve{\eta}_2)] \\
=\ &0,
\end{aligned}$$

where the last step follows from $c_H = \lambda_0\,[1-G(\breve{\eta}_2)]$.

In case (3)(a)(iv), note that the integrand of the objective function (43) is bounded above by

$$\begin{aligned}
&v_E\,\alpha_1(c) + \int_0^1 [(\lambda_0+\Delta)\, v(\eta)+\Delta]\,\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta) + \lambda_1 \int_0^1 v(\eta)\,\hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta) \\
\leq\ &v_E\,\alpha_1(c) + \int_{\bar{\eta}'}^1 [(\lambda_0+\Delta)\, v(\eta)+\Delta]\,\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta) + \lambda_1 \int_{\bar{\eta}}^1 v(\eta)\,\alpha_1(c)\,\mathrm{d}G(\eta) \\
=\ &\left[v_E + \lambda_1 \int_{\bar{\eta}}^1 v(\eta)\,\mathrm{d}G(\eta)\right] \alpha_1(c) + \int_{\bar{\eta}'}^1 [(\lambda_0+\Delta)\, v(\eta)+\Delta]\,\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta) \\
\leq\ &\int_{\bar{\eta}'}^1 [(\lambda_0+\Delta)\, v(\eta)+\Delta]\,\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta),
\end{aligned}$$

where the first inequality follows from the definitions of $\bar{\eta}$ and $\bar{\eta}'$ and the upper bound of $\hat{\alpha}_3(c,\eta)$, while the second inequality follows from the given condition $h(\bar{\eta}) \leq 0$. This upper bound is attained by committing to no conditional approval and adopting the myopic threshold ($\hat{\eta}_2 = \bar{\eta}'$). When $c_H \leq \underline{c}(\bar{\eta}')$, this policy enables the regulator to induce firms of all types to participate. Therefore, the mechanism is optimal under the given conditions.

To prove the optimality of cases (3)(b)(i) and (3)(b)(ii), define

$$\begin{aligned}
&\kappa \equiv -v(\acute{\eta}_2), \\
&\nu(c,\eta) \equiv (\lambda_0+\Delta)\,[v(\acute{\eta}_1)-v(\eta)]\, g(\eta)\, f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [0, \acute{\eta}_1],
\end{aligned}$$

$$\mu(c,\eta) \equiv (\lambda_0+\Delta)\,[v(\eta)-v(\acute{\eta}_1)]\,g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[\acute{\eta}_1,1],$$
$$\omega(c,\eta) \equiv \lambda_1\,[v(\acute{\eta}_2)-v(\eta)]\,g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[0,\acute{\eta}_2],$$
$$\xi(c,\eta) \equiv \lambda_1\,[v(\eta)-v(\acute{\eta}_2)]\,g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[\acute{\eta}_2,1],$$
$$\phi(c) \equiv \left\{-\lambda_1\int_{\acute{\eta}_2}^{1}[v(z)-v(\acute{\eta}_2)]\,g(z)\,\mathrm{d}z + v(\acute{\eta}_2) - v_E\right\} f(c), \quad \forall c\in[c_L,c_H],$$
$$\sigma(c) \equiv \left\{v(\acute{\eta}_2)\left[c+\frac{F(c)}{f(c)}\right] + (\lambda_0+\Delta)\int_{\acute{\eta}_1}^{1}[v(\eta)-v(\acute{\eta}_1)]\,g(\eta)\,\mathrm{d}\eta\right\} f(c), \quad \forall c\in\mathcal{C}_1,$$
$$\beta(c) \equiv -\sigma(c), \quad \forall c\in\mathcal{C}_2.$$

The nonnegativity of κ follows from the fact that $\acute{\eta}_2\le\bar{\eta}'\le\bar{\eta}$. The nonnegativity of ν, μ, ω, and ξ follows form their respective definitions directly. The nonnegativity of ϕ follows from Lemma EC.4(2).

In the first two possible cases, we have:

1. if $\underline{c}(\acute{\eta}_1)<c_L$, we know that β is nonnegative on $[c_L,c_H]$. In this case, we have $\mathcal{C}_1=\emptyset$ and $\mathcal{C}_2=[c_L,c_H]$; and
2. if $c_L\le\underline{c}(\acute{\eta}_1)\le c_H$, we know that σ and β are nonnegative on $[c_L,\underline{c}(\acute{\eta}_1)]$ and $[\underline{c}(\acute{\eta}_1),c_H]$, respectively. In this case, we have $\mathcal{C}_1=[c_L,\underline{c}(\acute{\eta}_1)]$ and $\mathcal{C}_2=[\underline{c}(\acute{\eta}_1),c_H]$.

Therefore, the nonnegativity of σ and β holds automatically.

Because

$$-\kappa-\int_{c_L}^{c_H}\phi(c)\,\mathrm{d}c-\int_{c_L}^{c_H}\int_{\acute{\eta}_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c=v_E,$$

$\forall(c,\eta)\in[c_L,c_H]\times[0,\acute{\eta}_1]$:

$$\lambda_0\,\phi(c)\,g(\eta)-\nu(c,\eta)+\lambda_0\,g(\eta)\int_{\acute{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z=[\lambda_0\,[v(\eta)-v_E]+\Delta\,[v(\eta)+1]]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\acute{\eta}_1,1]$:

$$\lambda_0\,\phi(c)\,g(\eta)+\mu(c,\eta)+\lambda_0\,g(\eta)\int_{\acute{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z=[\lambda_0\,[v(\eta)-v_E]+\Delta\,[v(\eta)+1]]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[0,\acute{\eta}_2]$:

$$\lambda_1\,\phi(c)\,g(\eta)-\omega(c,\eta)+\lambda_1\,g(\eta)\int_{\acute{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,[v(\eta)-v_E]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\acute{\eta}_2,1]$:

$$\lambda_1\,\phi(c)\,g(\eta)+\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\acute{\eta}_2}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,[v(\eta)-v_E]\,g(\eta)\,f(c),$$

$\forall c \in \mathcal{C}_1$:

$$-\int_{c_L}^{c} \phi(y)\,\mathrm{d}y - c\,\phi(c) + \sigma(c) - c\int_{\acute{\eta}_2}^{1} \xi(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\acute{\eta}_2}^{1} \xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y - \int_{\acute{\eta}_1}^{1} \mu(c,\eta)\,\mathrm{d}\eta = v_E\,[c\,f(c) + F(c)],$$

$\forall c \in \mathcal{C}_2$:

$$-\int_{c_L}^{c} \phi(y)\,\mathrm{d}y - c\,\phi(c) - \beta(c) - c\int_{\acute{\eta}_2}^{1} \xi(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\acute{\eta}_2}^{1} \xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y - \int_{\acute{\eta}_1}^{1} \mu(c,\eta)\,\mathrm{d}\eta = v_E\,[c\,f(c) + F(c)],$$

we have

$$\begin{aligned}
&\kappa\,[-\Pi(c_H)] + \int_{\mathcal{C}_1} \sigma(c)\,\tau(c)\,\mathrm{d}c + \int_{\mathcal{C}_2} \beta(c)\,[-\tau(c)]\,\mathrm{d}c + \int_{c_L}^{c_H}\int_{0}^{\acute{\eta}_1} \nu(c,\eta)\,[-\hat{\alpha}_2(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{\acute{\eta}_1}^{1} \mu(c,\eta)\,[\hat{\alpha}_2(c,\eta) - \tau(c)]\,\mathrm{d}\eta\,\mathrm{d}c + \int_{c_L}^{c_H}\int_{0}^{\acute{\eta}_2} \omega(c,\eta)\,[-\hat{\alpha}_3(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{\acute{\eta}_2}^{1} \xi(c,\eta)\left[\hat{\alpha}_3(c,\eta) - \Pi(c_H) - \int_{c}^{c_H} \tau(y)\,\mathrm{d}y - c\,\tau(c) + \int_{0}^{1} [\lambda_0\,\hat{\alpha}_2(c,z) + \lambda_1\,\hat{\alpha}_3(c,z)]\,\mathrm{d}G(z)\right]\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H} \phi(c)\left[-\Pi(c_H) - \int_{c}^{c_H} \tau(y)\,\mathrm{d}y - c\,\tau(c) + \lambda_0\int_{0}^{1} \hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta) + \lambda_1\int_{0}^{1} \hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta)\right]\mathrm{d}c \\
=\ &\left[-\kappa - \int_{c_L}^{c_H} \phi(c)\,\mathrm{d}c - \int_{c_L}^{c_H}\int_{\acute{\eta}_2}^{1} \xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\right]\Pi(c_H) \\
&+ \int_{c_L}^{c_H}\int_{0}^{\acute{\eta}_1}\left[\lambda_0\,\phi(c)\,g(\eta) - \nu(c,\eta) + \lambda_0\,g(\eta)\int_{\acute{\eta}_2}^{1} \xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{\acute{\eta}_1}^{1}\left[\lambda_0\,\phi(c)\,g(\eta) + \mu(c,\eta) + \lambda_0\,g(\eta)\int_{\acute{\eta}_2}^{1} \xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{0}^{\acute{\eta}_2}\left[\lambda_1\,\phi(c)\,g(\eta) - \omega(c,\eta) + \lambda_1\,g(\eta)\int_{\acute{\eta}_2}^{1} \xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{\acute{\eta}_2}^{1}\left[\lambda_1\,\phi(c)\,g(\eta) + \xi(c,\eta) + \lambda_1\,g(\eta)\int_{\acute{\eta}_1}^{1} \xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{\mathcal{C}_1}\left[-\int_{c_L}^{c} \phi(y)\,\mathrm{d}y - c\,\phi(c) + \sigma(c) - \int_{\acute{\eta}_1}^{1} \mu(c,\eta)\,\mathrm{d}\eta - c\int_{\acute{\eta}_2}^{1} \xi(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\acute{\eta}_2}^{1} \xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c \\
&+ \int_{\mathcal{C}_2}\left[-\int_{c_L}^{c} \phi(y)\,\mathrm{d}y - c\,\phi(c) - \beta(c) - \int_{\acute{\eta}_1}^{1} \mu(c,\eta)\,\mathrm{d}\eta - c\int_{\acute{\eta}_2}^{1} \xi(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\acute{\eta}_2}^{1} \xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c \\
=\ &v_E\,\Pi(c_H) + v_E\int_{c_L}^{c_H}\left[c + \frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\,\mathrm{d}c + \int_{c_L}^{c_H}\int_{0}^{1} [\lambda_0\,[v(\eta) - v_E] + \Delta\,[v(\eta) + 1]]\,\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c) \\
&+ \lambda_1\int_{c_L}^{c_H}\int_{0}^{1} [v(\eta) - v_E]\,\hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c).
\end{aligned}$$

Therefore, the objective function is bounded by $\int_{\mathcal{C}_1} \sigma(c)\,\mathrm{d}c$. Substituting our candidate solution into the objective function achieves the upper bound in the first two cases:

1. If $\underline{c}(\acute{\eta}_1) \le c_L$, we have $\hat{c} = c_L$ and the measure of $\mathcal{C}_1$ is 0. The regulator's expected payoff from our candidate policy and its upper bound are both 0.

2. If $c_L < \underline{c}(\acute{\eta}_1) \le c_H$, we have

$$\begin{aligned}
&\int_{c_L}^{\underline{c}(\acute{\eta}_1)} \sigma(c)\,\mathrm{d}c - F(\underline{c}(\acute{\eta}_1)) \int_{\acute{\eta}_1}^{1} [(\lambda_0+\Delta)\,v(\eta)+\Delta]\,\mathrm{d}G(\eta)\\
=&\int_{c_L}^{\underline{c}(\acute{\eta}_1)} \left\{ v(\acute{\eta}_2)\left[c+\frac{F(c)}{f(c)}\right] + (\lambda_0+\Delta)\int_{\acute{\eta}_1}^{1}[v(\eta)-v(\acute{\eta}_1)]\,g(\eta)\,\mathrm{d}\eta \right\} f(c)\,\mathrm{d}c\\
&- F(\underline{c}(\acute{\eta}_1)) \int_{\acute{\eta}_1}^{1} [(\lambda_0+\Delta)\,v(\eta)+\Delta]\,\mathrm{d}G(\eta)\\
=&\,[\lambda_0\,v(\acute{\eta}_2) - (\lambda_0+\Delta)\,v(\acute{\eta}_1) - \Delta]\,[1-G(\acute{\eta}_1)]\,F(\underline{c}(\acute{\eta}_1))\\
=&\,0,
\end{aligned}$$

where the last step follows from $\Psi_1(\acute{\eta}_1, \acute{\eta}_2) = 0$.

In case (3)(b)(iii), we set $\mathcal{C}_1 = [c_L, c_H]$ and $\mathcal{C}_2 = \emptyset$, and replace the thresholds $\acute{\eta}_1$ and $\acute{\eta}_2$ with $\breve{\eta}_1$ and $\eta_2(\breve{\eta}_1)$, respectively, where $\breve{\eta}_1$ is determined by $\underline{c}(\breve{\eta}_1) = c_H$, and $\eta_2(\eta_1)$ is obtained by solving $\Psi_1(\eta_1, \eta_2(\eta_1)) = 0$. In this case, the nonnegativity of κ, ν, μ, ω, and ξ follows from their definitions directly. The desired nonnegativity of ϕ follows from

$$h(\breve{\eta}_2) \le h(\acute{\eta}_2) \le 0.$$

The nonnegativity of σ follows from its monotonicity and

$$\sigma(c_H) = h_1(\breve{\eta}_1)\,f(c_H) \ge h_1(\acute{\eta}_1)\,f(c_H) = 0,$$

where the function h_1 is defined in (EC.24), and the inequality follows from the monotonicity of function h_1, as established in Lemma EC.4. Because all the multipliers remain nonnegative, the objective function is bounded by $\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c$, which is attained by our candidate solution, i.e.,

$$\begin{aligned}
&\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c - \int_{\breve{\eta}_1}^{1} [(\lambda_0+\Delta)\,v(\eta)+\Delta]\,\mathrm{d}G(\eta)\\
=&\int_{c_L}^{c_H} \left\{ (\lambda_0+\Delta)\int_{\breve{\eta}_1}^{1}[v(\eta)-v(\breve{\eta}_1)]\,g(\eta)\,\mathrm{d}\eta + \left[c+\frac{F(c)}{f(c)}\right]\left[v(\breve{\eta}_1)+\frac{\Delta}{\lambda_0}\,[v(\breve{\eta}_1)+1]\right] \right\} f(c)\,\mathrm{d}c\\
&-\int_{\breve{\eta}_1}^{1} [(\lambda_0+\Delta)\,v(\eta)+\Delta]\,\mathrm{d}G(\eta)\\
=&\,-(\lambda_0+\Delta)\,v(\breve{\eta}_1)\,[1-G(\breve{\eta}_1)] + c_H\left[v(\breve{\eta}_1)+\frac{\Delta}{\lambda_0}\,[v(\breve{\eta}_1)+1]\right] - \Delta\,[1-G(\breve{\eta}_1)]\\
=&\,0,
\end{aligned}$$

where the last step follows from $c_H = \lambda_0\,[1-G(\breve{\eta}_1)]$.

In case (3)(b)(iv), similar to case (3)(a)(iv), committing to no conditional approval and adopting the myopic threshold ($\hat{\eta}_1 = \bar{\eta}'$) enables the regulator to induce all firms to participate. Therefore, this mechanism is optimal under the given conditions.

It remains to prove the optimality of the last case. For (4)(a)(i) and (4)(a)(ii), define

$$
\begin{aligned}
&\kappa \equiv -v(\dot{\eta}_2),\\
&\nu(c,\eta) \equiv (\lambda_0+\Delta)\,[v(\dot{\eta}_1)-v(\eta)]\,g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[0,\dot{\eta}_1],\\
&\mu(c,\eta) \equiv (\lambda_0+\Delta)\,[v(\eta)-v(\dot{\eta}_1)]\,g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[\dot{\eta}_1,1],\\
&\omega(c,\eta) \equiv \lambda_1\,[v(\dot{\eta}_2)-v(\eta)]\,g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[0,\dot{\eta}_2],\\
&\xi(c,\eta) \equiv \lambda_1\,[v(\eta)-v(\dot{\eta}_2)]\,g(\eta)\,f(c), \quad \forall (c,\eta)\in[c_L,c_H]\times[\dot{\eta}_2,1],\\
&\psi(c) \equiv \left\{\lambda_1\int_{\dot{\eta}_2}^1 [v(z)-v(\dot{\eta}_2)]\,g(z)\,\mathrm{d}z - v(\dot{\eta}_2)+v_E\right\} f(c),\\
&\sigma(c) \equiv \left\{v(\dot{\eta}_2)\left[c+\frac{F(c)}{f(c)}\right] + (\lambda_0+\Delta)\int_{\dot{\eta}_1}^1 [v(\eta)-v(\dot{\eta}_1)]\,g(\eta)\,\mathrm{d}\eta\right.\\
&\qquad\qquad \left.+\lambda_1\int_{\dot{\eta}_2}^1 [v(z)-v(\dot{\eta}_2)]\,g(z)\,\mathrm{d}z - v(\dot{\eta}_2)+v_E\right\} f(c), \quad \forall c\in\mathcal{C}_1,\\
&\beta(c) \equiv -\sigma(c), \quad \forall c\in\mathcal{C}_2.
\end{aligned}
$$

The nonnegativity of κ follows from the fact that $\dot{\eta}_2 \le \bar{\eta}$. The nonnegativity of ν, μ, ω, and ξ follows form their respective definitions directly. The nonnegativity of ψ follows from Lemma EC.6(1).

In the first two possible cases, we have:

1. if $\bar{c}(\dot{\eta}_1,\dot{\eta}_2) < c_L$, we know that β is nonnegative on $[c_L,c_H]$. In this case, we have $\mathcal{C}_1=\emptyset$ and $\mathcal{C}_2=[c_L,c_H]$; and
2. if $c_L \le \bar{c}(\dot{\eta}_1,\dot{\eta}_2) \le c_H$, we know that σ and β are nonnegative on $[c_L,\bar{c}(\dot{\eta}_1,\dot{\eta}_2)]$ and $[\bar{c}(\dot{\eta}_1,\dot{\eta}_2),c_H]$, respectively. In this case, we have $\mathcal{C}_1=[c_L,\bar{c}(\dot{\eta}_1,\dot{\eta}_2)]$ and $\mathcal{C}_2=[\bar{c}(\dot{\eta}_1,\dot{\eta}_2),c_H]$.

Therefore, the nonnegativity of σ and β holds automatically.

Because

$$-\kappa + \int_{c_L}^{c_H}\psi(c)\,\mathrm{d}c - \int_{c_L}^{c_H}\int_{\dot{\eta}_2}^1 \xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c = v_E,$$

$\forall (c,\eta)\in[c_L,c_H]\times[0,\dot{\eta}_1]$:

$$-\lambda_0\,\psi(c)\,g(\eta) - \nu(c,\eta) + \lambda_0\,g(\eta)\int_{\dot{\eta}_2}^1 \xi(c,z)\,\mathrm{d}z = [\lambda_0\,[v(\eta)-v_E]+\Delta\,[v(\eta)+1]]\,g(\eta)\,f(c),$$

$\forall (c,\eta)\in[c_L,c_H]\times[\dot{\eta}_1,1]$:

$$-\lambda_0\,\psi(c)\,g(\eta) + \mu(c,\eta) + \lambda_0\,g(\eta)\int_{\dot{\eta}_2}^1 \xi(c,z)\,\mathrm{d}z = [\lambda_0\,[v(\eta)-v_E]+\Delta\,[v(\eta)+1]]\,g(\eta)\,f(c),$$

$\forall (c,\eta)\in[c_L,c_H]\times[0,\dot{\eta}_2]$:

$$-\lambda_1\,\psi(c)\,g(\eta) - \omega(c,\eta) + \lambda_1\,g(\eta)\int_{\dot{\eta}_2}^1 \xi(c,z)\,\mathrm{d}z = \lambda_1\,[v(\eta)-v_E]\,g(\eta)\,f(c),$$

$\forall(c,\eta)\in[c_L,c_H]\times[\dot\eta_2,1]$:

$$-\lambda_1\,\psi(c)\,g(\eta)+\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\dot\eta_2}^{1}\xi(c,z)\,\mathrm{d}z=\lambda_1\,[v(\eta)-v_E]\,g(\eta)\,f(c),$$

$\forall c\in[c_L,\hat c]$:

$$\int_{c_L}^{c}\psi(y)\,\mathrm{d}y+(c-1)\,\psi(c)+\sigma(c)-c\int_{\dot\eta_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\dot\eta_2}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y-\int_{\dot\eta_1}^{1}\mu(c,\eta)\,\mathrm{d}\eta=v_E\,[c\,f(c)+F(c)],$$

$\forall c\in[\hat c,c_H]$:

$$\int_{c_L}^{c}\psi(y)\,\mathrm{d}y+(c-1)\,\psi(c)-\beta(c)-c\int_{\dot\eta_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\dot\eta_2}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y-\int_{\dot\eta_1}^{1}\mu(c,\eta)\,\mathrm{d}\eta=v_E\,[c\,f(c)+F(c)],$$

we have

$$\begin{aligned}
&\kappa\,[-\Pi(c_H)]+\int_{\mathcal{C}_1}\sigma(c)\,\tau(c)\,\mathrm{d}c+\int_{\mathcal{C}_2}\beta(c)\,[-\tau(c)]\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{0}^{\dot\eta_1}\nu(c,\eta)\,[-\hat\alpha_2(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\dot\eta_1}^{1}\mu(c,\eta)\,[\hat\alpha_2(c,\eta)-\tau(c)]\,\mathrm{d}\eta\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{0}^{\dot\eta_2}\omega(c,\eta)\,[-\hat\alpha_3(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\dot\eta_2}^{1}\xi(c,\eta)\left[\hat\alpha_3(c,\eta)-\Pi(c_H)-\int_{c}^{c_H}\tau(y)\,\mathrm{d}y-c\,\tau(c)+\int_0^1[\lambda_0\,\hat\alpha_2(c,z)+\lambda_1\,\hat\alpha_3(c,z)]\,\mathrm{d}G(z)\right]\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\psi(c)\left[\Pi(c_H)+\int_{c}^{c_H}\tau(y)\,\mathrm{d}y+c\,\tau(c)-\lambda_0\int_0^1\hat\alpha_2(c,\eta)\,\mathrm{d}G(\eta)-\lambda_1\int_0^1\hat\alpha_3(c,\eta)\,\mathrm{d}G(\eta)-\tau(c)\right]\mathrm{d}c\\
=&\left[-\kappa+\int_{c_L}^{c_H}\psi(c)\,\mathrm{d}c-\int_{c_L}^{c_H}\int_{\dot\eta_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\right]\Pi(c_H)\\
&+\int_{c_L}^{c_H}\int_{0}^{\dot\eta_1}\left[-\lambda_0\,\psi(c)\,g(\eta)-\nu(c,\eta)+\lambda_0\,g(\eta)\int_{\dot\eta_2}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\dot\eta_1}^{1}\left[-\lambda_0\,\psi(c)\,g(\eta)+\mu(c,\eta)+\lambda_0\,g(\eta)\int_{\dot\eta_2}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{0}^{\dot\eta_2}\left[-\lambda_1\,\psi(c)\,g(\eta)-\omega(c,\eta)+\lambda_1\,g(\eta)\int_{\dot\eta_2}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat\alpha_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\dot\eta_2}^{1}\left[-\lambda_1\,\psi(c)\,g(\eta)+\xi(c,\eta)+\lambda_1\,g(\eta)\int_{\dot\eta_1}^{1}\xi(c,z)\,\mathrm{d}z\right]\hat\alpha_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{\mathcal{C}_1}\left[\int_{c_L}^{c}\psi(y)\,\mathrm{d}y+(c-1)\,\psi(c)+\sigma(c)-\int_{\dot\eta_1}^{1}\mu(c,\eta)\,\mathrm{d}\eta-c\int_{\dot\eta_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\dot\eta_2}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
&+\int_{\mathcal{C}_2}\left[\int_{c_L}^{c}\psi(y)\,\mathrm{d}y+(c-1)\,\psi(c)-\beta(c)-\int_{\dot\eta_1}^{1}\mu(c,\eta)\,\mathrm{d}\eta-c\int_{\dot\eta_2}^{1}\xi(c,\eta)\,\mathrm{d}\eta-\int_{c_L}^{c}\int_{\dot\eta_2}^{1}\xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c\\
=&\,v_E\,\Pi(c_H)+v_E\int_{c_L}^{c_H}\left[c+\frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\,\mathrm{d}c+\int_{c_L}^{c_H}\int_0^1[\lambda_0\,[v(\eta)-v_E]+\Delta\,[v(\eta)+1]]\,\hat\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c)\\
&+\lambda_1\int_{c_L}^{c_H}\int_0^1[v(\eta)-v_E]\,\hat\alpha_3(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c).
\end{aligned}$$

Therefore, the objective function is bounded by $\int_{\mathcal{C}_1}\sigma(c)\,\mathrm{d}c$. Substituting our candidate solution into the objective function achieves the upper bound in the first two cases, which follows from:

1. if $\bar{c}(\dot{\eta}_1, \dot{\eta}_2) \le c_L$, we have $\hat{c} = c_L$ and the measure of $\mathcal{C}_1$ is 0. The regulator's expected payoff from our candidate policy and its upper bound are both 0;
2. if $c_L < \bar{c}(\dot{\eta}_1, \dot{\eta}_2) \le c_H$, the regulator's expected payoff from our candidate policy attains this upper bound, i.e.,

$$\begin{aligned}
&\int_{c_L}^{\bar{c}(\dot{\eta}_1,\dot{\eta}_2)} \sigma(c)\,\mathrm{d}c - \left[v_E + \int_{\dot{\eta}_1}^{1} \left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta) + \lambda_1 \int_{\dot{\eta}_2}^{1} v(\eta)\,\mathrm{d}G(\eta)\right] F(\bar{c}(\dot{\eta}_1,\dot{\eta}_2)) \\
=\ & \{v(\dot{\eta}_2)\,[1+\lambda_0\,[1-G(\dot{\eta}_1)]+\lambda_1\,[1-G(\dot{\eta}_2)]] - \Delta\,[1-G(\dot{\eta}_1)] \\
& -(\lambda_0+\Delta)\,v(\dot{\eta}_1)\,[1-G(\dot{\eta}_1)] - \lambda_1\,v(\dot{\eta}_2)\,[1-G(\dot{\eta}_2)] - v(\dot{\eta}_2)\}\,F(\bar{c}(\dot{\eta}_1,\dot{\eta}_2)) \\
=\ & [\lambda_0\,v(\dot{\eta}_2) - \Delta - (\lambda_0+\Delta)\,v(\dot{\eta}_1)]\,[1-G(\dot{\eta}_1)]\,F(\bar{c}(\dot{\eta}_1,\dot{\eta}_2)) \\
=\ & 0,
\end{aligned}$$

   where the last step follows from $\Psi_1(\dot{\eta}_1, \dot{\eta}_2) = 0$.

In case (4)(a)(iii), we set $\mathcal{C}_1 = [c_L, c_H]$ and $\mathcal{C}_2 = \emptyset$, and replace the thresholds $\dot{\eta}_1$ and $\dot{\eta}_2$ with $\breve{\eta}_1$ and $\breve{\eta}_2$, respectively, where the two thresholds are determined by $\bar{c}(\breve{\eta}_1, \breve{\eta}_2) = c_H$ and $\Psi_1(\breve{\eta}_1, \breve{\eta}_2) = 0$. In this case, the nonnegativity of κ, ν, μ, ω, and ξ follows from their definitions directly. The nonnegativity of ψ follows as: if $h(\bar{\eta}) \le 0$, we have $h(\breve{\eta}_2) \ge h(\breve{\eta}_2) = 0$; otherwise, we have $h(\breve{\eta}) \ge h(\bar{\eta}) > 0$. The nonnegativity of σ follows from the fact that for any $c \in [c_L, c_H]$, we have

$$\sigma(c) \ge \sigma(c_H) = t_1(\breve{\eta}_2)\,f(c_H) \ge t_1(\dot{\eta}_2)\,f(c_H) = 0,$$

where the function t_1 is defined in (EC.25), and the inequality follows from the monotonicity of t_1, as established in Lemma EC.6. Because all the multipliers remain nonnegative, the objective function is bounded by $\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c$, which is attained by our candidate solution, i.e.,

$$\begin{aligned}
&\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c - \left[v_E + \int_{\breve{\eta}_1}^{1} \left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta) + \lambda_1 \int_{\breve{\eta}_2}^{1} v(\eta)\,\mathrm{d}G(\eta)\right] F(c_H) \\
=\ & c_H\,v(\breve{\eta}_2) + (\lambda_0+\Delta)\int_{\breve{\eta}_1}^{1} [v(\eta)-v(\breve{\eta}_1)]\,g(\eta)\,\mathrm{d}\eta + \lambda_1 \int_{\breve{\eta}_2}^{1} [v(z)-v(\breve{\eta}_2)]\,g(z)\,\mathrm{d}z - v(\breve{\eta}_2) + v_E \\
& - \left[v_E + \int_{\breve{\eta}_1}^{1} \left[(\lambda_0+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta) + \lambda_1 \int_{\breve{\eta}_2}^{1} v(\eta)\,\mathrm{d}G(\eta)\right] \\
=\ & c_H\,v(\breve{\eta}_2) - (\lambda_0+\Delta)\,v(\breve{\eta}_1)\,[1-G(\breve{\eta}_1)] - \lambda_1\,v(\breve{\eta}_2)\,[1-G(\breve{\eta}_2)] - v(\breve{\eta}_2) - \Delta\,[1-G(\breve{\eta}_1)] \\
=\ & [c_H - \lambda_1\,[1-G(\breve{\eta}_2)] - 1]\left[v(\breve{\eta}_1) + \frac{\Delta}{\lambda_0}\,[v(\breve{\eta}_1)+1]\right] - (\lambda_0+\Delta)\,v(\breve{\eta}_1)\,[1-G(\breve{\eta}_1)] - \Delta\,[1-G(\breve{\eta}_1)] \\
=\ & 0,
\end{aligned}$$

where the last equality follows from

$$c_H = 1 + \lambda_0\,[1-G(\breve{\eta}_1)] + \lambda_1\,[1-G(\breve{\eta}_2)]\,.$$

In case (4)(a)(iv), we replace the two thresholds $\dot{\eta}_1$ and $\dot{\eta}_2$ with $\bar{\eta}'$ and $\bar{\eta}$, respectively. The desired nonnegativity condition for ψ boils down to

$$\lambda_1 \int_{\bar{\eta}}^{1} v(z)\,\mathrm{d}G(z) + v_E \geq 0 \quad \Leftrightarrow \quad h(\bar{\eta}) \geq 0.$$

Suppose $h(\bar{\eta}) < 0$. We have

$$\check{c}_1 > \bar{c}(\check{\eta}_1, \check{\eta}_2) \geq 1 + \lambda_0\,[1 - G(\bar{\eta}')] + \lambda_1\,[1 - G(\bar{\eta})]$$

or

$$\check{c}_2 > \bar{c}(\mathring{\eta}_1, \mathring{\eta}_1) \geq 1 + (\lambda_0 + \lambda_1)\,\left[1 - G(\eta^{\dagger})\right] \geq 1 + (\lambda_0 + \lambda_1)\,[1 - G(\bar{\eta}')]\,,$$

which contradicts the condition $c_H \leq \bar{c}(\bar{\eta}', \bar{\eta})$. Therefore, we know that $h(\bar{\eta}) > 0$, guaranteeing the nonnegativity of ψ and σ, where the latter one follows from

$$\sigma(c) = \left\{ (\lambda_0 + \Delta) \int_{\bar{\eta}'}^{1} [v(\eta) - v(\bar{\eta}')]\; g(\eta)\,\mathrm{d}\eta + \underbrace{\lambda_1 \int_{\bar{\eta}}^{1} v(z)\,g(z)\,\mathrm{d}z + v_E}_{=h(\bar{\eta}) \geq 0} \right\} f(c) \geq 0.$$

The candidate primal solution attains the upper bound $\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c$, i.e.,

$$\begin{aligned}
&\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c - \left[v_E + \int_{\bar{\eta}'}^{1} [(\lambda_0 + \Delta)\,v(\eta) + \Delta]\,\mathrm{d}G(\eta) + \lambda_1 \int_{\bar{\eta}}^{1} v(\eta)\,\mathrm{d}G(\eta) \right] F(c_H) \\
&= (\lambda_0 + \Delta) \int_{\bar{\eta}'}^{1} [v(\eta) - v(\bar{\eta}')]\; g(\eta)\,\mathrm{d}\eta + \lambda_1 \int_{\bar{\eta}}^{1} v(z)\,g(z)\,\mathrm{d}z + v_E \\
&\quad - \left[v_E + \int_{\bar{\eta}'}^{1} [(\lambda_0 + \Delta)\,v(\eta) + \Delta]\,\mathrm{d}G(\eta) + \lambda_1 \int_{\bar{\eta}}^{1} v(\eta)\,\mathrm{d}G(\eta) \right] \\
&= -\,[(\lambda_0 + \Delta)\,v(\bar{\eta}') + \Delta]\;[1 - G(\bar{\eta}')] \\
&= 0,
\end{aligned}$$

where the last step follows from the definition of $\bar{\eta}'$.

It remains to prove the optimality in case (4)(b). For the first two subcases, define

$$\begin{aligned}
&\kappa \equiv -\frac{1}{\lambda_0}\,[(\lambda_0 + \Delta)\,v(\grave{\eta}_2) + \Delta]\,, \\
&\chi(c,\eta) \equiv (\lambda_0 + \Delta)\;[v(\grave{\eta}_2) - v(\eta)]\;g(\eta)\,f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [0, \grave{\eta}_2], \\
&\mu(c,\eta) = (\lambda_0 + \Delta)\;[v(\eta) - v(\grave{\eta}_2)]\;g(\eta)\,f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [\grave{\eta}_2, 1], \\
&\omega(c,\eta) = \left\{ (\lambda_0 + \lambda_1 + \Delta)\;[v(\grave{\eta}_2) - v(\eta)] + \frac{\lambda_1}{\lambda_0}\,\Delta\;[v(\grave{\eta}_2) + 1] \right\} g(\eta)\,f(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [0, \grave{\eta}_1], \\
&\xi(c,\eta) \equiv \begin{cases} -\left\{ (\lambda_0 + \lambda_1 + \Delta)\;[v(\grave{\eta}_2) - v(\eta)] + \frac{\lambda_1}{\lambda_0}\,\Delta\;[v(\grave{\eta}_2) + 1] \right\} g(\eta)\,f(c), & \forall (c,\eta) \in [c_L, c_H] \times [\grave{\eta}_1, \grave{\eta}_2], \\ \left\{ \lambda_1\;[v(\eta) - v(\grave{\eta}_2)] - \frac{\lambda_1}{\lambda_0}\,\Delta\;[v(\grave{\eta}_2) + 1] \right\} g(\eta)\,f(c), & \forall (c,\eta) \in [c_L, c_H] \times [\grave{\eta}_2, 1], \end{cases}
\end{aligned}$$

$$\psi(c) \equiv -\iota(\grave{\eta}_1, \grave{\eta}_2)\, f(c), \quad \forall c \in [c_L, c_H],$$
$$\sigma(c) \equiv \left\{ \left[v(\grave{\eta}_2) + \frac{\Delta}{\lambda_0} \left[v(\grave{\eta}_2) + 1 \right] \right] \left[c + \frac{F(c)}{f(c)} \right] + (\lambda_0 + \Delta) \int_{\grave{\eta}_2}^{1} \left[v(\eta) - v(\grave{\eta}_2) \right] g(\eta)\, \mathrm{d}\eta + \frac{\psi(c)}{f(c)} \right\} f(c), \quad \forall c \in \mathcal{C}_1,$$
$$\beta(c) \equiv -\sigma(c), \quad \forall c \in \mathcal{C}_2.$$

The nonnegativity of κ follows from the fact that $\grave{\eta}_2 \le \eta^\dagger \le \bar{\eta}'$. The nonnegativity of χ and μ follows from their respective definitions directly. The nonnegativity of ω and ξ follows from the relationship between $\grave{\eta}_1$ and $\grave{\eta}_2$ (i.e., $\Psi_2(\grave{\eta}_1, \grave{\eta}_2) = 0$, implying $\omega(c, \grave{\eta}_1) = \xi(c, \grave{\eta}_1) = 0$ for any $c \in [c_L, c_H]$) and their monotonicity in η. The nonnegativity of ψ follows from Lemma EC.6(2).

In the first two possible cases, we have:

1. if $\bar{c}(\grave{\eta}_1, \grave{\eta}_1) < c_L$, we know that β is nonnegative on $[c_L, c_H]$. In this case, we have $\mathcal{C}_1 = \emptyset$ and $\mathcal{C}_2 = [c_L, c_H]$; and
2. if $c_L \le \bar{c}(\grave{\eta}_1, \grave{\eta}_1) \le c_H$, we know that σ and β are nonnegative on $[c_L, \bar{c}(\grave{\eta}_1, \grave{\eta}_1)]$ and $[\bar{c}(\grave{\eta}_1, \grave{\eta}_1), c_H]$, respectively. In this case, we have $\mathcal{C}_1 = [c_L, \bar{c}(\grave{\eta}_1, \grave{\eta}_1)]$ and $\mathcal{C}_2 = [\bar{c}(\grave{\eta}_1, \grave{\eta}_1), c_H]$.

Therefore, the nonnegativity of σ and β holds automatically.

Because

$$-\kappa + \int_{c_L}^{c_H} \psi(c)\, \mathrm{d}c - \int_{c_L}^{c_H} \int_{\grave{\eta}_1}^{1} \xi(c, \eta)\, \mathrm{d}\eta\, \mathrm{d}c = v_E,$$

$\forall (c, \eta) \in [c_L, c_H] \times [0, \grave{\eta}_2]$:

$$-\lambda_0\, \psi(c)\, g(\eta) - \chi(c, \eta) + \lambda_0\, g(\eta) \int_{\grave{\eta}_1}^{1} \xi(c, z)\, \mathrm{d}z = \left[\lambda_0 \left[v(\eta) - v_E \right] + \Delta \left[v(\eta) + 1 \right] \right] g(\eta)\, f(c),$$

$\forall (c, \eta) \in [c_L, c_H] \times [\grave{\eta}_2, 1]$:

$$-\lambda_0\, \psi(c)\, g(\eta) + \mu(c, \eta) + \lambda_0\, g(\eta) \int_{\grave{\eta}_1}^{1} \xi(c, z)\, \mathrm{d}z = \left[\lambda_0 \left[v(\eta) - v_E \right] + \Delta \left[v(\eta) + 1 \right] \right] g(\eta)\, f(c),$$

$\forall (c, \eta) \in [c_L, c_H] \times [0, \grave{\eta}_1]$:

$$-\lambda_1\, \psi(c)\, g(\eta) - \omega(c, \eta) + \chi(c, \eta) + \lambda_1\, g(\eta) \int_{\grave{\eta}_1}^{1} \xi(c, z)\, \mathrm{d}z = \lambda_1 \left[v(\eta) - v_E \right] g(\eta)\, f(c),$$

$\forall (c, \eta) \in [c_L, c_H] \times [\grave{\eta}_1, \grave{\eta}_2]$:

$$-\lambda_1\, \psi(c)\, g(\eta) + \chi(c, \eta) + \xi(c, \eta) + \lambda_1\, g(\eta) \int_{\grave{\eta}_1}^{1} \xi(c, z)\, \mathrm{d}z = \lambda_1 \left[v(\eta) - v_E \right] g(\eta)\, f(c),$$

$\forall (c, \eta) \in [c_L, c_H] \times [\grave{\eta}_2, 1]$:

$$-\lambda_1\, \psi(c)\, g(\eta) + \xi(c, \eta) + \lambda_1\, g(\eta) \int_{\grave{\eta}_1}^{1} \xi(c, z)\, \mathrm{d}z = \lambda_1 \left[v(\eta) - v_E \right] g(\eta)\, f(c),$$

$\forall c \in [c_L, \hat{c}]$:

$$\int_{c_L}^{c} \psi(y)\,\mathrm{d}y + (c-1)\,\psi(c) + \sigma(c) - \int_{\dot{\eta}_2}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\dot{\eta}_1}^{1} \xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y - c\int_{\dot{\eta}_1}^{1} \xi(c,\eta)\,\mathrm{d}\eta = v_E\,[c\,f(c) + F(c)],$$

$\forall c \in [\hat{c}, c_H]$:

$$\int_{c_L}^{c} \psi(y)\,\mathrm{d}y + (c-1)\,\psi(c) - \beta(c) - \int_{\dot{\eta}_2}^{1} \mu(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\dot{\eta}_1}^{1} \xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y - c\int_{\dot{\eta}_1}^{1} \xi(c,\eta)\,\mathrm{d}\eta = v_E\,[c\,f(c) + F(c)],$$

we have

$$\begin{aligned}
&\kappa\,[-\Pi(c_H)] + \int_{\mathcal{C}_1} \sigma(c)\,\tau(c)\,\mathrm{d}c + \int_{\mathcal{C}_2} \beta(c)\,[-\tau(c)]\,\mathrm{d}c + \int_{c_L}^{c_H}\int_{0}^{\dot{\eta}_2} \chi(c,\eta)\,[\hat{\alpha}_3(c,\eta) - \hat{\alpha}_2(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{\dot{\eta}_2}^{1} \mu(c,\eta)\,[\hat{\alpha}_2(c,\eta) - \tau(c)]\,\mathrm{d}\eta\,\mathrm{d}c + \int_{c_L}^{c_H}\int_{0}^{\dot{\eta}_1} \omega(c,\eta)\,[-\hat{\alpha}_3(c,\eta)]\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{\dot{\eta}_1}^{1} \xi(c,\eta)\left[\hat{\alpha}_3(c,\eta) - \Pi(c_H) - \int_{c}^{c_H} \tau(y)\,\mathrm{d}y - c\,\tau(c) + \int_{0}^{1} [\lambda_0\,\hat{\alpha}_2(c,z) + \lambda_1\,\hat{\alpha}_3(c,z)]\,\mathrm{d}G(z)\right]\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H} \psi(c)\left[\Pi(c_H) + \int_{c}^{c_H} \tau(y)\,\mathrm{d}y + c\,\tau(c) - \lambda_0\int_{0}^{1} \hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta) - \lambda_1\int_{0}^{1} \hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta) - \tau(c)\right]\mathrm{d}c \\
=\ &\left[-\kappa + \int_{c_L}^{c_H} \psi(c)\,\mathrm{d}c - \int_{c_L}^{c_H}\int_{\dot{\eta}_1}^{1} \xi(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\right]\Pi(c_H) \\
&+ \int_{c_L}^{c_H}\int_{0}^{\dot{\eta}_2}\left[-\lambda_0\,\psi(c)\,g(\eta) - \chi(c,\eta) + \lambda_0\,g(\eta)\int_{\dot{\eta}_1}^{1} \xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{\dot{\eta}_2}^{1}\left[-\lambda_0\,\psi(c)\,g(\eta) + \mu(c,\eta) + \lambda_0\,g(\eta)\int_{\dot{\eta}_1}^{1} \xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{0}^{\dot{\eta}_1}\left[-\lambda_1\,\psi(c)\,g(\eta) - \omega(c,\eta) + \chi(c,\eta) + \lambda_1\,g(\eta)\int_{\dot{\eta}_1}^{1} \xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{\dot{\eta}_1}^{\dot{\eta}_2}\left[-\lambda_1\,\psi(c)\,g(\eta) + \chi(c,\eta) + \xi(c,\eta) + \lambda_1\,g(\eta)\int_{\dot{\eta}_1}^{1} \xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{c_L}^{c_H}\int_{\dot{\eta}_2}^{1}\left[-\lambda_1\,\psi(c)\,g(\eta) + \xi(c,\eta) + \lambda_1\,g(\eta)\int_{\dot{\eta}_1}^{1} \xi(c,z)\,\mathrm{d}z\right]\hat{\alpha}_3(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&+ \int_{\mathcal{C}_1}\left[\int_{c_L}^{c} \psi(y)\,\mathrm{d}y + (c-1)\,\psi(c) + \sigma(c) - \int_{\dot{\eta}_2}^{1} \mu(c,\eta)\,\mathrm{d}\eta - c\int_{\dot{\eta}_1}^{1} \xi(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\dot{\eta}_1}^{1} \xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c \\
&+ \int_{\mathcal{C}_2}\left[\int_{c_L}^{c} \psi(y)\,\mathrm{d}y + (c-1)\,\psi(c) - \beta(c) - \int_{\dot{\eta}_2}^{1} \mu(c,\eta)\,\mathrm{d}\eta - c\int_{\dot{\eta}_1}^{1} \xi(c,\eta)\,\mathrm{d}\eta - \int_{c_L}^{c}\int_{\dot{\eta}_1}^{1} \xi(y,\eta)\,\mathrm{d}\eta\,\mathrm{d}y\right]\tau(c)\,\mathrm{d}c \\
=\ &v_E\,\Pi(c_H) + v_E\int_{c_L}^{c_H}\left[c + \frac{F(c)}{f(c)}\right]\tau(c)\,f(c)\,\mathrm{d}c + \int_{c_L}^{c_H}\int_{0}^{1} [\lambda_0\,[v(\eta) - v_E] + \Delta\,[v(\eta) + 1]]\,\hat{\alpha}_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c) \\
&+ \lambda_1\int_{c_L}^{c_H}\int_{0}^{1} [v(\eta) - v_E]\,\hat{\alpha}_3(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c).
\end{aligned}$$

Therefore, the objective function is bounded by $\int_{\mathcal{C}_1} \sigma(c)\,\mathrm{d}c$. Substituting our candidate solution into the objective function achieves the upper bound in the first two cases, which follows from:

1. if $\bar{c}(\grave{\eta}_1, \grave{\eta}_1) \le c_L$, we have $\hat{c} = c_L$ and the measure of $\mathcal{C}_1$ is 0. The regulator's expected payoff from our candidate policy and its upper bound are both 0;
2. if $c_L < \bar{c}(\grave{\eta}_1, \grave{\eta}_1) \le c_H$, we have

$$\begin{aligned}
&\int_{c_L}^{\bar{c}(\grave{\eta}_1,\grave{\eta}_1)} \sigma(c)\,\mathrm{d}c - \left[v_E + \int_{\grave{\eta}_1}^{1} \left[(\lambda_0+\lambda_1+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)\right] F(\bar{c}(\grave{\eta}_1,\grave{\eta}_1)) \\
&= \left\{\bar{c}(\grave{\eta}_1,\grave{\eta}_1)\left[v(\grave{\eta}_2) + \frac{\Delta}{\lambda_0}\left[v(\grave{\eta}_2)+1\right]\right] + (\lambda_0+\Delta)\int_{\grave{\eta}_2}^{1}\left[v(\eta)-v(\grave{\eta}_2)\right] g(\eta)\,\mathrm{d}\eta\right. \\
&\quad \left. -\iota(\grave{\eta}_1,\grave{\eta}_2) - \left[v_E + \int_{\grave{\eta}_1}^{1}\left[(\lambda_0+\lambda_1+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)\right]\right\} F(\bar{c}(\grave{\eta}_1,\grave{\eta}_1)) \\
&= 0.
\end{aligned}$$

In case (4)(b)(iii), we set $\mathcal{C}_1 = [c_L, c_H]$ and $\mathcal{C}_2 = \emptyset$, and replace the thresholds $\acute{\eta}_1$ and $\acute{\eta}_2$ with $\breve{\eta}_1$ and $\breve{\eta}_2$, respectively, where the two thresholds are determined by $\bar{c}(\breve{\eta}_1, \breve{\eta}_1) = c_H$ and $\Psi_2(\breve{\eta}_1, \breve{\eta}_2) = 0$. In this case, the nonnegativity of κ, χ, μ, ω, and ξ follows from their definitions directly. Given the monotonicity of the function Ξ (defined by (EC.18)), the desired nonnegativity of ψ follows from the following argument:

- If $h(\bar{\eta}) \le 0$ and $\check{c}_1 > \bar{c}(\breve{\eta}_1, \breve{\eta}_2)$, we have

$$\iota(\breve{\eta}_1, \breve{\eta}_2) = -\frac{\Xi(\breve{\eta}_2)}{\lambda_0} \le -\frac{\Xi(\breve{\eta}_2)}{\lambda_0} \le = 0.$$

  The inequalities above follow from the fact that $\breve{\eta}_2 \le \check{\eta}_2$ and $\Xi(\breve{\eta}_2) \ge 0$, as established in Lemma EC.3.

- If $h(\bar{\eta}) \le 0$ and $\check{c}_2 > \bar{c}(\mathring{\eta}_1, \mathring{\eta}_1)$, we have

$$\iota(\breve{\eta}_1, \breve{\eta}_2) = -\frac{\Xi(\breve{\eta}_2)}{\lambda_0} \le -\frac{\Xi(\mathring{\eta}_2)}{\lambda_0} = \iota(\mathring{\eta}_1, \mathring{\eta}_2) = 0.$$

  The inequalities above follow from the fact that $\breve{\eta}_2 \le \mathring{\eta}_2$.

- If $h(\bar{\eta}) > 0$, we have

$$\iota(\breve{\eta}_1, \breve{\eta}_2) = -\frac{\Xi(\breve{\eta}_2)}{\lambda_0} \le -\frac{\Xi(\eta^\dagger)}{\lambda_0} \le 0.$$

  The inequalities above follow from the fact that $\breve{\eta}_2 \le \eta^\dagger$ and $\Xi(\eta^\dagger) = \lambda_0\, h(\eta^\dagger) \ge \lambda_0\, h(\bar{\eta}) > 0$.

The nonnegativity of σ follows from the fact that for any $c \in [c_L, c_H]$, we have

$$\sigma(c) \ge \sigma(c_H) = t_2(\breve{\eta}_2)\, f(c_H) \ge t_2(\grave{\eta}_2)\, f(c_H) = 0,$$

where the function t_2 is defined in (EC.26), and the inequality follows from the monotonicity of t_2, as established in Lemma EC.6. Because all the multipliers remain nonnegative, the objective function is bounded by $\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c$, which is attained by our candidate solution, i.e.,

$$\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c - \left[v_E + \int_{\breve{\eta}_1}^{1}\left[(\lambda_0+\lambda_1+\Delta)\,v(\eta)+\Delta\right]\mathrm{d}G(\eta)\right]$$

$$
\begin{aligned}
&= \frac{\Delta}{\lambda_0}\left[v(\breve{\eta}_2)+1\right]\left[c_H-\left[1+\lambda_1\left[1-G(\breve{\eta}_1)\right]\right]\right]+\left[c_H-(\lambda_0+\lambda_1+\Delta)\left[1-G(\breve{\eta}_1)\right]-1\right]v(\breve{\eta}_2)\\
&\quad -\Delta\left[1-G(\breve{\eta}_1)\right]\\
&= \Delta\left[v(\breve{\eta}_2)+1\right]\left[1-G(\breve{\eta}_1)\right]-\Delta\left[1-G(\breve{\eta}_1)\right]v(\breve{\eta}_2)-\Delta\left[1-G(\breve{\eta}_1)\right]\\
&= 0,
\end{aligned}
$$

where the second equality follows from

$$1+(\lambda_0+\lambda_1)\left[1-G(\breve{\eta}_1)\right]=c_H.$$

When $h(\bar{\eta})\le 0$ and $\breve{c}_2>\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)$, we have

$$c_H\ge\breve{c}_2>\bar{c}(\mathring{\eta}_1,\mathring{\eta}_1)\ge\bar{c}(\eta^\dagger,\eta^\dagger).$$

Therefore, it remains to prove cases (4)(b)(iv) and (4)(b)(v) for either (1) $h(\bar{\eta})\le 0$ and $\breve{c}_1>\bar{c}(\breve{\eta}_1,\breve{\eta}_2)$, or (2) $h(\bar{\eta})>0$.

The proof of case (4)(b)(iv) closely follows that of case (4)(a)(iv). The nonnegativity of ψ follows as: if $h(\bar{\eta})\le 0$ and $\breve{c}_1>\bar{c}(\breve{\eta}_1,\breve{\eta}_2)$, we have $h(\breve{\eta}_2)\ge h(\eta^\dagger)\ge h(\breve{\eta}_2)=0$; otherwise, we have $h(\breve{\eta})\ge h(\bar{\eta})>0$. The nonnegativity of σ follows from the fact that for any $c\in[c_L,c_H]$, we have

$$\sigma(c)\ge\sigma(c_H)=t_1(\breve{\eta}_2)\,f(c_H)\ge t_1(\eta^\dagger)\,f(c_H)\ge 0.$$

Since all the multipliers are nonnegative, optimality is established.

The proof of case (4)(b)(v) is identical to that of case (4)(a)(iv).

This completes the proof. □

Proof of Corollary 3. When $\Delta=0$, we have

$$\Psi_1(\eta_1,\eta_2)=0 \text{ or } \Psi_2(\eta_1,\eta_2)=0 \quad\Rightarrow\quad \eta_1=\eta_2.$$

Therefore, the optimal solution to the upper-bound problem (48) is reduced to

$$\hat{\alpha}_2(c,\eta)=\begin{cases}1,\ \forall(c,\eta)\in[c_L,\hat{c})\times[\hat{\eta},1],\\ 0,\ \text{otherwise},\end{cases}\qquad \hat{\alpha}_3(c,\eta)=\begin{cases}\hat{\alpha},\ \forall(c,\eta)\in[c_L,\hat{c})\times[\hat{\eta},1],\\ 0,\ \text{otherwise}.\end{cases}$$

Given any feasible mechanism $\{\hat{\tau}(c),\hat{\alpha}_1(c),\hat{\alpha}_2(c,\eta),\hat{\alpha}_3(c,\eta)\}$, the following mechanism $\{\hat{\tau}(c),a_1(c),a_2(c,\eta)\}$ is feasible for the original problem (13) and yields the same objective function value, where a_1 and a_2 are constructed as follows:

$$a_1(c)=\begin{cases}\frac{\hat{\alpha}_1(c)}{\hat{\tau}(c)},\ \text{if } \hat{\tau}(c)>0,\\ 0,\quad\ \text{if } \hat{\tau}(c)=0,\end{cases}\qquad a_2(c,\eta)=\begin{cases}\frac{\hat{\alpha}_2(c,\eta)}{\hat{\tau}(c)},\ \text{if } \hat{\tau}(c)>0,\\ 0,\quad\ \text{if } \hat{\tau}(c)=0.\end{cases}$$

This completes the proof. □

Proof of Proposition 7. By construction, we have

$$\int_0^1 \hat{\tau}(c)\,[\lambda_0+\lambda_1\,a_1'(c)]\;a_2'(c,\eta)\,\mathrm{d}G(\eta)=\int_0^1 [\lambda_0\,\hat{\alpha}_2(c,\eta)+\lambda_1\,\hat{\alpha}_3(c,\eta)]\,\mathrm{d}G(\eta),\quad \forall c\in[c_L,c_H].$$

Hence, under the following mechanism

$$\mathcal{M}_1\equiv\left\{\hat{\tau}(c)=\mathbb{1}_{c<\hat{c}},\;a_1'(c)=\hat{\alpha}\cdot\mathbb{1}_{c<\hat{c}},\;a_2'(c,\eta)=\mathbb{1}_{(c,\eta)\in[c_L,\hat{c}]\times[\eta_1,1]}\right\},$$

constraints (IC) and (IR) hold automatically. The remaining constraints (2), (3), and (4) hold trivially by their definitions. When $\eta_1>\eta_2$, by definition of η_2, we have

$$\int_{\eta_2}^1 [(\lambda_0+\lambda_1\,\hat{\alpha}+\Delta)\,v(\eta)+\Delta]\,\mathrm{d}G(\eta)\geq\int_{\eta_1}^1 [(\lambda_0+\lambda_1\,\hat{\alpha}+\Delta)\,v(\eta)+\Delta]\,\mathrm{d}G(\eta).$$

Furthermore, the testing threshold satisfies

$$\hat{c}\leq\hat{\alpha}+(\lambda_0+\lambda_1\,\hat{\alpha})\,[1-G(\eta_1)]<\hat{\alpha}+(\lambda_0+\lambda_1\,\hat{\alpha})\,[1-G(\eta_2)]=c_2,$$

implying that the regulator is better off under mechanism

$$\mathcal{M}_2\equiv\left\{\tau''(c)=\mathbb{1}_{c\leq c_2},\;a_1''(c)=\hat{\alpha}\cdot\mathbb{1}_{c\leq c_2},\;a_2''(c,\eta)=\mathbb{1}_{(c,\eta)\in[c_L,c_2]\times[\eta_2,1]}\right\}.$$

It is straightforward to verify the feasibility of $\mathcal{M}_2$. Because U_R^* is the optimal value of the objective function, the feasibility property implies that

$$\tilde{U}_R^*\leq U_R^*.$$

This completes the proof. □

EC.4.5. Proofs for Section 5: either conditional approval or leniency unavailable

In this section, we provide the optimal mechanisms in closed form when either condition approval ($a_1(c)=0$ for any $c\in[c_L,c_H]$) or leniency ($a_2(c,\eta)=\mathbb{1}_{\eta\in[\bar{\eta}',1]}$ for any $(c,\eta)\in[c_L,c_H]\times[0,1]$) is unavailable, for completeness.

With a slight abuse of notation, in this section, we introduce a new decision variable

$$\alpha_2(c,\eta)\equiv\tau(c)\,a_2(c,\eta),\quad\forall(c,\eta)\in[c_L,c_H]\times[0,1].$$

Both parties' payoff functions boil down to

$$\hat{\Pi}(c,c')=-c\,\tau(c')+\lambda_0\int_0^1\alpha_2(c',\eta)\,\mathrm{d}G(\eta)\tag{EC.27}$$

and

$$U_R(\tau,\alpha_2)=\int_{c_L}^{c_H}\int_0^1[(\lambda_0+\Delta)\,v(\eta)+\Delta]\;\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c).\tag{EC.28}$$

We redefine the regulator's utility function as

$$\tilde{v}(\eta) \equiv (\lambda_0 + \Delta)\, v(\eta) + \Delta, \quad \forall \eta \in [0,1],$$

and denote $\tilde{v}(0)$ and $\mathbb{E}[\tilde{v}(\eta)]$ by $\tilde{v}_L$ and $\tilde{v}_E$, respectively. With a slight abuse of notation, in this section, we define

$$\bar{\eta} \equiv \inf\{\eta \in [0,1] : \tilde{v}(\eta) \geq 0\}.$$

The regulator's problem can be formulated as follows:

$$\max_{\{\tau,\alpha_2\}\in\Omega_1'} U_R(\tau, \alpha_2), \tag{EC.29}$$

where the feasible set Ω_1' is determined by (IC), (IR), (2), and

$$0 \leq \alpha_2(c,\eta) \leq \tau(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [0,1]. \tag{EC.30}$$

The optimal solution to this restricted problem remains a "take-it-or-leave-it" offer, as stated in the following proposition.

Proposition EC.2. *Consider the optimization problem* (EC.29) *under Assumption 1. There exists a cost threshold* $\hat{c}' \in [c_L, c_H]$ *and a final approval threshold* $\hat{\eta}' \in [0,1]$ *such that it is optimal to request testing and grant final approval as follows:*

$$\tau^*(c) = \begin{cases} 1, \forall c \in [c_L, \hat{c}'), \\ 0, \forall c \in [\hat{c}', c_H], \end{cases} \quad \alpha_2^*(c,\eta) = \begin{cases} 1, \forall (c,\eta) \in [c_L, \hat{c}') \times [\hat{\eta}', 1], \\ 0, \textit{ otherwise.} \end{cases}$$

The values of $\hat{c}'$ *and* $\hat{\eta}'$ *depend on model parameters in the following cases.*

(1) If $\tilde{v}_L \frac{F(\underline{c}(0))}{f(\underline{c}(0))} + \lambda_0 \tilde{v}_E > 0$, *we have* $\hat{c}' = c_H$ *and* $\hat{\eta}' = \min\{\underline{c}^{-1}(c_H), \bar{\eta}\}$.

(2) If $\tilde{v}_L \frac{F(\underline{c}(0))}{f(\underline{c}(0))} + \lambda_0 \tilde{v}_E \leq 0$, *there exists a unique* $\tilde{\eta} \in [0, \bar{\eta}]$ *such that*

$$\tilde{v}(\tilde{\eta}) \frac{F(\underline{c}(\tilde{\eta}))}{f(\underline{c}(\tilde{\eta}))} + \lambda_0 \int_{\tilde{\eta}}^1 \tilde{v}(y)\, g(y)\, \mathrm{d}y = 0.$$

(a) If $c_L > \underline{c}(\tilde{\eta})$, *we have* $\hat{c}' = c_L$ *and* $\hat{\eta}' = 1$.

(b) If $c_L \leq \underline{c}(\tilde{\eta}) \leq c_H$, *we have* $\hat{c}' = \underline{c}(\tilde{\eta})$ *and* $\hat{\eta}' = \tilde{\eta}$.

(c) If $\underline{c}(\bar{\eta}) < c_H < \underline{c}(\tilde{\eta})$, *we have* $\hat{c}' = c_H$ *and* $\hat{\eta}' = \underline{c}^{-1}(c_H)$.

(d) If $c_H \leq \underline{c}(\bar{\eta})$, *we have* $\hat{c}' = c_H$ *and* $\hat{\eta}' = \bar{\eta}$.

Proof of Proposition EC.2. It is straightforward to verify that our candidate solutions in all the cases are feasible.

When $c_H \leq \underline{c}(\bar{\eta})$, our candidate solution attains the upper bound of the objective function $\int_{\bar{\eta}}^1 \tilde{v}(\eta)\, \mathrm{d}G(\eta)$, with all constraints holding automatically. In the remainder of the proof, we focus on scenarios where $c_H > \underline{c}(\bar{\eta})$.

Invoking envelope theorem, the incentive compatibility constraint can be replaced by (EC.4) and

$$c\,\tau(c) - \lambda_0 \int_0^1 \alpha_2(c,\eta)\,\mathrm{d}G(\eta) + \Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y = 0, \quad \forall c \in [c_L, c_H]. \tag{EC.31}$$

When $c_H > \underline{c}(\bar{\eta})$, to prove the optimality in cases 1 and 2(3), define

$$\begin{aligned}
&\kappa = -\tilde{v}(\underline{c}^{-1}(c_H)),\\
&\nu(c,\eta) = \lambda_0 \left[\tilde{v}(\underline{c}^{-1}(c_H)) - \tilde{v}(\eta)\right] f(c)\,g(\eta), && \forall (c,\eta) \in [c_L, c_H] \times [0, \underline{c}^{-1}(c_H)],\\
&\mu(c,\eta) = \lambda_0 \left[\tilde{v}(\eta) - \tilde{v}(\underline{c}^{-1}(c_H))\right] f(c)\,g(\eta), && \forall (c,\eta) \in [c_L, c_H] \times [\underline{c}^{-1}(c_H), 1],\\
&\phi(c) = -\tilde{v}(\underline{c}^{-1}(c_H))\,f(c), && \forall c \in [c_L, c_H],\\
&\sigma(c) = \left[\tilde{v}(\underline{c}^{-1}(c_H)) \left[(c - c_H) + \frac{F(c)}{f(c)}\right] + \lambda_0 \int_{\underline{c}^{-1}(c_H)}^1 \tilde{v}(y)\,g(y)\,\mathrm{d}y\right] f(c), && \forall c \in [c_L, c_H].
\end{aligned}$$

The monotonicity of $\underline{c}$ implies that

$$c_H > \underline{c}(\bar{\eta}) \quad \Leftrightarrow \quad \underline{c}^{-1}(c_H) < \bar{\eta},$$

guaranteeing the nonnegativity of κ. The nonnegativity of ν and μ follows from their definitions directly. Because $\sigma(c)/f(c)$ is decreasing, it suffices to verify the nonnegativity of $\sigma(c_H)$. When $\tilde{v}_L \frac{F(\underline{c}(0))}{f(\underline{c}(0))} + \lambda_0 \tilde{v}_E > 0$, we have

$$\sigma(c_H) = \underbrace{\left[\tilde{v}(\underline{c}^{-1}(c_H)) \frac{1}{f(c_H)} + \lambda_0 \int_{\underline{c}^{-1}(c_H)}^1 \tilde{v}(y)\,g(y)\,\mathrm{d}y\right]}_{>0} f(c_H) \geq 0.$$

When $\tilde{v}_L \frac{F(\underline{c}(0))}{f(\underline{c}(0))} + \lambda_0 \tilde{v}_E \leq 0$ and $c_H < \underline{c}(\tilde{\eta})$, we have $\underline{c}^{-1}(c_H) > \tilde{\eta}$, and hence,

$$\begin{aligned}
\sigma(c_H) &= \left[\tilde{v}(\underline{c}^{-1}(c_H)) \frac{F(c_H)}{f(c_H)} + \lambda_0 \int_{\underline{c}^{-1}(c_H)}^1 \tilde{v}(y)\,g(y)\,\mathrm{d}y\right] f(c_H)\\
&\geq \left[\tilde{v}(\tilde{\eta}) \frac{F(\underline{c}(\tilde{\eta}))}{f(\underline{c}(\tilde{\eta}))} + \lambda_0 \int_{\tilde{\eta}}^1 \tilde{v}(y)\,g(y)\,\mathrm{d}y\right] f(c_H)\\
&= 0,
\end{aligned}$$

where the inequality follows from the monotonicity of function

$$\Upsilon(\eta) \equiv \tilde{v}(\eta) \frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))} + \lambda_0 \int_\eta^1 \tilde{v}(y)\,g(y)\,\mathrm{d}y,$$

i.e.,

$$\Upsilon'(\eta) = \tilde{v}'(\eta) \frac{F(\underline{c}(\eta))}{f(\underline{c}(\eta))} - \lambda_0\,\tilde{v}(\eta)\,g(\eta) \underbrace{\frac{\mathrm{d}}{\mathrm{d}c}\left[c + \frac{F(c)}{f(c)}\right]}_{\geq 0 \text{ by Assumption 1}}\Bigg|_{c=\underline{c}(\eta)} \geq 0.$$

Because

$$\begin{aligned}
&\kappa\,[-\Pi(c_H)] + \int_{c_L}^{c_H} \phi(c)\,\left[c\,\tau(c) - \lambda_0 \int_0^1 \alpha_2(c,\eta)\,\mathrm{d}G(\eta) + \Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y\right]\mathrm{d}c \\
&\quad + \int_{c_L}^{c_H} \sigma(c)\,\tau(c)\,\mathrm{d}c - \int_{c_L}^{c_H} \int_0^{\underline{c}^{-1}(c_H)} \nu(c,\eta)\,\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c + \int_{c_L}^{c_H} \int_{\underline{c}^{-1}(c_H)}^1 \mu(c,\eta)\,[\alpha_2(c,\eta) - \tau(c)]\,\mathrm{d}\eta\,\mathrm{d}c \\
=\ &\left[-\kappa + \int_{c_L}^{c_H} \phi(c)\,\mathrm{d}c\right]\,\Pi(c_H) + \int_{c_L}^{c_H} \int_0^{\underline{c}^{-1}(c_H)} [-\lambda_0\,\phi(c)\,g(\eta) - \nu(c,\eta)]\,\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&\quad + \int_{c_L}^{c_H} \int_{\underline{c}^{-1}(c_H)}^1 [-\lambda_0\,\phi(c)\,g(\eta) + \mu(c,\eta)]\,\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c \\
&\quad + \int_{c_L}^{c_H} \left[\sigma(c) + c\,\phi(c) + \int_{c_L}^{c} \phi(y)\,\mathrm{d}y - \int_0^1 \mu(c,\eta)\,\mathrm{d}\eta\right]\,\tau(c)\,\mathrm{d}c \\
=\ &\lambda_0 \int_{c_L}^{c_H} \int_0^1 \tilde{v}(\eta)\,\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c),
\end{aligned}$$

the objective function is bounded by $\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c$. Substituting the decision variables with our candidate solution attains this upper bound, which follows from

$$\begin{aligned}
&\int_{c_L}^{c_H} \sigma(c)\,\mathrm{d}c - \lambda_0 \int_{\underline{c}^{-1}(c_H)}^1 \tilde{v}(\eta)\,\mathrm{d}G(\eta) \\
=\ &\int_{c_L}^{c_H} \left[\tilde{v}(\underline{c}^{-1}(c_H))\,\left[(c - c_H) + \frac{F(c)}{f(c)}\right] + \lambda_0 \int_{\underline{c}^{-1}(c_H)}^1 \tilde{v}(y)\,g(y)\,\mathrm{d}y\right]\,f(c)\,\mathrm{d}c - \lambda_0 \int_{\underline{c}^{-1}(c_H)}^1 \tilde{v}(\eta)\,\mathrm{d}G(\eta) \\
=\ &\tilde{v}(\underline{c}^{-1}(c_H)) \int_{c_L}^{c_H} [(c - c_H)\,f(c) + F(c)]\,\mathrm{d}c \\
=\ &0.
\end{aligned}$$

To prove the optimality in the remaining two subcases ((2)(a) and (2)(b)), define

$$\begin{aligned}
&\kappa = -\tilde{v}(\tilde{\eta}), \\
&\nu(c,\eta) = \lambda_0\,[\tilde{v}(\tilde{\eta}) - \tilde{v}(\eta)]\,f(c)\,g(\eta), && \forall (c,\eta) \in [c_L, c_H] \times [0, \tilde{\eta}], \\
&\mu(c,\eta) = \lambda_0\,[\tilde{v}(\eta) - \tilde{v}(\tilde{\eta})]\,f(c)\,g(\eta), && \forall (c,\eta) \in [c_L, c_H] \times [\tilde{\eta}, 1], \\
&\phi(c) = -\tilde{v}(\tilde{\eta})\,f(c), && \forall c \in [c_L, c_H], \\
&\sigma(c) = \left[\tilde{v}(\tilde{\eta})\,\left[[c - \underline{c}(\tilde{\eta})] + \frac{F(c)}{f(c)}\right] + \lambda_0 \int_{\tilde{\eta}}^1 \tilde{v}(y)\,g(y)\,\mathrm{d}y\right]\,f(c), && \forall c \in \mathcal{C}_1, \\
&\beta(c) = -\left[\tilde{v}(\tilde{\eta})\,\left[[c - \underline{c}(\tilde{\eta})] + \frac{F(c)}{f(c)}\right] + \lambda_0 \int_{\tilde{\eta}}^1 \tilde{v}(y)\,g(y)\,\mathrm{d}y\right]\,f(c), && \forall c \in \mathcal{C}_2.
\end{aligned}$$

When $c_L > \underline{c}(\tilde{\eta})$, we know that β is nonnegative on $[c_L, c_H]$. In this case, we have $\mathcal{C}_1 = \emptyset$ and $\mathcal{C}_2 = [c_L, c_H]$. When $c_L \le \underline{c}(\tilde{\eta}) \le c_H$, we know that σ and β are nonnegative on $[c_L, \underline{c}(\tilde{\eta})]$ and $[\underline{c}(\tilde{\eta}), c_H]$, respectively. In this case, we have $\mathcal{C}_1 = [c_L, \underline{c}(\tilde{\eta})]$ and $\mathcal{C}_2 = [\underline{c}(\tilde{\eta}), c_H]$.This guarantees the nonnegativity of all the multipliers (except ϕ, which corresponds to an equality constraint).

Because

$$\begin{aligned}
&\kappa\left[-\Pi(c_H)\right]+\int_{c_L}^{c_H}\phi(c)\left[c\,\tau(c)-\lambda_0\int_0^1\alpha_2(c,\eta)\,\mathrm{d}G(\eta)+\Pi(c_H)+\int_c^{c_H}\tau(y)\,\mathrm{d}y\right]\mathrm{d}c+\int_{\mathcal{C}_1}\sigma(c)\,\tau(c)\,\mathrm{d}c\\
&-\int_{\mathcal{C}_2}\beta(c)\,\tau(c)\,\mathrm{d}c-\int_{c_L}^{c_H}\int_0^{\tilde\eta}\nu(c,\eta)\,\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c+\int_{c_L}^{c_H}\int_{\tilde\eta}^1\mu(c,\eta)\,\left[\alpha_2(c,\eta)-\tau(c)\right]\mathrm{d}\eta\,\mathrm{d}c\\
=&\left[-\kappa+\int_{c_L}^{c_H}\phi(c)\,\mathrm{d}c\right]\Pi(c_H)+\int_{c_L}^{c_H}\int_0^{\tilde\eta}\left[-\lambda_0\,\phi(c)\,g(\eta)-\nu(c,\eta)\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{c_L}^{c_H}\int_{\tilde\eta}^1\left[-\lambda_0\,\phi(c)\,g(\eta)+\mu(c,\eta)\right]\alpha_2(c,\eta)\,\mathrm{d}\eta\,\mathrm{d}c\\
&+\int_{\mathcal{C}_1}\left[\sigma(c)+c\,\phi(c)+\int_{c_L}^{c}\phi(y)\,\mathrm{d}y-\int_0^1\mu(c,\eta)\,\mathrm{d}\eta\right]\tau(c)\,\mathrm{d}c\\
&+\int_{\mathcal{C}_2}\left[-\beta(c)+c\,\phi(c)+\int_{c_L}^{c}\phi(y)\,\mathrm{d}y-\int_0^1\mu(c,\eta)\,\mathrm{d}\eta\right]\tau(c)\,\mathrm{d}c\\
=&\,\lambda_0\int_{c_L}^{c_H}\int_0^1\tilde v(\eta)\,\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\,\mathrm{d}F(c),
\end{aligned}$$

the objective function is bounded by $\int_{\mathcal{C}_1}\sigma(c)\,\mathrm{d}c$.

1. In case (2)(a), the regulator's payoff generated by our candidate solution and its upper bound are both 0.
2. In case (2)(b), we have

$$\begin{aligned}
&\int_{c_L}^{\underline{c}(\tilde\eta)}\sigma(c)\,\mathrm{d}c-\lambda_0\,F(\underline{c}(\tilde\eta))\int_{\tilde\eta}^1\tilde v(\eta)\,\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\\
=&\int_{c_L}^{\underline{c}(\tilde\eta)}\left[\tilde v(\tilde\eta)\,\left[[c-\underline{c}(\tilde\eta)]+\frac{F(c)}{f(c)}\right]+\lambda_0\int_{\tilde\eta}^1\tilde v(y)\,g(y)\,\mathrm{d}y\right]f(c)\,\mathrm{d}c\\
&-\lambda_0\,F(\underline{c}(\tilde\eta))\int_{\tilde\eta}^1\tilde v(\eta)\,\alpha_2(c,\eta)\,\mathrm{d}G(\eta)\\
=&\,\tilde v(\tilde\eta)\int_{c_L}^{\underline{c}(\tilde\eta)}\left[[c-\underline{c}(\tilde\eta)]\,f(c)+F(c)\right]\mathrm{d}c\\
=&\,0,
\end{aligned}$$

   establishing the optimality.

This completes the proof. □

When leniency is unavailable, in which case we assume that the regulator will adopt the myopic threshold $\bar\eta'$ (defined in (55)), both parties' payoff functions are reduced to

$$\hat\Pi(c,c')=\alpha_1(c')-c\,\tau(c')+\left[\lambda_0\,\tau(c')+\lambda_1\,\alpha_1(c')\right]\bar G(\bar\eta')$$

and

$$U_R=\int_{c_L}^{c_H}\left[v_E\,\alpha_1(c)+\left[\lambda_0\,\tau(c)+\lambda_1\,\alpha_1(c)\right]\underbrace{\int_{\bar\eta'}^1 v(\eta)\,\mathrm{d}G(\eta)}_{\equiv z_1}+\tau(c)\,\underbrace{\Delta\int_{\bar\eta'}^1\left[v(\eta)+1\right]\mathrm{d}G(\eta)}_{\equiv z_2}\right]\mathrm{d}F(c),$$

where $\bar{G}$ denotes the complementary cumulative distribution function of G. By envelope theorem, we have

$$\alpha_1(c) = \frac{1}{1+\lambda_1 \bar{G}(\bar{\eta}')} \left[\Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y + (c - \lambda_0 \bar{G}(\bar{\eta}'))\,\tau(c) \right].$$

After eliminating α_1, the optimization problem can be reformulated as follows:

$$\max_{\{\tau,\alpha_1\}\in\Omega_2'} \frac{v_E + \lambda_1 z_1}{1+\lambda_1 \bar{G}(\bar{\eta}')}\,\Pi(c_H) + \int_{c_L}^{c_H} w(c)\,\tau(c)\,\mathrm{d}F(c), \tag{EC.32}$$

where the virtual valuation function w is given by

$$w(c) = \frac{v_E + \lambda_1 z_1}{1+\lambda_1 \bar{G}(\bar{\eta}')} \left[c + \frac{F(c)}{f(c)} - \lambda_0 \bar{G}(\bar{\eta}') \right] + \lambda_0 z_1 + z_2,$$

and the feasible set Ω_2' is determined by (2), (EC.4), (EC.6) and

$$0 \le \frac{1}{1+\lambda_1 \bar{G}(\bar{\eta}')} \left[\Pi(c_H) + \int_c^{c_H} \tau(y)\,\mathrm{d}y + (c - \lambda_0 \bar{G}(\bar{\eta}'))\,\tau(c) \right] \le \tau(c), \quad \forall (c,\eta) \in [c_L, c_H] \times [0,1]. \tag{EC.33}$$

Assumption 1 guarantees that the virtual valuation $w(c)$ is non-increasing in c when $v_E + \lambda_1 z_1$ is negative, and non-decreasing otherwise. Moreover, w changes its sign at most once as c increases in $[c_L, c_H]$. Therefore, when $v_E + \lambda_1 z_1 < 0$, w is nonnegative if $c < c_0$, and non-positive if $c_0 < c$, where the threshold c_0 is defined by

$$c_0 \equiv \sup\{c \in [c_L, c_H] : w(c) \ge 0\}. \tag{EC.34}$$

To characterize the optimal mechanism for the regulator's problem without leniency (EC.32), we make the following assumption.

Assumption EC.1. *The density function $f(c)$ is differentiable and non-increasing.*

Assumption EC.1 implies that the virtual valuation function $c + F(c)/f(c)$ is increasing.

The optimal mechanism is summarized in the following proposition.

Proposition EC.3. *Under Assumption EC.1, the optimal solution α_1^* and τ^* for the optimization problem* (EC.32) *is as follows:*

(1) if $v_E + \lambda_1 z_1 \ge 0$, then $\alpha_1^*(c) = 1$ and $\tau^*(c) = 1$ for any $c \in [c_L, c_H]$;

(2) if $v_E + \lambda_1 z_1 < 0$ and $c_H \le \lambda_0 \bar{G}(\bar{\eta}')$, then $\alpha_1^*(c) = 0$ and $\tau^*(c) = 1$ for any $c \in [c_L, c_H]$;

(3) if $v_E + \lambda_1 z_1 < 0$, $c_H > \lambda_0 \bar{G}(\bar{\eta}')$, and $c_0 \ge \lambda_0 \bar{G}(\bar{\eta}')$, then

$$\alpha_1^*(c) = \begin{cases} \dfrac{c_0 - \lambda_0 \bar{G}(\bar{\eta}')}{1+\lambda_1 \bar{G}(\bar{\eta}')}, & \forall c \in [c_L, c_0], \\ 0, & \forall c \in (c_0, c_H], \end{cases} \quad \tau^*(c) = \begin{cases} 1, & \forall c \in [c_L, c_0], \\ 0, & \forall c \in (c_0, c_H]; \end{cases}$$

(4) if $v_E + \lambda_1 z_1 < 0$, $c_H > \lambda_0 \bar{G}(\bar{\eta}')$, and $c_0 < \bar{G}(\bar{\eta}')$, then

$$\alpha_1^*(c) = 0, \ \forall c \in [c_L, c_H], \quad \tau^*(c) = \begin{cases} 1, & \forall c \in [c_L, \lambda_0 \bar{G}(\bar{\eta}')], \\ 0, & \forall c \in (\lambda_0 \bar{G}(\bar{\eta}'), c_H], \end{cases}$$

where c_0 is defined by (EC.34).

Proof of Proposition EC.3. In case (1), the coefficient of α_1 in the integrand of the objective function is nonnegative, i.e., $v_E + \lambda_1 z_1 \geq 0$. Therefore, it is optimal to set $\alpha_1^*(c) = 1$ for any $c \in [c_L, c_H]$. In this case, the regulator can induce firms of all types to participate, and hence, this policy is optimal.

In case (2), the coefficient of α_1 in the integrand of the objective function is negative, i.e., $v_E + \lambda_1 z_1 < 0$. Therefore, ignoring all constraints, it is optimal to set $\alpha_1^*(c) = 0$ for any $c \in [c_L, c_H]$. In this case, the regulator can still induce firms of all types to participate ($c_H \leq \lambda_0 \bar{G}(\bar{\eta}')$), and hence, this policy is optimal.

In case (3), we first ignore constraints (EC.4) and (EC.33), and solving the relaxed problem point-wise yields

$$\tau(c) = \begin{cases} 1, \, \forall c \in [c_L, c_0], \\ 0, \, \forall c \in (c_0, c_H], \end{cases}$$

satisfying (EC.4) automatically. Plugging the expression of τ into (EC.5) yields

$$\alpha_1(c) = \begin{cases} \dfrac{c_0 - \lambda_0 \bar{G}(\bar{\eta}')}{1 + \lambda_1 \bar{G}(\bar{\eta}')}, & \forall c \in [c_L, c_0], \\ 0, & \forall c \in (c_0, c_H]. \end{cases}$$

When $c_0 \geq \lambda_0 \bar{G}(\bar{\eta}')$, the feasibility constraint (EC.33) holds automatically, and thus the solution above is feasible.

In case (4), it is straightforward to verify that our candidate solution is feasible. When $c_L \leq c_0 < \lambda_0 \bar{G}(\bar{\eta}')$, define

$$\begin{aligned}
\phi &= -\frac{1}{c_2 - c_1} \int_{c_L}^{c_1} w(y) \, f(y) \, \mathrm{d}y - \frac{v_E + \lambda_1 z_1}{1 + \lambda_1 \bar{G}(\bar{\eta}')}, \\
\nu(c) &= \frac{1}{(c_2 - c)^2} \left[\int_{c_L}^{c} w(y) \, f(y) \, \mathrm{d}y + (c_2 - c) \, w(c) \, f(c) \right], && \forall c \in [c_L, c_1], \\
\sigma(c) &= \frac{1}{c_2 - c_1} \int_{c_L}^{c_1} w(y) \, f(y) \, \mathrm{d}y + w(c) \, f(c), && \forall c \in [c_1, c_2], \text{ and} \\
\mu(c) &= -\frac{1}{c_2 - c_1} \int_{c_L}^{c_1} w(y) \, f(y) \, \mathrm{d}y - w(c) \, f(c), && \forall c \in [c_2, c_H],
\end{aligned}$$

where $c_2 = \lambda_0 \bar{G}(\bar{\eta}')$ and c_1 is determined by

$$\int_{c_L}^{c_1} w(y) \, f(y) \, \mathrm{d}y + (c_2 - c_1) \, w(c_2) \, f(c_2) = 0.$$

We first verify the monotonicity of $w(c) \, f(c)$ on $[c_L, c_2]$. By definition, we have

$$w(c) \, f(c) = \frac{1}{1 + \lambda_1 \bar{G}(\bar{\eta}')} \left[\left[(\lambda_0 z_1 + z_2)(1 + \lambda_1 \bar{G}(\bar{\eta}')) + \underbrace{(v_E + \lambda_1 z_1)(c - \lambda_0 \bar{G}(\bar{\eta}'))}_{\text{nonnegative and decreasing in } c} \right] f(c) + \underbrace{(v_E + \lambda_1 z_1)}_{<0} F(c) \right].$$

Under Assumption EC.1, we know that $w(c)\,f(c)$ is decreasing in c.

Then we prove the unique existence of c_1 on $[c_L, c_2]$. Let

$$\psi(c) \equiv \int_{c_L}^{c} w(y)\,f(y)\,\mathrm{d}y + (c_2 - c)\,w(c_2)\,f(c_2), \quad \forall c \in [c_L, c_2].$$

Because

$$\begin{aligned}
&\psi'(c) = w(c)\,f(c) - w(c_2)\,f(c_2) \geq 0,\\
&\psi(c_L) = (c_2 - c_L)\,w(c_2)\,f(c_2) < 0, \text{ and}\\
&\psi(c_2) = \int_{c_L}^{c_2} w(y)\,f(y)\,\mathrm{d}y > 0,
\end{aligned}$$

the unique existence of c_1 is guaranteed. The first inequality follows from the monotonicity of $w(c)\,f(c)$, and the last inequality follows from the fact that under the mechanism granting no conditional approval, the regulator's expected payoff is strictly positive.

Next, we will prove that ϕ, ν, σ, and μ are all nonnegative. By definition of c_1, we have

$$\phi = (\lambda_0\,z_1 + z_2)\,f(c_2) - \frac{\overbrace{v_E + \lambda_1\,z_1}^{<0}}{1 + \lambda_1\,\bar{G}(\bar{\eta}')}\,[1 - F(c_2)] \geq 0.$$

The nonnegativity of ν boils down to

$$h(c) \equiv \int_{c_L}^{c} w(y)\,f(y)\,\mathrm{d}y + (c_2 - c)\,w(c)\,f(c) \geq 0, \quad \forall c \in [c_L, c_1].$$

Taking derivative yields

$$h'(c) = (c_2 - c)\,\frac{\mathrm{d}}{\mathrm{d}c}\,(w(c)\,f(c)) \leq 0, \quad \forall c \in [c_L, c_1],$$

where the inequality follows from the monotonicity of $w(c)\,f(c)$. Therefore, the desired property is reduced to

$$h(c_1) = \int_{c_L}^{c_1} w(y)\,f(y)\,\mathrm{d}y + (c_2 - c_1)\,w(c_1)\,f(c_1) \geq 0,$$

which follows from

$$\int_{c_L}^{c_1} w(y)\,f(y)\,\mathrm{d}y + (c_2 - c_1)\,w(c_1)\,f(c_1) \geq \int_{c_L}^{c_1} w(y)\,f(y)\,\mathrm{d}y + (c_2 - c_1)\,w(c_2)\,f(c_2) = 0.$$

For any $c \in [c_1, c_2]$ and $c \in [c_2, c_H]$, we have

$$\sigma'(c) = \frac{\mathrm{d}}{\mathrm{d}c}\,(w(c)\,f(c)) \leq 0$$

and

$$\mu(c) \geq \mu(c_2),$$

respectively. The first and second inequalities follow from the monotonicity of $w(c)\,f(c)$ on $c \in [c_L, c_2]$ and

$$\begin{aligned}
\mu(c) \geq \mu(c_2) \;&\Leftrightarrow\; w(c)\,f(c) \leq w(c_2)\,f(c_2) \\
&\Leftrightarrow\; \left[(\lambda_0\,z_1 + z_2)\,(1+\lambda_1\,\bar{G}(\bar{\eta}')) + (v_E + \lambda_1\,z_1)\,(c - \lambda_0\,\bar{G}(\bar{\eta}'))\right]\,f(c) + (v_E+\lambda_1\,z_1)\,F(c) \\
&\leq \left[(\lambda_0\,z_1 + z_2)\,(1+\lambda_1\,\bar{G}(\bar{\eta}')) + (v_E + \lambda_1\,z_1)\,(c_2 - \lambda_0\,\bar{G}(\bar{\eta}'))\right]\,f(c_2) + (v_E+\lambda_1\,z_1)\,F(c_2)
\end{aligned}$$

for any $c \in [c_2, c_H]$, where the last step follows from

$$v_E + \lambda_1\,z_1 < 0, \quad f(c) \leq f(c_2), \quad \text{and} \quad c - \lambda_0\,\bar{G}(\bar{\eta}') > 0 = c_2 - \lambda_0\,\bar{G}(\bar{\eta}').$$

The desired nonnegativity conditions are automatically implied by $\sigma(c_2) = \mu(c_2) = 0$.

Because

$$\int_{c_L}^{c} \nu(y) = \frac{1}{c_2 - c}\int_{c_L}^{c} w(y)\,f(y)\,\mathrm{d}y, \quad \forall c \in [c_L, c_1],$$

we have

$$\begin{aligned}
&-\phi - \int_{c_L}^{c_1} \nu(y)\,\mathrm{d}y = \frac{v_E + \lambda_1\,z_1}{1+\lambda_1\,\bar{G}(\bar{\eta}')}, \\
&(c_2 - c)\,\nu(c) - \int_{c_L}^{c} \nu(y)\,\mathrm{d}y = w(c)\,f(c), && \forall c \in [c_L, c_1], \\
&\sigma(c) - \int_{c_L}^{c_1} \nu(y)\,\mathrm{d}y = w(c)\,f(c), && \forall c \in [c_1, c_2], \text{ and} \\
&-\mu(c) - \int_{c_L}^{c_1} \nu(y)\,\mathrm{d}y = w(c)\,f(c), && \forall c \in [c_2, c_H].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\phi\,[-\Pi(c_H)] - \int_{c_L}^{c_1}\left[\Pi(c_H) + \int_{c}^{c_H} \tau(y)\,\mathrm{d}y + (c - \lambda_0\,\bar{G}(\bar{\eta}'))\,\tau(c)\right]\,\nu(c)\,\mathrm{d}c \\
&+ \int_{c_1}^{c_2} \tau(c)\,\sigma(c)\,\mathrm{d}c - \int_{c_2}^{c_H} \tau(c)\,\mu(c)\,\mathrm{d}c \\
=&\left[-\phi - \int_{c_L}^{c_1} \nu(c)\,\mathrm{d}c\right]\,\Pi(c_H) + \int_{c_L}^{c_1}\left[(c_2 - c)\,\nu(c) - \int_{c_L}^{c} \nu(y)\,\mathrm{d}y\right]\,\tau(c)\,\mathrm{d}c \\
&+ \int_{c_1}^{c_2}\left[\sigma(c) - \int_{c_L}^{c_1} \nu(y)\,\mathrm{d}y\right]\,\tau(c)\,\mathrm{d}c + \int_{c_2}^{c_H}\left[-\mu(c) - \int_{c_L}^{c_1} \nu(y)\,\mathrm{d}y\right]\,\tau(c)\,\mathrm{d}c \\
=&\frac{v_E + \lambda_1\,z_1}{1+\lambda_1\,\bar{G}(\bar{\eta}')}\,\Pi(c_H) + \int_{c_L}^{c_H} w(c)\,\tau(c)\;\mathrm{d}F(c).
\end{aligned}$$

Therefore, the objective function is bounded by $\int_{c_1}^{c_2} \sigma(c)\,\mathrm{d}c$. Substituting our candidate primal solution into the objective function yields $\int_{c_L}^{c_2} w(y)\,f(y)\,\mathrm{d}y$, which is the same value as the upper bound, i.e.,

$$\int_{c_1}^{c_2} \sigma(y)\,\mathrm{d}y = \int_{c_1}^{c_2}\left[\frac{1}{c_2 - c_1}\int_{c_L}^{c_1} w(z)\,f(z)\,\mathrm{d}z + w(y)\,f(y)\right]\mathrm{d}y$$

$$
\begin{aligned}
&= \int_{c_L}^{c_1} w(y)\, f(y)\, \mathrm{d}y + \int_{c_1}^{c_2} w(y)\, f(y)\, \mathrm{d}y \\
&= \int_{c_L}^{c_2} w(y)\, f(y)\, \mathrm{d}y.
\end{aligned}
$$

This establishes the optimality of our candidate solution.

When $c_0 = -\infty$, the virtual valuation function w is always non-positive on $[c_L, c_H]$. Moreover, the coefficient of $\Pi(c_H)$ is also negative; hence, it is evident that the optimal mechanism is to not request testing and to grant no conditional approval.

This completes the proof. □