The ESG Rating Game: Deviation, Disagreement, and Greenwashing*

Qinhan Duan† Yilun Song‡ Ruodu Wang§ Jingping Yang¶ and Ruixun Zhang‖

March 17, 2026

Abstract

Environmental, social, and governance (ESG) ratings are pivotal in sustainable investing, informing portfolio construction and regulatory compliance. We develop a game-theoretic principal–agent model involving ESG-conscious investors and competing rating agencies. Investors impose ESG-related portfolio constraints by selecting ratings from different agencies, while agencies strategically issue ESG ratings based on a trade-off between financial incentives related to market share and reputational risks from deviating from ESG fundamentals. The interaction forms an extensive-form game with imperfect information, and we characterize its Stackelberg stable equilibrium. Our analysis shows that 1) agencies may strategically deviate ESG ratings from fundamentals under certain conditions, and such deviations tend to align ratings with asset returns when investors have strong green preferences; 2) asymmetric aversion for deviation generate rating disagreement between multiple agencies even when they have access to the same ESG fundamentals, potentially allowing one agency to monopolize green investors; 3) the magnitude of rating deviation governs a trade-off between investors' pecuniary utility and portfolio's fundamental greenness; and 4) firm-level greenwashing exacerbates rating deviation and further distorts the trade-off between utility and greenness. These results rationalize several empirical findings in the ESG literature and provide insights into designing mechanisms that encourage more informative ESG ratings.

Keywords: Environmental, Social, and Governance (ESG); Rating Deviation; Greenwashing; Principal-Agent Framework; Imperfect Information; Stackelberg Stable Equilibrium.

*Ruixun Zhang acknowledges research funding from the National Natural Science Foundation of China (12271013,72342004) and the National Key R&D Program of China (2022YFA1007900). Jingping Yang acknowledges research funding from the National Natural Science Foundation of China (12471445,12071016). Ruodu Wang acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2024-03728) and Canada Research Chairs (CRC-2022-00141).

†Department of Financial Mathematics, Peking University, China. 2301110079@pku.edu.cn

‡School of Insurance, University of International Business and Economics, China. 2193167815@qq.com

§Department of Statistics and Actuarial Science, University of Waterloo, Canada. wang@uwaterloo.ca

¶Laboratory for Mathematical Economics and Quantitative Finance, and Department of Financial Mathematics, Peking University, China. yangjp@math.pku.edu.cn

‖School of Mathematical Sciences, Center for Statistical Science, Laboratory for Mathematical Economics and Quantitative Finance, and National Engineering Laboratory for Big Data Analysis and Applications, Peking University, China. zhangruixun@pku.edu.cn. Corresponding author.

Contents

1 Introduction

Background and Motivation. The surge of sustainable investing has made environmental, social, and governance (ESG) considerations a central theme in asset management (Hartzmark and Sussman, 2019; Starks, 2023).[1] ESG investing aims to achieve short-term competitive financial returns while managing long-term environmental and social risks (Krueger et al., 2020). Recent theoretical and empirical studies highlight the growing complexity of integrating ESG factors into financial decision-making (Pástor et al., 2021, 2022; Pedersen et al., 2021; Berg et al., 2022; Zerbib, 2022; Lo and Zhang, 2024, 2025).

ESG rating agencies play a crucial role in the sustainable investing ecosystem by providing measures that investors use to screen assets, construct portfolios, and meet regulatory requirements (Berg et al., 2021, 2022; De Angelis et al., 2023). In addition, similar to the credit rating industry, ESG rating agencies compete to gain market share as they serve as information intermediaries in financial markets (Bolton et al., 2012; Christensen et al., 2022).

Despite its importance, delivering reliable and accurate ESG ratings faces several fundamental challenges. First, ESG ratings do not have a widely accepted ground truth, making their validation challenging and leaving rating agencies with considerable discretion in their behavior. In contrast to credit ratings, which can be assessed ex post through realized default events (Piccolo and Shapiro, 2022), ESG ratings contain more subjective interpretations by design.

Second, ESG ratings show substantial *disagreement* across different rating agencies. Both Chatterji et al. (2016) and Berg et al. (2022) find low correlations in the ESG ratings, even after adjusting for definitional differences, a phenomenon they call rating disagreement or divergence. Moreover, Christensen et al. (2022) show that greater ESG disclosure may paradoxically amplify rather than mitigate rating disagreement, reinforcing the challenge of achieving rating consistency.

Third, rating agencies' financial objectives and incentives linked to index licensing may lead to revealed ESG ratings that *deviate* from the most truthful assessment of a firm's sustainability performance—sometimes referred to as the ESG fundamental (Agrawal et al., 2025). This can lead to ESG ratings that tend to align with stock returns and negatively affect the informativeness of

[1] Sustainable investment approaches reached US$16.7 trillion in 2024, representing about 27% of the total fund market and increasing by nearly US$5.5 trillion over the previous two years (Global Sustainable Investment Alliance, 2025).

their assessments (Larcker et al., 2022). For example, Berg et al. (2021) document that revisions in historical ESG scores may materially alter firm classifications and empirical conclusions about ESG performance.

Finally, corporate greenwashing further exacerbates the above difficulties (Cartellier et al., 2025). Kim et al. (2012) provide empirical evidence that firms with higher corporate social responsibility scores may engage in greater earnings management, suggesting that firms selectively disclose favorable sustainability activities while masking unfavorable aspects. Such selective disclosure increases the difficulty for ESG rating agencies in delivering consistent and accurate assessments, thereby complicating efforts to promote sustainable investment through reliable ratings.

Our Work and Contribution. In this paper, we develop a game-theoretic principal–agent model involving ESG-conscious investors and competing rating agencies to characterize the emergence of ESG rating deviation and disagreement, as well as the impact of these phenomena and greenwashing on investor utility and the portfolio's fundamental greenness.

We consider a world with exogenously determined asset prices. In our model, ESG-conscious investors purchase ESG ratings from competing ESG rating agencies to evaluate firms' sustainability profiles, consistent with the prevalence of investor-pays arrangements in the ESG rating market (Lovo and Olivier, 2025). Investors impose ESG-related portfolio constraints, but cannot observe the ESG fundamentals that are discovered by and accessible only to rating agencies. Rating agencies strategically issue ESG ratings based on a trade-off between financial incentives related to market share and reputational risks from deviating from the ESG fundamentals. Motivated by empirical evidence that ESG score updates are discrete and infrequent (Avramov et al., 2022), we model agencies' rating decisions as a simultaneous-move extensive-form game with imperfect information (Osborne, 2004). Because investors allocate capital toward the most favorable ratings, agencies' objective functions contain discontinuities, in which case the standard Nash equilibrium fails to exist. Moreover, when ESG ratings are publicly disclosed, they cannot be easily revised, introducing a commitment effect. This effect makes each agency's strategic choice forward-looking rather than purely simultaneous. To incorporate this economic feature and to guarantee equilibrium existence, we adopt the Stackelberg stable equilibrium (SSE) to capture strategic anticipations and ensure the existence of equilibrium (Mailath and Samuelson, 2006; Heller and Winter, 2016).

Our contributions to the literature on ESG ratings and sustainable investing are multi-fold. First, we characterize the emergence of rating deviation from ESG fundamentals under competition. Deviation arises when an agency's financial incentives exceed its risks from deviation, with the less deviation-averse agency issuing ratings farther away from the ESG fundamental. Moreover, agencies deviate in a way that induces a *positive* correlation between asset returns and issued ESG scores, when investors have a strong green preference. These results explain the empirical findings that raters with stronger index licensing incentives tend to issue higher ESG ratings for firms with better stock returns, as documented in Berg et al. (2021) and Agrawal et al. (2025).

Second, we show how rating disagreement naturally emerges from asymmetric agency aversion toward rating deviations. While this result is consistent with observed low correlations among ESG ratings across providers (Chatterji et al., 2016; Billio et al., 2021; Berg et al., 2022), we highlight a different reason that ratings can disagree—it is due to differences in incentives rather than different rating methodologies, even when they have access to the same ESG fundamentals.[2] Furthermore, disagreement allows the agency with a lower aversion to monopolize the market for green investors, echoing empirical findings that raters with high index licensing incentives capture larger asset management flows and fee revenues (Agrawal et al., 2025).

Third, our framework highlights a trade-off between investors' pecuniary utility and portfolio's fundamental greenness. In the absence of rating deviation from ESG fundamentals, green investors achieve higher portfolio greenness at the cost of lower pecuniary utility. On the other hand, a high level of rating deviation allows green investors to achieve utility levels comparable to unconstrained investors, but reduces portfolio's fundamental greenness. This trade-off arises from the ESG constraint in the investor's optimization problem, and illustrates the profound impact of the behavior of ESG rating agencies on investor utility and collective environmental outcomes. It is also consistent with the implicit trade-off embedded in preference-based models such as Pástor et al. (2021) and Pedersen et al. (2021), where investors' non-pecuniary sustainability motives lead them to accept lower expected returns in exchange for greener portfolios.

Finally, we demonstrate how firm-level greenwashing exacerbates rating deviation and affects the trade-off between utility and greenness. Greenwashing by firms whose stocks have large weights

[2]The literature has highlighted the different information sources and methodologies as reasons for ESG rating disagreement (Berg et al., 2022).

in the portfolio boosts investor utility, while greenwashing by small-weight firms improves portfolio greenness. These dynamics align with findings on how selective disclosure under low-transparency environments distorts welfare outcomes (Wu et al., 2020).

Related Work. Empirical studies reveal that ESG ratings exhibit substantial disagreement across agencies due to methodological differences (Chatterji et al., 2016; Berg et al., 2022), and may even deviate from firms' ESG fundamentals as evidenced by historical revisions (Berg et al., 2021). Recent evidence further links rating deviations to strategic considerations, such as index licensing incentives, rather than pure measurement noise (Agrawal et al., 2025). Our model provides a theoretical explanation by endogenizing both rating deviation and disagreement as strategic outcomes of agency competition under imperfect information.

Several studies develop equilibrium models of green investing. Pástor et al. (2021) and Pedersen et al. (2021) show that ESG preferences affect asset prices and expected returns. Extensions incorporating ESG rating uncertainty (Avramov et al., 2022) and climate-specific firm-level considerations (De Angelis et al., 2023) further highlight the informational frictions in sustainable investing. Lo and Zhang (2024, 2025) and Lo et al. (2024) propose a quantitative framework for optimally managing impact portfolios. While these studies take ESG ratings as given, our work differs by modeling the strategic behavior of ESG rating agencies and showing how their competition under imperfect information generates market-level distortions observed in ESG investing.

Our work is closely related to Azarmsa and Shapiro (2024), who study competition between ESG rating agencies acquiring noisy information across multiple unrelated ESG categories. They also investigate how greenwashing incentives shift raters' equilibrium behavior, thereby reducing the informativeness of ratings. While their model focuses on the structure of information acquisition and its effect on rating disagreement, our framework differs by directly modeling rating deviation relative to the ESG fundamental, and examining how rating deviation systematically affects investor utility, portfolio greenness, and correlations between asset returns and ESG scores. We further show that greenwashing exacerbates rating deviation and its impact.

Our analysis also connects more broadly to the literature on strategic information provision by rating agencies. In credit rating models, asymmetric information between investors, issuers, and rating agencies plays a central role, and typically investors use ratings to infer the underlying quality

or value of investments (DeMarzo and Duffie, 1999; DeMarzo, 2005; Guo et al., 2025). Bolton et al. (2012) model the game between competing credit rating agencies and study agency behavior under competitive pressures. Our framework differs from the credit rating setting fundamentally because investors use ESG ratings primarily to achieve portfolio sustainability targets, rather than to infer the ESG fundamental or other information about investment quality.

Outline. Section 2 introduces our model, including the market setting and the objectives of investors and competing rating agencies. Section 3 formulates the ESG rating game, defines the SSE, and characterizes equilibrium behaviors. Section 4 discusses model implications, including rating deviation, rating disagreement, the trade-off between investor utility and portfolio's fundamental greenness, and the impact of greenwashing. Section 5 concludes. All proofs are provided in the Appendix.

Notational Convention. All vectors and matrices are defined over the n-dimensional real Euclidean space $\mathbb{R}^n$. Scalars are denoted by italic letters (e.g., x), column vectors by boldface lower-case letters (e.g., $\boldsymbol{x}$), and matrices by boldface upper-case letters (e.g., $\boldsymbol{A}$). The transpose of a vector or matrix $\boldsymbol{x}$ is denoted by $\boldsymbol{x}'$. The identity matrix is denoted by $\boldsymbol{I}$, and $\mathbf{1}$ (respectively, $\mathbf{0}$) denotes a vector of ones (zeros) of appropriate dimension. The standard basis vector $\boldsymbol{e}_i \in \mathbb{R}^n$ is defined as the n-dimensional vector whose i-th entry is 1 and all other entries are 0, for $1 \leq i \leq n$. For a differentiable scalar function $f : \mathbb{R}^n \to \mathbb{R}$ we write $\nabla f(\boldsymbol{x})$ for its gradient and $\nabla^2 f(\boldsymbol{x})$ for its Hessian. The indicator function of a statement S is denoted by $\mathbb{I}_S$, which equals 1 if S holds and 0 otherwise. Unless specified otherwise, $\|\cdot\|$ denotes the Euclidean norm and $\langle \cdot, \cdot \rangle$ denotes the associated inner product.

2 The Model for ESG Ratings

Our model focuses on the strategic behavior of ESG rating agencies and their influence on the financial market. There are three key components of the model: investable assets (stocks) in the market, many ESG-conscious investors, and two competing ESG rating agencies that provide ESG ratings of stocks for those investors. Based on ex-ante beliefs about the ESG fundamental, the two rating agencies issue ESG ratings of n stocks, and each ESG-conscious investor chooses one set

of ratings to construct portfolios that meet a preset ESG target. This section describes the three components in detail.

2.1 Asset Return and Mean–Variance Optimal Portfolio

Consider a financial market with an ESG-neutral risk-free asset and $n \geq 3$ stocks, whose returns in excess of the risk-free rate are modeled by a random vector $\boldsymbol{r} = (r_1, \ldots, r_n)'$. We write

$$\boldsymbol{r} = \boldsymbol{\mu} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$ represents the expected returns, and $\boldsymbol{\varepsilon}$ denotes a zero-mean random vector with a positive definite covariance matrix $\boldsymbol{\Sigma}$ that captures the risk and correlation structure among the stock returns. We treat $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as exogenous and known to the investors and ESG rating agencies. In our model, the distributions of asset returns are exogenously determined and do not change with investors' behavior. While this is certainly an approximation of reality, we make this simplified modeling choice to focus our study on the strategic interaction between investors and rating agencies, rather than the equilibrium effects of asset prices, which are well covered by the literature such as Pástor et al. (2021), Pedersen et al. (2021), and Zerbib (2022).

Investors construct portfolios from these n stocks, and we assume that investors can choose any $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)' \in \mathbb{R}^n$, meaning that short-selling and unlimited borrowing are allowed. The (excess) return of the portfolio is therefore given by $R(\boldsymbol{\omega}) = \boldsymbol{\omega}'\boldsymbol{r}$. The mean-variance utility function of the investor is given by

$$u(\boldsymbol{\omega}) = \mathbb{E}[R(\boldsymbol{\omega})] - \frac{\gamma}{2} \operatorname{Var}[R(\boldsymbol{\omega})] = \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}, \tag{2.1}$$

where $\gamma > 0$ represents the investor's subjective level of risk aversion. It is easy to see that the optimal portfolio weights for an investor with the mean-variance utility u are given by $\gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. We refer to this portfolio as the mean–variance optimal (MVO) portfolio, which traces its origin back to Markowitz (1952).

Definition 1. We define the *unit mean–variance optimal portfolio* by

$$\boldsymbol{\omega}_M = (\omega_{M,1}, \ldots, \omega_{M,n})' := \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|}.$$

Any portfolio $\boldsymbol{\omega}$ whose weights are proportional to $\boldsymbol{\omega}_M$ is referred to as an *MVO portfolio*.

Throughout our analysis, we impose the following assumption:

Assumption 1. The stock market satisfies the condition $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1} > 0$, and $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ is not proportional to $\mathbf{1}$, where $\mathbf{1}$ is the n-dimensional vector of ones.

Assumption 1 ensures that the market offers a strictly positive risk premium, indicating that the systematic risk is priced in the market. Consequently, the allocation to the unit MVO portfolio, given by $\boldsymbol{\omega}_M'\mathbf{1}$, is strictly positive. Moreover, we suppose that the MVO portfolio is not an equally weighted portfolio.

2.2 Representative Investor's Portfolio with Green Preference

We refer to investors with green preferences as ESG-conscious investors. To integrate sustainability considerations, our model explicitly incorporates ESG criteria through an ESG rating vector $\boldsymbol{g} \in \mathbb{R}^n$. We write g_i for the ESG score of the firm i, which captures its assessed ESG performance and guides portfolio allocations aligned with sustainability objectives. In practice, investors rely on ESG scores published by rating agencies as sustainability signals used in portfolio construction. The rating vector $\boldsymbol{g}$ serves as the standardized and contractible signal on which screening rules, reporting requirements, and investment mandates are based.

A representative ESG-conscious investor constructs stock portfolios to maximize her mean-variance utility while adhering to specific ESG constraints. Each investor is characterized by a risk aversion parameter $\gamma > 0$, and an ESG target $t \in \mathbb{R}$, representing the minimum acceptable weighted ESG score for their portfolio. In this subsection, we consider that the market adopts a single ESG rating system, denoted by the vector $\boldsymbol{g}$. The ESG level of each asset is computed directly from the rating adopted in the market.

Definition 2. A representative investor is called an *ESG-conscious investor* with ESG target t, if

the portfolio weight $\boldsymbol{\omega}$ satisfies the ESG constraint

$$\frac{\boldsymbol{\omega}'\boldsymbol{g}}{\boldsymbol{\omega}'\mathbf{1}} \geq t. \tag{2.2}$$

The constraint (2.2) ensures that the weighted average ESG score of the portfolio meets or exceeds the investor's ESG target t, which is consistent with how typical ESG passive funds are designed at major asset managers. [3] Hence, the ESG-conscious investor determines her optimal portfolio given the ESG score $\boldsymbol{g}$ by solving the following optimization problem

$$\begin{aligned} &\max_{\boldsymbol{\omega}\in\mathbb{R}^n} u(\boldsymbol{\omega}) = \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}, \\ &\text{s.t.} \quad \frac{\boldsymbol{\omega}'\boldsymbol{g}}{\boldsymbol{\omega}'\mathbf{1}} \geq t. \end{aligned} \tag{2.3}$$

While some studies model sustainable investors as deriving non-pecuniary utility from holding greener assets, we represent investors' sustainability preferences through an ESG constraint. The constraint-based formulation (2.3) is consistent with common responsible investment practices.[4] From a theoretical perspective, the above optimization problem is a quadratic program with linear constraints, which admits an explicit solution.

Proposition 1. *Under Assumption 1, the optimization problem* (2.3) *admits the explicit solution*

$$\boldsymbol{\omega}^* = \boldsymbol{\omega}(\gamma, t, \boldsymbol{g}) := \begin{cases} \gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \lambda^*\gamma^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{g} - t\mathbf{1}), & \text{if } \dfrac{\boldsymbol{\omega}_M'\boldsymbol{g}}{\boldsymbol{\omega}_M'\mathbf{1}} < t; \\ \gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, & \text{if } \dfrac{\boldsymbol{\omega}_M'\boldsymbol{g}}{\boldsymbol{\omega}_M'\mathbf{1}} \geq t, \end{cases} \tag{2.4}$$

where

$$\lambda^* = \lambda(t, \boldsymbol{g}) := \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(t\mathbf{1} - \boldsymbol{g})}{(\boldsymbol{g} - t\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g} - t\mathbf{1})}.$$

[3] For example, the MSCI ESG Fund Ratings Methodology (MSCI, 2025) explains that they adopt a weighted average method for calculating fund ESG scores. In practice, investors can use these passive funds as an instrument to access a broad index of green assets, and then build their actual portfolios of the green fund and the risk-free asset depending on their own risk preferences.

[4] Empirical evidence shows that investors mainly use ESG information for screening and risk management purposes rather than to express non-pecuniary preferences (Amel-Zadeh and Serafeim, 2018). Consistently, survey data indicate that exclusionary screening and best-in-class selection represent about 47% and 37% of responsible investment strategies, respectively (Eccles et al., 2017). In practice, it is difficult to choose the right weight for non-pecuniary utility, but much easier to interpret an overall ESG threshold for the portfolio. Nonetheless, incorporating a non-pecuniary ESG preference as an additive utility term (Pástor et al., 2021; Pedersen et al., 2021) is mathematically equivalent to imposing an ESG constraint with a different multiplier, so the constraint-based approach preserves the essential trade-off between expected return and sustainability preference.

The corresponding utility for the optimal $\boldsymbol{\omega}^*$ *is*

$$u(\boldsymbol{\omega}^*) = \begin{cases} \frac{1}{2\gamma}\Big(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \lambda^*\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(t\mathbf{1}-\boldsymbol{g})\Big), & \text{if } \dfrac{\boldsymbol{\omega}_M'\boldsymbol{g}}{\boldsymbol{\omega}_M'\mathbf{1}} < t; \\ \frac{1}{2\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, & \text{if } \dfrac{\boldsymbol{\omega}_M'\boldsymbol{g}}{\boldsymbol{\omega}_M'\mathbf{1}} \geq t. \end{cases} \tag{2.5}$$

Remark 1. The solution $\boldsymbol{\omega}^*$ represents the optimal portfolio allocation of the ESG-conscious investor. When $\boldsymbol{\omega}_M'\boldsymbol{g}/\boldsymbol{\omega}_M'\mathbf{1} \geq t$, $\boldsymbol{\omega}^*$ coincides with the classical MVO asset allocation $\gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. Otherwise, the MVO portfolio $\gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ is insufficient to meet the investor's requirement and the asset allocation is adjusted by

$$\boldsymbol{\omega}^* \;=\; \underbrace{\gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}_{\text{MVO portfolio}} \;+\; \underbrace{\lambda^*\gamma^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{g}-t\mathbf{1})}_{\text{ESG correction}}.$$

This decomposition has a clear interpretation. The direction of the ESG correction, $\boldsymbol{\Sigma}^{-1}(\boldsymbol{g}-t\mathbf{1})$, is the steepest way (in the risk metric induced by $\boldsymbol{\Sigma}$) to increase the weighted average ESG score of the portfolio $\boldsymbol{\omega}'\boldsymbol{g}/\boldsymbol{\omega}'\mathbf{1}$. It tilts weights toward assets with $g_i > t$ and away from those with $g_i < t$. The magnitude $\lambda^* > 0$ is chosen so that the greenness constraint binds exactly: the numerator $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(t\mathbf{1}-\boldsymbol{g})$ is the greenness shortfall of the MVO portfolio measured in the same risk metric, while the denominator $(\boldsymbol{g}-t\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g}-t\mathbf{1})$ is the squared risk-adjusted norm of the constraint direction. Hence λ^* is the precise scaling required to move the MVO portfolio $\gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ on the ESG boundary $\{\boldsymbol{\omega} : \boldsymbol{\omega}'\boldsymbol{g} = t\}$.

Based on the criteria derived in Proposition 1, which specifies whether the MVO portfolio satisfies the investor's ESG target t, the optimal choice of an ESG-conscious investor depends on the relationship between the ESG ratings of all stocks and the investor's ESG target. This relationship is driven primarily by the ESG target t, rather than the investor's level of risk aversion γ. As a result, investors are categorized based on their ESG target t, not by their risk preferences. Inspired by (2.5), we define the utility of ESG-conscious investors as follows:

Definition 3. The *utility of an ESG-conscious investor* with ESG target t and rating scores $\boldsymbol{g}$ is defined by

$$U(t,\boldsymbol{g}) := \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2} - \frac{\big(\max\{\boldsymbol{\omega}_M'(t\mathbf{1}-\boldsymbol{g}),0\}\big)^2}{(\boldsymbol{g}-t\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g}-t\mathbf{1})}. \tag{2.6}$$

It can be verified from Proposition 1 that if an ESG-conscious investor has risk aversion level γ, the investor's mean-variance utility at the optimal portfolio $\boldsymbol{\omega}^*$ can be expressed as $u(\boldsymbol{\omega}^*) = U(t, \boldsymbol{g}) \cdot \|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2/(2\gamma)$. The last term in (2.6) shows how ESG considerations impact the optimal portfolio choice. Specifically, this term represents the utility loss incurred by adhering to an ESG target t, leading to a possible deviation from the MVO portfolio. If it equals zero, then the investor chooses the MVO portfolio and the utility reaches the global maximum $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2$.

2.3 Rating Methodology of Competing ESG Rating Agencies

In the previous discussion, we defined investors' mean-variance utility and analyzed their investment decisions based on a single ESG rating. We now introduce the ESG rating agencies into the market model. Consider two competing ESG rating agencies, Agency A and Agency B. Each agency issues its own ESG ratings, denoted as $\boldsymbol{g}_A = (g_{A,1}, \ldots, g_{A,n})'$ and $\boldsymbol{g}_B = (g_{B,1}, \ldots, g_{B,n})'$, respectively. To facilitate meaningful comparisons across different agencies, we normalize the scores such that their average is zero, i.e., $\boldsymbol{g}_A'\mathbf{1} = \boldsymbol{g}_B'\mathbf{1} = 0$. In our model, we consider investors who must choose between Agency A and Agency B as their ESG provider to highlight the strategic competition between rating agencies.

Our discussion proceeds in two parts. First, we introduce the concept of an ESG fundamental as a benchmark. Second, we define the utility structure for rating agencies to illustrate the incentives they derive from issuing ESG ratings.

ESG Fundamental. We follow Agrawal et al. (2025) to introduce the concept of *ESG fundamental* that represents the objective sustainability performance of a firm, which is distinguished from observed agency-issued ratings. In our model, we denote by $\tilde{\boldsymbol{g}}_T \in \mathbb{R}^n$ the ESG fundamental, with $\tilde{g}_{T,i}$ capturing the objective ESG quality of the firm i.

We impose the following assumption on the ESG fundamental.

Assumption 2. The ESG fundamental $\tilde{\boldsymbol{g}}_T$ is observable to ESG rating agencies but unobservable to ESG-conscious investors. Moreover, it satisfies the normalization condition

$$\tilde{\boldsymbol{g}}_T'\mathbf{1} = \tilde{g}_{T,1} + \cdots + \tilde{g}_{T,n} = 0.$$

Assumption 2 implies that ESG rating agencies have exclusive access to ESG fundamentals through their research and analysis, but they are not available to investors. We normalize the ESG fundamentals so that they are comparable across different agencies.

We next introduce the concept of intrinsic ESG benchmark derived from the weighted average ESG score of the MVO portfolio, which serves as an objective threshold to categorize investors based on their ESG preferences.

Definition 4. The *intrinsic ESG benchmark* of market, denoted as e_T, is defined by

$$e_T := \frac{\boldsymbol{\omega}_M' \tilde{\boldsymbol{g}}_T}{\boldsymbol{\omega}_M' \mathbf{1}} = \frac{\omega_{M,1}\tilde{g}_{T,1} + \cdots + \omega_{M,n}\tilde{g}_{T,n}}{\omega_{M,1} + \cdots + \omega_{M,n}}. \tag{2.7}$$

An ESG-conscious investor is referred to as a *green investor* if her ESG target $t > e_T$, and a *brown investor* if $t \leq e_T$.

From (2.7), the intrinsic ESG benchmark e_T encapsulates the overall ESG performance of the MVO portfolio, providing a natural and objective criterion for distinguishing investors as green or brown based on ESG considerations. Specifically, green investors set ESG targets above this benchmark, implying that their investment decisions are more likely to be influenced by ESG factors. Although the other investors are called brown, they may still be ESG-conscious investors, but simply with ESG targets lower than this benchmark. As a result, their optimal portfolio allocation are not affected by the ESG target because the ESG target $t \leq e_T$ is satisfied by the MVO portfolio based on Proposition 1.

Utility of competing agencies. The utility of ESG rating agencies consists of two components: a non-pecuniary deviation penalty component, and a pecuniary market share component. We next analyze each component in detail.

First, the deviation penalty reflects the reputational cost of misalignment with firms' truthful sustainability performance. Motivated by empirical evidence that ESG rating revisions are linked to concerns over historical accuracy and reputational credibility (Berg et al., 2021), we measure deviation using the Euclidean distance between the issued ratings and the ESG fundamentals. The

deviation penalty for each agency $j \in \{\text{A,B}\}$ is defined by

$$N_j(\boldsymbol{g}_j; \alpha_j) = \frac{\alpha_j}{2} \|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_T\|^2 = \frac{\alpha_j}{2} \sum_{i=1}^{n} (g_{j,i} - \tilde{g}_{T,i})^2, \tag{2.8}$$

where α_A and α_B are parameters representing the aversion toward deviation. From the expression (2.8), a smaller aversion parameter α_j indicates lower aversion for deviation, allowing more aggressive rating adjustments, whereas a larger α_j signifies higher aversion, emphasizing alignment with the ESG fundamental.

Second, the market share reflects the financial incentives received by agencies. Notice that the deviation penalty (2.8) is not directly observable to investors, since investors do not know the ESG fundamental. Accordingly, investors are not able to form a preference for one agency over the other simply based on their reputational penalties. Therefore, each investor aiming to maximize their utility $U(t, \boldsymbol{g})$ as defined in expression (2.6), selects the agency whose ESG ratings provide them with the greatest utility. Although investors may respond more smoothly in practice, we use a discrete response rule here for simplicity, and this does not affect the key implications of the model. Specifically, an ESG-conscious investor with target t chooses Agency A with probability

$$P_A(t; \boldsymbol{g}_A, \boldsymbol{g}_B) = \begin{cases} 1, & \text{if } U(t, \boldsymbol{g}_A) > U(t, \boldsymbol{g}_B); \\ 0.5, & \text{if } U(t, \boldsymbol{g}_A) = U(t, \boldsymbol{g}_B); \\ 0, & \text{if } U(t, \boldsymbol{g}_A) < U(t, \boldsymbol{g}_B). \end{cases} \tag{2.9}$$

Correspondingly, the probability of choosing Agency B is

$$P_B(t; \boldsymbol{g}_A, \boldsymbol{g}_B) = \begin{cases} 0, & \text{if } U(t, \boldsymbol{g}_A) > U(t, \boldsymbol{g}_B); \\ 0.5, & \text{if } U(t, \boldsymbol{g}_A) = U(t, \boldsymbol{g}_B); \\ 1, & \text{if } U(t, \boldsymbol{g}_A) < U(t, \boldsymbol{g}_B). \end{cases} \tag{2.10}$$

Since two agencies are competing, the probability of choosing Agency B is simply the complement, i.e., $P_B(t; \boldsymbol{g}_A, \boldsymbol{g}_B) = 1 - P_A(t; \boldsymbol{g}_A, \boldsymbol{g}_B)$.

Each agency's market share is defined as the expected probability that an ESG-conscious in-

vestor selects that agency, taken with respect to the distribution of investors' ESG targets. Denote by π the probability distribution of the target t, characterizing the distribution of ESG-conscious investors. For $j \in \{\text{A,B}\}$, define the market share of Agency j as

$$M_j^{\pi}(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) := \mathbb{E}_{t\sim\pi}\left[P_j(t; \boldsymbol{g}_A, \boldsymbol{g}_B)\right], \tag{2.11}$$

where $\boldsymbol{g}_{-j}$ denotes the rating issued by agency j's rival. (2.11) quantifies the fraction of investors adopting the ESG ratings of Agency j. Consequently, an agency seeking to attract more investors should issue ESG ratings that confer higher expected utility than those offered by its competitors.

Notice that

$$M_A^{\pi}(\boldsymbol{g}_A, \boldsymbol{g}_B) + M_B^{\pi}(\boldsymbol{g}_B, \boldsymbol{g}_A) = 1,$$

so it ensures full market coverage. It is also worth emphasizing that allowing for an additional group of investors who consult both agencies would not affect our main results. This is because such investors can be modeled as contributing a common constant to each agency's client base, so that the market shares are only subject to a common affine transformation, and the implications of our model remain unchanged.

Based on the definitions of deviation penalties and market shares for the two agencies, we define the utility of the competing ESG rating agencies as follows.

Definition 5. The *utility of rating agency* for agency $j \in \{\text{A,B}\}$ is defined as:

$$V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) := M_j^{\pi}(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) - N_j(\boldsymbol{g}_j; \alpha_j) = \mathbb{E}_{t\sim\pi}[P_j(t; \boldsymbol{g}_A, \boldsymbol{g}_B)] - \frac{\alpha_j}{2}\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_T\|^2, \tag{2.12}$$

where $\boldsymbol{g}_{-j}$ denotes the rating issued by agency j's rival.

3 ESG Rating Game

In this section, we construct the ESG rating game and characterize the equilibrium behavior of the two rating agencies.

3.1 Definition of Stackelberg Stable Equilibrium

The two competing rating agencies aim to maximize their respective utilities. This decision involves balancing the trade-off between increasing market share and the accuracy of ESG scores. We can model this situation as an extensive-form game with imperfect information, where the two players are Agency A and Agency B. The timing of the game with competition unfolds as follows. First, both agencies choose their ratings simultaneously without observing the rival's choice. Then, the ratings $\boldsymbol{g}_A$ and $\boldsymbol{g}_B$ are publicly issued to investors. In the end, investors choose a preferred rating agency based on maximizing their own utilities.

Define the action set $G = \{\boldsymbol{g} \in \mathbb{R}^n : \boldsymbol{g}'\mathbf{1} = 0\}$, which represents the set of possible ESG ratings that each agency can choose. Following standard definitions in game theory (e.g., Osborne (2004)), the rating profile $(\boldsymbol{g}_A, \boldsymbol{g}_B)$ forms a Nash equilibrium (NE) if for $j \in \{\text{A,B}\}$,

$$V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) \geq V_j(\boldsymbol{g}'_j, \boldsymbol{g}_{-j}),\ \forall\ \boldsymbol{g}'_j \in G.$$

This formulation fails to adequately describe strategic outcomes in this setting for two main reasons.

First, the NE framework does not yield a behaviorally meaningful solution in our setting. When π is continuous with density, the utility of rating agency V_j in Definition 5 is continuous in $\boldsymbol{g}$. The quadratic reputational penalty of V_j renders strategies with arbitrarily large norms suboptimal. This allows the game to be equivalently analyzed on a compact and convex subset of G. It follows from Glicksberg's theorem (Glicksberg, 1952) that a mixed strategy Nash equilibrium exists, but it is difficult to characterize in explicit form. When π is discrete, the market share M_j^π is discontinuous in $\boldsymbol{g}$. It makes V_j discontinuous due to the discrete structure of M_j^π. This discontinuity implies that the supremum in the best response condition may not be attained.

Second, the annual update cycle gives any agency the chance to announce a rating before its rival acts, turning that announcement into a credible commitment, as once a rating is disclosed it cannot be easily revised or withdrawn. A simultaneous-move NE that appears stable could therefore be overturned when this option is taken into account. Moreover, we are able to characterize the necessary and sufficient conditions for the existence of a traditional NE, which we formally derive in Section 3.2.

As a result, we replace NE with a Stackelberg stable equilibrium (SSE), in which no agency

can increase its payoff by facing the rival's (approximate) best response. The SSE extends the NE to environments where actions become credible commitments. While the NE captures mutual best responses under simultaneous moves, the SSE refines this concept by ensuring stability to unilateral early announcements, where no agency can improve its payoff by committing to its rating before the rival. This refinement therefore captures the forward-looking behavior in ESG-rating competition, and provides a natural bridge between the simultaneous-move structure of the NE and the commitment-driven nature of practical ESG disclosures.

Definition 6. Given $\varepsilon \geq 0$, the *ε-best response set* for agency $j \in \{\text{A,B}\}$ is defined as

$$\mathrm{BR}_j^{\varepsilon}(\boldsymbol{g}_{-j}) := \left\{ \boldsymbol{g}_j \in G \;\middle|\; V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) \geq \sup_{\tilde{\boldsymbol{g}}_j \in G} V_j(\tilde{\boldsymbol{g}}_j, \boldsymbol{g}_{-j}) - \varepsilon \right\}. \tag{3.1}$$

A rating profile $(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*)$ forms a *Stackelberg stable equilibrium (SSE)* if for each $j \in \{\text{A,B}\}$, the following inequalities hold:

$$V_j(\boldsymbol{g}_j^*, \boldsymbol{g}_{-j}^*) \geq \lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_{-j} \in \mathrm{BR}_{-j}^{\varepsilon}(\boldsymbol{g}_j)} V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}),\ \forall\ \boldsymbol{g}_j \in G. \tag{3.2}$$

Remark 2. The supremum over $\boldsymbol{g}_j \in G$ of the right-hand side of (3.2) is the Stackelberg value of Agency j in the sense of Mailath and Samuelson (2006): it is the highest payoff an agency can secure by first fixing a rating and then facing the rival's worst response. A rating profile is therefore Stackelberg stable precisely when no unilateral early announcement can improve its outcome.

3.2 Characterizing the Aversion Parameters

As explained above, the model's parameters include the distribution of ESG-conscious investors π, and the aversion parameters α_A and α_B for agencies toward deviation. To solve the SSE, we first introduce tractable assumptions on π that still capture realistic market conditions, and then relate the two parameter classes by characterizing the aversion parameters.

We consider two types of investors with targets e (green) and b (brown), satisfying $b < e_T < e$. Both groups are part of the ESG-conscious investors of the market, but their roles differ. The proportion of green investors in the market is denoted by $p \in (0, 1]$. Hence, the distribution of ESG targets is given by $\pi(\{e\}) = p$ and $\pi(\{b\}) = 1 - p$. Green investors share a common target e,

reflecting regulatory or market standards that push ESG performance above the intrinsic market benchmark e_T. In our terminology, brown investors are those who have an ESG constraint already satisfied by the ESG fundamental $\tilde{\boldsymbol{g}}_T$; this does not necessarily imply that they do not have any ESG preferences.

For agencies, the utility function (2.12) gives a clear trade-off: deviating from the ESG fundamental helps to attract green investors but incurs a reputational penalty. This penalty is scaled by the agency's aversion parameter α_j. The following proposition motivates us to formalize this trade-off and provide an operational characterization of α_j.

Proposition 2. *Under Assumptions 1 and 2, for $s \geq e_T$, the optimization problem*

$$\min_{\boldsymbol{g}\in\mathbb{R}^n} \|\boldsymbol{g} - \tilde{\boldsymbol{g}}_T\|^2 \quad s.t. \quad \frac{\boldsymbol{\omega}_M'\boldsymbol{g}}{\boldsymbol{\omega}_M'\mathbf{1}} \geq s \quad and \quad \boldsymbol{g}'\mathbf{1} = 0 \tag{3.3}$$

has a unique solution

$$\boldsymbol{g}_s^* = \tilde{\boldsymbol{g}}_T + (s - e_T)\boldsymbol{g}_M, \tag{3.4}$$

where $\boldsymbol{g}_M$ is an $n \times 1$ adjusted vector given by

$$\boldsymbol{g}_M = \frac{n(\boldsymbol{\omega}_M'\mathbf{1})}{n - (\boldsymbol{\omega}_M'\mathbf{1})^2}\boldsymbol{\omega}_M - \frac{(\boldsymbol{\omega}_M'\mathbf{1})^2}{n - (\boldsymbol{\omega}_M'\mathbf{1})^2}\mathbf{1}. \tag{3.5}$$

Proposition 2 gives the minimum degree of deviation between the ratings and the ESG fundamental for achieving the ESG target s. Specifically, we have

$$\|\boldsymbol{g}_s^* - \tilde{\boldsymbol{g}}_T\|^2 = (s - e_T)^2\|\boldsymbol{g}_M\|^2 = (s - e_T)^2\frac{n(\boldsymbol{\omega}_M'\mathbf{1})^2}{n - (\boldsymbol{\omega}_M'\mathbf{1})^2}. \tag{3.6}$$

From (3.6), the deviation $\|\boldsymbol{g}_s^* - \tilde{\boldsymbol{g}}_T\|^2$ increases monotonically with s. The result in Proposition 2 can be interpreted as a linear deviation from the ESG fundamental $\tilde{\boldsymbol{g}}_T$ along the direction of $\boldsymbol{g}_M$, with an intensity proportional to $s - e_T$. Economically, this means that the optimal rating adjustment moves away from the ESG fundamental in a fixed direction determined by the MVO portfolio $\boldsymbol{\omega}_M$. The magnitude of this deviation provides a direct measure of the market's demand for greener ratings. To further illustrate the implication of Proposition 2, the following example considers a single-agency case where the trade-off between market share gains and reputational costs becomes

explicit.

Example 1. Consider a single agency that caters exclusively to investors who require the MVO portfolio weighted ESG score to be at least s. Assume that the utility of this agency is

$$V(\boldsymbol{g}) = q \cdot \mathbb{I}_{\{\boldsymbol{g}:\ \boldsymbol{\omega}_M'\boldsymbol{g}/\boldsymbol{\omega}_M'\mathbf{1} \geq s\}} - \frac{\alpha}{2}\|\boldsymbol{g} - \tilde{\boldsymbol{g}}_T\|^2,$$

where $q > 0$ represents the profit from serving this clientele (that is, the market share payoff), and $\alpha > 0$ is the agency's aversion parameter. If $s \leq e_T$, the ESG fundamental $\tilde{\boldsymbol{g}}_T$ satisfies the constraint exactly, so the agency issues $\tilde{\boldsymbol{g}}_T$ and earns the full profit q without incurring any reputational cost. If $s > e_T$, the ESG fundamental no longer meets the target, yielding $V(\tilde{\boldsymbol{g}}_T) = 0$. By Proposition 2, the least-cost rating that attains the target is $\boldsymbol{g}_s^*$. Choosing the vector $\boldsymbol{g}_s^*$ is profitable if and only if

$$\alpha \leq \frac{2q}{\|\boldsymbol{g}_s^* - \tilde{\boldsymbol{g}}_T\|^2} = \frac{2q}{(s - e_T)^2}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right].$$

When the above inequality fails, the agency issues the ESG fundamental and earns zero. Otherwise, the agency issues $\boldsymbol{g}_s^*$ and acquires utility $V(\boldsymbol{g}_s^*) = q - \frac{\alpha}{2}\|\boldsymbol{g}_s^* - \tilde{\boldsymbol{g}}_T\|^2 > 0$.

Based on the above example, we formalize the trade-off of rating agencies through the concept of a green threshold for the aversion parameter. Before proceeding, recall that e denotes the greenness level targeted by green investors, and $\boldsymbol{g}_e^*$ represents the ESG rating profile that an agency would optimally issue when aiming to attract investors with ESG target e. Then we introduce the parameter $\alpha_{e,p}$ to quantify the agency's aversion threshold, balancing the incentive to capture market share p against the penalty from deviating from the ESG fundamental $\tilde{\boldsymbol{g}}_T$.

Definition 7. The *green threshold* for the aversion parameter is defined by

$$\alpha_{e,p} := \frac{p}{\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2}. \tag{3.7}$$

To facilitate the equilibrium analysis, the level of an agency's aversion parameter α is classified as low, medium, or high based on its relationship to the green threshold $\alpha_{e,p}$. Table 1 summarizes this classification.

Agencies with high aversion have little incentive to misreport; medium aversion allows moderate

Table 1: Classification of aversion parameter for rating deviation

Aversion Type	**The condition for α**
Low Aversion	$\alpha \leq \alpha_{e,p}$
Medium Aversion	$\alpha_{e,p} < \alpha \leq 2\alpha_{e,p}$
High Aversion	$\alpha > 2\alpha_{e,p}$

adjustments; low aversion leads agencies to inflate ratings substantially for capturing market shares. It is worth noting that this classification is relative: as the threshold $\alpha_{e,p}$ increases, a given level of α might correspond to a lower aversion, since agencies become less averse to rating deviations.

Based on this classification of aversion parameters, we establish a necessary and sufficient condition for the existence of a Nash equilibrium in our game model.

Proposition 3. *Under Assumptions 1 and 2, an NE* $(\boldsymbol{g}_A^{\#}, \boldsymbol{g}_B^{\#})$ *exists if and only if both agencies have low aversion, and the issued ratings in NE are uniquely given by* $\boldsymbol{g}_A^{\#} = \boldsymbol{g}_B^{\#} = \boldsymbol{g}_e^*$.

The proof of Proposition 3 builds on Lemma A.1, whose statement and proof are provided in Appendix A.3. With Lemma A.1 established, we present the proof of Proposition 3 in Appendix A.4. Proposition 3 shows that an NE exists only when both agencies have low aversion and are willing to substantially deviate from the ESG fundamental to attract investors. Intuitively, each investor chooses the rating that maximizes her utility, implying that green investors are attracted to the agency issuing the greener rating. In a continuous rating space, each agency has an incentive to slightly adjust its rating toward investors' preferred greenness to capture additional market share. When at least one agency does not have low aversion, this competitive dynamic prevents best responses from stabilizing, and thus no NE exists. When both agencies have low aversion, however, their optimal deviations coincide at $\boldsymbol{g}_e^*$, restoring stability and yielding a unique equilibrium.

In contrast, SSE accommodates a broader range of aversion configurations and ensures equilibrium existence even when NE fails to hold. As will be shown later, the threshold parameter $\alpha_{e,p}$ plays a crucial role in determining the equilibrium outcomes, since different ranges of aversion parameters relative to $\alpha_{e,p}$ lead to distinct types of equilibria.

3.3 Rating Behavior in Stackelberg Stable Equilibrium

We now derive the SSE of the game for ESG rating agencies. Our analysis focuses on the assumption that the distribution of ESG targets is given by $\pi(\{e\}) = p \in (0, 1]$ and $\pi(\{b\}) = 1 - p$. Therefore, the threshold $\alpha_{e,p}$ remains constant. The distinct equilibrium emerges depending on the different values of α_A and α_B. In Section 4, we will allow (e, p) to vary within their admissible ranges to study further implications.

Theorem 1. *Under Assumptions 1 and 2, an SSE $(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*)$ always exists and is characterized for each agency $j \in \{A,B\}$ as follows:*

(1) If $\alpha_j/\alpha_{-j} > 1/2$, then Agency j's issued rating $\boldsymbol{g}_j^*$ in equilibrium is given by

$$\boldsymbol{g}_j^* = \begin{cases} \tilde{\boldsymbol{g}}_T, & \alpha_j \text{ is medium or high,} \\ \boldsymbol{g}_e^*, & \alpha_j \text{ is low.} \end{cases} \tag{3.8}$$

(2) If $\alpha_j/\alpha_{-j} \leq 1/2$, then Agency j's issued rating $\boldsymbol{g}_j^*$ in equilibrium is given by

$$\boldsymbol{g}_j^* = \begin{cases} \boldsymbol{g}_e^*, & \alpha_j \text{ is low or medium,} \\ \boldsymbol{g}^{**}(\alpha_{-j}), & \alpha_j \text{ is high,} \end{cases} \tag{3.9}$$

where $\boldsymbol{g}^{**}(\alpha_{-j})$ is a solution of

$$\max_{\boldsymbol{g} \in G} U(e, \boldsymbol{g}) \quad s.t. \quad \frac{\alpha_{-j}}{2} \|\boldsymbol{g} - \tilde{\boldsymbol{g}}_T\|^2 \leq p, \ \boldsymbol{g}'\mathbf{1} = 0. \tag{3.10}$$

Here α_{-j} is the aversion parameter of agency j's rival.

Remark 3. Because ESG scores are essentially percentile ranks and the ratings are sometimes normalized with respect to their distribution, it is fair to consider ESG scores with a normalization constraint. This difference does not change the key implications of our model. In particular, if the normalization constraint is convex with respect to the ESG rating $\boldsymbol{g}$, such as the condition that the ESG rating of each firm is within $[-1, 1]$, the proof of Lemma A.1 still holds. It follows that there still exists $\boldsymbol{g}_e^*$ and $\boldsymbol{g}^{**}$ solving Problems (3.3) and (3.10) with the additional normalization constraint, while $\boldsymbol{g}_e^*$ becomes implicit under the constraint $g_{j,i} \in [-1, 1]$, $j = A, B$, $i = 1, \ldots, n$.

Based on Lemma A.1, the SSE still has the same form as that in Theorem 1, where $\boldsymbol{g}_e^*$ and $\boldsymbol{g}^{**}$ are both implicit. Hence, we can consider the model without normalization for simplicity when analyzing the implications of equilibrium behaviors.

Theorem 1 characterizes the agencies' equilibrium rating strategies, with the proof given in Appendix A.6. Its derivation relies first on Lemma A.1 and then on Lemma A.2 in Appendix A.5. Lemma A.1 characterizes the optimizer $\boldsymbol{g}^{**}$ in terms of a scalar parameter s. To apply this result to the equilibrium problem (3.10) in Theorem 1, we link the parameter s to the condition

$$\frac{\alpha_{-j}}{2}\|\boldsymbol{g}_s^* - \tilde{\boldsymbol{g}}_T\|^2 = p.$$

Substituting the expression (3.6) into this relationship yields the closed-form solution

$$s = e_T + \sqrt{\frac{2p}{\alpha_{-j}}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}. \tag{3.11}$$

This establishes the form of the equilibrium rating $\boldsymbol{g}^{**}(\alpha_{-j})$ in case (2) of Theorem 1. Thus, in case (2), the agency's equilibrium rating $\boldsymbol{g}_j^*$ is given by the corresponding optimizer $\boldsymbol{g}^{**}$ obtained from Lemma A.1.

This agency's equilibrium rating strategy involves solving an optimization problem constrained by the opponent's aversion level. The dependence of the constraint in problem (3.10) on the rival's parameter α_{-j} reflects the strategic nature of the equilibrium: when the rival is highly averse (large α_{-j}), it avoids aggressive green deviation and thus poses little competitive pressure. In this case, Agency j just needs to inflate its rating enough to remain more appealing than its rival. Hence, Agency j's optimal deviation is endogenously determined by the opponent's aversion parameter.

A key concept emerging from the equilibrium structure in Theorem 1 is the notion of *relative aversion*, defined as the ratio α_j/α_{-j} for agency $j \in \{\text{A,B}\}$. The classification of relative aversion into high or low critically shapes the agency's equilibrium behavior. Specifically:

- **Low relative aversion**: $\alpha_j/\alpha_{-j} \leq 1/2$, meaning that the rival agency imposes stricter penalties for deviation relative to Agency j;
- **High relative aversion**: $\alpha_j/\alpha_{-j} > 1/2$, meaning that the rival agency has comparable or weaker alignment incentives relative to Agency j.

Now we discuss the equilibrium results of Theorem 1. Figure 1 illustrates the issued ESG ratings in SSE across varying parameter values. It explicitly illustrates the equilibrium ESG ratings as functions of the aversion parameters α_A and α_B, shown on the horizontal and vertical axes, respectively. While the parameter space technically divides into six regions based on aversion classifications, symmetry between the two agencies allows us to consolidate the analysis into four representative regions, where equilibrium strategies follow the structure characterized in Theorem 1. Specifically, we summarize the equilibrium behavior in four regions as follows.

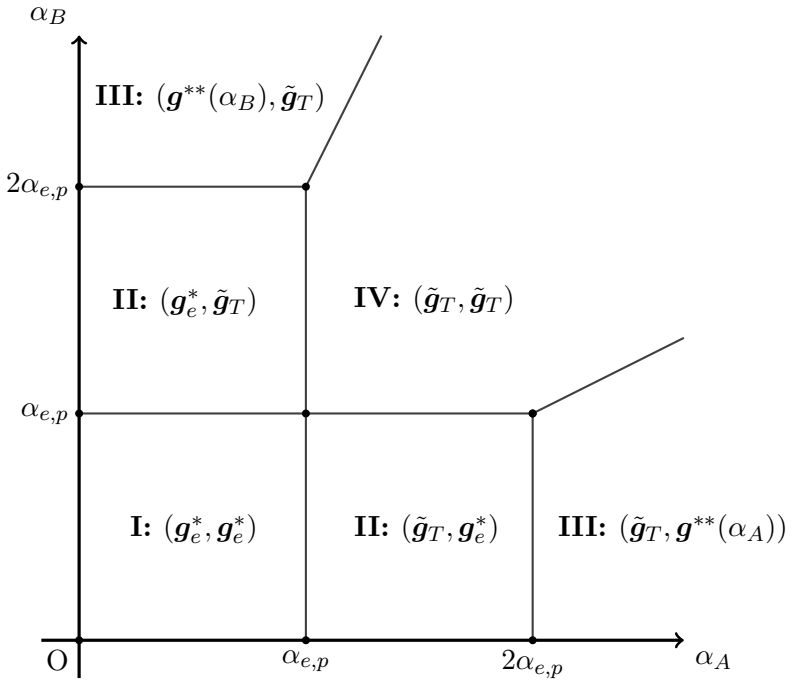

Figure 1: The issued ESG ratings in SSE with (α_A, α_B).

Region I : Both agencies have low aversion. In equilibrium, each issues the same rating $\boldsymbol{g}_e^*$.

Region II : One agency has low aversion, and the other has medium aversion. In equilibrium, the low aversion agency issues $\boldsymbol{g}_e^*$, while the medium aversion agency issues the ESG fundamental $\tilde{\boldsymbol{g}}_T$.

Region III: One agency has low relative aversion and the other has high aversion. In equilibrium, the high aversion agency issues $\tilde{\boldsymbol{g}}_T$, while the other agency issues $\boldsymbol{g}^{**}(\alpha_{-j})$, determined by the rival's aversion α_{-j}.

Region IV: Neither agency has low aversion or low relative aversion. In equilibrium, both agencies issue the ESG fundamental $\tilde{\boldsymbol{g}}_T$.

Our treatment of boundary cases is consistent with the definitions of high/medium/low aversion and high/low relative aversion. For each region boundary, the equilibrium is assigned according to the region with the smaller region index. For example, the boundary between **Region II** and **Region III** is treated as **Region II**. These boundaries may involve abrupt changes in equilibrium behavior, primarily due to discontinuities in market share allocation.

From Proposition 3, NE exists only in **Region I** and the issued ratings are both $\boldsymbol{g}_e^*$. Our proposed SSE is consistent with NE when it exists, but it accounts for a wider set of parameters.

4 Implications of Equilibrium Behaviors

This section discusses the equilibrium implications of our model and connects them to the empirical literature on ESG. First, we identify conditions that lead to ESG rating deviation and disagreement. Second, we investigate the trade-off between investors' (pecuniary) utility and the portfolio's fundamental greenness. Finally, we further evaluate the impact of greenwashing on these effects.

By symmetry, we assume that the aversion parameters satisfy $\alpha_A > \alpha_B$, meaning that Agency A has a higher aversion than Agency B. In this setting, based on the equilibrium behaviors in Theorem 1, Agency A adopts a conservative strategy for rating accuracy, while Agency B takes a bold stance to capture the market.

4.1 ESG Rating Deviation

We analyze conditions under which ESG rating deviation emerges and tends to grow, and derive the cross-sectional association between the issued ESG ratings and the asset return to understand how rating agencies deviate their ratings.

As shown in Figure 1, rating deviation occurs in all regions except **Region IV**. We specify the conditions for the emergence of deviation and characterize the degree of deviation as follows.

Proposition 4. *Suppose $\alpha_A > \alpha_B$. Rating deviation emerges under the following conditions.*

(1) When $\alpha_B/\alpha_A \in (1/2, 1)$, for $j \in \{A,B\}$, Agency j deviates from the ESG fundamental if and only if

$$e \leq e_T + \sqrt{\frac{p}{\alpha_j}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}. \tag{4.1}$$

When (4.1) holds, the rating $\boldsymbol{g}_j^* = \boldsymbol{g}_e^* = \tilde{\boldsymbol{g}}_T + (e - e_T)\boldsymbol{g}_M$ is issued, and the degree of deviation equals

$$\|\boldsymbol{g}_j^* - \tilde{\boldsymbol{g}}_T\|^2 = (e - e_T)^2 \frac{n(\boldsymbol{\omega}_M'\mathbf{1})^2}{n - (\boldsymbol{\omega}_M'\mathbf{1})^2}. \tag{4.2}$$

(2) When $\alpha_B/\alpha_A \in (0, 1/2]$, Agency B always deviates, while Agency A deviates when (4.1) holds for $j = A$. When Agency A deviates, the rating $\boldsymbol{g}_A^* = \boldsymbol{g}_e^*$ is issued, and the norm of deviation is given by (4.2). On the other hand, Agency B is supposed to issue $\boldsymbol{g}_B^* = \boldsymbol{g}_s^{**}$, where $s = \min\{e, e_T + \sqrt{\frac{2p}{\alpha_A}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}\}$, and the corresponding degree of deviation is given by

$$\|\boldsymbol{g}_B^* - \tilde{\boldsymbol{g}}_T\|^2 = \min\left\{(e - e_T)^2 \frac{n(\boldsymbol{\omega}_M'\mathbf{1})^2}{n - (\boldsymbol{\omega}_M'\mathbf{1})^2}, \frac{2p}{\alpha_A}\right\}. \tag{4.3}$$

Remark 4. By (3.6) and (3.7), the condition (4.1) is equivalent to $\alpha_j \leq \alpha_{e,p}$.

Proposition 4 shows that, for Agency A, rating deviation occurs if and only if the aversion α_A is low; for Agency B, rating deviation occurs if and only if the aversion α_B is low or the relative aversion is low ($\alpha_B/\alpha_A \leq 1/2$). In other words, rating deviations tend to occur when the aversion or relative aversion is low, because the agencies with low aversion prioritize attracting green investors over rating accuracy.

Given fixed agency aversion parameters, Proposition 4 further indicates that rating deviations become more likely either when the green target e is lower or when the proportion of green investors p is higher. Specifically, the condition (4.1) shows that lower green targets increase the likelihood of deviations. This is because smaller targets are easier and less costly for agencies to satisfy their sustainability requirements. From (4.2) and (4.3), the magnitude of deviation under these conditions is typically smaller. On the other hand, a higher proportion of green investors amplifies market competition, motivating agencies to deviate ratings more aggressively to capture this larger market segment. When agency B has low relative aversion, the degree of deviation grows with the proportion of green investors, yet it remains constrained by an upper limit for a fixed green target according to (4.3).

Next, we study how rating deviation can generate a positive association between issued ESG ratings and firms' stock returns. Khan et al. (2016) find that the stocks with good ESG ratings on the material issues have better returns than those with poor ESG ratings on the material issues. Berg et al. (2021) also document that firms with higher stock returns tend to receive upward revisions in historical ESG scores, suggesting that ESG ratings are influenced by financial performance. Our model provides a theoretical explanation for this pattern by characterizing when rating deviation results in a positive link between ESG scores and stock returns. The following proposition describes the conditions of positive association.

Proposition 5. *Suppose $\alpha_A > \alpha_B$, and the aversion parameters satisfy the conditions in* **Region I–II**, *i.e.,* $\max\{\alpha_A/2, \alpha_B\} \leq \alpha_{e,p}$. *Then the issued ESG ratings by Agency B with lower aversion are given by $\boldsymbol{g}_B^* = \boldsymbol{g}_e^*$, and the cross-sectional association $\boldsymbol{\mu}'\boldsymbol{g}_B^*$ satisfies*

$$\boldsymbol{\mu}'\boldsymbol{g}_B^* = \boldsymbol{\mu}'\tilde{\boldsymbol{g}}_T + (e - e_T)\frac{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1} \cdot (\boldsymbol{\mu} - \bar{\mu}\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})}{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\mu} - (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2},$$

where $\bar{\mu} = \boldsymbol{\mu}'\mathbf{1}/n$. The cross-sectional association $\boldsymbol{\mu}'\boldsymbol{g}_B^*$ is positive if and only if

$$e > e_T - \frac{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\mu} - (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})}\boldsymbol{\mu}'\tilde{\boldsymbol{g}}_T.$$

Proposition 5 demonstrates that in **Region I–II**, when the ESG target e exceeds a certain threshold, the issued rating selected by green investors is positively associated with stock returns, even if the underlying ESG fundamentals negatively relate to returns. In other words, the agency with a low aversion has the incentive to issue a rating that tends to align with stock returns.

4.2 ESG Rating Disagreement

We now formalize the conditions for persistent disagreement in ESG ratings, defined as agencies that systematically issue different ESG ratings (Chatterji et al., 2016; Berg et al., 2022), i.e., $\boldsymbol{g}_A^* \neq \boldsymbol{g}_B^*$ in equilibrium. By Theorem 1, the emergence of ESG rating disagreement occurs in **Region II–III** where the issued rating of the agency with lower aversion deviates and the other is just the ESG fundamental. The following proposition identifies critical thresholds in these parameters.

Proposition 6. *When $\alpha_A > \alpha_B$, a necessary condition for the existence of rating disagreement is*

$$e > e_T + \sqrt{\frac{p}{\alpha_A}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}.$$

Under this condition, rating disagreement arises if either (i) $e \leq e_T + \sqrt{\frac{p}{\alpha_B}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}$, or (ii) $\alpha_A \geq 2\alpha_B$. Furthermore, the extent of rating disagreement increases as the rating deviation e becomes larger.

This proposition reflects a moral hierarchy between agencies. Rating disagreement arises when Agency A issues the ESG fundamental, while Agency B deviates the rating. Crucially, rating disagreement is equivalent to market segmentation driven by investor choice. In other words, if the ESG rating disagreement is found, B captures all green investors and half of the others, securing a dominant majority share $(1+p)/2$, while A retains the residual fraction $(1-p)/2$. This bifurcation occurs because B has a lower aversion, allowing ratings that better cater to green investors' preferences, even at the cost of deviating from the ESG fundamental.

4.3 Trade-off between Investor Utility and Portfolio's Fundamental Greenness

We focus on the greenness of green investors (who choose ESG rating $\boldsymbol{g}_B^*$ issued by Agency B in equilibrium) because only their portfolio allocations are directly affected by ESG constraints, providing the channel through which ESG ratings affect market-level greenness. We distinguish two notions of greenness—the fundamental greenness and the observed greenness—of the portfolio held by green investors as:

$$G_F(\boldsymbol{\omega}) := \frac{\boldsymbol{\omega}'\tilde{\boldsymbol{g}}_T}{\boldsymbol{\omega}'\mathbf{1}} \qquad \text{and} \qquad G_O(\boldsymbol{\omega}) := \frac{\boldsymbol{\omega}'\boldsymbol{g}_B^*}{\boldsymbol{\omega}'\mathbf{1}}. \tag{4.4}$$

By construction, G_F is the weighted average of the ESG fundamental under the portfolio $\boldsymbol{\omega}$, measuring fundamental greenness. G_O is the weighted average of the issued ESG rating, which is observable for investors.

In the following, we compare the green investor's utility and fundamental greenness in **Region II** across two scenarios: (i) portfolios optimized under the ESG fundamental $\tilde{\boldsymbol{g}}_T$ as a hypothetical (unattainable) benchmark and (ii) portfolios optimized under the issued ESG ratings $\boldsymbol{g}_B^*$ as the

reality.

Proposition 7. *Suppose that the ESG target e, the fraction of green investors p, and the aversion parameters α_A and α_B satisfy the conditions in* **Region II**, *i.e., $\alpha_B \leq \alpha_{e,p} < \alpha_A \leq 2\alpha_{e,p}$. Then the green investor chooses the ESG rating $\boldsymbol{g}_e^*$ issued by Agency B, and we have the following results:*

(1) The pecuniary utility of the green investor under the issued ESG rating $\boldsymbol{g}_e^*$ with a high deviation is

$$U(e, \boldsymbol{g}_e^*) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2},$$

and the pecuniary utility of the green investor under the ESG fundamental $\tilde{\boldsymbol{g}}_T$ without deviation is

$$U(e, \tilde{\boldsymbol{g}}_T) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2} - \frac{(\boldsymbol{\omega}_M'\mathbf{1})^2(e - e_T)^2}{(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)} < U(e, \boldsymbol{g}_e^*). \tag{4.5}$$

(2) The observed greenness under the portfolio $\boldsymbol{\omega}_e^* = \boldsymbol{\omega}(\gamma, e, \boldsymbol{g}_e^*)$ is

$$G_O(\boldsymbol{\omega}_e^*) = e,$$

the fundamental greenness under the portfolio $\boldsymbol{\omega}_e^* = \boldsymbol{\omega}(\gamma, e, \boldsymbol{g}_e^*)$ is

$$G_F(\boldsymbol{\omega}_e^*) = e_T,$$

and the fundamental greenness under the portfolio $\boldsymbol{\omega}_T^* = \boldsymbol{\omega}(\gamma, e, \tilde{\boldsymbol{g}}_T)$ is

$$G_F(\boldsymbol{\omega}_T^*) = e > G_F(\boldsymbol{\omega}_e^*). \tag{4.6}$$

From (4.5) and (4.6) in Proposition 7, we conclude that

$$\Big(U(e, \boldsymbol{g}_e^*) - U(e, \tilde{\boldsymbol{g}}_T)\Big) \cdot \Big(G_F(\boldsymbol{\omega}_e^*) - G_F(\boldsymbol{\omega}_T^*)\Big) < 0.$$

This inequality reveals a systematic trade-off: the allocation that maximizes utility under the issued ratings generally does not maximize fundamental greenness, and vice versa. This trade-off is driven by rating deviation. Proposition 7 also shows that a greater rating deviation leads to a

higher fundamental greenness discrepancy for the green investor. This phenomenon occurs because a higher ESG target e leads to a larger rating deviation, which means that the ESG rating of the popular assets in the MVO portfolio becomes unreliably high. As a result, the discrepancy in fundamental greenness between the two scenarios increases.

4.4 Effect of Firm-level Greenwashing

The recent literature has documented the practice of greenwashing, where certain companies market their products or services as environmentally friendly or sustainable when they are not, or when their claims are exaggerated or misleading (Baker et al., 2024). In our framework, the ESG fundamental is $\tilde{\boldsymbol{g}}_T$ and we assume that each firm i greenwashes by a level of δ_i. The normalized adjustment observed by rating agencies is the post-greenwashing ESG fundamental

$$\tilde{\boldsymbol{g}}_{T,\mathrm{wa}} := \tilde{\boldsymbol{g}}_T + \sum_{i=1}^{n} \delta_i \left(\boldsymbol{e}_i - \frac{1}{n}\mathbf{1} \right),$$

which incorporates the normalization ensuring that Assumption 2 still holds for $\tilde{\boldsymbol{g}}_{T,\mathrm{wa}}$. As a result, the firm i's ESG rating increases if and only if

$$\delta_i > \frac{1}{n} \sum_{j=1}^{n} \delta_j.$$

For instance, if only firm i engages in greenwashing by a level of δ_i, then its ESG rating increases by $(n-1)\delta_i/n$ in $\tilde{\boldsymbol{g}}_{T,\mathrm{wa}}$, while the ESG ratings of all other firms decrease by δ_i/n.

The presence of greenwashing requires an extension of our baseline model. We first characterize equilibrium behaviors under greenwashing, and then examine the corresponding rating deviation as well as the trade-off between investor utility and the portfolio's fundamental greenness.

Equilibrium Behavior under Greenwashing. To characterize the equilibrium behavior, we define the post-greenwashing ESG benchmark as

$$e_{T,\mathrm{wa}} = \frac{\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}}{\boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \mathbf{1}}.$$

Proposition A.1 characterizes the equilibrium ratings issued by rating agencies, which clarifies the implications of different realizations of $e_{T,\text{wa}}$ for the equilibrium ratings. This result is in parallel to the equilibrium structure described in Theorem 1. The detailed formulation of Proposition A.1 and its proof are deferred to Appendix A.11. Note that

$$\begin{aligned} e_{T,\text{wa}} - e_T &= \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\left(\tilde{\boldsymbol{g}}_{T,\text{wa}} - \tilde{\boldsymbol{g}}_T\right)}{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} \\ &= \sum_{i=1}^{n} \delta_i \left(\frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_i}{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} - \frac{1}{n} \right) \\ &= \sum_{i=1}^{n} \frac{\delta_i}{n} \left(\frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_i}{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}/n} - 1 \right). \end{aligned} \tag{4.7}$$

By (4.7), if the stock of firm i has a large weight in the MVO portfolio (higher than the average weight, i.e., $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_i > \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}/n$), greenwashing by firm i makes the difference $e_{T,\text{wa}} - e_T$ higher. Similarly, if the stock of firm i has a small weight in the MVO portfolio (lower than the average weight, i.e., $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_i < \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}/n$), greenwashing by firm i makes the difference $e_{T,\text{wa}} - e_T$ lower.

We refer to the difference $|e_{T,\text{wa}} - e_T|$ as the magnitude of greenwashing. Building on Proposition A.1, the model yields three cases depending on the value of $|e_{T,\text{wa}} - e_T|$, which provide the theoretical foundation for our subsequent analysis of the economic implications of greenwashing.

If the magnitude of greenwashing $|e_{T,\text{wa}} - e_T|$ is not large such that $b \leq e_{T,\text{wa}} < e$ still holds, we can directly apply Theorem 1 by treating $e_{T,\text{wa}}$ as e_T.

If the magnitude of greenwashing $|e_{T,\text{wa}} - e_T|$ is large enough so that $e_{T,\text{wa}} \geq e$, both ESG targets e and b are attained by the MVO portfolio $\boldsymbol{\omega}_M$ and the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\text{wa}}$. In this case, the two agencies just issue the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\text{wa}}$ after greenwashing and the optimal portfolios of both investors are the MVO portfolio.

If the magnitude of greenwashing $|e_{T,\text{wa}} - e_T|$ is large enough so that $e_{T,\text{wa}} < b$, the ESG target b of the brown investor is not attained by the MVO portfolio $\boldsymbol{\omega}_M$ and the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\text{wa}}$.In this case, we derive the equilibrium behavior in Proposition 8, and the rating profiles in equilibrium are shown in the proof in Appendix A.11.

The role of greenwashing in rating deviation. In order to analyze how the magnitude of greenwashing $|e_{T,\text{wa}} - e_T|$ affects rating deviation, we denote by

$$\Delta \boldsymbol{g} = \frac{\tilde{\boldsymbol{g}}_{T,\text{wa}} - \tilde{\boldsymbol{g}}_T}{e_{T,\text{wa}} - e_T}$$

the normalized variation of the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\text{wa}}$, and define the set $\mathcal{D}$ of all possible normalized variations by

$$\mathcal{D} = \left\{ \Delta \boldsymbol{g} \in \mathbb{R}^n : \text{ There exists an ESG rating } \tilde{\boldsymbol{g}}_{T,\text{wa}}, \text{ s.t. } \Delta \boldsymbol{g} = \frac{\tilde{\boldsymbol{g}}_{T,\text{wa}} - \tilde{\boldsymbol{g}}_T}{e_{T,\text{wa}} - e_T} \right\}.$$

Given that any arbitrary vector $\Delta \boldsymbol{g} \in \mathcal{D}$ satisfies the conditions $\mathbf{1}'\Delta \boldsymbol{g} = 0$ and $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\Delta \boldsymbol{g} = 1$, we can characterize the set $\mathcal{D}$ as:

$$\mathcal{D} = \left\{ \Delta \boldsymbol{g} \in \mathbb{R}^n : \ \mathbf{1}'\Delta \boldsymbol{g} = 0, \ \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\Delta \boldsymbol{g} = 1 \right\}.$$

Now we show the relationship between the magnitude of greenwashing $|e_{T,\text{wa}} - e_T|$ and rating deviation in the following proposition.

Proposition 8. *Suppose $\alpha_A > \alpha_B$. Fix any normalized variation $\Delta \boldsymbol{g} \in \mathcal{D}$ and any magnitude of greenwashing $|e_{T,\text{wa}} - e_T| \geq 0$, and define the corresponding post-greenwashing ESG fundamental by $\tilde{\boldsymbol{g}}_{T,\text{wa}} = \tilde{\boldsymbol{g}}_T + (e_{T,\text{wa}} - e_T)\Delta \boldsymbol{g}$. Let $(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*)$ denote the issued rating profile in SSE under $\tilde{\boldsymbol{g}}_{T,\text{wa}}$.*

(1) Assume that $\frac{\alpha_A}{2}\|\tilde{\boldsymbol{g}}_{T,\text{wa}} - \tilde{\boldsymbol{g}}_T\|^2 \notin \{p, 1-p\}$, then the issued ESG ratings $\boldsymbol{g}_A^*$ and $\boldsymbol{g}_B^*$ always deviate from the ESG fundamental $\tilde{\boldsymbol{g}}_T$.

(2) When $\alpha_B/\alpha_A \in (1/2, 1)$, Agency j deviates from the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\text{wa}}$ for $j \in \{A,B\}$ if and only if

$$\max\left\{b, e - \sqrt{\frac{p}{\alpha_j}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}\right\} \leq e_{T,\text{wa}} < e \quad \text{or} \quad b - \sqrt{\frac{1-p}{\alpha_j}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]} \leq e_{T,\text{wa}} < b. \tag{4.8}$$

(3) When $\alpha_B/\alpha_A \in (0, 1/2]$, Agency A deviates from the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\text{wa}}$ if (4.8) holds for $j = A$, and Agency B deviates from $\tilde{\boldsymbol{g}}_{T,\text{wa}}$ if and only if $e_{T,\text{wa}} < e$.

Proposition 8 shows how greenwashing shapes rating deviations. The issued ESG rating deviates from the post-greenwashing fundamental $\tilde{\boldsymbol{g}}_{T,\mathrm{wa}}$ only when the intrinsic ESG benchmark $e_{T,\mathrm{wa}}$ falls into the threshold regions positioned immediately below the ESG targets e or b.

Moreover, when greenwashing is mild, the post-greenwashing ESG benchmark remains close to the benchmark without greenwashing, only small strategic deviations are needed and the overall rating deviation stays limited. In contrast, when greenwashing is more severe, the post-greenwashing fundamental itself can already be far from the ESG fundamental, so the rating deviation remains large even though there is no extra deviation for issuing the post-greenwashing fundamental in equilibrium. As a result, under appropriate conditions, a larger magnitude of greenwashing $|e_{T,\mathrm{wa}} - e_T|$ leads to a greater rating deviation of the issued rating from the true ESG fundamental $\tilde{\boldsymbol{g}}_T$.

The effect of greenwashing on utility-greenness trade-off. In order to derive the effects of the magnitude of greenwashing on investor utilities and fundamental greenness, we consider a more specific case with some additional technical assumptions in the following proposition. The proof is shown in Appendix A.13, where we give the formulations of investor utility and fundamental greenness. In Appendix A.14, we show that the technical condition (4.9) is not very restrictive.

Proposition 9. *Assume that the post-greenwashing ESG fundamental is* $\tilde{\boldsymbol{g}}_{T,\mathrm{wa}} = \tilde{\boldsymbol{g}}_T + (e_{T,\mathrm{wa}} - e_T)\boldsymbol{g}_M$*, where* $\boldsymbol{g}_M$ *is defined in* (3.4)*, and suppose*

$$(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M \geq (e - e_T)\boldsymbol{g}_M'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M. \tag{4.9}$$

Then we have the following results:

(1) When $\alpha_B/\alpha_A \in (1/2, 1)$*, for both types of investors, utility is non-decreasing on* $e_{T,\mathrm{wa}}$*, and fundamental greenness is non-increasing on* $e_{T,\mathrm{wa}}$*.*

(2) When $\alpha_B/\alpha_A \in (0, 1/2]$ *and*

$$e_{T,\mathrm{wa}} \geq e - \sqrt{\frac{2p}{\alpha_A}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}, \tag{4.10}$$

for both types of investors, utility is non-decreasing on $e_{T,\mathrm{wa}}$*, and fundamental greenness is non-increasing on* $e_{T,\mathrm{wa}}$*.*

Proposition 9 shows that greenwashing by firms whose stocks have a large weight in the MVO portfolio increases $e_{T,\mathrm{wa}}$, i.e., the weighted ESG score of the MVO portfolio becomes higher. This allows the investor with ESG target e to invest more in the stocks with high returns and relatively low ESG scores, thereby increasing utility while reducing fundamental greenness. On the contrary, greenwashing by firms with a small weight in the MVO portfolio lowers $e_{T,\mathrm{wa}}$, i.e., the weighted ESG score of the MVO portfolio becomes lower. This forces the investor to hold more high-ESG but low-return stocks, which reduces utility and increases fundamental greenness. As a result, greenwashing has opposite impacts on utility and fundamental greenness for investors, which distorts the trade-off between utility and fundamental greenness for investors. Particularly, if firms with low weights in the MVO portfolio engage heavily enough in greenwashing, the post-greenwashing ESG benchmark $e_{T,\mathrm{wa}}$ becomes lower than the ESG target b, and then both agencies issue the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\mathrm{wa}}$. In this case, the ESG target b of the brown investor is not attained by the MVO portfolio, resulting in a utility loss for the brown investor.

5 Conclusion

We model the interaction between an ESG-conscious investor and two competing ESG rating agencies within a principal–agent framework. Our framework aims to capture the intrinsic market forces behind key empirical phenomena in the ESG literature, including rating deviation and disagreement. By characterizing agencies' aversion to deviation as low, medium, or high, and examining their relative differences, we identify four types of equilibrium behavior.

We show that ESG ratings deviate from the ESG fundamentals when at least one agency's aversion or the relative aversion is low. Furthermore, if investors' ESG preference is sufficiently strong, issued ESG ratings can become positively correlated with asset returns, even when the underlying ESG fundamental is negatively correlated with returns. Rating disagreement emerges when one agency has a low (or relatively low) aversion while the other does not. The lower aversion agency monopolizes the market of green investors. In addition, rating deviation leads to a trade-off between investors' pecuniary utility and the portfolio's fundamental greenness.

Moreover, we analyze the effects of firm-level greenwashing, interpreted as distortions to the ESG fundamentals observed by rating agencies. Greenwashing systematically exacerbates rating

deviation, and distorts the utility-greenness trade-off. Notably, when firms with low weights in the MVO portfolio engage heavily in greenwashing, the intrinsic benchmark used for ESG screening can fall substantially. In such cases, all ESG-conscious investors would be forced into a less efficient allocation in order to satisfy the ESG constraint, leading to a higher pecuniary utility loss.

These findings carry implications for regulators and investors. For regulators, the model underscores the need for penalties that effectively deter agencies from issuing strategically deviated ratings. It is also important that the intensity of enforcement remains consistent across agencies. For investors who do not have full visibility into firms' return and risk characteristics, our analysis shows that rating disagreement can be informative, as differences between agencies' ratings may partially reflect risk-adjusted return information.

References

Agrawal, S., Liu, L. Y., Rajgopal, S., Sridharan, S. A., Yan, Y., and Yohn, T. L. (2025). ESG ratings of ESG index providers. Columbia Business School Research Paper No. 4468531. Retrieved from `https://ssrn.com/abstract=4468531`.

Amel-Zadeh, A. and Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal*, 74(3):87–103.

Avramov, D., Cheng, S., Lioui, A., and Tarelli, A. (2022). Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics*, 145(2):642–664.

Azarmsa, E. and Shapiro, J. D. (2024). The market for ESG ratings. SSRN Electronic Journal. Retrieved from `https://ssrn.com/abstract=4236912`.

Baker, A. C., Larcker, D. F., McClure, C. G., Saraph, D., and Watts, E. M. (2024). Diversity washing. *Journal of Accounting Research*, 62(5):1661–1709.

Berg, F., Fabisik, K., and Sautner, Z. (2021). Is history repeating itself? The (un)predictable past of ESG ratings. SSRN Electronic Journal. Retrieved from `https://ssrn.com/abstract=3806084`.

Berg, F., Koelbel, J. F., and Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6):1315–1344.

Billio, M., Costola, M., Hristova, I., Latino, C., and Pelizzon, L. (2021). Inside the ESG ratings: (dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, 28(5):1426–1445.

Bolton, P., Freixas, X., and Shapiro, J. (2012). The credit ratings game. *The Journal of Finance*, 67(1):85–111.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK.

Cartellier, F., Tankov, P., and Zerbib, O. D. (2025). Can investors curb greenwashing? *Journal of Economic Dynamics and Control*, 180:105195.

Chatterji, A. K., Durand, R., Levine, D. I., and Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37(8):1597–1614.

Christensen, D. M., Serafeim, G., and Sikochi, A. (2022). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review*, 97(1):147–175.

De Angelis, T., Tankov, P., and Zerbib, O. D. (2023). Climate impact investing. *Management Science*, 69(12):7669–7692.

DeMarzo, P. and Duffie, D. (1999). A liquidity-based model of security design. *Econometrica*, 67(1):65–99.

DeMarzo, P. M. (2005). The pooling and tranching of securities: A model of informed intermediation. *Review of Financial Studies*, 18(1):1–35.

Eccles, R. G., Kastrapeli, M. D., and Potter, S. J. (2017). How to integrate ESG into investment decision-making: Results of a global survey of institutional investors. *Journal of Applied Corporate Finance*, 29(4):125–133.

Glicksberg, I. L. (1952). A further generalization of the kakutani fixed point theorem. *Proceedings of the American Mathematical Society*, 3(1):170–174.

Global Sustainable Investment Alliance (2025). Global sustainable investment review 2024. Retrieved from https://www.gsi-alliance.org.

Guo, N., Kou, S., Wang, B., and Wang, R. (2025). A theory of credit rating criteria. *Management Science*, 71(4):3583–3599.

Hartzmark, S. M. and Sussman, A. B. (2019). Do investors value sustainability? a natural experiment examining ranking and fund flows. *The Journal of Finance*, 74(6):2789–2837.

Heller, Y. and Winter, E. (2016). Rule rationality. *International Economic Review*, 57(3):997–1026.

Khan, M., Serafeim, G., and Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *The Accounting Review*, 91(6):1697–1724.

Kim, Y., Park, M. S., and Wier, B. (2012). Is earnings quality associated with corporate social responsibility? *The Accounting Review*, 87(3):761–796.

Krueger, P., Sautner, Z., and Starks, L. T. (2020). The importance of climate risks for institutional investors. *The Review of Financial Studies*, 33(3):1067–1111.

Larcker, D. F., Pomorski, L., Tayan, B., and Watts, E. M. (2022). ESG ratings: A compass without direction. Rock Center for Corporate Governance at Stanford University Working Paper. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4179647.

Lo, A. W., Wu, L., Zhang, R., and Zhao, C. (2024). Optimal impact portfolios with general dependence and marginals. *Operations Research*.

Lo, A. W. and Zhang, R. (2024). Quantifying the impact of impact investing. *Management Science*, 70(10):7161–7186.

Lo, A. W. and Zhang, R. (2025). Performance attribution for portfolio constraints. *Management Science*, 71(9):7537–7559.

Lovo, S. and Olivier, J. (2025). Who should pay for ESG ratings? *Review of Finance*.

Mailath, G. J. and Samuelson, L. (2006). *Repeated games and reputations: Long-run relationships*. Oxford University Press.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

MSCI (2025). ESG fund ratings methodology. Retrieved from `https://www.msci.com/documents/1296102/34424357/MSCI+ESG+Fund+Ratings+Methodology.pdf`.

Osborne, M. J. (2004). *An introduction to game theory*. Oxford University Press.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2):550–571.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2022). Dissecting green returns. *Journal of Financial Economics*, 146(2):403–424.

Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, 142(2):572–597.

Piccolo, A. and Shapiro, J. (2022). Credit ratings and market information. *The Review of Financial Studies*, 35(10):4425–4473.

Starks, L. T. (2023). Presidential address: Sustainable finance and ESG issues—value versus values. *The Journal of Finance*, 78(4):1837–1872.

Wu, Y., Zhang, K., and Xie, J. (2020). Bad greenwashing, good greenwashing: Corporate social responsibility and information transparency. *Management Science*, 66(7):3095–3112.

Zerbib, O. D. (2022). A sustainable capital asset pricing model (s-capm): Evidence from environmental integration and sin stock exclusion. *Review of Finance*, 26(6):1345–1388.

A Proofs

A.1 Proof of Proposition 1

Let $D = \{\boldsymbol{\omega} : \boldsymbol{\omega}'\boldsymbol{g}/\boldsymbol{\omega}'\mathbf{1} \geq t\}$ be the admissible set. Our goal is to maximize $u(\boldsymbol{\omega}) = \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}$ for $\boldsymbol{\omega} \in D$. We first consider the unconstrained maximization of $u(\boldsymbol{\omega})$ over $\mathbb{R}^n$.

For $\boldsymbol{\omega} \in \mathbb{R}^n$, the gradient and Hessian of utility $u(\boldsymbol{\omega})$ are given by

$$\nabla u = \boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\boldsymbol{\omega}, \quad \nabla^2 u = -\gamma\boldsymbol{\Sigma}.$$

Since $\boldsymbol{\Sigma}$ is positive definite, the Hessian is negative definite, which implies that u is strictly concave. Setting $\nabla u = \mathbf{0}$ yields the unique global maximizer

$$\boldsymbol{\omega}_o = \gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},$$

and the maximum value is

$$u(\boldsymbol{\omega}_o) = \boldsymbol{\omega}_o'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}_o'\boldsymbol{\Sigma}\boldsymbol{\omega}_o = \frac{1}{2\gamma}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}.$$

Now we turn to the constrained problem, where $\boldsymbol{\omega} \in D$. We distinguish two cases:

(i) When $\boldsymbol{\omega}_M'\boldsymbol{g}/\boldsymbol{\omega}_M'\mathbf{1} \geq t$: It follows from Assumption 1 that $\boldsymbol{\omega}_M'\mathbf{1} > 0$. In this case, we have $\boldsymbol{\omega}_M'(\boldsymbol{g} - t\mathbf{1}) \geq 0$. Recall that

$$\boldsymbol{\omega}_M = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|} = \frac{\gamma}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|}\boldsymbol{\omega}_o,$$

so we have $\boldsymbol{\omega}_o'(\boldsymbol{g} - t\mathbf{1}) \geq 0$, and then $\boldsymbol{\omega}_o \in D$. As a result, the utility $u(\boldsymbol{\omega})$ attains the global maximum at $\boldsymbol{\omega}^* = \boldsymbol{\omega}_o = \gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

(ii) When $\boldsymbol{\omega}_M'\boldsymbol{g}/\boldsymbol{\omega}_M'\mathbf{1} < t$: In this case, we can similarly get that $\boldsymbol{\omega}_o'(\boldsymbol{g} - t\mathbf{1}) < 0$ and $\boldsymbol{\omega}_o \notin D$, so the unconstrained maximizer is infeasible. Note that

$$D = \{\boldsymbol{\omega} :\ \boldsymbol{\omega}'\mathbf{1} > 0,\ (\boldsymbol{g} - t\mathbf{1})'\boldsymbol{\omega} \geq 0\} \ \cup\ \{\boldsymbol{\omega} :\ \boldsymbol{\omega}'\mathbf{1} < 0,\ (\boldsymbol{g} - t\mathbf{1})'\boldsymbol{\omega} \leq 0\},$$

so the admissible set D is the union of two convex sets. Since the utility $u(\boldsymbol{\omega})$ is concave, the optimization problem (2.3) attains its maximum on the boundary of D. Therefore, the problem reduces to

$$\max_{\boldsymbol{\omega}}(\boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}) \quad \text{s.t.} \quad \boldsymbol{\omega}'\boldsymbol{g} = t\boldsymbol{\omega}'\mathbf{1}.$$

To solve this, we introduce its Lagrangian function

$$\mathcal{L}(\boldsymbol{\omega}, \lambda) = \boldsymbol{\omega}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} + \lambda\boldsymbol{\omega}'(\boldsymbol{g} - t\mathbf{1}).$$

Taking first-order conditions with respect to $\boldsymbol{\omega}$ and λ, we obtain

$$\nabla_{\boldsymbol{\omega}}\mathcal{L} = \boldsymbol{\mu} - \gamma\boldsymbol{\Sigma}\boldsymbol{\omega} + \lambda(\boldsymbol{g} - t\mathbf{1}) = \mathbf{0}, \tag{A.1}$$

and

$$\nabla_{\lambda}\mathcal{L} = \boldsymbol{\omega}'(\boldsymbol{g} - t\mathbf{1}) = 0. \tag{A.2}$$

Solving (A.1) for $\boldsymbol{\omega}$ yields

$$\boldsymbol{\omega} = \gamma^{-1}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \lambda\boldsymbol{\Sigma}^{-1}(\boldsymbol{g} - t\mathbf{1})\big). \tag{A.3}$$

Substituting (A.3) into (A.2) gives

$$\lambda^* = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(t\mathbf{1} - \boldsymbol{g})}{(\boldsymbol{g} - t\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g} - t\mathbf{1})}.$$

By (A.3), the optimal solution $\boldsymbol{\omega}^*$ is given by

$$\boldsymbol{\omega}^* = \gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \lambda^*\gamma^{-1}\boldsymbol{\Sigma}^{-1}(\boldsymbol{g} - t\mathbf{1}).$$

By substituting the above $\boldsymbol{\omega}^*$ into the utility $u(\boldsymbol{\omega})$, we can derive that

$$u(\boldsymbol{\omega}^*) = \frac{1}{2\gamma}\left(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(t\mathbf{1} - \boldsymbol{g}))^2}{(\boldsymbol{g} - t\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g} - t\mathbf{1})}\right) = \frac{1}{2\gamma}\left(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \lambda^*\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(t\mathbf{1} - \boldsymbol{g})\right).$$

Combining two cases, we get the expressions (2.4) and (2.5) and complete the proof.

A.2 Proof of Proposition 2

It is obvious that $\boldsymbol{g}_s^* = \tilde{\boldsymbol{g}}_T$ when $s = e_T$. We assume that $s > e_T$ in the following. By Assumption 1, the objective function is strictly convex and the constraints are affine, so the Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient for optimality (see Boyd and Vandenberghe, 2004). Let $\lambda_1 \geq 0$ and $\lambda_2 \in \mathbb{R}$ be the Lagrange multipliers associated with the inequality and equality constraints, respectively. The Lagrangian is

$$\mathcal{L}(\boldsymbol{g}, \lambda_1, \lambda_2) = \frac{1}{2}\|\boldsymbol{g} - \tilde{\boldsymbol{g}}_T\|^2 - \lambda_1(\boldsymbol{\omega}_M'\boldsymbol{g} - s\boldsymbol{\omega}_M'\mathbf{1}) + \lambda_2\boldsymbol{g}'\mathbf{1}.$$

Differentiating the Lagrangian with respect to $\boldsymbol{g}$ yields the first-order condition

$$\nabla_g \mathcal{L} = \boldsymbol{g} - \tilde{\boldsymbol{g}}_T - \lambda_1\boldsymbol{\omega}_M + \lambda_2\mathbf{1}.$$

Then, the KKT conditions are

$$\begin{cases} \boldsymbol{g} - \tilde{\boldsymbol{g}}_T - \lambda_1\boldsymbol{\omega}_M + \lambda_2\mathbf{1} = \mathbf{0}, \\ \boldsymbol{\omega}_M'\boldsymbol{g} - s\,\boldsymbol{\omega}_M'\mathbf{1} \geq 0, \quad \boldsymbol{g}'\mathbf{1} = 0, \\ \lambda_1 \geq 0, \quad \lambda_1(\boldsymbol{\omega}_M'\boldsymbol{g} - s\,\boldsymbol{\omega}_M'\mathbf{1}) = 0. \end{cases} \tag{A.4}$$

Let $(\boldsymbol{g}_s^*, \lambda_1^*, \lambda_2^*)$ be the solution of the system (A.4). Thus,

$$\boldsymbol{g}_s^* = \tilde{\boldsymbol{g}}_T + \lambda_1^*\boldsymbol{\omega}_M - \lambda_2^*\mathbf{1}. \tag{A.5}$$

It follows from $s > e_T = \boldsymbol{\omega}_M'\tilde{\boldsymbol{g}}_T/\boldsymbol{\omega}_M'\mathbf{1}$ that $\boldsymbol{\omega}_M'\tilde{\boldsymbol{g}}_T - s\boldsymbol{\omega}_M'\mathbf{1} < 0$. If $\lambda_1^* = 0$, we have $\boldsymbol{g}_s^* = \tilde{\boldsymbol{g}}_T + \lambda_2^*\mathbf{1}$. Under Assumption 2, we have both $\tilde{\boldsymbol{g}}_T'\mathbf{1} = 0$ and $\boldsymbol{g}_s^{*\prime}\mathbf{1} = 0$, so it follows that $\lambda_2^* = 0$ and $\boldsymbol{g}_s^* = \tilde{\boldsymbol{g}}_T$, which contradicts with $\boldsymbol{\omega}_M'\boldsymbol{g}_s^* - s\,\boldsymbol{\omega}_M'\mathbf{1} \geq 0$. This implies that $\lambda_1^* > 0$ and $\boldsymbol{\omega}_M'\boldsymbol{g}_s^* = s\boldsymbol{\omega}_M'\mathbf{1}$.
Combining these with (A.5), the system (A.4) converts into the linear system for λ_1^* and λ_2^* as follows,

$$\begin{cases} \boldsymbol{\omega}_M'\tilde{\boldsymbol{g}}_T + \lambda_1^*\,\boldsymbol{\omega}_M'\boldsymbol{\omega}_M - \lambda_2^*\,\boldsymbol{\omega}_M'\mathbf{1} - s\,\boldsymbol{\omega}_M'\mathbf{1} = 0, \\ \tilde{\boldsymbol{g}}_T'\mathbf{1} + \lambda_1^*\,\boldsymbol{\omega}_M'\mathbf{1} - \lambda_2^*\,\mathbf{1}'\mathbf{1} = 0. \end{cases} \tag{A.6}$$

Solving the above system, we obtain that

$$\begin{cases} \lambda_1^* &= \frac{n(\boldsymbol{\omega}_M'\mathbf{1})}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}s - \frac{n}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}\boldsymbol{\omega}_M'\tilde{\boldsymbol{g}}_T + \frac{(\boldsymbol{\omega}_M'\mathbf{1})}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}\tilde{\boldsymbol{g}}_T'\mathbf{1}, \\ \lambda_2^* &= \frac{(\boldsymbol{\omega}_M'\mathbf{1})^2}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}s - \frac{(\boldsymbol{\omega}_M'\mathbf{1})}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}\boldsymbol{\omega}_M'\tilde{\boldsymbol{g}}_T + \frac{1}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}\tilde{\boldsymbol{g}}_T'\mathbf{1}. \end{cases}$$

By $\tilde{\boldsymbol{g}}_T'\mathbf{1} = 0$ and $e_T = \boldsymbol{\omega}_M'\tilde{\boldsymbol{g}}_T/\boldsymbol{\omega}_M'\mathbf{1}$, we have

$$\begin{aligned} \lambda_1^* &= s\frac{n(\boldsymbol{\omega}_M'\mathbf{1})}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2} - e_T\frac{n(\boldsymbol{\omega}_M'\mathbf{1})}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2} + \frac{\tilde{\boldsymbol{g}}_T'\mathbf{1}}{n}\frac{n(\boldsymbol{\omega}_M'\mathbf{1})}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2} \\ &= (s-e_T)\cdot\frac{n(\boldsymbol{\omega}_M'\mathbf{1})}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}, \end{aligned}$$

and

$$\begin{aligned} \lambda_2^* &= s\frac{(\boldsymbol{\omega}_M'\mathbf{1})^2}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2} - e_T\frac{(\boldsymbol{\omega}_M'\mathbf{1})^2}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2} + \frac{\tilde{\boldsymbol{g}}_T'\mathbf{1}}{n}\frac{n}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2} \\ &= (s-e_T)\cdot\frac{(\boldsymbol{\omega}_M'\mathbf{1})^2}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}. \end{aligned}$$

Substituting those into (A.5), we have

$$\boldsymbol{g}_s^* = \tilde{\boldsymbol{g}}_T + (s-e_T)\cdot\frac{n(\boldsymbol{\omega}_M'\mathbf{1})}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}\boldsymbol{\omega}_M - (s-e_T)\cdot\frac{(\boldsymbol{\omega}_M'\mathbf{1})^2}{n-(\boldsymbol{\omega}_M'\mathbf{1})^2}\mathbf{1}.$$

By the expression (3.5) of $\boldsymbol{g}_M$, it simplifies to (3.4). Therefore, we derive the unique solution g_s^* and complete the proof.

A.3 Lemma A.1 and its Proof

Lemma A.1. *Under Assumptions 1 and 2, for fixed $s \in [e_T, e]$, there exists a solution $\boldsymbol{g}_s^{**}$ for the optimization problem*

$$\min_{\boldsymbol{g}\in\mathbb{R}^n} \frac{(\boldsymbol{\omega}_M'(\boldsymbol{g}-e\mathbf{1}))^2}{(\boldsymbol{g}-e\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g}-e\mathbf{1})} \quad s.t. \quad \|\boldsymbol{g}-\tilde{\boldsymbol{g}}_T\|^2 \le \|\boldsymbol{g}_s^*-\tilde{\boldsymbol{g}}_T\|^2 \quad and \quad \boldsymbol{g}'\mathbf{1}=0. \tag{A.7}$$

In addition, for any solution $\boldsymbol{g}_s^{**}$, we have

$$\|\boldsymbol{g}_s^{**}-\tilde{\boldsymbol{g}}_T\|^2 = \|\boldsymbol{g}_s^*-\tilde{\boldsymbol{g}}_T\|^2. \tag{A.8}$$

Proof. When $s = e$, the problem (A.7) has the solution $\boldsymbol{g}_e^*$ with the global minimum value zero. It suffices to consider the case that $e_T \leq s < e$. In order to prove the existence of the solution for (A.7), we aim to convert the problem into a convex program by an appropriate change of variables. Recall $\boldsymbol{\Sigma}$ is positive definite, indicating that $\boldsymbol{\Sigma}^{-\frac{1}{2}}$ is as well. Define $\boldsymbol{\alpha}_1 = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}$ and $\boldsymbol{\alpha}_2 = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{1}$. Let $\boldsymbol{Q} = (\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2, \ldots, \hat{\boldsymbol{\alpha}}_n) \in \mathbb{R}^{n\times n}$ be an orthonormal basis obtained by Gram–Schmidt with

$$\hat{\boldsymbol{\alpha}}_1 = \frac{\boldsymbol{\alpha}_1}{\|\boldsymbol{\alpha}_1\|},\ \hat{\boldsymbol{\alpha}}_2 = \frac{\boldsymbol{\alpha}_2 - \langle\hat{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\rangle\hat{\boldsymbol{\alpha}}_1}{\|\boldsymbol{\alpha}_2 - \langle\hat{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\rangle\hat{\boldsymbol{\alpha}}_1\|}.$$

Then

$$\boldsymbol{\alpha}_1'\boldsymbol{Q} = (\|\boldsymbol{\alpha}_1\|, 0, \ldots, 0) = \|\boldsymbol{\alpha}_1\|\boldsymbol{e}_1', \tag{A.9}$$

and

$$\boldsymbol{\alpha}_2'\boldsymbol{Q} = (\langle\hat{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\rangle, \langle\hat{\boldsymbol{\alpha}}_2, \boldsymbol{\alpha}_2\rangle, 0, \ldots, 0) = \langle\hat{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\rangle\boldsymbol{e}_1' + \langle\hat{\boldsymbol{\alpha}}_2, \boldsymbol{\alpha}_2\rangle\boldsymbol{e}_2'. \tag{A.10}$$

We introduce the orthogonal change of variables

$$\boldsymbol{z} = (z_1, z_2, \ldots, z_n)' := \boldsymbol{Q}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{g} - e\mathbf{1}).$$

In particular, let $\boldsymbol{z}^* = \boldsymbol{Q}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{g}_s^* - e\mathbf{1})$. Since $\boldsymbol{Q}$ is orthonormal and $\boldsymbol{\Sigma}^{-1/2}$ is invertible, the mapping $\boldsymbol{g} \mapsto \boldsymbol{z}$ is bijective and affine, with inverse

$$\boldsymbol{g} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}\,\boldsymbol{z} + e\mathbf{1}. \tag{A.11}$$

By $\boldsymbol{\alpha}_1 = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ and (A.9), we have $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Q} = \|\boldsymbol{\alpha}_1\|\boldsymbol{e}_1'$. Hence,

$$\begin{aligned}
\boldsymbol{\omega}_M'(\boldsymbol{g} - e\mathbf{1}) &= \frac{(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})'(\boldsymbol{g} - e\mathbf{1})}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|} \\
&= \frac{(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})'\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{g} - e\mathbf{1})}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|} \\
&= \frac{(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{z}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|} = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{Q}\boldsymbol{z}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|} \\
&= \frac{\|\boldsymbol{\alpha}_1\|}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|}\boldsymbol{e}_1'\boldsymbol{z} = \frac{\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\|}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|}z_1.
\end{aligned} \tag{A.12}$$

On the other hand,

$$\begin{aligned}(\boldsymbol{g}-e\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g}-e\mathbf{1}) &= (\boldsymbol{g}-e\mathbf{1})'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{g}-e\mathbf{1}) \\ &= \boldsymbol{z}'\boldsymbol{z} = \|\boldsymbol{z}\|^2.\end{aligned} \tag{A.13}$$

Combining (A.12) and (A.13), it follows that

$$\frac{(\boldsymbol{\omega}_M'(\boldsymbol{g}-e\mathbf{1}))^2}{(\boldsymbol{g}-e\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g}-e\mathbf{1})} = \frac{\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\|^2}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2}\cdot\frac{z_1^2}{\|\boldsymbol{z}\|^2}. \tag{A.14}$$

By $\boldsymbol{\alpha}_2 = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{1}$, (A.10) and (A.11), we also have

$$\begin{aligned}\boldsymbol{g}'\mathbf{1} &= \boldsymbol{z}'\boldsymbol{Q}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{1} + e\mathbf{1}'\mathbf{1} \\ &= \boldsymbol{z}'\boldsymbol{Q}\boldsymbol{\alpha}_2 + e\cdot n \\ &= \boldsymbol{\alpha}_2'\boldsymbol{Q}\boldsymbol{z} + ne \\ &= \langle\hat{\boldsymbol{\alpha}}_1,\boldsymbol{\alpha}_2\rangle z_1 + \langle\hat{\boldsymbol{\alpha}}_2,\boldsymbol{\alpha}_2\rangle z_2 + ne.\end{aligned} \tag{A.15}$$

From (A.11), (A.14) and (A.15), the optimization problem (A.7) is equivalent to

$$\begin{aligned}&\min_{\boldsymbol{z}\in\mathbb{R}^n}\ \frac{z_1^2}{\|\boldsymbol{z}\|^2} \\ &\text{s.t.}\quad \|\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{z} + e\mathbf{1} - \tilde{\boldsymbol{g}}_T\|^2 \le \|\boldsymbol{g}_s^* - \tilde{\boldsymbol{g}}_T\|^2 \quad\text{and}\quad \langle\hat{\boldsymbol{\alpha}}_1,\boldsymbol{\alpha}_2\rangle z_1 + \langle\hat{\boldsymbol{\alpha}}_2,\boldsymbol{\alpha}_2\rangle z_2 + ne = 0.\end{aligned} \tag{A.16}$$

For $s\in[e_T,e)$, let the feasible set with respect to $\boldsymbol{z}$ be

$$Z_s = \big\{\boldsymbol{z} : \|\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{z} + e\mathbf{1} - \tilde{\boldsymbol{g}}_{\boldsymbol{T}}\|^2 \le \|\boldsymbol{g}_s^* - \tilde{\boldsymbol{g}}_T\|^2,\ \langle\hat{\boldsymbol{\alpha}}_1,\boldsymbol{\alpha}_2\rangle z_1 + \langle\hat{\boldsymbol{\alpha}}_2,\boldsymbol{\alpha}_2\rangle z_2 + ne = 0\big\}.$$

Under Assumption 1,

$$\frac{\boldsymbol{\omega}_M'\boldsymbol{g}}{\boldsymbol{\omega}_M'\mathbf{1}} \ge s \Leftrightarrow \boldsymbol{\omega}_M'(\boldsymbol{g}-e\mathbf{1}) \ge (s-e)\boldsymbol{\omega}_M'\mathbf{1}.$$

By (A.12), we have

$$\frac{\boldsymbol{\omega}_M'\boldsymbol{g}}{\boldsymbol{\omega}_M'\mathbf{1}} \ge s \Leftrightarrow z_1 \ge \frac{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|}{\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\|}(s-e)\boldsymbol{\omega}_M'\mathbf{1}. \tag{A.17}$$

According to Proposition 2, it follows from (A.11), (A.15) and (A.17) that $\boldsymbol{z}^*$ is the unique solution of the optimization scheme

$$\begin{aligned}&\min_{\boldsymbol{z}=(z_1,\ldots,z_n)'} \|\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{z} + e\mathbf{1} - \tilde{\boldsymbol{g}}_T\|^2\\ &\text{s.t.} \quad z_1 \geq \frac{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|}{\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\|}(s-e)\boldsymbol{\omega}_M'\mathbf{1} \quad \text{and} \quad \langle\hat{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\rangle z_1 + \langle\hat{\boldsymbol{\alpha}}_2, \boldsymbol{\alpha}_2\rangle z_2 + ne = 0,\end{aligned}$$

and the minimum value is exactly $\|\boldsymbol{g}_s^* - \tilde{\boldsymbol{g}}_T\|^2$. This implies

$$Z_s \subset \left\{\boldsymbol{z} = (z_1, \ldots, z_n)' : z_1 \leq \frac{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|}{\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\|}(s-e)\boldsymbol{\omega}_M'\mathbf{1}\right\}.$$

Since $s < e$ and $\boldsymbol{z}^* \in Z_s$, the set Z_s is a nonempty convex subset of $\mathbb{R}^n$. In particular, for any $\boldsymbol{z} \in Z_s$, we have $z_1 < 0$. By rewriting (A.16), the optimization problem becomes

$$\max_{\boldsymbol{z}}(z_1^2 + z_2^2 + \cdots + z_n^2)/z_1^2 \quad \text{with} \quad \boldsymbol{z} \in Z_s.$$

It follows that the function $(z_1^2+z_2^2+\cdots+z_n^2)/z_1^2$ is convex on the feasible set Z_s. Hence, (A.16) is a convex optimization problem whose KKT conditions are satisfied and for which a maximizer exists. Equivalently, problem (A.7) admits a solution, and the maximum is attained on the boundary of the convex closed set, which yields (A.8).

A.4 Proof of Proposition 3

Sufficiency: Suppose that both agencies have low aversion, that is, $\alpha_A \leq \alpha_{e,p}$ and $\alpha_B \leq \alpha_{e,p}$. Recall that a rating profile $(\boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#)$ is a Nash equilibrium if, for each agency $j \in \{\text{A,B}\}$,

$$V_j(\boldsymbol{g}_j^\#, \boldsymbol{g}_{-j}^\#) \;\geq\; V_j(\boldsymbol{g}_j', \boldsymbol{g}_{-j}^\#) \quad \text{for all } \boldsymbol{g}_j' \in G. \tag{A.18}$$

Our goal is to prove that $\boldsymbol{g}_A^\# = \boldsymbol{g}_B^\# = \boldsymbol{g}_e^*$ is an equilibrium profile. Because the two agencies are ex ante symmetric, it suffices to verify the condition (A.18) for Agency A when Agency B chooses $\boldsymbol{g}_B^\# = \boldsymbol{g}_e^*$; the argument for Agency B is similar.

Fix $\boldsymbol{g}_B^\# = \boldsymbol{g}_e^*$, then it follows that $\boldsymbol{\omega}_M'\boldsymbol{g}_B^\#/\boldsymbol{\omega}_M'\mathbf{1} = e$. From the definition of the market share (2.11) with expressions (2.9) and (2.10), any $\boldsymbol{g}_A \in G$ by Agency A can at most get the half market, i.e.,

$M_j^\pi(\boldsymbol{g}_A, \boldsymbol{g}_B^\#) \le 1/2$. The equality holds when $\boldsymbol{\omega}_M'\boldsymbol{g}_A/\boldsymbol{\omega}_M'\mathbf{1} \ge e$, and by Proposition 2, in such case we have $\|\boldsymbol{g}_A - \tilde{\boldsymbol{g}}_T\|^2 \ge \|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2$. Hence

$$V_A(\boldsymbol{g}_e^*, \boldsymbol{g}_B^\#) \ \ge \ V_A(\boldsymbol{g}_A, \boldsymbol{g}_B^\#) \quad \text{for} \quad \boldsymbol{\omega}_M'\boldsymbol{g}_A/\boldsymbol{\omega}_M'\mathbf{1} \ge e.$$

For $\boldsymbol{g}_A \in G$ satisfying $\boldsymbol{\omega}_M'\boldsymbol{g}_A/\boldsymbol{\omega}_M'\mathbf{1} < e$, Agency A cannot attract the green investor and from $\pi(\{b\}) = 1 - p$ we have $M_j^\pi(\boldsymbol{g}_A, \boldsymbol{g}_B^\#) \le (1-p)/2$, so it suffices to prove $V_A(\boldsymbol{g}_e^*, \boldsymbol{g}_e^*) \ge V_A(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_e^*)$. Recall that $\alpha_{e,p} = p/\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2$. When both agencies have low aversion ($\alpha_A, \alpha_B \le \alpha_{e,p}$), we get $\alpha_A\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2 \le p$ and then

$$V_A(\boldsymbol{g}_e^*, \boldsymbol{g}_e^*) = \frac{1}{2} - \frac{\alpha_A}{2}\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2 \ge \frac{1-p}{2} = V_A(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_e^*).$$

Hence, it can be concluded that $\boldsymbol{g}_A^\# = \boldsymbol{g}_B^\# = \boldsymbol{g}_e^*$ constitutes an NE.

Necessity: Suppose $(\boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#)$ is an NE, then our goal is to prove that both agencies have low aversion, that is $\alpha_A, \alpha_B \le \alpha_{e,p}$. Fix either agency $j \in \{\mathrm{A}, \mathrm{B}\}$ and consider the unilateral deviation in which j issues the ESG fundamental $\tilde{\boldsymbol{g}}_T$ in place of its equilibrium strategy $\boldsymbol{g}_j^\#$. Note that $b \le e_T < e$, so

$$\boldsymbol{\omega}_M'\tilde{\boldsymbol{g}}_T = e_T\boldsymbol{\omega}_M'\mathbf{1} \ge b\,\boldsymbol{\omega}_M'\mathbf{1}.$$

Combining it with (2.6), we have

$$U(b, \tilde{\boldsymbol{g}}_T) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2} = \max_{\boldsymbol{g}\in G} U(b, \boldsymbol{g}). \tag{A.19}$$

Then $U(b, \tilde{\boldsymbol{g}}_T) \ge U(b, \boldsymbol{g}_{-j}^\#)$, which implies that brown investors (with ESG target b) choose agency j with probability at least $1/2$. It follows from $\pi(\{b\}) = 1 - p$ that $M_j^\pi(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_{-j}^\#) \ge (1-p)/2$. By the definition of NE and the agency utility in (2.12),

$$M_j^\pi(\boldsymbol{g}_j^\#, \boldsymbol{g}_{-j}^\#) \ge V_j(\boldsymbol{g}_j^\#, \boldsymbol{g}_{-j}^\#) \ge V_j(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_{-j}^\#) = M_j^\pi(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_{-j}^\#) \ge \frac{1-p}{2}, \tag{A.20}$$

where the equality holds because issuing $\tilde{\boldsymbol{g}}_T$ incurs no penalty (the penalty term vanishes at $\tilde{\boldsymbol{g}}_T$).

Since $M_A^\pi(\boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#) + M_B^\pi(\boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#) = 1$, without loss of generality, we assume

$$\frac{1-p}{2} \le M_A^\pi(\boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#) \le \frac{1}{2}.$$

By definition (2.11), the market share of Agency A is given by

$$M_A^\pi(\boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#) = (1-p)P_A(b, \boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#) + pP_A(e, \boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#), \tag{A.21}$$

At the NE $(\boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#)$, write

$$P_A(t) = P_A\big(t; \boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#\big), \quad t \in \{b, e\},$$

so $P_A(t) \in \{0, 1/2, 1\}$. Then (A.21) has the expression

$$M_A^\pi(\boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#) = \begin{cases} 1-p, & \text{if} \quad P_A(b) = 1,\ P_A(e) = 0, \\ p, & \text{if} \quad P_A(b) = 0,\ P_A(e) = 1, \\ \frac{1-p}{2}, & \text{if} \quad P_A(b) = \frac{1}{2},\ P_A(e) = 0, \\ \frac{p}{2}, & \text{if} \quad P_A(b) = 0,\ P_A(e) = \frac{1}{2}, \\ \frac{1}{2}, & \text{if} \quad P_A(b) = \frac{1}{2},\ P_A(e) = \frac{1}{2}. \end{cases} \tag{A.22}$$

In the following, we examine each case individually to determine whether an NE may exist. We will show that only the final case admits an NE, and that in this case for $j \in \{\text{A,B}\}$ we have $\alpha_j \le \alpha_{e,p}$.

(i) When $M_A^\pi(\boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#) = 1 - p$ **:** In this case,

$$P_B(e; \boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#) = 1, \qquad P_B(b; \boldsymbol{g}_A^\#, \boldsymbol{g}_B^\#) = 0.$$

From (A.19), we have

$$U(b, \boldsymbol{g}_B^\#) < U(b, \boldsymbol{g}_A^\#) \le U(b, \tilde{\boldsymbol{g}}_T).$$

Let Agency A replace $\boldsymbol{g}_A^{\#}$ with $\tilde{\boldsymbol{g}}_T$. Since $N_A(\tilde{\boldsymbol{g}}_T;\alpha_A)=0$,

$$V_A(\tilde{\boldsymbol{g}}_T,\boldsymbol{g}_B^{\#})=M_A^{\pi}(\tilde{\boldsymbol{g}}_T,\boldsymbol{g}_B^{\#})\geq 1-p=M_A^{\pi}(\boldsymbol{g}_A^{\#},\boldsymbol{g}_B^{\#})\geq V_A(\boldsymbol{g}_A^{\#},\boldsymbol{g}_B^{\#}).$$

But NE requires $V_A(\boldsymbol{g}_A^{\#},\boldsymbol{g}_B^{\#})\geq V_A(\tilde{\boldsymbol{g}}_T,\boldsymbol{g}_B^{\#})$, so we obtain that $\boldsymbol{g}_A^{\#}=\tilde{\boldsymbol{g}}_T$. By (3.6), $\|\boldsymbol{g}_k^*-\tilde{\boldsymbol{g}}_T\|^2$ is continuous with respect to k and $\|\boldsymbol{g}_{e_T}^*-\tilde{\boldsymbol{g}}_T\|^2=0$. There exists $k\in(e_T,e]$ such that

$$\left\|\boldsymbol{g}_k^*-\tilde{\boldsymbol{g}}_T\right\|^2<\frac{1-p}{\alpha_B}.$$

Thus, we obtain that

$$\begin{aligned}V_B(\boldsymbol{g}_k^*,\boldsymbol{g}_A^{\#})&=M_B^{\pi}(\boldsymbol{g}_k^*,\tilde{\boldsymbol{g}}_T)-N_B(\boldsymbol{g}_k^*;\alpha_B)\\&=\frac{1-p}{2}+p-\frac{\alpha_j}{2}\left\|\boldsymbol{g}_k^*-\tilde{\boldsymbol{g}}_T\right\|^2\\&>p=M_B(\boldsymbol{g}_B^{\#},\boldsymbol{g}_A^{\#})\\&\geq\ V_B(\boldsymbol{g}_B^{\#},\boldsymbol{g}_A^{\#}),\end{aligned}$$

which contradicts $(\boldsymbol{g}_A^{\#},\boldsymbol{g}_B^{\#})$ being an NE.

(ii) When $M_A^{\pi}(\boldsymbol{g}_A^{\#},\boldsymbol{g}_B^{\#})=p$: We can get the contradiction with similar analysis in **case (i)**.

(iii) When $M_A^{\pi}(\boldsymbol{g}_A^{\#},\boldsymbol{g}_B^{\#})=(1-p)/2$: In this case, the choice profile is

$$P_A(b)=\tfrac{1}{2},\quad P_B(b)=\tfrac{1}{2},\quad P_B(e)=1.$$

It follows that the inequality (A.20) takes the equality sign and $\boldsymbol{g}_A^{\#}=\tilde{\boldsymbol{g}}_T$. Then, $U(e;\boldsymbol{g}_B^{\#})>U(e;\tilde{\boldsymbol{g}}_T)$ and thus $\boldsymbol{g}_B^{\#}\neq\tilde{\boldsymbol{g}}_T$. By Lemma A.1 and (A.8), there exists $k\in(e_T,e]$ such that

$$\|\boldsymbol{g}_k^{**}-\tilde{\boldsymbol{g}}_T\|^2<\|\boldsymbol{g}_B^{\#}-\tilde{\boldsymbol{g}}_T\|^2,$$

where $\boldsymbol{g}_k^{**}$ is derived from the solution of the optimization problem (A.7) in Lemma A.1. Consequently,

$$\begin{aligned} V_B(\boldsymbol{g}_k^{**}, \boldsymbol{g}_A^{\#}) &= M_B^{\pi}(\boldsymbol{g}_k^{**}, \tilde{\boldsymbol{g}}_T) - N_B(\boldsymbol{g}_k^{**}; \alpha_B) \\ &= \frac{1+p}{2} - \frac{\alpha_B}{2}\|\boldsymbol{g}_k^{**} - \tilde{\boldsymbol{g}}_T\|^2 \\ &> \frac{1+p}{2} - \frac{\alpha_B}{2}\|\boldsymbol{g}_B^{\#} - \tilde{\boldsymbol{g}}_T\|^2 \\ &= M_B^{\pi}(\boldsymbol{g}_B^{\#}, \boldsymbol{g}_A^{\#}) - N_B(\boldsymbol{g}_B^{\#}; \alpha_B) \\ &= V_B(\boldsymbol{g}_B^{\#}, \boldsymbol{g}_A^{\#}). \end{aligned} \tag{A.23}$$

The inequality (A.23) contradicts $(\boldsymbol{g}_A^{\#}, \boldsymbol{g}_B^{\#})$ being an NE.

(iv) When $M_A^{\pi}(\boldsymbol{g}_A^{\#}, \boldsymbol{g}_B^{\#}) = p/2$: In this case, $P_A(e) = P_B(e) = 1/2$, then $U(e, \boldsymbol{g}_A^{\#}) = U(e, \boldsymbol{g}_B^{\#})$. Since the utility $U(e, \boldsymbol{g})$ is a continuous function with respect to $\boldsymbol{g}$, it requires that $\boldsymbol{g}_A^{\#}$ must be a maximizer of $U(e, \cdot)$. In fact, if $U(e, \boldsymbol{g}_A^{\#})$ doesn't reach the maximum of utility, then there exists a feasible revision $\tilde{\boldsymbol{g}}_A$ such that

$$U(e, \tilde{\boldsymbol{g}}_A) > U(e, \boldsymbol{g}_A^{\#}) = U(e, \boldsymbol{g}_B^{\#})$$

and

$$\|\tilde{\boldsymbol{g}}_A - \boldsymbol{g}_A^{\#}\|^2 < \frac{p}{\alpha_A}.$$

Therefore, we can obtain that $P_A(e; \tilde{\boldsymbol{g}}_A, \boldsymbol{g}_B^{\#})$ and thus $M_A^{\pi}(\tilde{\boldsymbol{g}}_A, \boldsymbol{g}_B^{\#}) \geq p$. Using the triangle inequality for the Euclidean norm, we obtain

$$\begin{aligned} V_A(\tilde{\boldsymbol{g}}_A, \boldsymbol{g}_B^{\#}) &= M_A^{\pi}(\tilde{\boldsymbol{g}}_A, \boldsymbol{g}_B^{\#}) - N_A(\tilde{\boldsymbol{g}}_A; \alpha_A) \\ &\geq p - \frac{\alpha_A}{2}\|\tilde{\boldsymbol{g}}_A - \tilde{\boldsymbol{g}}_T\|^2 \\ &\geq p - \frac{\alpha_A}{2}\|\boldsymbol{g}_A^{\#} - \tilde{\boldsymbol{g}}_T\|^2 - \frac{\alpha_A}{2}\|\tilde{\boldsymbol{g}}_A - \boldsymbol{g}_A^{\#}\|^2 \\ &= M_A^{\pi}(\boldsymbol{g}_A^{\#}, \boldsymbol{g}_B^{\#}) - N_A(\boldsymbol{g}_A^{\#}; \alpha_A) + \frac{p}{2} - \frac{\alpha_A}{2}\|\tilde{\boldsymbol{g}}_A - \boldsymbol{g}_A^{\#}\|^2 \\ &> V_A(\boldsymbol{g}_A^{\#}, \boldsymbol{g}_B^{\#}). \end{aligned} \tag{A.24}$$

However, (A.24) contradicts NE because at equilibrium $V_A(\boldsymbol{g}_A^{\#}, \boldsymbol{g}_B^{\#}) \geq V_A(\tilde{\boldsymbol{g}}_A, \boldsymbol{g}_B^{\#})$.
As a result, $\boldsymbol{g}_A^{\#}$ maximizes $U(e, \cdot)$. By (2.6),

$$U(e, \boldsymbol{g}_A^{\#}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2},$$

meaning that $\boldsymbol{\omega}_M'(e\mathbf{1} - \boldsymbol{g}_A^{\#}) \leq 0$. Since $b < e$, we also have $\boldsymbol{\omega}_M'(b\mathbf{1} - \boldsymbol{g}_A^{\#}) \leq 0$, i.e., $U(b, \boldsymbol{g}_A^{\#})$ attains its maximum. Therefore, $P_A(e) \geq \frac{1}{2}$ and $P_A(b) \geq \frac{1}{2}$, indicating $M_A^{\pi}(\boldsymbol{g}_A^{\#}, \boldsymbol{g}_B^{\#}) \geq \frac{1}{2}$. It contradicts $M_A^{\pi}(\boldsymbol{g}_A^{\#}, \boldsymbol{g}_B^{\#}) = p/2$ for $p < 1$. The discussion for $p = 1$ leaves to **case (v)**.

(v) When $M_A^{\pi}(\boldsymbol{g}_A^{\#}, \boldsymbol{g}_B^{\#}) = 1/2$: In this case, we also have $P_A(e) = P_B(e) = 1/2$. Following the same logic in **case (iv)**, NE requires both $\boldsymbol{g}_A^{\#}$ and $\boldsymbol{g}_B^{\#}$ maximizes $U(e, \cdot)$. Then for $j \in \{\text{A,B}\}$,

$$\boldsymbol{\omega}_M'\boldsymbol{g}_j^{\#} \geq e\boldsymbol{\omega}_M'\mathbf{1},$$

and

$$V_j(\boldsymbol{g}_j^{\#}, \boldsymbol{g}_{-j}^{\#}) = \frac{1}{2} - \frac{\alpha_j}{2}\|\boldsymbol{g}_A^{\#} - \tilde{\boldsymbol{g}}_T\|^2.$$

By Proposition 2, the optimal rating is uniquely attained when

$$\boldsymbol{g}_A^{\#} = \boldsymbol{g}_B^{\#} = \boldsymbol{g}_e^*. \tag{A.25}$$

Using (3.7) and the inequality $V_j(\boldsymbol{g}_j^{\#}, \boldsymbol{g}_{-j}^{\#}) \geq V_j(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_{-j}^{\#})$, we can conclude that $\alpha_j \leq \alpha_{e,p}$, i.e., both agencies have low aversion.

Uniqueness: The necessity analysis shows that an NE exists when (A.25). Here, the uniqueness is guaranteed by Proposition 2. In summary, we have completed the proof.

A.5 Lemma A.2 and its Proof

Lemma A.2. *The function $s \mapsto U(s, \boldsymbol{g}_s^{**})$ is an one-to-one strictly increasing continuous function from $[e_T, e]$ to $[U(e, \tilde{\boldsymbol{g}}_T), U(e, \boldsymbol{g}_e^*)]$. In particular, when $s = e_T$, the optimization problem* (A.7) *has the unique solution $\boldsymbol{g}_s^{**} = \boldsymbol{g}_{e_T}^* = \tilde{\boldsymbol{g}}_T$; when $s = e$, we can conclude that $\boldsymbol{g}_s^{**} = \boldsymbol{g}_e^*$ and $U(e, \boldsymbol{g}_e^*)$ arrives at the maximum value $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2$.*

Proof. For any $s_1, s_2 \in [e_T, e]$ with $s_1 < s_2$, consider the feasible sets

$$\mathcal{G}(s_i) = \left\{ \boldsymbol{g} \in G : \|\boldsymbol{g} - \tilde{\boldsymbol{g}}_T\|^2 \leq \|\boldsymbol{g}^*_{s_i} - \tilde{\boldsymbol{g}}_T\|^2 \right\}, \quad i = 1, 2.$$

Since $\|\boldsymbol{g}^*_{s_1} - \tilde{\boldsymbol{g}}_T\|^2 < \|\boldsymbol{g}^*_{s_2} - \tilde{\boldsymbol{g}}_T\|^2$, we have $\mathcal{G}(s_1) \subseteq \mathcal{G}(s_2)$. By Lemma A.1, the optimal solution $\boldsymbol{g}^{**}_s$ is attained on the boundary

$$\|\boldsymbol{g}^{**}_s - \tilde{\boldsymbol{g}}_T\|^2 = \|\boldsymbol{g}^*_s - \tilde{\boldsymbol{g}}_T\|^2,$$

and minimizes the objective function $\boldsymbol{g} \mapsto \frac{(\boldsymbol{\omega}'_M(\boldsymbol{g} - e\mathbf{1}))^2}{(\boldsymbol{g} - e\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g} - e\mathbf{1})}$. Therefore, as the feasible set expands continuously with s, the optimal value $U(e, \boldsymbol{g}^{**}_s)$ strictly increases, implying continuity, monotonicity and injectivity of $s \mapsto U(e, \boldsymbol{g}^{**}_s)$. It remains to verify the two endpoint cases: $s = e_T$ and $s = e$.

(i) When $s = e_T$: In this case, $\boldsymbol{g}^*_{e_T} = \tilde{\boldsymbol{g}}_T$ from Lemma A.1. It implies that the unique optimizer of (A.7) is attained at $\boldsymbol{g}^{**}_s = \boldsymbol{g}^*_{e_T} = \tilde{\boldsymbol{g}}_T$, and the corresponding value of utility at $s = e_T$ is $U(e, \tilde{\boldsymbol{g}}_T)$.

(ii) When $s = e$: It follows from $\boldsymbol{\omega}'_M(e\mathbf{1} - \boldsymbol{g}^*_e) = 0$ that $\boldsymbol{g}^*_e$ maximizes $U(e, \boldsymbol{g})$ over $\boldsymbol{g} \in \mathbb{R}^n$. Since $\boldsymbol{g}^*_e \in \mathcal{G}(s)$, the utility $U(s, \boldsymbol{g})$ is attained at $\boldsymbol{g}^{**}_s = \boldsymbol{g}^*_e$ and the maximal value is

$$U(e, \boldsymbol{g}^*_e) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2},$$

which completes the proof.

A.6 Proof of Theorem 1

Sufficiency: Without loss of generality, we assume that $\alpha_A > \alpha_B$.

To prove that $(\boldsymbol{g}^*_A, \boldsymbol{g}^*_B)$ described in Theorem 1 constitutes an SSE, it suffices to show that it satisfies the equilibrium condition (3.2) in Definition 6. Next, we consider the two cases $\alpha_B/\alpha_A \in (1/2, 1)$ and $\alpha_B/\alpha_A \in (0, 1/2]$.

Case I: $\alpha_B/\alpha_A \in (1/2, 1)$**.**

(i) $\alpha_B < \alpha_A \leq \alpha_{e,p}$.

In the following, we verify that $(\boldsymbol{g}_e^*, \boldsymbol{g}_e^*)$ is an SSE. For $j \in \{\text{A,B}\}$ and $\boldsymbol{g} \in G$, we have

$$V_j(\boldsymbol{g}, \boldsymbol{g}) = \frac{1}{2} - \frac{\alpha_j}{2}\|\boldsymbol{g} - \tilde{\boldsymbol{g}}_T\|^2. \tag{A.26}$$

Based on Definition 6, it suffices to prove that for arbitrary $\boldsymbol{g}_j \in G$,

$$\lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_{-j} \in \mathrm{BR}_{-j}^{\varepsilon}(\boldsymbol{g}_j)} V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) \le V_j(\boldsymbol{g}_e^*, \boldsymbol{g}_e^*) = \frac{1}{2} - \frac{\alpha_j}{2}\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2.$$

We prove the above inequality by considering $\boldsymbol{g}_j$ in the following two scenarios.

(a) $\boldsymbol{\omega}_M' \boldsymbol{g}_j \ge e\boldsymbol{\omega}_M' \mathbf{1}$.

It follows from Proposition 2 that

$$\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_T\|^2 \ge \|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2. \tag{A.27}$$

Since $\alpha_{-j} \le \alpha_{e,p}$, similar to the sufficiency part of Proposition 3, we have

$$\boldsymbol{g}_e^* = \operatorname*{argmax}_{\boldsymbol{g}_{-j} \in G} V_{-j}(\boldsymbol{g}_{-j}, \boldsymbol{g}_j),$$

then $\mathrm{BR}_{-j}^{\varepsilon}(\boldsymbol{g}_j)$ is a neighborhood of $\boldsymbol{g}_e^*$. Therefore, we obtain that

$$\lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_{-j} \in \mathrm{BR}_{-j}^{\varepsilon}(\boldsymbol{g}_j)} V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) \le V_j(\boldsymbol{g}_j, \boldsymbol{g}_e^*) = \frac{1}{2} - \frac{\alpha_j}{2}\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_T\|^2 \le \frac{1}{2} - \frac{\alpha_j}{2}\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2.$$

(b) $\boldsymbol{\omega}_M' \boldsymbol{g}_j < e\boldsymbol{\omega}_M' \mathbf{1}$.

In this case, we first show that the following statement holds. For sufficient small $\varepsilon > 0$,

$$\exists\, s_\varepsilon \in [e_T, e) \text{ s.t. } \boldsymbol{g}_{s_\varepsilon}^{**} \in \mathrm{BR}_{-j}^{\varepsilon}(\boldsymbol{g}_j) \;\text{ and }\; M_j^{\pi}(\boldsymbol{g}_j, \boldsymbol{g}_{s_\varepsilon}^{**}) \le \frac{1-p}{2}. \tag{A.28}$$

By Proposition 1 we obtain that $U(e, \boldsymbol{g}_j) < U(e, \boldsymbol{g}_e^*)$. It follows from Lemma A.2 that there exists $s \in [e_T, e)$ such that

$$U(e, \boldsymbol{g}_j) < U(e, \boldsymbol{g}_s^{**}). \tag{A.29}$$

For Agency $-j$, by (A.29) we have

$$M^{\pi}_{-j}(\boldsymbol{g}^{**}_s, \boldsymbol{g}_j) \geq \frac{1+p}{2}.$$

Then we have

$$V_{-j}(\boldsymbol{g}^{**}_s, \boldsymbol{g}_j) = M^{\pi}_{-j}(\boldsymbol{g}^{**}_s, \boldsymbol{g}_j) - N_{-j}(\boldsymbol{g}^{**}_s; \alpha_{-j}) \geq \frac{1+p}{2} - \frac{\alpha_{-j}}{2}\|\boldsymbol{g}^{**}_s - \tilde{\boldsymbol{g}}_T\|^2.$$

Moreover, since $\alpha_{-j} \leq \alpha_{e,p}$, we also have

$$V_{-j}(\boldsymbol{g}^{**}_s, \boldsymbol{g}_j) \geq \frac{1+p}{2} - \frac{\alpha_{-j}}{2}\|\boldsymbol{g}^{**}_s - \tilde{\boldsymbol{g}}_T\|^2 > \frac{1}{2}.$$

This implies that for sufficient small $\varepsilon > 0$,

$$\sup_{\boldsymbol{g}_{-j} \in G} V_{-j}(\boldsymbol{g}_{-j}, \boldsymbol{g}_j) - \varepsilon > \frac{1}{2}.$$

Therefore, $\mathrm{BR}^{\varepsilon}_{-j}(\boldsymbol{g}_j)$ is a subset of $\{\boldsymbol{g} : U(e, \boldsymbol{g}) > U(e, \boldsymbol{g}_j)\}$. Then there exists an appropriate $K_\varepsilon > 0$ satisfying

$$\mathrm{BR}^{\varepsilon}_{-j}(\boldsymbol{g}_j) = \{\boldsymbol{g} : U(e, \boldsymbol{g}) > U(e, \boldsymbol{g}_j)\} \cap \{\boldsymbol{g} : \|\boldsymbol{g} - \tilde{\boldsymbol{g}}_T\|^2 \leq K_\varepsilon\}. \tag{A.30}$$

Using Lemma A.2, we conclude that (A.28) holds.

Hence,

$$\min_{\boldsymbol{g}_{-j} \in \mathrm{BR}^{\varepsilon}_{-j}(\boldsymbol{g}_j)} V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) \leq V_j(\boldsymbol{g}_j, \boldsymbol{g}^{**}_{s_\varepsilon}) \leq M^{\pi}_j(\boldsymbol{g}_j, \boldsymbol{g}^{**}_{s_\varepsilon}) \leq \frac{1-p}{2}. \tag{A.31}$$

This implies

$$\lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_{-j} \in \mathrm{BR}^{\varepsilon}_{-j}(\boldsymbol{g}_j)} V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) \leq \frac{1-p}{2} \leq \frac{1}{2} - \frac{\alpha_j}{2}\|\boldsymbol{g}^*_e - \tilde{\boldsymbol{g}}_T\|^2.$$

Therefore, based on Definition 6, the rating profile $(\boldsymbol{g}^*_e, \boldsymbol{g}^*_e)$ is an SSE.

(ii) $\alpha_B \leq \alpha_{e,p} < \alpha_A \leq 2\alpha_{e,p}$.

In the following, we verify that $(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}^*_e)$ is an SSE, i.e., $\boldsymbol{g}^*_A = \tilde{\boldsymbol{g}}_T, \boldsymbol{g}^*_B = \boldsymbol{g}^*_e$.

For Agency A: When $\boldsymbol{g}_A$ satisfies $\boldsymbol{\omega}_M'\boldsymbol{g}_A < e\boldsymbol{\omega}_M'\mathbf{1}$, it is similar to (A.31) that

$$\min_{\boldsymbol{g}_B \in \mathrm{BR}_B^{\varepsilon}(\boldsymbol{g}_A)} V_A(\boldsymbol{g}_A, \boldsymbol{g}_B) \leq \frac{1-p}{2}, \tag{A.32}$$

for sufficient small $\varepsilon > 0$.

When $\boldsymbol{g}_A$ satisfies $\boldsymbol{\omega}_M'\boldsymbol{g}_A \geq e\boldsymbol{\omega}_M'\mathbf{1}$, similar to the sufficiency part of Proposition 3, it follows from $\alpha_B \leq \alpha_{e,p}$ that $BR_B^{\varepsilon}(\boldsymbol{g}_A)$ is the neighborhood of $\boldsymbol{g}_e^*$. From $\alpha_{e,p} < \alpha_A$, we have

$$V_A(\boldsymbol{g}_A, \boldsymbol{g}_e^*) \leq V_A(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_e^*) = \frac{1-p}{2}.$$

Thus, we also have

$$\min_{\boldsymbol{g}_B \in \mathrm{BR}_B^{\varepsilon}(\boldsymbol{g}_A)} V_A(\boldsymbol{g}_A, \boldsymbol{g}_B) \leq V_A(\boldsymbol{g}_A, \boldsymbol{g}_e^*) \leq \frac{1-p}{2}. \tag{A.33}$$

Combining (A.32) and (A.33), it follows that for any $\boldsymbol{g}_A \in G$,

$$\lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_B \in \mathrm{BR}_B^{\varepsilon}(\boldsymbol{g}_A)} V_A(\boldsymbol{g}_A, \boldsymbol{g}_B) \leq \frac{1-p}{2} = V_A(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_e^*).$$

For Agency B: When $\boldsymbol{g}_B$ satisfies $\boldsymbol{\omega}_M'\boldsymbol{g}_B \geq e\boldsymbol{\omega}_M'\mathbf{1}$, (A.27) holds for j =B. In addition, for sufficient small $\varepsilon > 0$ and any $\boldsymbol{g}_A \in \mathrm{BR}_A^{\varepsilon}(\boldsymbol{g}_B)$, we have $M_A^{\pi}(\boldsymbol{g}_A, \boldsymbol{g}_B) \geq (1-p)/2$, and $M_B^{\pi}(\boldsymbol{g}_B, \boldsymbol{g}_A) \leq (1+p)/2$. Then we have

$$V_B(\boldsymbol{g}_B, \boldsymbol{g}_A) \leq \frac{1+p}{2} - \frac{\alpha_B}{2}\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2.$$

When $\boldsymbol{g}_B$ satisfies $\boldsymbol{\omega}_M'\boldsymbol{g}_B < e\boldsymbol{\omega}_M'\mathbf{1}$, similar to (A.30), it follows from $\alpha_A \leq 2\alpha_{e,p}$ that for sufficient small $\varepsilon > 0$, there exists an appropriate $K_\varepsilon > 0$ satisfying

$$\mathrm{BR}_A^{\varepsilon}(\boldsymbol{g}_B) = \{\boldsymbol{g} : U(e, \boldsymbol{g}) > U(e, \boldsymbol{g}_B)\} \cap \{\boldsymbol{g} : \|\boldsymbol{g} - \tilde{\boldsymbol{g}}_T\|^2 \leq K_\varepsilon\}.$$

Then we have

$$\min_{\boldsymbol{g}_A \in \mathrm{BR}_A^{\varepsilon}(\boldsymbol{g}_B)} V_B(\boldsymbol{g}_B, \boldsymbol{g}_A) \leq \frac{1-p}{2} < \frac{1+p}{2} - \frac{\alpha_B}{2}\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2.$$

Consequently, for any $\boldsymbol{g}_B \in G$ we have

$$\lim_{\varepsilon\to 0}\min_{\boldsymbol{g}_A\in \mathrm{BR}_A^{\varepsilon}(\boldsymbol{g}_B)} V_B(\boldsymbol{g}_B,\boldsymbol{g}_A) \le \frac{1+p}{2} - \frac{\alpha_B}{2}\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2 = V_B(\boldsymbol{g}_e^*, \tilde{\boldsymbol{g}}_T).$$

Therefore, based on Definition 6, the rating profile $(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_e^*)$ is an SSE.

(iii) $2\alpha_{e,p} < \alpha_B < \alpha_A$.

We verify that $(\tilde{\boldsymbol{g}}_T, \tilde{\boldsymbol{g}}_T)$ is an SSE. From (A.26), we only need to prove that for $\boldsymbol{g}_j \in G$,

$$\lim_{\varepsilon\to 0}\min_{\boldsymbol{g}_{-j}\in \mathrm{BR}_{-j}^{\varepsilon}(\boldsymbol{g}_j)} V_j(\boldsymbol{g}_j,\boldsymbol{g}_{-j}) \le V_j(\tilde{\boldsymbol{g}}_T, \tilde{\boldsymbol{g}}_T) = \frac{1}{2}. \tag{A.34}$$

We prove it by considering $\boldsymbol{g}_j$ in three scenarios.

(a) $N_j(\boldsymbol{g}_j;\alpha_j) \ge p/2$.

In this case, $V_j(\boldsymbol{g}_j, \tilde{\boldsymbol{g}}_T) = (1+p)/2 - N_j(\boldsymbol{g}_j;\alpha_j) \le 1/2$, which implies (A.34).

(b) $N_j(\boldsymbol{g}_j;\alpha_j) = 0$.

In this case, we know that $\boldsymbol{g}_j = \tilde{\boldsymbol{g}}_T$. It is straightforward that for any $\boldsymbol{g}_{-j} \neq \tilde{\boldsymbol{g}}_T$, $V_j(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}_{-j}) = (1-p)/2$, which implies (A.34).

(c) $0 < N_j(\boldsymbol{g}_j;\alpha_j) < p/2$.

It follows from $0 < N_j(\boldsymbol{g}_j;\alpha_j) < p/2$ that $\boldsymbol{g}_j \neq \tilde{\boldsymbol{g}}_T$ and $\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_T\|^2 < p/\alpha_j$. Note that $\alpha_j > \alpha_{e,p} = p/\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2$, so we have $\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_T\|^2 < p/\alpha_j < \|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2$. By Proposition 1 we obtain that $U(e, \boldsymbol{g}_j) < U(e, \boldsymbol{g}_e^*)$.

Let $\delta_j = \alpha_j/\alpha_{-j}$, then $\delta_j > 1/2$. It follows from Lemma A.2 that there exists $s \in [e_T, e)$ such that $U(e, \boldsymbol{g}_j) < U(e, \boldsymbol{g}_s^{**})$ and

$$\|\boldsymbol{g}_s^{**} - \tilde{\boldsymbol{g}}_T\|^2 \le \|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_T\|^2 + \frac{(2-\delta_j^{-1})p}{\alpha_{-j}}. \tag{A.35}$$

By (A.35), we obtain that

$$V_{-j}(\boldsymbol{g}_s^{**}, \boldsymbol{g}_j) = \frac{1+p}{2} - \frac{\alpha_{-j}}{2}\|\boldsymbol{g}_s^{**} - \tilde{\boldsymbol{g}}_T\|^2 \ge \frac{1+p}{2} - \frac{\alpha_{-j}}{2}\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_T\|^2 - \frac{(2-\delta_j^{-1})p}{2}. \tag{A.36}$$

Note that

$$\frac{\alpha_{-j}}{2}\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_T\|^2 = \frac{\alpha_{-j}}{\alpha_j}N_j(\boldsymbol{g}_j;\alpha_j) = \delta_j^{-1}N_j(\boldsymbol{g}_j;\alpha_j) < \frac{\delta_j^{-1}p}{2},$$

so from (A.36) we have

$$V_{-j}(\boldsymbol{g}_s^{**},\boldsymbol{g}_j) > \frac{1-p}{2} = V_{-j}(\tilde{\boldsymbol{g}}_T,\boldsymbol{g}_j^*).$$

Similarly, we can conclude that $\mathrm{BR}_{-j}^{\varepsilon}(\boldsymbol{g}_j)$ has form (A.30) for sufficient small $\varepsilon > 0$. As a result, we can also get (A.31), implying (A.34).

Case II: $\alpha_B/\alpha_A \in (0, 1/2]$**.** It is enough to focus on the scenario $2\alpha_{e,p} < \alpha_B < \alpha_A$, corresponding to (iii) in **Case I**, and the remaining cases can be treated in an analogous way to **Case I** and therefore are omitted. When $2\alpha_{e,p} < \alpha_B < \alpha_A$, we verify that $(\tilde{\boldsymbol{g}}_T, \boldsymbol{g}^{**}(\alpha_A))$ is an SSE.

For Agency A: Our goal is to get the inequality:

$$\lim_{\varepsilon\to 0}\min_{\boldsymbol{g}_B\in\mathrm{BR}_B^{\varepsilon}(\boldsymbol{g}_A)} V_A(\boldsymbol{g}_A,\boldsymbol{g}_B) \le V_A(\tilde{\boldsymbol{g}}_T,\boldsymbol{g}^{**}(\alpha_A)) = \frac{1-p}{2}. \tag{A.37}$$

We prove it by considering $\boldsymbol{g}_A$ in three scenarios.

(a) $N_A(\boldsymbol{g}_A;\alpha_A) \ge p$.

In this case, $V_A(\boldsymbol{g}_A,\tilde{\boldsymbol{g}}_T) = (1+p)/2 - N_A(\boldsymbol{g}_A;\alpha_A) \le (1-p)/2$, which implies (A.37).

(b) $N_A(\boldsymbol{g}_A;\alpha_A) = 0$.

In this case, we know that $\boldsymbol{g}_A = \tilde{\boldsymbol{g}}_T$. It is straightforward that for any $\boldsymbol{g}_B \neq \tilde{\boldsymbol{g}}_T$, $V_A(\tilde{\boldsymbol{g}}_T,\boldsymbol{g}_B) = (1-p)/2$, which implies (A.37).

(c) $0 < N_A(\boldsymbol{g}_A;\alpha_A) < p$.

Since $\alpha_A > 2\alpha_{e,p}$ and $N_A(\boldsymbol{g}_A;\alpha_A) < p$, we have

$$\|\boldsymbol{g}_A - \tilde{\boldsymbol{g}}_T\|^2 < \frac{2p}{\alpha_A} < \|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2.$$

By Proposition 1 we obtain that $U(e,\boldsymbol{g}_A) < U(e,\boldsymbol{g}_e^*)$. Let $\delta_A = \alpha_A/\alpha_B$, then $\delta_A > 2$. Similar to the analysis of part (c) in **Case I** (iii), we also derive (A.37).

For Agency B: By definition of $\boldsymbol{g}^{**}(\alpha_A)$ in Theorem 1, we have

$$N_B(\boldsymbol{g}^{**}(\alpha_A);\alpha_B) = \frac{\alpha_B}{\alpha_A}\cdot p \le \frac{p}{2}.$$

Thus,

$$V_B(\boldsymbol{g}^{**}(\alpha_A),\tilde{\boldsymbol{g}}_T) = \frac{1+p}{2} - \frac{\alpha_B}{\alpha_A}\cdot p \ge \frac{1}{2}.$$

Our goal is to show that, for any $\boldsymbol{g}_B \in G$,

$$\lim_{\varepsilon\to 0}\min_{\boldsymbol{g}_A\in \mathrm{BR}_A^{\varepsilon}(\boldsymbol{g}_B)} V_B(\boldsymbol{g}_B,\boldsymbol{g}_A) \le V_B(\boldsymbol{g}^{**}(\alpha_A),\tilde{\boldsymbol{g}}_T) = \frac{1+p}{2} - \frac{\alpha_B}{\alpha_A}\cdot p. \tag{A.38}$$

It therefore suffices to focus on the case $N_B(\boldsymbol{g}_B;\alpha_B) < p\alpha_B/\alpha_A$, since for the remaining cases the inequality (A.38) follows directly. Note that $\alpha_A \ge 2\alpha_{e,p}$, it follows from $N_B(\boldsymbol{g}_B;\alpha_B) < p\alpha_B/\alpha_A$ that

$$\|\boldsymbol{g}_B - \tilde{\boldsymbol{g}}_T\|^2 < \frac{2p}{\alpha_A} < \|\boldsymbol{g}_e - \tilde{\boldsymbol{g}}_T\|^2.$$

By Proposition 1 we obtain that $U(e,\boldsymbol{g}_B) < U(e,\boldsymbol{g}_e^*)$. Similar to the analysis of part (c) in **Case I** (iii), we can also derive that for sufficiently small $\varepsilon > 0$,

$$\min_{\boldsymbol{g}_A\in \mathrm{BR}_A^{\varepsilon}(\boldsymbol{g}_B)} V_B(\boldsymbol{g}_B,\boldsymbol{g}_A) \le \frac{1-p}{2},$$

implying (A.38).

In summary, the above arguments establish the sufficiency part of the proof.

Necessity: Suppose that $(\boldsymbol{g}_A^*,\boldsymbol{g}_B^*)$ constitutes an SSE. Without loss of generality, we assume that $M_A^{\pi}(\boldsymbol{g}_A^*,\boldsymbol{g}_B^*) \le 1/2$. Then, following the same analysis as in the necessity proof of Proposition 3, we can also obtain (A.20) and (A.22). In the same way, we proceed to characterize the corresponding SSE under the four scenarios in (A.22).

(i) When $M_A^{\pi}(\boldsymbol{g}_A^{\#},\boldsymbol{g}_B^{\#}) = 1-p$ or p: This situation can be analogous to the proof of **case (i)** and **case (ii)** in Proposition 3 to get a contradiction.

(ii) When $M_A^\pi(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*) = (1-p)/2$: From (A.20), Agency A has to issue the ESG fundamental $\tilde{\boldsymbol{g}}_T$, then $V_A(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*) = (1-p)/2$. Therefore, the SSE indicates that

$$\lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_B \in \mathrm{BR}_B^\varepsilon(\boldsymbol{g}_A)} V_A(\boldsymbol{g}_A, \boldsymbol{g}_B) \leq \frac{1-p}{2}, \ \forall \boldsymbol{g}_A \in G. \tag{A.39}$$

Take $\boldsymbol{g}_A = \boldsymbol{g}_e^*$ in (A.39), then we have:

$$\frac{1-p}{2} \geq \lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_B \in \mathrm{BR}_B^\varepsilon(\boldsymbol{g}_e^*)} V_A(\boldsymbol{g}_e^*, \boldsymbol{g}_B) \geq \min_{\boldsymbol{g}_B \in G} V_A(\boldsymbol{g}_e^*, \boldsymbol{g}_B).$$

Since $U(e, \boldsymbol{g}_e^*)$ attains the maximum of utility from Lemma A.2, we have $M_A^\pi(\boldsymbol{g}_e^*, \boldsymbol{g}_B) \geq 1/2$ for any $\boldsymbol{g}_B \in G$. Recall that

$$V_A(\boldsymbol{g}_e^*, \boldsymbol{g}_B) = M_A^\pi(\boldsymbol{g}_e^*, \boldsymbol{g}_B) - \frac{\alpha_A}{2} \|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2,$$

so we get the inequality

$$\alpha_A \geq \frac{p}{\|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2} = \alpha_{e,p}.$$

On the other hand, for Agency B, we also have

$$\lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_A \in \mathrm{BR}_A^\varepsilon(\boldsymbol{g}_B)} V_B(\boldsymbol{g}_B, \boldsymbol{g}_A) \leq \frac{1+p}{2} - \frac{\alpha_B}{2} \|\boldsymbol{g}_B^* - \tilde{\boldsymbol{g}}_T\|^2, \ \forall \boldsymbol{g}_B \in G. \tag{A.40}$$

Take $\boldsymbol{g}_B = \boldsymbol{g}_e^*$ in (A.40), and notice that $\alpha_A \geq \alpha_{e,p}$, which means that $\mathrm{BR}_A^\varepsilon(\boldsymbol{g}_e^*)$ is a small neighborhood of $\tilde{\boldsymbol{g}}_T$ with sufficient small $\varepsilon > 0$. In this sense, we get

$$\lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_A \in \mathrm{BR}_A^\varepsilon(\boldsymbol{g}_B)} V_B(\boldsymbol{g}_e^*, \boldsymbol{g}_A) = V_B(\boldsymbol{g}_e^*, \tilde{\boldsymbol{g}}_T) = \frac{1+p}{2} - \frac{\alpha_B}{2} \|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2,$$

and then

$$\|\boldsymbol{g}_B^* - \tilde{\boldsymbol{g}}_T\|^2 \leq \|\boldsymbol{g}_e^* - \tilde{\boldsymbol{g}}_T\|^2.$$

Hence, there exists $s \in (e_T, e]$ such that $\|\boldsymbol{g}_s^{**} - \tilde{\boldsymbol{g}}_T\|^2 = \|\boldsymbol{g}_B^* - \tilde{\boldsymbol{g}}_T\|^2$. Taking $\boldsymbol{g}_s^{**}$ in (A.40), we conclude that $\mathrm{BR}_A^\varepsilon(\boldsymbol{g}_s^{**})$ should be a small neighborhood of $\tilde{\boldsymbol{g}}_T$ with a sufficiently small $\varepsilon > 0$. That is to say, for any $\eta > 0$, $\boldsymbol{g}_{s+\eta}^{**} \notin \mathrm{BR}_A^\varepsilon(\boldsymbol{g}_s^{**})$. For Agency A, since $\boldsymbol{g}_{s+\eta}^{**}$ is not a $\varepsilon-$best response for the rival's rating $\boldsymbol{g}_k^{**}$, we have:

$$N_A(\boldsymbol{g}_s^{**}; \alpha_A) > p.$$

If $s < e$, it holds when $\alpha_A/\alpha_B \geq 2$; otherwise, since $\boldsymbol{g}_B^* = \boldsymbol{g}_e^*$, we conclude that the Agency B has low aversion. Those facts are consistent with the equilibrium characterization in (3.8) and (3.9)

(iii) When $M_A^{\pi}(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*) = p/2$: This situation can be analogous to the proof of **case (iv)** in Proposition 3 to rule out the possibility of the existence of SSE when $p < 1$.

(iv) When $M_A^{\pi}(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*) = 1/2$: In this situation, if both agencies have low aversion, then SSE holds for the unique NE, as proved in case (v) of Proposition 3. Otherwise, we are left with the case where neither agency has low aversion nor low relative aversion (so we only need to consider parameter configurations lying in **Region III** of Figure 1). By the sufficiency analysis of (iii) in **Case I**, it follows that $(\tilde{\boldsymbol{g}}_T, \tilde{\boldsymbol{g}}_T)$ is an SSE in the present setting and that (A.34) holds. That is, in an SSE we must have

$$V_A(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*) = V_B(\boldsymbol{g}_B^*, \boldsymbol{g}_A^*) = \frac{1}{2},$$

which implies that the only possible equilibrium profile is given by $\boldsymbol{g}_A = \boldsymbol{g}_B = \tilde{\boldsymbol{g}}_T$.
In summary, we establish the necessity part and complete the proof.

A.7 Proof of Proposition 4

(1) Based on (3.8) in Theorem 1, if $\alpha_B/\alpha_A \in (1/2, 1)$, Agency j issues the rating $\boldsymbol{g}_j^* \neq \tilde{\boldsymbol{g}}_T$ if and only if Agency j has low aversion, i.e., $\alpha_j \leq \alpha_{e,p}$. From Remark 4, we obtain that $\boldsymbol{g}_j^* \neq \tilde{\boldsymbol{g}}_T$ if and only if

$$e \leq e_T + \sqrt{\frac{p}{\alpha_j}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}.$$

When this condition holds, it follows from Theorem 1 that the rating $\boldsymbol{g}_j^* = \boldsymbol{g}_e^* = \tilde{\boldsymbol{g}}_T + (e - e_T)\boldsymbol{g}_M$ is issued, whose deviation norm is given by

$$\|\boldsymbol{g}_j^* - \tilde{\boldsymbol{g}}_T\|^2 = (e - e_T)^2 \boldsymbol{g}_M'\boldsymbol{g}_M = (e - e_T)^2 \frac{n(\boldsymbol{\omega}_M'\mathbf{1})^2}{n - (\boldsymbol{\omega}_M'\mathbf{1})^2}. \tag{A.41}$$

(2) Based on (3.9) in Theorem 1, if $\alpha_B/\alpha_A \in (0, 1/2]$, Agency B always deviates from the ESG fundamental, while Agency A deviates if and only if Agency A has low aversion, i.e., $\alpha_A \leq \alpha_{e,p}$.

From Remark 4, $\alpha_A \leq \alpha_{e,p}$ is equivalent to

$$e \leq e_T + \sqrt{\frac{p}{\alpha_A}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}$$

When this condition holds, the rating $\boldsymbol{g}_A^* = \boldsymbol{g}_e^* = \tilde{\boldsymbol{g}}_T + (e - e_T)\boldsymbol{g}_M$ is issued, whose norm of deviation is given by (A.41) for $j = A$. From Theorem 1, Agency B issues $\boldsymbol{g}_e^*$ when

$$e \leq e_T + \sqrt{\frac{2p}{\alpha_A}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]},$$

and issues $\boldsymbol{g}_r^{**}$ when

$$e > e_T + \sqrt{\frac{2p}{\alpha_A}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]},$$

where $r = e_T + \sqrt{\frac{2p}{\alpha_A}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}$. Combining both scenarios, Agency B issues rating $\boldsymbol{g}_B^* = \boldsymbol{g}_s^{**}$, where $s = \min\{e, r\}$. From Lemma A.1 and (3.11), the deviation norm of Agency B is given by

$$\|\boldsymbol{g}_B^* - \tilde{\boldsymbol{g}}_T\|^2 = (s - e_T)^2 \boldsymbol{g}_M'\boldsymbol{g}_M = \min\left\{(e - e_T)^2 \frac{n(\boldsymbol{\omega}_M'\mathbf{1})^2}{n - (\boldsymbol{\omega}_M'\mathbf{1})^2}, \frac{2p}{\alpha_A}\right\}.$$

A.8 Proof of Proposition 5

Based on Theorem 1, we know that Agency B issues $\boldsymbol{g}_B^* = \boldsymbol{g}_e^*$ in **Region I–II**. According to the expressions (3.4) and (3.5), the cross-sectional association $\boldsymbol{\mu}'\boldsymbol{g}_B^*$ is given by

$$\begin{aligned}
\boldsymbol{\mu}'\boldsymbol{g}_B^* &= \boldsymbol{\mu}'\left(\tilde{\boldsymbol{g}}_T + (e - e_T)\boldsymbol{g}_M\right) \\
&= \boldsymbol{\mu}'\tilde{\boldsymbol{g}}_T + (e - e_T)(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})'\boldsymbol{g}_M \\
&= \boldsymbol{\mu}'\tilde{\boldsymbol{g}}_T + (e - e_T)\frac{(\boldsymbol{\omega}_M'\mathbf{1})}{n - (\boldsymbol{\omega}_M'\mathbf{1})^2}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})'\left(n \cdot \boldsymbol{\omega}_M - \boldsymbol{\omega}_M'\mathbf{1} \cdot \mathbf{1}\right) \\
&= \boldsymbol{\mu}'\tilde{\boldsymbol{g}}_T + (e - e_T)\frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\mu} - (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})'\left(n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{1} \cdot \boldsymbol{\Sigma}^{-1}\mathbf{1}\right) \\
&= \boldsymbol{\mu}'\tilde{\boldsymbol{g}}_T + (e - e_T)\frac{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\mu} - (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})'\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \bar{\mu}\boldsymbol{\Sigma}^{-1}\mathbf{1}\right) \\
&= \boldsymbol{\mu}'\tilde{\boldsymbol{g}}_T + (e - e_T)\frac{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1} \cdot (\boldsymbol{\mu} - \bar{\mu}\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})}{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\mu} - (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}.
\end{aligned}$$

The second equality is because $\boldsymbol{g}_M'\mathbf{1} = 0$. By Cauchy's inequality, we have

$$n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\mu} - (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2 = \mathbf{1}'\mathbf{1}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2 > 0.$$

Therefore, the cross-sectional association $\boldsymbol{\mu}'\boldsymbol{g}_B^*$ is positive if and only if

$$e > e_T - \frac{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\mu} - (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}{n\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1})}\boldsymbol{\mu}'\tilde{\boldsymbol{g}}_T.$$

A.9 Proof of Proposition 6

Based on (3.9) and (3.8) in Theorem 1, we know that rating disagreement $\boldsymbol{g}_A^* \neq \boldsymbol{g}_B^*$ arises if and only if Agency B deviates and Agency A issues the ESG fundamental in **Region II-III**. As such, the necessary and sufficient condition of rating disagreement is that $\alpha_B \leq \alpha_{e,p} < \alpha_A$ or $\alpha_A > 2\max\{\alpha_B, \alpha_{e,p}\}$. Then the necessity condition of rating disagreement is $\alpha_A > \alpha_{e,p}$. From Remark 4 it is equivalent to

$$e > e_T + \sqrt{\frac{p}{\alpha_A}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}.$$

Under this condition, it follows that rating disagreement arises if either $\alpha_B \leq \alpha_{e,p}$ or $\alpha_A > 2\alpha_B$ holds, i.e.,

$$e \leq e_T + \sqrt{\frac{p}{\alpha_B}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]} \quad \text{or} \quad \alpha_A > 2\alpha_B.$$

Based on Theorem 1, we know that, when rating disagreement arises, the norm of rating disagreement is equal to the norm of Agency B's rating deviation, i.e., $\|\boldsymbol{g}_A^* - \boldsymbol{g}_B^*\|^2 = \|\boldsymbol{g}_B^* - \tilde{\boldsymbol{g}}_T\|^2$. Then rating disagreement tends to grow when rating deviation becomes larger.

A.10 Proof of Proposition 7

It follows from Theorem 1 that Agency A issues the ESG fundamental $\tilde{\boldsymbol{g}}_T$ and Agency B issues the ESG rating $\boldsymbol{g}_e^*$. In this case, the green investor chooses the ESG rating $\boldsymbol{g}_e^*$ issued by Agency B. Then we prove the remaining results in Proposition 7.

(1) Based on (2.5) in Proposition 1, the pecuniary utility of the green investor under the ESG

rating $\boldsymbol{g}_e^*$ is given by

$$u(e, \boldsymbol{g}_e^*) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2} - \frac{\left(\max\{\boldsymbol{\omega}_M'(e\mathbf{1} - \boldsymbol{g}_e^*), 0\}\right)^2}{(e\mathbf{1} - \boldsymbol{g}_e^*)'\boldsymbol{\Sigma}^{-1}(e1 - \boldsymbol{g}_e^*)} = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2}.$$

Based on (2.5) in Proposition 1, under the ESG fundamental $\tilde{\boldsymbol{g}}_T$, the pecuniary utility of the green investor is given by

$$u(e, \tilde{\boldsymbol{g}}_T) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2} - \frac{\left(\max\{\boldsymbol{\omega}_M'(e\mathbf{1} - \tilde{\boldsymbol{g}}_T), 0\}\right)^2}{(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e1 - \tilde{\boldsymbol{g}}_T)} = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2} - \frac{\big(e - e_T)^2(\boldsymbol{\omega}_M'\mathbf{1})^2}{(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e1 - \tilde{\boldsymbol{g}}_T)}.$$

(2) Based on (2.4) and (2.5) in Proposition 1, we know that the optimal investment strategy $\boldsymbol{\omega}_e^*$ is $\boldsymbol{\omega}_e^* = \boldsymbol{\omega}(\gamma, e, \boldsymbol{g}_e^*) = \gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$.

By the definition of G_F and G_O in (4.4), we get

$$G_F(\boldsymbol{\omega}_e^*) = \frac{\boldsymbol{\omega}_e^{*\prime}\tilde{\boldsymbol{g}}_T}{\boldsymbol{\omega}_e^{*\prime}\mathbf{1}} = \frac{\gamma^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{g}}_T}{\gamma^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} = e_T, \tag{A.42}$$

and

$$G_O(\boldsymbol{\omega}_e^*) = \frac{\boldsymbol{\omega}_e^{*\prime}\boldsymbol{g}_e^*}{\boldsymbol{\omega}_e^{*\prime}\mathbf{1}} = \frac{\gamma^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_e^*}{\gamma^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} = e. \tag{A.43}$$

Under the ESG fundamental $\tilde{\boldsymbol{g}}_T$, based on (2.4) in Proposition 1 the optimal investment strategy of the green investor is

$$\boldsymbol{\omega}_T^* = \boldsymbol{\omega}(\gamma, e, \tilde{\boldsymbol{g}}_T) = \gamma^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)}{\gamma(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)}\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T).$$

Hence, we have

$$\begin{aligned}
\boldsymbol{\omega}_T^{*\prime}\mathbf{1} &= \gamma^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1} - \gamma^{-1}\frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)\mathbf{1}'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)}{(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)} \\
&= \gamma^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1} - \gamma^{-1}\frac{(e - e_T)\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}\cdot\mathbf{1}'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)}{(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)} \\
&= \gamma^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}\frac{(e_T\mathbf{1} - \tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)}{(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1} - \tilde{\boldsymbol{g}}_T)}.
\end{aligned}$$

By Definition (4.4), we get

$$
\begin{aligned}
G_F(\boldsymbol{\omega}_T^*) &= \frac{\boldsymbol{\omega}_T^{*\prime}\tilde{\boldsymbol{g}}_T}{\boldsymbol{\omega}_T^{*\prime}\mathbf{1}} \\
&= \frac{\gamma^{-1}}{\boldsymbol{\omega}_T^{*\prime}\mathbf{1}}\left[\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{g}}_T - \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)\tilde{\boldsymbol{g}}_T'\boldsymbol{\Sigma}^{-1}(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)}{(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)}\right] \\
&= \frac{\gamma^{-1}}{\boldsymbol{\omega}_T^{*\prime}\mathbf{1}}\left[e_T\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1} - \frac{(e-e_T)\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}\cdot\tilde{\boldsymbol{g}}_T'\boldsymbol{\Sigma}^{-1}(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)}{(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)}\right] \\
&= \frac{\gamma^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\boldsymbol{\omega}_T^{*\prime}\mathbf{1}}\left[e_T - \frac{(e-e_T)\tilde{\boldsymbol{g}}_T'\boldsymbol{\Sigma}^{-1}(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)}{(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)}\right] \\
&= \frac{\gamma^{-1}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\boldsymbol{\omega}_T^{*\prime}\mathbf{1}}\frac{(e\cdot e_T\mathbf{1}-e\tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)}{(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)'\boldsymbol{\Sigma}^{-1}(e\mathbf{1}-\tilde{\boldsymbol{g}}_T)} \\
&= e.
\end{aligned}
\tag{A.44}
$$

From (A.42) and (A.44), we have $G_R(\boldsymbol{\omega}_T^*) = e > e_T = G_R(\boldsymbol{\omega}_e^*)$.

A.11 Proposition A.1 and its proof

Proposition A.1. Under Assumption 1, if the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\mathrm{wa}}$ satisfies $\tilde{\boldsymbol{g}}_{T,\mathrm{wa}}'\mathbf{1} = 0$, then there exists an SSE $(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*)$. Let

$$
\boldsymbol{g}_{e,\mathrm{wa}}^* := \tilde{\boldsymbol{g}}_{T,\mathrm{wa}} + (e - e_{T,\mathrm{wa}})\boldsymbol{g}_M, \qquad \boldsymbol{g}_{b,\mathrm{wa}}^* := \tilde{\boldsymbol{g}}_{T,\mathrm{wa}} + (b - e_{T,\mathrm{wa}})\boldsymbol{g}_M.
$$

For each agency $j \in \{\mathrm{A},\mathrm{B}\}$, the equilibrium ratings of each agency are characterized as follows:

(1) If $\alpha_j/\alpha_{-j} > 1/2$, then Agency j's issued rating $\boldsymbol{g}_j^*$ in equilibrium is given by

$$
\boldsymbol{g}_j^* = \begin{cases}
\boldsymbol{g}_{e,\mathrm{wa}}^*, & e_{T,\mathrm{wa}} \geq e - \sqrt{\dfrac{p}{\alpha_j}\left[\dfrac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \dfrac{1}{n}\right]} \quad \text{and} \quad b \leq e_{T,\mathrm{wa}} < e, \\
\boldsymbol{g}_{b,\mathrm{wa}}^*, & b - \sqrt{\dfrac{1-p}{\alpha_j}\left[\dfrac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \dfrac{1}{n}\right]} \leq e_{T,\mathrm{wa}} < b, \\
\tilde{\boldsymbol{g}}_{T,\mathrm{wa}}, & \text{otherwise.}
\end{cases}
\tag{A.45}
$$

(2) If $\alpha_j/\alpha_{-j} \le 1/2$, then Agency j's issued rating $\boldsymbol{g}_j^*$ in equilibrium is given by

$$\boldsymbol{g}_j^* = \begin{cases} \boldsymbol{g}_{e,\mathrm{wa}}^*, & e_{T,\mathrm{wa}} \ge e - \sqrt{\dfrac{2p}{\alpha_{-j}}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]} \quad \text{and} \quad b \le e_{T,\mathrm{wa}} < e, \\ \boldsymbol{g}_{\mathrm{wa}}^{**}(\alpha_{-j}, p), & b \le e_{T,\mathrm{wa}} < e - \sqrt{\dfrac{2p}{\alpha_{-j}}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}, \\ \boldsymbol{g}_{b,\mathrm{wa}}^*, & b - \sqrt{\dfrac{(1-p)}{\alpha_{-j}}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]} \le e_{T,\mathrm{wa}} < b, \\ \boldsymbol{g}_{\mathrm{wa}}^{**}(\alpha_{-j}, 1-p), & b - \sqrt{\dfrac{2(1-p)}{\alpha_{-j}}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]} \le e_{T,\mathrm{wa}} < b - \sqrt{\dfrac{(1-p)}{\alpha_{-j}}\left[\frac{1}{(\boldsymbol{\omega}_M'\mathbf{1})^2} - \frac{1}{n}\right]}, \\ \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}, & \text{otherwise.} \end{cases} \tag{A.46}$$

where, for $q \in \{p, 1-p\}$, $\boldsymbol{g}_{\mathrm{wa}}^{**}(\alpha_{-j}, q)$ is a solution of

$$\max_{\boldsymbol{g}\in G} U(e, \boldsymbol{g}) \quad \text{s.t.} \quad \frac{\alpha_{-j}}{2}\|\boldsymbol{g} - \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}\|^2 \le q, \ \boldsymbol{g}'\mathbf{1} = 0.$$

Proof. To prove Proposition A.1, we derive the issued ratings in equilibrium and calculate the norm of deviation by considering the two cases $\alpha_B/\alpha_A \in (1/2, 1)$ and $\alpha_B/\alpha_A \in (0, 1/2]$.

Case I: $\alpha_B/\alpha_A \in (1/2, 1)$**.**

In this case, we consider the following three ranges of $e_{T,\mathrm{wa}}$: $e_{T,\mathrm{wa}} \ge e$, $e_{T,\mathrm{wa}} \in [b, e)$ and $e_{T,\mathrm{wa}} \le b$.

(i) $e_{T,\mathrm{wa}} \ge e$.

Based on Proposition 1, for any $s \le e$ and any $\boldsymbol{g} \in G$, we have the fact that

$$U(s, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\|^2} \ge U(s, \boldsymbol{g}). \tag{A.47}$$

Combining (2.9) and (A.47), for any $\boldsymbol{g}_A$, $\boldsymbol{g}_B \in G$, we have

$$P_A(e; \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}, \boldsymbol{g}_B) \ge P_A(e; \boldsymbol{g}_A, \boldsymbol{g}_B), \quad P_B(e; \boldsymbol{g}_A, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}) \ge P_B(e; \boldsymbol{g}_A, \boldsymbol{g}_B). \tag{A.48}$$

Based on (2.12) and (A.48), for $j \in \{A, B\}$ and $\boldsymbol{g}_j$, $\boldsymbol{g}_{-j} \in G$, we have

$$\begin{aligned} V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}) &= M_j^{\pi}(\boldsymbol{g}_A, \boldsymbol{g}_B) - N_j(\boldsymbol{g}_j; \alpha_j) \\ &= pP_j(e; \boldsymbol{g}_A, \boldsymbol{g}_B) + (1-p)P_j(e; \boldsymbol{g}_A, \boldsymbol{g}_B) - \frac{\alpha_j}{2}\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}\|^2 \\ &\leq pP_j(e; \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}, \boldsymbol{g}_{-j}) + (1-p)P_j(e; \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}, \boldsymbol{g}_{-j}) \\ &= V_j(\tilde{\boldsymbol{g}}_{T,\mathrm{wa}}, \boldsymbol{g}_{-j}). \end{aligned}$$

By the above inequality, for $j \in \{A, B\}$ and $\boldsymbol{g}_j \in G$, we get

$$\begin{aligned} V_j(\tilde{\boldsymbol{g}}_{T,\mathrm{wa}}, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}) &\geq V_j(\boldsymbol{g}_j, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}) \\ &= \lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_{-j} \in \mathrm{BR}_{-j}^{\varepsilon}(\boldsymbol{g}_j)} V_j(\boldsymbol{g}_j, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}) \\ &\geq \lim_{\varepsilon \to 0} \min_{\boldsymbol{g}_{-j} \in \mathrm{BR}_{-j}^{\varepsilon}(\boldsymbol{g}_j)} V_j\big(\boldsymbol{g}_j, \boldsymbol{g}_{-j}\big). \end{aligned} \tag{A.49}$$

Combining (A.49) and (3.2) in Definition 6, $\boldsymbol{g}_A^* = \boldsymbol{g}_B^* = \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}$ formulates an SSE.The necessity is similar to the proof of Theorem 1 in Section A.6.

(ii) $b < e_{T,\mathrm{wa}} < e$.

In this case, because $e_{T,\mathrm{wa}}$ satisfies the assumption $e_{T,\mathrm{wa}} \in [b, e)$, we directly apply the equilibrium behaviors in Theorem 1 by considering the three cases $e_{T,\mathrm{wa}} \geq e - \sqrt{\frac{p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$, $e_{T,\mathrm{wa}} \in \left[e - \sqrt{\frac{p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}, e - \sqrt{\frac{p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}\right)$ and $e_{T,\mathrm{wa}} < e - \sqrt{\frac{p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

(a) $e_{T,\mathrm{wa}} \geq e - \sqrt{\frac{p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (3.7), we treat $e_{T,\mathrm{wa}}$ as the intrinsic ESG benchmark and know that the post-greenwashing threshold $\alpha_{e,p}^{\mathrm{wa}}$ is

$$\alpha_{e,p}^{\mathrm{wa}} = \frac{1}{(e - e_{T,\mathrm{wa}})^2} \frac{n - (\boldsymbol{\omega}_M' \mathbf{1})^2}{n(\boldsymbol{\omega}_M' \mathbf{1})^2}, \tag{A.50}$$

and $\alpha_B < \alpha_A \leq \alpha_{e,p}^{\mathrm{wa}}$, which shows that α_A and α_B are both low. Based on (3.8) in Case (1) of Theorem 1, the issued ESG ratings are $\boldsymbol{g}_A^* = \boldsymbol{g}_B^* = \boldsymbol{g}_{e,\mathrm{wa}}^*$.

(b) $e - \sqrt{\frac{p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}} \leq e_{T,\mathrm{wa}} < e - \sqrt{\frac{p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.50), we get $\alpha_B \leq \alpha_{e,p}^{\mathrm{wa}} < \alpha_A < 2\alpha_{e,p}^{\mathrm{wa}}$, which shows that α_A is medium

and α_B is low. Based on (3.8) in Case (1) of Theorem 1, the issued ESG ratings are $\boldsymbol{g}_A^* = \tilde{\boldsymbol{g}}_{T,\text{wa}}$, $\boldsymbol{g}_B^* = \boldsymbol{g}_{e,\text{wa}}^*$.

(c) $e_{T,\text{wa}} < e - \sqrt{\frac{p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.50), we get $\alpha_{e,p}^{\text{wa}} < \alpha_B < \alpha_A$, which shows that α_A and α_B are both medium or high. Based on (3.8) in Case (1) of Theorem 1, the issued ESG ratings are $\boldsymbol{g}_A^* = \boldsymbol{g}_B^* = \tilde{\boldsymbol{g}}_{T,\text{wa}}$.

(iii) $e_{T,\text{wa}} < b$.

In this case, we treat b as another ESG target lower than e and calculate the lower post-greenwashing threshold

$$\alpha_{b,p}^{\text{wa}} = \frac{1}{(b - e_{T,\text{wa}})^2} \frac{n - (\boldsymbol{\omega}_M' \mathbf{1})^2}{n(\boldsymbol{\omega}_M' \mathbf{1})^2}. \tag{A.51}$$

We derive the issued ratings in equilibrium and calculate the norm of deviation by considering the three cases $e_{T,\text{wa}} \geq b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$, $e_{T,\text{wa}} \in \left[b - \sqrt{\frac{1-p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}, b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}\right)$ and $b < b - \sqrt{\frac{1-p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

(a) $e_{T,\text{wa}} \geq b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.51), we get $\alpha_{e,p}^{\text{wa}} < \alpha_B < \alpha_A < \alpha_{b,p}^{\text{wa}}$, and we prove that the issued ESG ratings $\boldsymbol{g}_A^* = \boldsymbol{g}_B^* = \boldsymbol{g}_{b,\text{wa}}^*$ formulate an SSE.

For $j \in \{\text{A,B}\}$ and any $\boldsymbol{g}_j \neq \boldsymbol{g}_{b,\text{wa}}^*$, based on Lemma A.1, we have the fact that either $\boldsymbol{\omega}_M' \boldsymbol{g}_j < b\ \boldsymbol{\omega}_M' \mathbf{1}$ or for any $\varepsilon > 0$, there exists $\boldsymbol{g}_{-j}^* \in G$ such that $\boldsymbol{\omega}_M' \boldsymbol{g}_{-j}^* \geq b\ \boldsymbol{\omega}_M' \mathbf{1}$, $U(e, \boldsymbol{g}_{-j}^*) > U(e, \boldsymbol{g}_j)$, and $\|\boldsymbol{g}_{-j}^* - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 < \|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 + \varepsilon$. Similar to the proof in **Case I** (i), we calculate the utilities of the agencies and know that

$$V_j(\boldsymbol{g}_j, \boldsymbol{g}_{-j}^*) \leq \frac{1-p}{2} - \frac{\alpha_j}{2}(b - e_{T,\text{wa}})^2 \boldsymbol{g}_M' \boldsymbol{g}_M < V_j(\boldsymbol{g}_{b,\text{wa}}^*, \boldsymbol{g}_{b,\text{wa}}^*).$$

Based on (3.2) in Definition 6, the rating profile $(\boldsymbol{g}_{b,\text{wa}}^*, \boldsymbol{g}_{b,\text{wa}}^*)$ is an SSE.

(b) $b - \sqrt{\frac{1-p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}} \leq \tilde{\boldsymbol{g}}_{T,\text{wa}} < b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.51), we get $\alpha_{e,p}^{\text{wa}} < \alpha_B < \alpha_{b,p}^{\text{wa}} < \alpha_A < 2\alpha_{b,p}^{\text{wa}}$. In the following, we prove that, in an SSE, the rating of Agency A is supposed to be $\boldsymbol{g}_A^* = \tilde{\boldsymbol{g}}_{T,\text{wa}}$, and the rating of Agency B is supposed to be $\boldsymbol{g}_B^* = g_{b,\text{wa}}^*$.

For Agency A, when $\boldsymbol{g}_A \neq \tilde{\boldsymbol{g}}_{T,\text{wa}}$, based on Lemma A.1, we have the fact that either $\|\boldsymbol{g}_A - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 > 2/\alpha_A$ or for any $\varepsilon > 0$, there exists $\tilde{\boldsymbol{g}}_B \in G$ such that $U(b, \tilde{\boldsymbol{g}}_B) \geq U(b, \boldsymbol{g}_A)$, $U(e, \tilde{\boldsymbol{g}}_B) > U(e, \boldsymbol{g}_A)$, and $\|\tilde{\boldsymbol{g}}_B - \tilde{\boldsymbol{g}}_{T,\text{wa}}\| \leq \|\boldsymbol{g}_A - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 + \varepsilon$. Similar to the proof in **Case I** (i), we calculate the utilities of the agencies and know that

$$V(\boldsymbol{g}_A, \tilde{\boldsymbol{g}}_B) \leq \frac{1-p}{2} - \frac{\alpha_A}{2}\|\boldsymbol{g}_A - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 \leq 0 = V_A(\tilde{\boldsymbol{g}}_{T,\text{wa}}, \boldsymbol{g}^*_{b,\text{wa}}).$$

Based on (3.2) in Definition 6, in an SSE, the rating of Agency A is supposed to be $\boldsymbol{g}^*_A = \tilde{\boldsymbol{g}}_{T,\text{wa}}$.

For Agency B, when $\boldsymbol{g}_B \neq \boldsymbol{g}^*_{b,\text{wa}}$, we just choose $\tilde{\boldsymbol{g}}_A = \tilde{\boldsymbol{g}}_{T,\text{wa}}$ or $\tilde{\boldsymbol{g}}_A = \boldsymbol{g}^*_{b,\text{wa}}$. Similar to the proof in **Case I** (i), we have the fact that

$$V_B(\boldsymbol{g}_B, \tilde{\boldsymbol{g}}_{T,\text{wa}}) \leq 1 - \frac{\alpha_B}{2}\|\boldsymbol{g}^*_{b,\text{wa}} - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 = V_B(\boldsymbol{g}^*_{b,\text{wa}}, \tilde{\boldsymbol{g}}_{T,\text{wa}}).$$

Based on (3.2) in Definition 6, in an SSE, the rating of Agency B is supposed to be $\boldsymbol{g}^*_B = g^*_{b,\text{wa}}$.

Therefore, based on Definition 6, the rating profile $(\tilde{\boldsymbol{g}}_{T,\text{wa}}, \boldsymbol{g}^*_{b,\text{wa}})$ is an SSE.

(c) $e_{T,\text{wa}} < b - \sqrt{\frac{1-p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.51), We get $\alpha^{\text{wa}}_{b,p} < \alpha_B < \alpha_A$. And we prove that the rating profile $(\boldsymbol{g}^*_A, \boldsymbol{g}^*_B) = (\tilde{\boldsymbol{g}}_{T,\text{wa}}, \tilde{\boldsymbol{g}}_{T,\text{wa}})$ is an SSE.

For $j \in \{\text{A, B}\}$ and any $\boldsymbol{g}_j \neq \tilde{\boldsymbol{g}}_{T,\text{wa}}$, based on Lemma A.1, we have the fact that either $\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 \geq 1/\alpha_j$ or there exists $\boldsymbol{g}^*_{-j} \in G$ such that $\tilde{\boldsymbol{g}}_{-j} \in G$ such that $U(b, \tilde{\boldsymbol{g}}_{-j}) \geq U(b, \boldsymbol{g}_j)$, $U(e, \tilde{\boldsymbol{g}}_{-j}) \geq U(e, \boldsymbol{g}_j)$, and $\|\tilde{\boldsymbol{g}}_{-j} - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 < \|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 + \varepsilon$. Similar to the proof in **Case I** (i), we calculate the utilities of the agencies and know that

$$V_j(\boldsymbol{g}_j, \tilde{\boldsymbol{g}}_{-j}) = \frac{1}{2} - \frac{\alpha_j}{2}\|\boldsymbol{g}_j - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 < \frac{1}{2} = V_j(\tilde{\boldsymbol{g}}_{T,\text{wa}}, \tilde{\boldsymbol{g}}_{T,\text{wa}}).$$

Based on (3.2) in Definition 6, the rating profile $(\tilde{\boldsymbol{g}}_{T,\text{wa}}, \tilde{\boldsymbol{g}}_{T,\text{wa}})$ is an SSE.

Case II: $\alpha_B/\alpha_A \in (0, 1/2]$**.**

In this case, we consider the following three ranges of $e_{T,\text{wa}}$: $e_{T,\text{wa}} \geq e$, $e_{T,\text{wa}} \in [b, e)$ and $e_{T,\text{wa}} \leq b$.

(i) $e_{T,\text{wa}} \geq e$.

Based on Theorem 1, similar to the proof in **Case I** (i), the rating profile $(\tilde{\boldsymbol{g}}_{T,\text{wa}}, \tilde{\boldsymbol{g}}_{T,\text{wa}})$ is an SSE.

(ii) $b \leq e_{T,\text{wa}} < e$.

In this case, because $e_{T,\text{wa}}$ satisfies the assumption $e_{T,\text{wa}} \in [b, e)$, we directly apply the equilibrium behaviors in Theorem 1 by considering the three cases $e_{T,\text{wa}} \geq e - \sqrt{\frac{p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$, $e_{T,\text{wa}} \in \left[e - \sqrt{\frac{2p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}, e - \sqrt{\frac{p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}\right)$ and $e_{T,\text{wa}} < e - \sqrt{\frac{2p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

(a) $e_{T,\text{wa}} \geq e - \sqrt{\frac{p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.50), we treat $e_{T,\text{wa}}$ as the intrinsic ESG benchmark and know that the post-greenwashing threshold $\alpha_{e,p}^{\text{wa}}$ is

$$\alpha_{e,p}^{\text{wa}} = (e - e_{T,\text{wa}})^2 \frac{n(\boldsymbol{\omega}_M' \mathbf{1})^2}{n - (\boldsymbol{\omega}_M' \mathbf{1})^2},$$

and $\alpha_B < \alpha_A < \alpha_{e,p}^{\text{wa}}$. Based on (3.9) in Case (2) of Theorem 1, the issued ESG ratings are $\boldsymbol{g}_A^* = \boldsymbol{g}_B^* = \boldsymbol{g}_{e,\text{wa}}^*$.

(b) $e - \sqrt{\frac{2p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}} \leq e_{T,\text{wa}} < e - \sqrt{\frac{p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.50), we get $\alpha_B < \alpha_{e,p}^{\text{wa}} < \alpha_A < 2\alpha_{e,p}^{\text{wa}}$. Based on (3.9) in Case (2) of Theorem 1, the issued ESG ratings are $\boldsymbol{g}_A^* = \tilde{\boldsymbol{g}}_{T,\text{wa}}$, $\boldsymbol{g}_B^* = \boldsymbol{g}_{e,\text{wa}}^*$.

(c) $e_{T,\text{wa}} < e - \sqrt{\frac{2p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.50), we get $\alpha_A \geq 2\alpha_B$, $\alpha_A > 2\alpha_{e,p}^{\text{wa}}$. Based on (3.9) in Case (2) of Theorem 1, the issued ESG ratings are $\boldsymbol{g}_A^* = \tilde{\boldsymbol{g}}_{T,\text{wa}}$, $\boldsymbol{g}_B^* = \boldsymbol{g}_{\text{wa}}^{**}(\alpha_A, p)$.

(iii) $e_{T,\text{wa}} < b$.

In this case, we treat b as another ESG target lower than e and calculate the lower post-greenwashing threshold

$$\alpha_{b,p}^{\text{wa}} = (b - e_{T,\text{wa}})^2 \frac{n(\boldsymbol{\omega}_M' \mathbf{1})^2}{n - (\boldsymbol{\omega}_M' \mathbf{1})^2}.$$

(a) $e_{T,\text{wa}} \geq b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.51), we get $\alpha_{e,p}^{\text{wa}} < \alpha_B < \alpha_A < \alpha_{b,p}^{\text{wa}}$. Similar to the proof in **Case I** (iii), we have the fact that the issued ESG ratings $\boldsymbol{g}_A^* = \boldsymbol{g}_B^* = \boldsymbol{g}_{b,\text{wa}}^*$ formulate an SSE.

(b) $b - \sqrt{\frac{2(1-p)}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}} \le e_{T,\text{wa}} < b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.51), we get $\alpha_{e,p}^{\text{wa}} < \alpha_B < \alpha_{b,p}^{\text{wa}} < \alpha_A < 2\alpha_{b,p}^{\text{wa}}$. Similar to the proof in **Case I** (iii), we have the fact that the rating profile $(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*) = (\tilde{\boldsymbol{g}}_{T,\text{wa}}, \boldsymbol{g}_{b,\text{wa}}^*)$ is an SSE.

(c) $e_{T,\text{wa}} < b - \sqrt{\frac{2(1-p)}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

According to (A.51), we get $\alpha_A \ge 2\alpha_B$, $\alpha_A > 2\alpha_{b,p}^{\text{wa}}$. Similar to the proof in **Case I** (iii), we have the fact that the rating profile $(\boldsymbol{g}_A^*, \boldsymbol{g}_B^*) = (\tilde{\boldsymbol{g}}_{T,\text{wa}}, \boldsymbol{g}_{\text{wa}}^{**}(\alpha_A, 1-p))$ is an SSE.

A.12 Proof of Proposition 8

Based on the equilibrium behaviors given in Proposition A.1, we prove the results (1)–(3) in Proposition 8.

(1) Based on (A.45) and (A.46) in Proposition A.1, we know that the equilibrium behaviors are always within $\{\tilde{\boldsymbol{g}}_{T,\text{wa}}, \boldsymbol{g}_{e,\text{wa}}^*, \boldsymbol{g}_{b,\text{wa}}^*, \boldsymbol{g}_{\text{wa}}^{**}(\alpha_A, p), \boldsymbol{g}_{\text{wa}}^{**}(\alpha_A, 1-p)\}$. The corresponding rating deviations of the equilibrium behaviors $\tilde{\boldsymbol{g}}_{T,\text{wa}}$, $\boldsymbol{g}_{e,\text{wa}}^*$ and $\boldsymbol{g}_{b,\text{wa}}^*$ are given by

$$\begin{cases} \|\tilde{\boldsymbol{g}}_{T,\text{wa}} - \tilde{\boldsymbol{g}}_T\|^2 = (e_{T,\text{wa}} - e_T)^2 \Delta \boldsymbol{g}' \Delta \boldsymbol{g}, \\ \|\boldsymbol{g}_{e,\text{wa}}^* - \tilde{\boldsymbol{g}}_T\|^2 = \|(e - e_{T,\text{wa}})\boldsymbol{g}_M + (e_{T,\text{wa}} - e_T)\Delta \boldsymbol{g} + \tilde{\boldsymbol{g}}_T - \tilde{\boldsymbol{g}}_T\|^2 \\ \qquad\qquad = (e - e_T)^2 \boldsymbol{g}_M' \boldsymbol{g}_M + (e_{T,\text{wa}} - e_T)^2 (\Delta \boldsymbol{g} - \boldsymbol{g}_M)'(\Delta \boldsymbol{g} - \boldsymbol{g}_M), \\ \|\boldsymbol{g}_{b,\text{wa}}^* - \tilde{\boldsymbol{g}}_T\|^2 = \|(b - e_{T,\text{wa}})\boldsymbol{g}_M + (e_{T,\text{wa}} - e_T)\Delta \boldsymbol{g} + \tilde{\boldsymbol{g}}_T - \tilde{\boldsymbol{g}}_T\|^2 \\ \qquad\qquad = (b - e_T)^2 \boldsymbol{g}_M' \boldsymbol{g}_M + (e_{T,\text{wa}} - e_T)^2 (\Delta \boldsymbol{g} - \boldsymbol{g}_M)'(\Delta \boldsymbol{g} - \boldsymbol{g}_M). \end{cases} \tag{A.52}$$

Moreover, based on Proposition A.1, we know

$$\frac{\alpha_A}{2} \|\boldsymbol{g}_{\text{wa}}^{**}(\alpha_A, p) - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 = p \quad \text{and} \quad \frac{\alpha_A}{2} \|\boldsymbol{g}_{\text{wa}}^{**}(\alpha_A, 1-p) - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 = 1 - p.$$

Because of the above equality and the assumption $\alpha_A/2 \cdot \|\tilde{\boldsymbol{g}}_{T,\text{wa}} - \tilde{\boldsymbol{g}}_T\|^2 \notin \{p, 1-p\}$, we have

the following results based on the triangle inequality,

$$\begin{cases} \|\boldsymbol{g}^{**}_{\text{wa}}(\alpha_A, p) - \tilde{\boldsymbol{g}}_T\|^2 \geq \left| \|\tilde{\boldsymbol{g}}_{T,\text{wa}} - \tilde{\boldsymbol{g}}_T\|^2 - \|\boldsymbol{g}^{**}_{\text{wa}}(\alpha_A, p) - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 \right| > 0, \\ \|\boldsymbol{g}^{**}_{\text{wa}}(\alpha_A, 1-p) - \tilde{\boldsymbol{g}}_T\|^2 \geq \left| \|\tilde{\boldsymbol{g}}_{T,\text{wa}} - \tilde{\boldsymbol{g}}_T\|^2 - \|\boldsymbol{g}^{**}_{\text{wa}}(\alpha_A, 1-p) - \tilde{\boldsymbol{g}}_{T,\text{wa}}\|^2 \right| > 0. \end{cases} \tag{A.53}$$

Therefore, based on (A.52) and (A.53), rating deviations are all positive because $e_{T,\text{wa}} \neq e_T$.

(2) Based on (A.45) in case (1) of Proposition A.1, we have the fact that there exists a rating deviation from the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\text{wa}}$ if and only if at least one agency issues the ESG rating $\boldsymbol{g}^*_{e,\text{wa}}$ or $\boldsymbol{g}^*_{e,\text{wa}}$, which is equivalent to

$$\max\left\{b, e - \sqrt{\frac{p}{\alpha_j}\left[\frac{1}{(\boldsymbol{\omega}'_M\mathbf{1})^2} - \frac{1}{n}\right]}\right\} \leq e_{T,\text{wa}} < e \quad \text{or} \quad b - \sqrt{\frac{1-p}{\alpha_j}\left[\frac{1}{(\boldsymbol{\omega}'_M\mathbf{1})^2} - \frac{1}{n}\right]} \leq e_{T,\text{wa}} < b.$$

(3) Based on (A.46) in case (2) of Proposition A.1, we know that the equilibrium behavior $\boldsymbol{g}^*_B$ of Agency B deviates from the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\text{wa}}$ if and only if $e_{T,\text{wa}} < e$, and both agencies issue $\tilde{\boldsymbol{g}}_{T,\text{wa}}$ if $e_{T,\text{wa}} \geq e$. Hence, there exists rating deviation from the post-greenwashing ESG fundamental $\tilde{\boldsymbol{g}}_{T,\text{wa}}$ if and only if $e_{T,\text{wa}} < e$.

A.13 Proof of Proposition 9

Based on Proposition A.1, the equilibrium behaviors are always within $\{\tilde{\boldsymbol{g}}_{T,\text{wa}}, \boldsymbol{g}^{**}_{e,\text{wa}}, \boldsymbol{g}^{**}_{b,\text{wa}}\}$ in Case (1), and within $\{\tilde{\boldsymbol{g}}_{T,\text{wa}}, \boldsymbol{g}^{**}_{e,\text{wa}}\}$ under condition (4.10). Therefore, we only prove the results in Case (1) by showing the investors' optimal portfolios and utilities in the following ranges of $e_{T,\text{wa}}$: $e_{T,\text{wa}} \geq e$, $e_{T,\text{wa}} \in [b, e)$ and $e_{T,\text{wa}} \leq b$.

(i) $e_{T,\text{wa}} \geq e$.

Based on (2.4) and (2.5) in Proposition 1, the investors' optimal portfolios are given by

$$\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}}) = \frac{1}{2\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \quad \boldsymbol{\omega}^*(\gamma, b, \tilde{\boldsymbol{g}}_{T,\text{wa}}) = \frac{1}{2\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu};$$

the investors' optimal utilities are given by

$$U(e, \tilde{\boldsymbol{g}}_{T,\text{wa}}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mu}{4\gamma}, \quad U(b, \tilde{\boldsymbol{g}}_{T,\text{wa}}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mu}{4\gamma}.$$

Based on (4.4), the fundamental greenness of the green investor is given by

$$G_F\left(\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}})\right) = e_T, \quad G_F\left(\boldsymbol{\omega}^*(\gamma, b, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}})\right) = e_T.$$

(ii) $b \le e_{T,\mathrm{wa}} < e$.

Based on Proposition 8, we derive the investors' optimal portfolios and utilities by considering the two cases $e_{T,\mathrm{wa}} \ge e - \sqrt{\frac{p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$ and $e_{T,\mathrm{wa}} < e - \sqrt{\frac{p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

(a) $e_{T,\mathrm{wa}} \ge e - \sqrt{\frac{p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

Based on (2.4) and (2.5) in Proposition 1, the investors' optimal portfolios are given by

$$\boldsymbol{\omega}^*(\gamma, e, \boldsymbol{g}^*_{e,\mathrm{wa}}) = \frac{1}{2\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \quad \boldsymbol{\omega}^*(\gamma, b, \boldsymbol{g}^*_{e,\mathrm{wa}}) = \frac{1}{2\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu};$$

the investors' optimal utilities are given by

$$U(e, \boldsymbol{g}^*_{e,\mathrm{wa}}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mu}{4\gamma}, \quad U(b, \boldsymbol{g}^*_{e,\mathrm{wa}}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mu}{4\gamma}.$$

Based on (4.4), the fundamental greenness of the green investor is given by

$$G_F\left(\boldsymbol{\omega}^*(\gamma, e, \boldsymbol{g}^{**}_{e,\mathrm{wa}})\right) = e_T, \quad G_F\left(\boldsymbol{\omega}^*(\gamma, b, \boldsymbol{g}^{**}_{e,\mathrm{wa}})\right) = e_T.$$

(b) $e_{T,\mathrm{wa}} < e - \sqrt{\frac{p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

Based on (2.4) and (2.5) in Proposition 1, the investors' optimal portfolios are given by

$$\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}) = \frac{1}{2\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\mathbf{1}(e - e_{T,\mathrm{wa}}) \cdot \boldsymbol{\Sigma}^{-1}\left(\tilde{\boldsymbol{g}}_T + (e_{T,\mathrm{wa}} - e_T)\boldsymbol{g}_M - e\mathbf{1}\right)}{2\gamma\left(e\mathbf{1} - \tilde{\boldsymbol{g}}_T - (e_{T,\mathrm{wa}} - e_T)\boldsymbol{g}_M\right)'\boldsymbol{\Sigma}^{-1}\left(e\mathbf{1} - \tilde{\boldsymbol{g}}_T - (e_{T,\mathrm{wa}} - e_T)\boldsymbol{g}_M\right)},$$
$$\boldsymbol{\omega}^*(\gamma, b, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}) = \frac{1}{2\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu};$$

the investors' optimal utilities are given by

$$U(e, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mu}{4\gamma} - \frac{(\boldsymbol{\mu}\boldsymbol{\Sigma}^{-1}\mathbf{1})^2(e - e_{T,\mathrm{wa}})^2}{2\gamma\left(e\mathbf{1} - \tilde{\boldsymbol{g}}_T - (e_{T,\mathrm{wa}} - e_T)\boldsymbol{g}_M\right)'\boldsymbol{\Sigma}^{-1}\left(e\mathbf{1} - \tilde{\boldsymbol{g}}_T - (e_{T,\mathrm{wa}} - e_T)\boldsymbol{g}_M\right)},$$
$$U(b, \tilde{\boldsymbol{g}}_{T,\mathrm{wa}}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mu}{4\gamma}.$$

The utility $U(e, \boldsymbol{g}_A)$ is increasing in $e_{T,\text{wa}}$ because the denominator is an quadratic function of $1/(e_T - e_{T,\text{wa}})$.

Based on (4.4), the fundamental greenness of the green investor is given by

$$G_F\left(\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})\right) = \frac{\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})'\tilde{\boldsymbol{g}}_{T,\text{wa}}}{\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})'\mathbf{1}} - \frac{\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})'\boldsymbol{g}_M(e_{T,\text{wa}} - e_T)}{\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})'\mathbf{1}},$$

$$G_F\left(\boldsymbol{\omega}^*(\gamma, b, \tilde{\boldsymbol{g}}_{T,\text{wa}})\right) = e_T.$$

Based on Proposition 7, we know

$$\frac{\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})'\tilde{\boldsymbol{g}}_{T,\text{wa}}}{\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})'\mathbf{1}} = e,$$

which implies

$$\begin{aligned}
&G_F\left(\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})\right) \\
&= e - \frac{\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})'\boldsymbol{g}_M(e_{T,\text{wa}} - e_T)}{\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})'\mathbf{1}} \\
&= e - (e_{T,\text{wa}} - e_T)\frac{\boldsymbol{\omega}_M'\boldsymbol{g}_M + \lambda^*\left[(e_{T,\text{wa}} - e_T)\boldsymbol{g}_M'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M + \boldsymbol{g}_M'\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{g}}_T - e\mathbf{1})\right]}{\boldsymbol{\omega}_M'\mathbf{1} + \lambda^*\left[(e_{T,\text{wa}} - e_T)\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M + \mathbf{1}'\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{g}}_T - e\mathbf{1})\right]} \\
&= e - (e_{T,\text{wa}} - e_T)\frac{\boldsymbol{\omega}_M'\mathbf{1} + \lambda^*\left[(e_{T,\text{wa}} - e_T)\boldsymbol{g}_M'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M + \boldsymbol{g}_M'\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{g}}_T - e\mathbf{1})\right]}{\boldsymbol{\omega}_M'\mathbf{1} + \lambda^*\left[(e_{T,\text{wa}} - e_T)\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M + \mathbf{1}'\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{g}}_T - e\mathbf{1})\right]}.
\end{aligned} \tag{A.54}$$

Based on the results of the utilities, we know that λ^* is increasing of $e_{T,\text{wa}}$, and the fraction

$$\frac{(e_{T,\text{wa}} - e_T)\boldsymbol{g}_M'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M + \boldsymbol{g}_M'\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{g}}_T - e\mathbf{1})}{(e_{T,\text{wa}} - e_T)\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M + \mathbf{1}'\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{g}}_T - e\mathbf{1})}$$

is also increasing of $e_{T,\text{wa}}$ because

$$\boldsymbol{g}_M'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M - \mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M = \frac{n(\boldsymbol{\omega}_M'\mathbf{1})^2}{(n-k^2)^2}\left[\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1} - \frac{n + \boldsymbol{\omega}_M'\mathbf{1}}{\boldsymbol{\omega}_M'\mathbf{1}}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M + \frac{n}{\boldsymbol{\omega}_M'\mathbf{1}}\boldsymbol{g}_M'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M\right] \geq 0.$$

Thus, the fundamental greenness $G_F\left(\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})\right)$ of the green investor is decreasing of $e_{T,\text{wa}}$.

(iii) $e_{T,\text{wa}} < b$.

Based on Proposition 8, we derive the investors' optimal portfolios and utilities by considering

the three cases $e_{T,\text{wa}} \geq b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$, $e_{T,\text{wa}} \in [b - \sqrt{\frac{1-p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}, b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}})$ and $e_{T,\text{wa}} < b - \sqrt{\frac{1-p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

(a) $e_{T,\text{wa}} \geq b - \sqrt{\frac{1-p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

Based on (2.4) and (2.5) in Proposition 1, the investors' optimal portfolios are given by

$$\boldsymbol{\omega}^*(\gamma, e, \boldsymbol{g}^*_{b,\text{wa}}) = \frac{1}{2\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}(e-b)\cdot\boldsymbol{\Sigma}^{-1}\left(\tilde{\boldsymbol{g}}_T + (b-e_T)\boldsymbol{g}_M - e\mathbf{1}\right)}{2\gamma\left(e\mathbf{1} - \tilde{\boldsymbol{g}}_T - (b-e_T)\boldsymbol{g}_M\right)'\boldsymbol{\Sigma}^{-1}\left(e\mathbf{1} - \tilde{\boldsymbol{g}}_T - (b-e_T)\boldsymbol{g}_M\right)},$$
$$\boldsymbol{\omega}^*(\gamma, b, \boldsymbol{g}^*_{b,\text{wa}}) = \frac{1}{2\gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu};$$

the investors' optimal utilities are given by

$$U(e, \boldsymbol{g}^*_{b,\text{wa}}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mu}{4\gamma} - \frac{(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^2(e-b)^2}{2\gamma\left(e\mathbf{1} - \tilde{\boldsymbol{g}}_T - (b-e_T)\boldsymbol{g}_M\right)'\boldsymbol{\Sigma}^{-1}\left(e\mathbf{1} - \tilde{\boldsymbol{g}}_T - (b-e_T)\boldsymbol{g}_M\right)},$$
$$U(b, \boldsymbol{g}^*_{b,\text{wa}}) = \frac{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mu}{4\gamma}.$$

For the fundamental greenness, because the optimal portfolios are both independent of $e_{T,\text{wa}}$, the fundamental greenness of the investors remains constant when $e_{T,\text{wa}}$ varies in the range.

(b) $b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}} \leq e_{T,\text{wa}} < b - \sqrt{\frac{1-p}{\alpha_B \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

Based on (2.4) and (2.5) in Proposition 1, the investors' optimal portfolios the investors' optimal portfolios have the same formulation as the portfolios $\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})$ and $\gamma^{-1}\boldsymbol{\Sigma}^{-1}\mu/2$ in (ii) of case (2), which implies that the utilities and the fundamental greenness of the investors have the same property as the utilities and the fundamental greenness of the investors in (ii) of case (2).

(c) $e_{T,\text{wa}} < b - \sqrt{\frac{1-p}{\alpha_A \boldsymbol{g}_M' \boldsymbol{g}_M}}$.

Based on (2.4) and (2.5) in Proposition 1, the investors' optimal portfolios have the same formulation as the portfolio $\boldsymbol{\omega}^*(\gamma, e, \tilde{\boldsymbol{g}}_{T,\text{wa}})$ in (ii) of case (2), which implies that the utilities and the fundamental greenness of the investors have the same property as the utility and the fundamental greenness of the green investor in (ii) of case (2).

A.14 Additional Analysis of Condition (4.9)

Noting that $\mathbf{1}'\boldsymbol{g}_M = 0$ and $\boldsymbol{\omega}_M'\boldsymbol{g}_M = \boldsymbol{\omega}_M'\mathbf{1}$, we know

$$\mathbf{1}'(\tilde{\boldsymbol{g}}_T - e_T\boldsymbol{g}_M) = \boldsymbol{\omega}_M'(\tilde{\boldsymbol{g}}_T - e_T\boldsymbol{g}_M) = 0.$$

Hence, $\tilde{\boldsymbol{g}}_T - e_T\boldsymbol{g}_M$ has the following representation

$$\tilde{\boldsymbol{g}}_T - e_T\boldsymbol{g}_M = y_T\boldsymbol{\beta}_T,$$

where $y_T \in \mathbb{R}$ and $\boldsymbol{\beta}_T$ satisfies $\mathbf{1}'\boldsymbol{\beta}_T = \boldsymbol{\omega}_M'\boldsymbol{\beta}_T = 0$. Because the condition (4.9) is equivalent to

$$(\tilde{\boldsymbol{g}}_T - e_T\boldsymbol{g}_M)'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M \leq e(\mathbf{1} - \boldsymbol{g}_M)'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M,$$

we have the fact that, for any $\boldsymbol{\beta}_T$ satisfying $\mathbf{1}'\boldsymbol{\beta}_T = \boldsymbol{\omega}_M'\boldsymbol{\beta}_T = 0$ and $\boldsymbol{\beta}_T'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M > 0$, the condition (4.9) holds if and only if

$$y_T \leq \frac{e(\mathbf{1} - \boldsymbol{g}_M)'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M}{\boldsymbol{\beta}_T'\boldsymbol{\Sigma}^{-1}\boldsymbol{g}_M},$$

which is not very restrictive.