A Data-Driven Approach to High-Volume Recruitment: Application to Student Admission

Lilun Du Qing Li

School of Business and Management

Hong Kong University of Science and Technology

Clear Water Bay, Kowloon

Hong Kong

dulilun@ust.hk • imqli@ust.hk

Proof of Theorem 1

Let $|.|_2$ denote the Frobenius norm of a vector or a matrix. The theorem can be proved in the following three steps.

STEP ONE: As $\{x_{i1}, i = 1, \dots, m\}$ are independent and identically distributed with mean μ_{x_1} and finite variance $\sigma^2_{x_1}$, we have $E(\widehat{\mu}_{x_1}) = \mu_{x_1}$ and $\mathrm{var}(\widehat{\mu}_{x_1}) = \sigma^2_{x_1}/m$. Hence, $\widehat{\mu}_{x_1} - \mu_{x_1} = O_p(\frac{1}{\sqrt{m}})$. To prove the convergence of $\widehat{\sigma}^2_{x_1}$, we reformulate it as

$$\widehat{\sigma}^2_{x_1} = \frac{1}{m-1}(x_{11}, \dots, x_{m1})\mathbf{M}(x_{11}, \dots, x_{m1})^\top,$$

where $\mathbf{M} = \mathbf{I}_{m\times m} - \frac{1}{m}\mathbf{1}_{m\times 1}\mathbf{1}^\top_{m\times 1}$ is the idempotent matrix that $\mathbf{M} = \mathbf{M}^2$. Because $\mathbf{M}$ is a symmetric and real matrix, it is orthogonally diagonalizable and has m linearly independent eigenvectors. Its eigenvalues are 0 (with multiplicity 1) and 1 (with multiplicity m-1). Letting $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_{m-1}$ be the eigenvectors with eigenvalue 1, then $z_i = (x_{11}, \dots, x_{m1})\boldsymbol{v}_i$ are independent random variables with mean zero and variance $\sigma^2_{x_1}$ (Horn and Johnson 1990). Therefore,

$$\widehat{\sigma}^2_{x_1} - \sigma^2_{x_1} = \frac{1}{m-1}\sum_{i=1}^{m-1} z_i^2 - \sigma^2_{x_1} = O_p(\frac{1}{\sqrt{m-1}}).$$

The proof of Equation (11) is completed.

STEP TWO: From model (5), standard results from linear regression show that

$$\widehat{\beta}_0 - \beta_0 = O_p(\frac{1}{\sqrt{m_1}}), \widehat{\beta}_1 - \beta_1 = O_p(\frac{1}{\sqrt{m_1}}), \text{ and } \widehat{\sigma}^2_{x_2|x_1} - \sigma^2_{x_2|x_1} = O_p(\frac{1}{\sqrt{m_1}}).$$

As a result, one can obtain

$$\widehat{\beta}_1^2 - \beta_1^2 = (\widehat{\beta}_1 + \beta_1)(\widehat{\beta}_1 - \beta_1) = (\widehat{\beta}_1 - \beta_1)^2 + 2\beta_1(\widehat{\beta}_1 - \beta_1) = O_p(\frac{1}{m_1}) + O_p(\frac{1}{\sqrt{m_1}}) = O_p(\frac{1}{\sqrt{m_1}}).$$

Using the above results and Equation (6), we can show the convergence of $\widehat{\mu}_{x_2}$ as follows:

$$\begin{aligned}
\widehat{\mu}_{x_2} - \mu_{x_2} &= (\widehat{\beta}_0 - \beta_0) + (\widehat{\beta}_1 \widehat{\mu}_{x_1} - \beta_1 \mu_{x_1}) \\
&= (\widehat{\beta}_0 - \beta_0) + (\widehat{\mu}_{x_1} - \mu_{x_1})(\widehat{\beta}_1 - \beta_1) + \mu_{x_1}(\widehat{\beta}_1 - \beta_1) + \beta_1(\widehat{\mu}_{x_1} - \mu_{x_1}) \\
&= O_p(\frac{1}{\sqrt{m_1}}) + O_p(\frac{1}{\sqrt{m m_1}}) + O_p(\frac{1}{\sqrt{m_1}}) + O_p(\frac{1}{\sqrt{m}}) = O_p(\frac{1}{\sqrt{m_1}}).
\end{aligned}$$

Similarly, $\widehat{\sigma}_{x_2}^2 - \sigma_{x_2}^2$ can be bounded by

$$\begin{aligned}
\widehat{\sigma}_{x_2}^2 - \sigma_{x_2}^2 &= (\widehat{\beta}_1^2 - \beta_1^2)(\widehat{\sigma}_{x_1}^2 - \sigma_{x_1}^2) + \sigma_{x_1}^2(\widehat{\beta}_1^2 - \beta_1^2) + \beta_1^2(\widehat{\sigma}_{x_1}^2 - \sigma_{x_1}^2) + (\widehat{\sigma}_{x_2|x_1}^2 - \sigma_{x_2|x_1}^2) \\
&= O_p(\frac{1}{\sqrt{m m_1}}) + O_p(\frac{1}{\sqrt{m_1}}) + O_p(\frac{1}{\sqrt{m}}) + O_p(\frac{1}{\sqrt{m_1}}) = O_p(\frac{1}{\sqrt{m_1}}).
\end{aligned}$$

For the term $\widehat{\sigma}_{x_1,x_2}$, similar arguments lead to

$$\widehat{\sigma}_{x_1,x_2} - \sigma_{x_1,x_2} = (\widehat{\sigma}_{x_1}^2 - \sigma_{x_1}^2)(\widehat{\beta}_1 - \beta_1) + \beta_1(\widehat{\sigma}_{x_1}^2 - \sigma_{x_1}^2) + \sigma_{x_1}^2(\widehat{\beta}_1 - \beta_1) = O_p(\frac{1}{\sqrt{m_1}}).$$

This completes the proof of Equation (12).

STEP THREE: According to model (7), we have

$$\widehat{\gamma}_0 - \gamma_0 = O_p(\frac{1}{\sqrt{m_3}}), |\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}|_2 = O_p(\frac{1}{\sqrt{m_3}}), \text{ and } \widehat{\sigma}_{y|\boldsymbol{x}}^2 - \sigma_{y|\boldsymbol{x}}^2 = O_p(\frac{1}{\sqrt{m_3}}).$$

This implies

$$|\widehat{\boldsymbol{\gamma}}|_2^2 - |\boldsymbol{\gamma}|_2^2 = \left||\widehat{\boldsymbol{\gamma}}|_2 - |\boldsymbol{\gamma}|_2\right|^2 + 2|\boldsymbol{\gamma}|_2 \times \left||\widehat{\boldsymbol{\gamma}}|_2 - |\boldsymbol{\gamma}|_2\right| = O_p(\frac{1}{m_3}) + O_p(\frac{1}{\sqrt{m_3}}) = O_p(\frac{1}{\sqrt{m_3}}).$$

By Equation (10), $\widehat{\mu}_y$ can be decomposed as

$$\begin{aligned}
\widehat{\mu}_y - \mu_y &= (\widehat{\gamma}_0 - \gamma_0) + (\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}} - \boldsymbol{\mu}_{\boldsymbol{x}})^\top(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + (\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}} - \boldsymbol{\mu}_{\boldsymbol{x}})^\top \boldsymbol{\gamma} + \mu_{\boldsymbol{x}}^\top(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \\
&\leq O_p(\frac{1}{\sqrt{m_3}}) + |\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}} - \boldsymbol{\mu}_{\boldsymbol{x}}|_2 \times |\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}|_2 + |\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}} - \boldsymbol{\mu}_{\boldsymbol{x}}|_2 \times |\boldsymbol{\gamma}|_2 + |\boldsymbol{\mu}_{\boldsymbol{x}}|_2 \times |\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}|_2 \\
&= O_p(\frac{1}{\sqrt{m_3}}) + O_p(\frac{1}{\sqrt{m_1}}) \times O_p(\frac{1}{\sqrt{m_3}}) + O_p(\frac{1}{\sqrt{m_1}}) + O_p(\frac{1}{\sqrt{m_3}}) = O_p(\frac{1}{\sqrt{m_3}}).
\end{aligned}$$

Now, we turn to analyze the term $\widehat{\sigma}_y^2$. Derivations yield that

$$\begin{aligned}
\widehat{\sigma}_y^2 - \sigma_y^2 &= (\widehat{\sigma}_{y|\boldsymbol{x}}^2 - \sigma_{y|\boldsymbol{x}}^2) + \widehat{\boldsymbol{\gamma}}^\top \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\gamma} \\
&= O_p(\frac{1}{\sqrt{m_3}}) + \widehat{\boldsymbol{\gamma}}^\top(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} - \boldsymbol{\Sigma}_{\boldsymbol{x}})\widehat{\boldsymbol{\gamma}} + \widehat{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{x}} \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\gamma}
\end{aligned}$$

$$
\begin{aligned}
&\leq O_p(\frac{1}{\sqrt{m_3}}) + O_p(\frac{1}{\sqrt{m_1}}) + \lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{x}})\Big||\widehat{\boldsymbol{\gamma}}|_2^2 - |\boldsymbol{\gamma}|_2^2\Big| \\
&= O_p(\frac{1}{\sqrt{m_3}}) + O_p(\frac{1}{\sqrt{m_1}}) + O_p(\frac{1}{\sqrt{m_3}}) = O_p(\frac{1}{\sqrt{m_3}}),
\end{aligned}
$$

where $\lambda_{\max}(\boldsymbol{\Sigma}_{\boldsymbol{x}})$ is the largest eigenvalue of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$. Finally, we show that $\widehat{\sigma}_{y,x_1}$ and $\widehat{\sigma}_{y,x_2}$ converge to their respective true values in probability. Let $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ be the estimators of γ_1 and γ_2, respectively. Thus, $\widehat{\sigma}_{y,x_1} = \widehat{\sigma}_{x_1}^2\widehat{\gamma}_1 + \widehat{\sigma}_{x_1,x_2}\widehat{\gamma}_2$ converges to σ_{y,x_1} at the rate of $O_p(\frac{1}{\sqrt{m_3}})$, that is $|\widehat{\sigma}_{y,x_1} - \sigma_{y,x_1}| = O_p(\frac{1}{\sqrt{m_3}})$. Analogously, $|\widehat{\sigma}_{y,x_2} - \sigma_{y,x_2}| = O_p(\frac{1}{\sqrt{m_3}})$. Equation (13) hence follows.

Proof of Theorem 2 For ease of presentation, denote $a_m = O(b_m)$ if there exists a finite constant C such that $|a_m| \leq C|b_m|$, for all m. From Theorem 1, we have

$$
|\widehat{\boldsymbol{\mu}}_{\boldsymbol{x},y} - \boldsymbol{\mu}_{\boldsymbol{x},y}|_2 = O_p(\frac{1}{\sqrt{m_3}}), \text{ and } |\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x},y} - \boldsymbol{\Sigma}_{\boldsymbol{x},y}|_2 = O_p(\frac{1}{\sqrt{m_3}}), \tag{A.1}
$$

where $\widehat{\boldsymbol{\mu}}_{\boldsymbol{x},y}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x},y}$ are the mean and variance estimators obtained via the regression-based moment estimation approach. To facilitate the derivation, we transform the vector $(x_1, x_2, y)^\top$ to $\boldsymbol{\omega} = (x_1, x_1 + x_2, y)$ using the transformation matrix $\mathbf{M}_1 = \{x_{ij}\}_{3\times 3}$, where $x_{ij} = 1$ for $(i,j) = \{(1,1),(2,2),(3,3),(2,1)\}$ and $x_{ij} = 0$ otherwise. As a result, $\boldsymbol{\omega}$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{\omega}} = \mathbf{M}_1\boldsymbol{\mu}_{\boldsymbol{x},y}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\omega}} = \mathbf{M}_1\boldsymbol{\Sigma}_{\boldsymbol{x},y}\mathbf{M}_1^\top$. Denote by $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\omega}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\omega}}$ their corresponding estimators. By (A.1), we can have $|\widehat{\boldsymbol{\mu}}_{\boldsymbol{\omega}} - \boldsymbol{\mu}_{\boldsymbol{\omega}}|_2 = O_p(\frac{1}{\sqrt{m_3}})$ and $|\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\omega}} - \boldsymbol{\Sigma}_{\boldsymbol{\omega}}|_2 = O_p(\frac{1}{\sqrt{m_3}})$.

To show the consistency of $\widehat{\Pr}(y > \widehat{c}_3 \mid \{x_1 < \widehat{c}_1\} \cup \{x_1 > \widehat{c}_1, x_1 + x_2 < \widehat{c}_2\})$, we first use the multivariate normal distribution to derive a specific form for $\Pr(y > c_3 \mid \{x_1 < c_1\} \cup \{x_1 > c_1, x_1 + x_2 < c_2\})$, that is

$$
\begin{aligned}
&\Pr(y > c_3|\{x_1 < c_1\} \cup \{x_1 > c_1, x_1 + x_2 < c_2\}) \\
=\ & \frac{\Pr(y > c_3) - \Pr(y > c_3, x_1 + x_2 > c_2, x_1 > c_1)}{1 - \Pr(x_1 + x_2 > c_2, x_1 > c_1)} \\
=\ & \frac{\Pr(\omega_3 > c_3) - \Pr(\omega_3 > c_3, \omega_2 > c_2, \omega_1 > c_1)}{1 - \Pr(\omega_2 > c_2, \omega_1 > c_1)} \\
=\ & \frac{\int_{c_3}^{\infty} \phi_1(\omega_3; \mu_3, \sigma_3)d\omega_3 - \int_{c_3}^{\infty}\int_{c_2}^{\infty}\int_{c_1}^{\infty} \phi_{123}(\boldsymbol{\omega}; \boldsymbol{\mu}_{\boldsymbol{\omega}}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})d\boldsymbol{\omega}}{1 - \int_{c_2}^{\infty}\int_{c_1}^{\infty} \phi_{12}(\omega_1, \omega_2; \boldsymbol{\mu}_{\boldsymbol{\omega}}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})d\omega_1 d\omega_2},
\end{aligned} \tag{A.2}
$$

where ϕ_1, ϕ_{12}, and ϕ_{123} are the density functions of ω_3, $(\omega_1, \omega_2)^\top$, and $\boldsymbol{\omega}$, respectively. From (A.2), $\widehat{\Pr}(y > \widehat{c}_3 \mid \{x_1 < \widehat{c}_1\} \cup \{x_1 > \widehat{c}_1, x_1 + x_2 < \widehat{c}_2\})$ can be similarly formulated, with $\boldsymbol{\mu}_{\boldsymbol{\omega}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\omega}}$, and $(c_1, c_2, c_3)^\top$ replaced by $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\omega}}$, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\omega}}$, and $(\widehat{c}_1, \widehat{c}_2, \widehat{c}_3)^\top$, respectively. As the density functions of normal random variables are exponentially smooth, it immediately follows from the continuous mapping theorem (see Theorem 2.3 in Van der Vaart 1998) that

$$\left|\widehat{\Pr}(y > \widehat{c}_3 \mid \{x_1 < \widehat{c}_1\} \cup \{x_1 > \widehat{c}_1, x_1 + x_2 < \widehat{c}_2\}) - \Pr(y > c_3 | \{x_1 < c_1\} \cup \{x_1 > c_1, x_1 + x_2 < c_2\})\right|$$

$$\leq O(|\widehat{\boldsymbol{\mu}}_{\boldsymbol{\omega}} - \boldsymbol{\mu}_{\boldsymbol{\omega}}|_2 + |\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\omega}} - \boldsymbol{\Sigma}_{\boldsymbol{\omega}}|_2 + |\widehat{c}_1 - c_1| + |\widehat{c}_2 - c_2| + |\widehat{c}_3 - c_3|)$$
$$= O_p(\frac{1}{\sqrt{m_3}}) + O(|\widehat{c}_1 - c_1| + |\widehat{c}_2 - c_2| + |\widehat{c}_3 - c_3|).$$

Now we derive the convergence rate of $\widehat{c}_1$, $\widehat{c}_2$, and $\widehat{c}_3$. From Equations (17) and (18), $\widehat{c}_i - c_i = O(|\widehat{\boldsymbol{\mu}}_{\boldsymbol{\omega}} - \boldsymbol{\mu}_{\boldsymbol{\omega}}|_2 + |\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\omega}} - \boldsymbol{\Sigma}_{\boldsymbol{\omega}}|_2) = O_p(\frac{1}{\sqrt{m_3}})$, for $i = 1, 2$. Similarly, from Equation (19), we have

$$|\widehat{c}_3 - c_3| = O(|\widehat{\boldsymbol{\mu}}_{\boldsymbol{\omega}} - \boldsymbol{\mu}_{\boldsymbol{\omega}}|_2 + |\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\omega}} - \boldsymbol{\Sigma}_{\boldsymbol{\omega}}|_2) + O(|\widehat{\varphi}_0 - \varphi_0| + |\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}|_2).$$

Under the logistic regression model (20), one can show that $O(|\widehat{\varphi}_0 - \varphi_0| + |\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}|_2) = O_p(\frac{1}{\sqrt{m_2}})$. This completes the proof of Theorem 2.

Proof of Theorem 3 Throughout the proofs, we use the generic notation $f_X(\cdot)$ to denote the density function of any generic random variable X. We shall prove that under the random yield assumption, c_3 as a function of $\boldsymbol{\theta}$ achieves its maximum when $\boldsymbol{\theta}$ is proportional to $\boldsymbol{\gamma}$. For ease of notation, we reformulate the models (5) and (7) as

$$x_2 = \beta_0 + \beta_1 x_1 + \widetilde{x}_2, y = (\gamma_0 + \gamma_2\beta_0) + (\gamma_1 + \gamma_2\beta_1)x_1 + \gamma_2\widetilde{x}_2 + \widetilde{y}, \tag{A.3}$$

where x_1, $\widetilde{x}_2$, and $\widetilde{y}$ are independent. Let $\gamma_0^* = \gamma_0 + \gamma_2\beta_0$, $\gamma_1^* = \gamma_1 + \gamma_2\beta_1$.

Under the random yield assumption, given q_1, q_2, and q_3, c_3 as a function of $\boldsymbol{\theta}$ can be sequentially solved from (17) to (19)

$$q_1 = \Pr(x_1 > c_1), q_2 q_1 = \Pr(\boldsymbol{x}^\top\boldsymbol{\theta} > c_2, x_1 > c_1), q_3 q_2 q_1 = \Pr(y > c_3, \mathbf{x}^\top\boldsymbol{\theta} > c_2, x_1 > c_1).$$

Without loss of generality, we assume $\theta_2 > 0$. Taking derivatives with respect to $\boldsymbol{\theta}$ for the above second equation leads to

$$\int_{c_1}^{\infty} (\frac{\partial c_2}{\partial \theta_1} - x_1) f_{\widetilde{x}_2}(l(x_1)) f_{x_1}(x_1) dx_1 = 0, \int_{c_1}^{\infty} (\frac{\partial c_2}{\partial \theta_2}\theta_2 - c_2 + \theta_1 x_1) f_{\widetilde{x}_2}(l(x_1)) f_{x_1}(x_1) dx_1 = 0,$$

where $l(x_1) = \frac{c_2 - \theta_0^* - \theta_1^* x_1}{\theta_2}$ with $\theta_0^* = \beta_0\theta_2$ and $\theta_1^* = \theta_1 + \beta_1\theta_2$. Accordingly, c_2 as a function of $\boldsymbol{\theta}$ can be solved via

$$\frac{\partial c_2}{\partial \theta_1} = \frac{\int_{c_1}^{\infty} x_1 f_{\widetilde{x}_2}(l(x_1)) f_{x_1}(x_1) dx_1}{\int_{c_1}^{\infty} f_{\widetilde{x}_2}(l(x_1)) f_{x_1}(x_1) dx_1}, \frac{\partial c_2}{\partial \theta_2}\theta_2 - c_2 = -\theta_1 \times \frac{\int_{c_1}^{\infty} x_1 f_{\widetilde{x}_2}(l(x_1)) f_{x_1}(x_1) dx_1}{\int_{c_1}^{\infty} f_{\widetilde{x}_2}(l(x_1)) f_{x_1}(x_1) dx_1}.$$

Similarly, c_3 as a function of $\boldsymbol{\theta}$ can be expressed analogously as

$$\begin{aligned}
&-\frac{\partial c_3}{\partial \theta_1}\int_{c_1}^{\infty}\int_{l(x_1)}^{\infty} f_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*x_1-\gamma_2\widetilde{x}_2)f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)d\widetilde{x}_2dx_1\\
=&\int_{c_1}^{\infty}(\frac{\partial c_2}{\partial \theta_1}-x_1)/\theta_2\{1-F_{\widetilde{y}}(k(\theta_1,\theta_2))\}f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)dx_1 \text{ and}
\end{aligned}$$

$$\begin{aligned}
&-\frac{\partial c_3}{\partial \theta_2}\int_{c_1}^{\infty}\int_{l(x_1)}^{\infty} f_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*x_1-\gamma_2\widetilde{x}_2)f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)d\widetilde{x}_2dx_1\\
=&\int_{c_1}^{\infty}(\frac{\partial c_2}{\partial \theta_2}\theta_2-c_2+\theta_1x_1)/\theta_2^2\{1-F_{\widetilde{y}}(k(\theta_1,\theta_2))\}f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)dx_1,
\end{aligned}$$

where $k(\theta_1,\theta_2)=c_3-\gamma_0^*-\gamma_1^*x_1-\gamma_2\frac{c_2-\theta_0^*-\theta_1^*x_1}{\theta_2}$ and $F_{\widetilde{y}}(\cdot)$ is the cumulative distribution function of $\widetilde{y}$ in (A.3). Consequently, $\frac{\partial c_3}{\partial \theta_1}$ is proportional to (up to a negative constant)

$$\begin{aligned}
&\int_{c_1}^{\infty} x_1 f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)dx_1\int_{c_1}^{\infty}\{1-F_{\widetilde{y}}(k(\theta_1,\theta_2))\}f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)dx_1\\
&-\int_{c_1}^{\infty} f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)dx_1\int_{c_1}^{\infty}x_1\{1-F_{\widetilde{y}}(k(\theta_1,\theta_2))\}f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)dx_1. \qquad \text{(A.4)}
\end{aligned}$$

Let $\widetilde{f}_{x_1}(x_1)\equiv f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)/\int_{c_1}^{\infty}f_{\widetilde{x}_2}(l(x_1))f_{x_1}(x_1)dx_1$ be a density function. Then, (A.4) can be written in terms of expectation as

$$E_{\widetilde{f}_{x_1}}[x_1]E_{\widetilde{f}_{x_1}}[\{1-F_{\widetilde{y}}(k(\theta_1,\theta_2))\}]-E_{\widetilde{f}_{x_1}}[x_1\{1-F_{\widetilde{y}}(k(\theta_1,\theta_2))\}]. \qquad \text{(A.5)}$$

Note that $k(\theta_1,\theta_2)$ is independent of x_1 if $\gamma_2\frac{\theta_1}{\theta_2}-\gamma_1=0$. The condition $\gamma_2\frac{\theta_1}{\theta_2}-\gamma_1=0$ implies that $\boldsymbol{\theta}$ is proportional to $\boldsymbol{\gamma}$, denoted by $\boldsymbol{\theta}\propto\boldsymbol{\gamma}$. Thus, when $\boldsymbol{\theta}\propto\boldsymbol{\gamma}$, (A.5) $=0$ and $\frac{\partial c_3}{\partial\theta_1}=0$. The same conclusion holds for $\frac{\partial c_3}{\partial\theta_2}$, that is,

$$\frac{\partial c_3}{\partial \theta_2}\propto\theta_1\times\Big\{E_{\widetilde{f}_{x_1}}[x_1]E_{\widetilde{f}_{x_1}}[\{1-F_{\widetilde{y}}(k(\theta_1,\theta_2))\}]-E_{\widetilde{f}_{x_1}}[x_1\{1-F_{\widetilde{y}}(k(\theta_1,\theta_2))\}]\Big\},$$

which equals zero when $\boldsymbol{\theta}\propto\boldsymbol{\gamma}$.

For any given $\theta_2>0$, it holds that $k(\theta_1,\theta_2)$ is an increasing function of x_1 when $\gamma_2\frac{\theta_1}{\theta_2}>\gamma_1$. Hence, by Chebyshev's sum inequality, (A.5) >0. On the other hand, (A.5) <0 when $\gamma_2\frac{\theta_1}{\theta_2}<\gamma_1$ for any given $\theta_2>0$. By condition $\gamma_2/\theta_2>0$, we have $\gamma_2>0$ when $\theta_2>0$. Thus, it holds that for any given $\theta_2>0$, $c_3(\theta_1,\theta_2)$ achieves its maximum when $\theta_1=\gamma_1\theta_2/\gamma_2$. Now consider the function $c_3(\gamma_1\theta_2/\gamma_2,\theta_2)$. Taking derivative with respect to θ_2 yields that

$$\frac{dc_3(\gamma_1\theta_2/\gamma_2,\theta_2)}{d\theta_2}=\frac{\gamma_1}{\gamma_2}\times\frac{\partial c_3(\theta_1,\theta_2)}{\partial\theta_1}\bigg|_{\theta_1=\gamma_1\theta_2/\gamma_2}+\frac{\partial c_3(\gamma_1\theta_2/\gamma_2,\theta_2)}{\partial\theta_2}=0.$$

As a result, $c_3(\boldsymbol{\theta})$ achieves its maximum when $\boldsymbol{\theta}\propto\boldsymbol{\gamma}$.

Now we verify the error function achieves its minimum when $\boldsymbol{\theta} \propto \boldsymbol{\gamma}$. Under the random yield assumption, the error rate can be simplified as

$$\Pr(y > c_3 \mid \{x_1 < c_1\} \cup \{x_1 > c_1, \boldsymbol{x}^\top \boldsymbol{\theta} < c_2\}) = \frac{\Pr(y > c_3)}{1 - q_1 q_2} - \frac{q_1 q_2 q_3}{1 - q_1 q_2},$$

which is an decreasing function c_3. This together with the result for $c_3(\boldsymbol{\theta})$ yields that the error rate achieves its minimum when $\boldsymbol{\theta} \propto \boldsymbol{\gamma}$.

Proof of Theorem 4 To simplify the presentation, we first prove the result when $\boldsymbol{\theta} = \mathbf{1}$. Recall that the error rate is defined as $\Pr(y > c_3 \mid \{x_1 < c_1\} \cup \{x_1 > c_1, x_1 + x_2 < c_2\})$. Because the intake target Q is given, a fixed percentage of shortlisted candidates must be accepted. This amounts to fixing $q_1 q_2$ when solving c_2 from $\Pr(x_1 > c_1, x_1 + x_2 > c_2) = q_1 q_2$. From the logistic regression model (20), c_3 can be solved directly from $\Pr(y > c_3 \mid \delta = 1, x_1 + x_2 > c_2, x_1 > c_1) = q_3$, given c_1 and c_2.

To show that $\Pr(y > c_3 \mid \{x_1 < c_1\} \cup \{x_1 > c_1, x_1 + x_2 < c_2\})$ is an increasing function with respect to c_1, we first verify that $\frac{\partial c_3}{\partial c_1} < 0$.

According to the condition $\Pr(x_1 + x_2 > c_2, x_1 > c_1) = q_1 q_2$, it holds that

$$\int_{c_1}^{\infty} \int_{c_2 - \beta_0 - (1+\beta_1)x_1}^{\infty} f_{\widetilde{x}_2}(\widetilde{x}_2) f_{x_1}(x_1) d\widetilde{x}_2 dx_1 = q_1 q_2,$$

where $\widetilde{x}_2$ is defined in (A.3). By taking the first-order derivative of the above equation with respect to c_1, we obtain

$$\frac{\partial c_2}{\partial c_1} = -\frac{\int_{c_2 - \beta_0 - (1+\beta_1)c_1}^{\infty} f_{\widetilde{x}_2}(\widetilde{x}_2) d\widetilde{x}_2 \times f_{x_1}(c_1)}{\int_{c_1}^{\infty} f_{\widetilde{x}_2}(c_2 - \beta_0 - (1+\beta_1)x_1) f_{x_1}(x_1) dx_1} < 0. \tag{A.6}$$

(A.6) implies that c_2 decreases when c_1 increases. Moreover, (A.6) is equivalent to

$$\int_{-\infty}^{c_1} f_{\widetilde{x}_2}(c_2 - \beta_0 - (1+\beta_1)x_1)(1+\beta_1) dx_1 \times f_{x_1}(c_1) = -\frac{\partial c_2}{\partial c_1} \int_{c_1}^{\infty} f_{\widetilde{x}_2}(c_2 - \beta_0 - (1+\beta_1)x_1) f_{x_1}(x_1) dx_1,$$

which equals

$$\begin{aligned} & \int_0^{\infty} f_{\widetilde{x}_2}(c_2 - \beta_0 - (1+\beta_1)(-x_1 + c_1))(1+\beta_1) dx_1 \times f_{x_1}(c_1) \\ = \; & -\frac{\partial c_2}{\partial c_1} \int_0^{\infty} f_{\widetilde{x}_2}(c_2 - \beta_0 - (1+\beta_1)(x_1 + c_1)) f_{x_1}(x_1 + c_1) dx_1, \end{aligned} \tag{A.7}$$

by changing the lower and upper limits of the integral.

For notational simplicity, model (20) is reparameterized by

$$\Pr(\delta = 1 \mid x_1, x_2) = \frac{1}{1 + \exp\{-\varphi_0^* - \varphi_1^* x_1 - \varphi_2 \widetilde{x}_2\}}$$

with $\varphi_0^* = \varphi_0 + \varphi_2\beta_0$ and $\varphi_1^* = \varphi_1 + \varphi_2\beta_1$. Then the condition that $\Pr(y > c_3 \mid \delta = 1, x_1 + x_2 > c_2, x_1 > c_1) = q_3$ can be expressed as

$$\int_{c_1}^{\infty}\int_{c_2-\beta_0-(1+\beta_1)x_1}^{\infty}[1-F_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*x_1-\gamma_2\widetilde{x}_2)]\frac{1}{1+\exp\{-\varphi_0^*-\varphi_1^*x_1-\varphi_2\widetilde{x}_2\}}f_{\widetilde{x}_2}(\widetilde{x}_2)f_{x_1}(x_1)d\widetilde{x}_2dx_1$$

$$= q_3\int_{c_1}^{\infty}\int_{c_2-\beta_0-(1+\beta_1)x_1}^{\infty}\frac{1}{1+\exp\{-\varphi_0^*-\varphi_1^*x_1-\varphi_2\widetilde{x}_2\}}f_{\widetilde{x}_2}(\widetilde{x}_2)f_{x_1}(x_1)d\widetilde{x}_2dx_1, \qquad \text{(A.8)}$$

where $\widetilde{y}$ is defined in (A.3). Taking derivatives from both sides with respect to c_1 results in

$$\begin{aligned}
&-\frac{\partial c_3}{\partial c_1}\int_{c_1}^{\infty}\int_{c_2-\beta_0-(1+\beta_1)x_1}^{\infty}f_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*x_1-\gamma_2\widetilde{x}_2)\frac{1}{1+\exp\{-\varphi_0^*-\varphi_1^*x_1-\varphi_2\widetilde{x}_2\}}f_{\widetilde{x}_2}(\widetilde{x}_2)f_{x_1}(x_1)d\widetilde{x}_2dx_1\\
&=\int_{c_2-\beta_0-(1+\beta_1)c_1}^{\infty}[1-q_3-F_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*c_1-\gamma_2\widetilde{x}_2)]\frac{1}{1+\exp\{-\varphi_0^*-\varphi_1^*c_1-\varphi_2\widetilde{x}_2\}}f_{\widetilde{x}_2}(\widetilde{x}_2)d\widetilde{x}_2\times f_{x_1}(c_1)\\
&\quad+\frac{\partial c_2}{\partial c_1}\int_{c_1}^{\infty}[1-q_3-F_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*x_1-\gamma_2\{c_2-\beta_0-(1+\beta_1)x_1\})]\\
&\quad\times\frac{1}{1+\exp\{-\varphi_0^*-\varphi_1^*x_1-\varphi_2(c_2-\beta_0-(1+\beta_1)x_1)\}}f_{\widetilde{x}_2}(c_2-\beta_0-(1+\beta_1)x_1)f_{x_1}(x_1)dx_1.
\end{aligned}$$

The right hand side can be derived as

$$\begin{aligned}
&\int_{-\infty}^{c_1}[1-q_3-F_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*c_1-\gamma_2\{c_2-\beta_0-(1+\beta_1)x_1\})]\\
&\frac{1}{1+\exp\{-\varphi_0^*-\varphi_1^*c_1-\varphi_2\{c_2-\beta_0-(1+\beta_1)x_1\}\}}f_{\widetilde{x}_2}(c_2-\beta_0-(1+\beta_1)x_1)(1+\beta_1)dx_1\times f_{x_1}(c_1)\\
&\quad+\frac{\partial c_2}{\partial c_1}\int_{c_1}^{\infty}[1-q_3-F_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*x_1-\gamma_2\{c_2-\beta_0-(1+\beta_1)x_1\})]\\
&\quad\times\frac{1}{1+\exp\{-\varphi_0^*-\varphi_1^*x_1-\varphi_2(c_2-\beta_0-(1+\beta_1)x_1)\}}f_{\widetilde{x}_2}(c_2-\beta_0-(1+\beta_1)x_1)f_{x_1}(x_1)dx_1\\
&=\int_0^{\infty}A(x_1)f_{\widetilde{x}_2}(c_2-\beta_0-(1+\beta_1)(-x_1+c_1))(1+\beta_1)dx_1\times f_{x_1}(c_1)\\
&\quad+\frac{\partial c_2}{\partial c_1}\int_0^{\infty}B(x_1)f_{\widetilde{x}_2}(c_2-\beta_0-(1+\beta_1)(x_1+c_1))f_{x_1}(x_1+c_1)dx_1, \qquad \text{(A.9)}
\end{aligned}$$

where

$$\begin{aligned}
A(x_1) &= [1-q_3-F_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*c_1-\gamma_2\{c_2-\beta_0-(1+\beta_1)(-x_1+c_1)\})]\\
&\quad\times\frac{1}{1+\exp\{-\varphi_0^*-\varphi_1^*c_1-\varphi_2\{c_2-\beta_0-(1+\beta_1)(-x_1+c_1)\}\}},\\
B(x_1) &= [1-q_3-F_{\widetilde{y}}(c_3-\gamma_0^*-\gamma_1^*(x_1+c_1)-\gamma_2\{c_2-\beta_0-(1+\beta_1)(x_1+c_1)\})]\\
&\quad\times\frac{1}{1+\exp\{-\varphi_0^*-\varphi_1^*(x_1+c_1)-\varphi_2\{c_2-\beta_0-(1+\beta_1)(x_1+c_1)\}\}}.
\end{aligned}$$

Note that $1-q_3-F_{\widetilde{y}}(\cdot) > 0$ by the condition $\Pr(y > c_3 \mid \delta = 1, x_1 > c_1, x_1 + x_2 > c_2) = q_3$. Thus, $A(x) > 0$ and $B(x) > 0$. In addition, both $1-q_3-F_{\widetilde{y}}(\cdot)$ and $\frac{1}{1+\exp(\cdot)}$ are decreasing functions. We then verify that

$$c_3-\gamma_0^*-\gamma_1^*c_1-\gamma_2\{c_2-\beta_0-(1+\beta_1)(-x_1+c_1)\} \le c_3-\gamma_0^*-\gamma_1^*(x_1+c_1)-\gamma_2\{c_2-\beta_0-(1+\beta_1)(x_1+c_1)\},$$

if and only if $\gamma_1 \le \gamma_2(2+\beta_1)$. Similarly,

$$-\varphi_0^*-\varphi_1^*c_1-\varphi_2\{c_2-\beta_0-(1+\beta_1)(-x_1+c_1)\} \le -\varphi_0^*-\varphi_1^*(x_1+c_1)-\varphi_2\{c_2-\beta_0-(1+\beta_1)(x_1+c_1)\},$$

if and only if $\varphi_1 \le \varphi_2(2+\beta_1)$. Thus, $A(x) \ge B(x)$. This together with (A.7) yields that (A.9) is larger than zero. As a result, we have $\frac{\partial c_3}{\partial c_1} < 0$.

Now we prove that the error rate function is increasing with respect to the cut-off c_1. It is straightforward to verify that the error function is proportional to $\Pr(y > c_3) - \Pr(y > c_3, x_1 + x_2 > c_2, x_1 > c_1)$. Similar to (A.8), it follows that

$$\begin{aligned}
&\Pr(y > c_3) - \Pr(y > c_3, x_1 + x_2 > c_2, x_1 > c_1) \\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[1 - F_{\widetilde{y}}(c_3 - \gamma_0^* - \gamma_1^* x_1 - \gamma_2\widetilde{x}_2)]f_{\widetilde{x}_2}(\widetilde{x}_2)f_{x_1}(x_1)d\widetilde{x}_2dx_1 \\
&-\int_{c_1}^{\infty}\int_{c_2-\beta_0-(1+\beta_1)x_1}^{\infty}[1 - F_{\widetilde{y}}(c_3 - \gamma_0^* - \gamma_1^* x_1 - \gamma_2\widetilde{x}_2)]f_{\widetilde{x}_2}(\widetilde{x}_2)f_{x_1}(x_1)d\widetilde{x}_2dx_1.
\end{aligned}$$

Taking derivatives with respect to c_1, we have

$$\begin{aligned}
&-\frac{\partial c_3}{\partial c_1}\int\int_{\{x_1<c_1\}\cup\{\widetilde{x}_2<c_2-\beta_0-(1+\beta_1)x_1\}} f_{\widetilde{y}}(c_3 - \gamma_0^* - \gamma_1^* x_1 - \gamma_2\widetilde{x}_2)f_{\widetilde{x}_2}(\widetilde{x}_2)f_{x_1}(x_1)d\widetilde{x}_2dx_1 \\
&+\Big\{\int_{c_2-\beta_0-(1+\beta_1)c_1}^{\infty}[1 - F_{\widetilde{y}}(c_3 - \gamma_0^* - \gamma_1^* c_1 - \gamma_2\widetilde{x}_2)]f_{\widetilde{x}_2}(\widetilde{x}_2)d\widetilde{x}_2 \times f_{x_1}(c_1) \\
&+\frac{\partial c_2}{\partial c_1}\int_{c_1}^{\infty}[1 - F_{\widetilde{y}}(c_3 - \gamma_0^* - \gamma_1^* x_1 - \gamma_2\{c_2 - \beta_0 - (1+\beta_1)x_1\})]f_{\widetilde{x}_2}(c_2 - \beta_0 - (1+\beta_1)x_1)f_{x_1}(x_1)dx_1\Big\}.
\end{aligned}$$

By the result of (A.9), the second term is non-negative. This together with the result that $\frac{\partial c_3}{\partial c_1} < 0$ yields that the error function is increasing with respect to c_1, thus decreasing for q_1.

For the general case $\{\boldsymbol{x}^\top\boldsymbol{\theta} > c_2, x_1 > c_1\}$, we can similarly show that the error rate $\Pr(y > c_3 \mid \{x_1 < c_1\} \cup \{x_1 > c_1, \boldsymbol{x}^\top\boldsymbol{\theta} < c_2\})$ is increasing in c_1 under conditions: $\gamma_1 \le \gamma_2(2\theta_1/\theta_2 + \beta_1)$ and $\varphi_1 \le \varphi_2(2\theta_1/\theta_2 + \beta_1)$.