SUPPLEMENT: Simulation Results and Proofs

This supplement accompanies the main document,

"Selection-Bias-Adjusted Inference for the Bivariate Normal Distribution under Soft-Threshold Sampling"

Joseph B. Lang

Dept of Statistics and Actuarial Science, Univ. of Iowa, Jan. 22, 2025

In this supplement, Section 1 reports simulation results and Section 2 gives proofs of the main document's Results 1-10.

1. SIMULATION RESULTS: FIGURES 2-9 AND TABLE S.1

To assess the reasonableness of the estimation and prediction approaches of this paper, we carried out a small-scale simulation based on the course score data in Table 1 of the main document. Specifically, we assume that N students initially enroll and the data $(\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{t})$ are realizations of $(\underline{\mathbb{X}}_1, \underline{\mathbb{X}}_2, \underline{\mathbb{X}}_3, \underline{\mathbb{T}})$, which has IID components each with the same multivariate Normal distribution. The observed data $(\underline{x}_1[\underline{s}], \underline{x}_2[\underline{s}], \underline{x}_3[\underline{s}])$ are determined by $\underline{s}$, a realization of $\underline{\mathbb{S}}$, where $(\underline{\mathbb{S}} = \underline{s}) = (\underline{\mathbb{T}}[\underline{s}] > \theta_0,\ \underline{\mathbb{T}}[\underline{s}^c] \leqslant \theta_0)$. In the simulation, we varied values such as N, π, ρ_{12}, and ψ, and set other parameters to values similar to those in the last three rows of Table 1; for example, we set $\mu_1 = 75$, $\mu_2 = 71$, $\sigma_1 = 9$, and $\sigma_2 = 11$.

The first simulation uses $N = 520, \pi = 0.8, \rho = 0.6,\ \psi = 0.8$, and $x_1 = 46$. The results are displayed in Figure 1 of the main paper. The next Figures 2-9, below, graphically display simulation results for a variety of other settings. See the description of Figure 1 in the main paper for interpretations of these graphs. Finally, the simulation results are displayed in tabular form in Table S.1, and summarized.

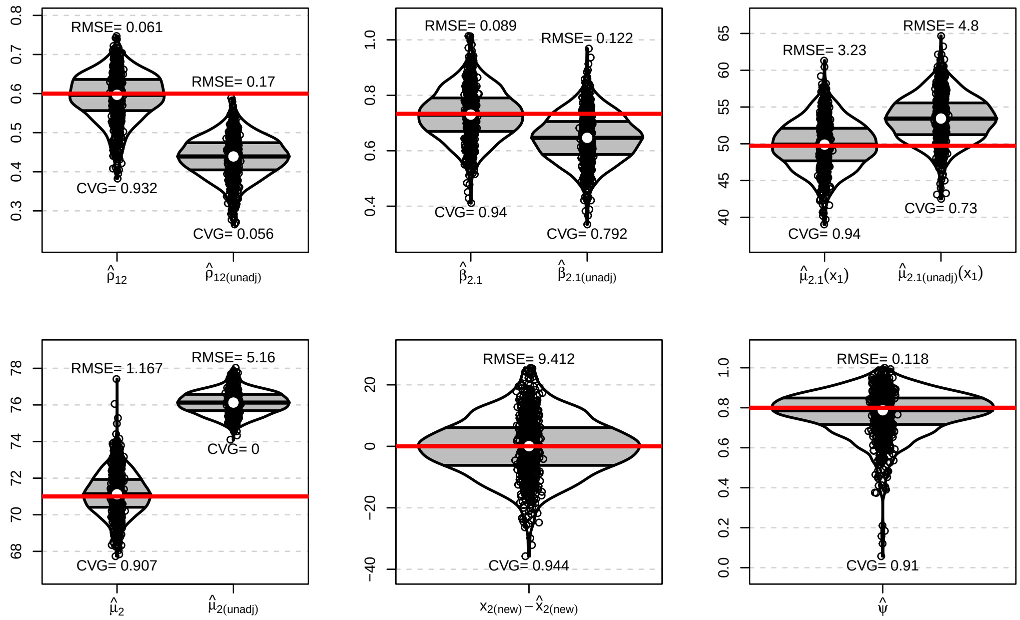

Figure 2. Simulation Results ($N = 520, \pi = 0.5, \rho_{12} = 0.6, \psi = 0.8, x_1 = 46, n_{sim} = 1000$).

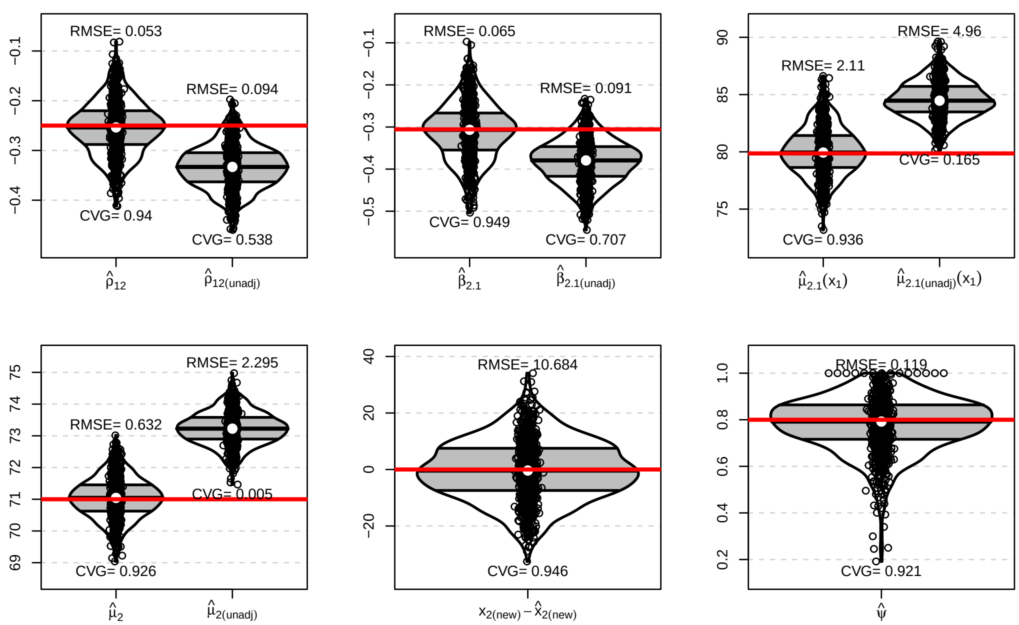

Figure 3. Simulation Results ($N = 520, \pi = 0.8, \rho_{12} = -0.25, \psi = 0.8, x_1 = 46, n_{sim} = 1000$).

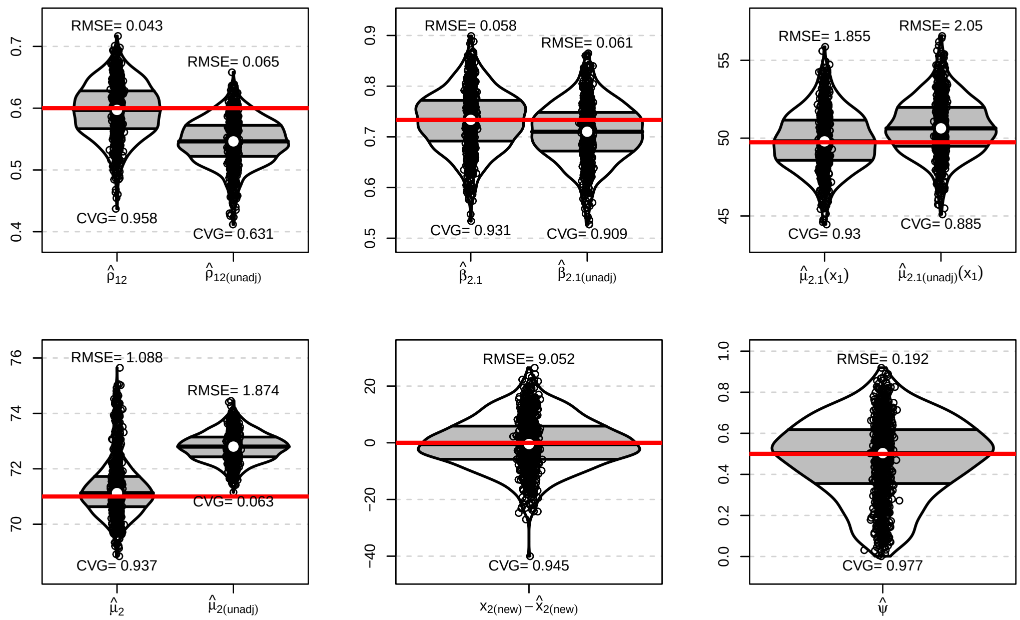

Figure 4. Simulation Results ($N = 520, \pi = 0.8, \rho_{12} = 0.6, \psi = 0.5, x_1 = 46, n_{sim} = 1000$)**.**

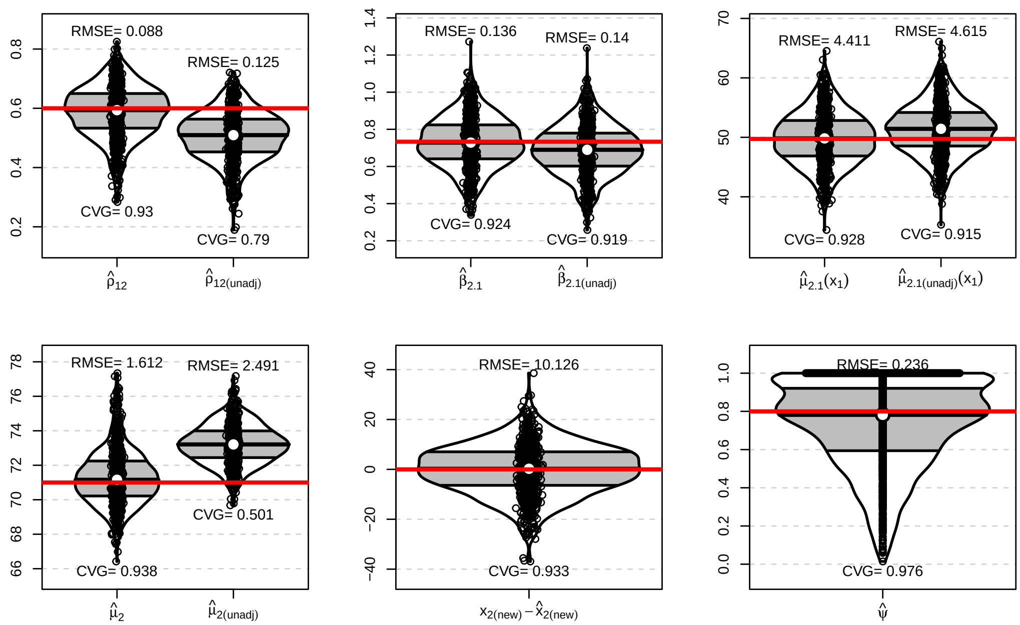

Figure 5. Simulation Results ($N = 100, \pi = 0.8, \rho_{12} = 0.6, \psi = 0.8, x_1 = 46, n_{sim} = 1000$)**.**

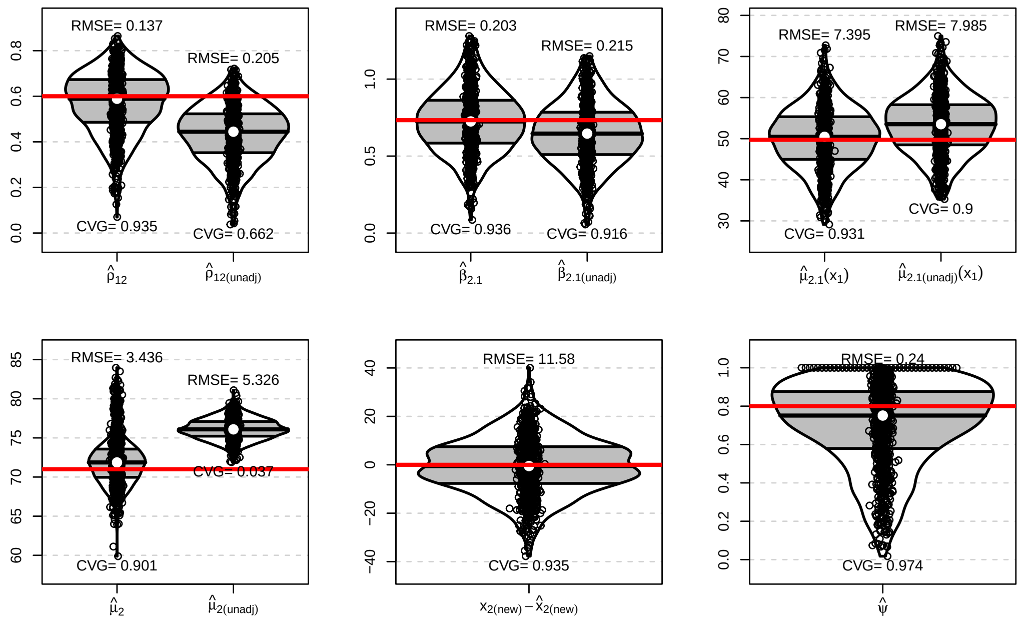

Figure 6. Simulation Results $(N = 100, \pi = 0.5, \rho_{12} = 0.6, \psi = 0.9, x_1 = 46, n_{sim} = 1000)$**.**

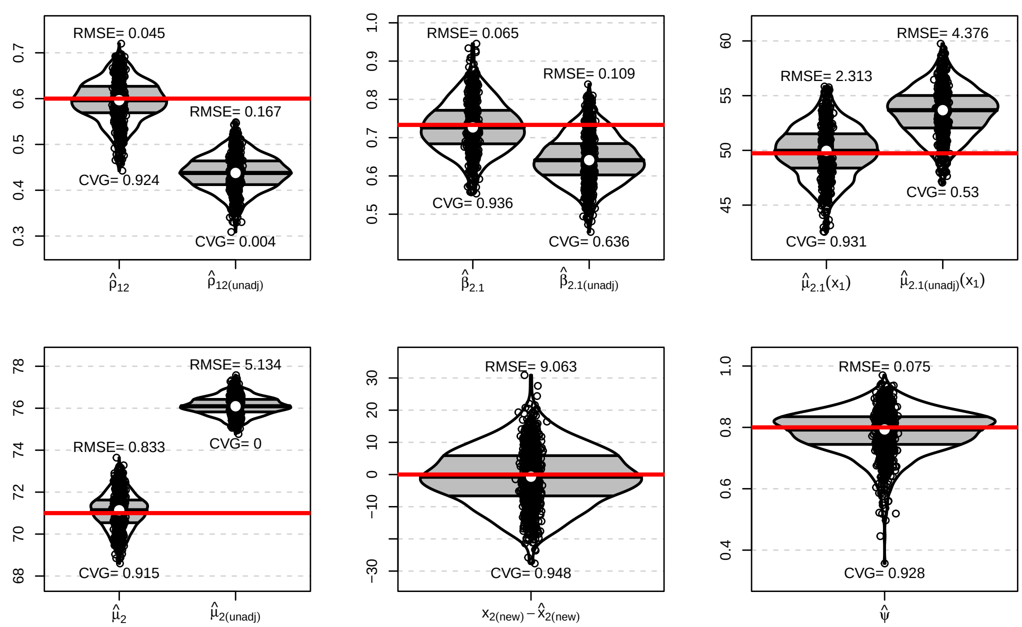

Figure 7. Simulation Results $(N = 1000, \pi = 0.5, \rho_{12} = 0.6, \psi = 0.8, x_1 = 46, n_{sim} = 1000)$**.**

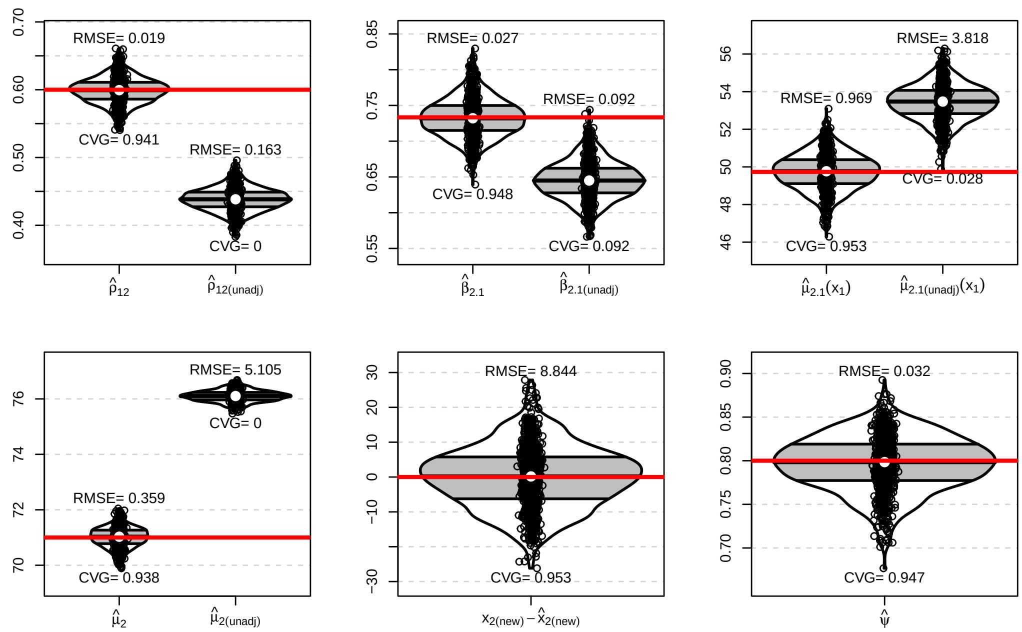

Figure 8. Simulation Results ($N = 5000, \pi = 0.5, \rho_{12} = 0.6, \psi = 0.9, x_1 = 46, n_{sim} = 1000$).

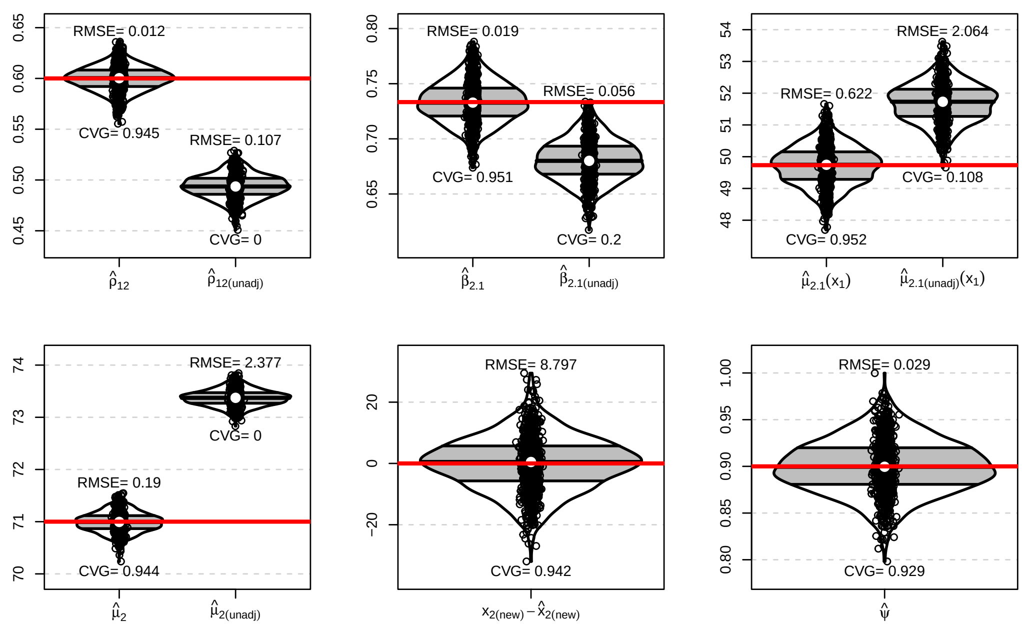

Figure 9. Simulation Results ($N = 5000, \pi = 0.8, \rho_{12} = 0.6, \psi = 0.9, x_1 = 46, n_{sim} = 1000$).

Table S.1 below summarizes the results corresponding to Figures 1-9. This table shows that the bias and variability of the adjusted estimators, as measured by the root mean squared error (RMSE(adj)) shrinks as the sample size N increases. This finding corroborates the consistency results for the adjusted estimators in the paper. The table (see last column) also shows that the nominal 95% bootstrap confidence intervals cover approximately 95% of the time, as expected by the results in the paper. Of course, as expected, Table S.1 also shows how poorly the unadjusted estimators can perform in terms of RMSE and confidence interval coverage. Note that the RMSE(unadj) and coverage CVG(unadj) values for the ψ estimator are not included because that parameter does not arise in the unadjusted model.

Table S.1. Monte Carlo Simulation Results.

					Estimator Statistics ($nsim = 1000$)			
Estimand	N	π	ρ_{12}	ψ	RMSE(unadj)	CVG(unadj)	RMSE(adj)	CVG(adj)
ρ_{12}	100	0.5	0.6	0.9	0.205	0.662	0.137	0.935
	100	0.8	0.6	0.8	0.125	0.790	0.088	0.930
	520	0.5	0.6	0.8	0.170	0.056	0.061	0.932
	520	0.8	−0.25	0.5	0.094	0.538	0.053	0.940
	520	0.8	0.6	0.5	0.065	0.631	0.043	0.958
	520	0.8	0.6	0.8	0.099	0.227	0.039	0.944
	1000	0.5	0.6	0.8	0.167	0.004	0.045	0.924
	5000	0.5	0.6	0.9	0.163	0.000	0.019	0.941
	5000	0.8	0.6	0.9	0.107	0.000	0.012	0.945
$\beta_{2\|1}$	100	0.5	0.6	0.9	0.215	0.916	0.203	0.936
	100	0.8	0.6	0.8	0.140	0.919	0.136	0.924
	520	0.5	0.6	0.8	0.122	0.792	0.089	0.940
	520	0.8	−0.25	0.5	0.091	0.707	0.065	0.949
	520	0.8	0.6	0.5	0.061	0.909	0.058	0.931
	520	0.8	0.6	0.8	0.072	0.873	0.058	0.948
	1000	0.5	0.6	0.8	0.109	0.636	0.065	0.936
	5000	0.5	0.6	0.9	0.092	0.092	0.027	0.948
	5000	0.8	0.6	0.9	0.056	0.200	0.019	0.951
ψ	100	0.5	0.6	0.9			0.240	0.974
	100	0.8	0.6	0.8			0.236	0.976
	520	0.5	0.6	0.8			0.118	0.910
	520	0.8	−0.25	0.5			0.119	0.921
	520	0.8	0.6	0.5			0.192	0.977
	520	0.8	0.6	0.8			0.113	0.941
	1000	0.5	0.6	0.8			0.075	0.928
	5000	0.5	0.6	0.9			0.032	0.947
	5000	0.8	0.6	0.9			0.029	0.929

2. PROOFS OF RESULTS 1-10

In these proofs, it is assumed that the model of equation (2) in the paper holds. For this model, $\mathbb{X}_j = \mu_j + \sigma_j \mathbb{Z}_j, j = 1, 2, 3$ and $\mathbb{T} = \gamma_0 + \gamma_1 \mathbb{X}_3 + \mathbb{E} \equiv \mu_T + \sigma_T \mathbb{Z}_T$, where $\mu_T = \gamma_0 + \gamma_1 \mu_3$, $\sigma_{TT} = \gamma_1^2 \sigma_{33} + \sigma^2$, and $(\mathbb{Z}_1, \mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_T)$ is a four-variate Normal with standard Normal marginals. It is also assumed that the Normally distributed $\mathbb{E}$ is independent of the $\mathbb{X}$ variables and has mean 0 and variance σ^2. Let $\rho_{ij} \equiv corr(\mathbb{X}_i, \mathbb{X}_j) = corr(\mathbb{Z}_i, \mathbb{Z}_j)$ for $i, j \in \{1, 2, 3\}$ and let $\rho_{3T} \equiv corr(\mathbb{X}_3, \mathbb{T}) = corr(\mathbb{Z}_3, \mathbb{Z}_T)$. Finally, as in the paper, let $\sigma_{jj} \equiv \sigma_j^2$.

The proofs of Results 1-10 will make use of the following lemma.

Lemma. For this model, (i) $\rho_{3T} = \gamma_1 \sigma_3 / \sigma_T$ and (ii) $\rho_{iT} = \rho_{i3} \rho_{3T}$, for $i = 1, 2$.

Proof of (i). Note that $cov(\mathbb{X}_3, \mathbb{T}) = cov(\mathbb{X}_3, \gamma_1 \mathbb{X}_3 + E) = \gamma_1 \sigma_{33}$, $var(\mathbb{X}_3) = \sigma_{33}$ and $var(\gamma_1 \mathbb{X}_3 + E) = var(\mathbb{T}) = \gamma_1^2 \sigma_{33} + \sigma^2 = \sigma_{TT}$. Hence, $\rho_{3T} = \gamma_1 \sigma_{33} / (\sigma_3 \sigma_T) = \gamma_1 \sigma_3 / \sigma_T$. QED(i).

Proof of (ii). Note that $\rho_{iT} \equiv corr(\mathbb{X}_i, \mathbb{T}) = corr(\mathbb{X}_i, \gamma_1 \mathbb{X}_3 + E) = \gamma_1 cov(\mathbb{X}_i, \mathbb{X}_3) / (\sigma_i \sigma_T) = \gamma_1 \rho_{i3} \sigma_i \sigma_3 / (\sigma_i \sigma_T) = \rho_{i3} (\gamma_1 \sigma_3 / \sigma_T) = \rho_{i3} \rho_{3T}$. QED(ii).

In what follows, let $\theta \equiv (\theta_0 - \mu_T)/\sigma_T$, $\lambda \equiv \phi(\theta)/\Phi(-\theta)$, $\delta \equiv \lambda(\lambda - \theta)$, and $\psi \equiv \rho_{3T}^2$.

Proof of Result 1. Note that $\mathbb{T} = \mu_T + \sigma_T \mathbb{Z}_T$, where $\mathbb{Z}_T \sim \mathbb{Z} \equiv N(0, 1)$. It follows that $\pi \equiv P(\mathbb{T} > \theta_T) = P(\mathbb{Z}_T > (\theta_T - \mu_T)/\sigma_T) = P(\mathbb{Z} > \theta) = \Phi(-\theta)$. QED(1).

Proof of Result 2. For convenience, let $\rho \equiv \rho_{3T}$. Write $skew_{3(t)} = Num/Den$. The numerator of the skew parameter is $Num \equiv E[(\mathbb{X}_3 - E(\mathbb{X}_3 | \mathbb{T} > \theta_0))^3 | \mathbb{T} > \theta_0]$. Here, $E(\mathbb{X}_3 | \mathbb{T} > \theta_0) = E(\mu_3 + \sigma_3 \mathbb{Z}_3 | \mathbb{Z}_T > \theta) = \mu_3 + \sigma_3 \lambda \rho$, by identity (6) in the paper. Thus, $Num = \sigma_3^3 [E(Y - \lambda \rho)^3]$, where $Y \sim \mathbb{Z}_3 | (\mathbb{Z}_T > \theta)$. Expanding and using identities (6), (8), and (14) in the paper, a little algebra shows that $Num = \sigma_3^3 \lambda \rho^3 (\theta^2 - 1 - 3\lambda\theta + 2\lambda^2)$ or because $\rho^3 = \mathrm{sgn}(\rho) \psi^{3/2}$, we have

$$Num = \sigma_3^3 \mathrm{sgn}(\rho) \lambda \psi^{3/2} (\theta^2 - 1 - 3\lambda\theta + 2\lambda^2).$$

The denominator of the skew satisfies $Den^{2/3} \equiv var(\mathbb{X}_3 | \mathbb{T} > \theta_0) = \sigma_{33} var(\mathbb{Z}_3 | \mathbb{Z}_T > \theta) = \sigma_{33}(1 - \delta \rho^2)$, by identity (8) in the paper. Again, noting that $\psi = \rho^2$ and $\rho \equiv \rho_{3T}$, it follows that

$$skew_{3(t)} = Num/Den = \mathrm{sgn}(\rho_{3\mathrm{T}}) \lambda (\theta^2 - 1 - 3\lambda\theta + 2\lambda^2) \left(\frac{\psi}{1 - \delta\psi} \right)^{3/2},$$

as was to be shown. Finally, the paper argues that all the multiplicative contributions in $skew_{3(t)}$ are positive except for $\mathrm{sgn}(\rho_{3\mathrm{T}})$. It follows that $\mathrm{sgn}(skew_{3(t)}) = \mathrm{sgn}(\rho_{3T})$. QED(2).

Proof of Result 3. $\rho_{12(t)} \equiv corr(\mathbb{X}_1, \mathbb{X}_2 | \mathbb{T} > \theta_0) = corr(\mathbb{Z}_1, \mathbb{Z}_2 | \mathbb{Z}_T > \theta) = \dfrac{\rho_{12} - \delta\rho_{1T}\rho_{2T}}{\sqrt{(1-\delta\rho_{1T}^2)}\sqrt{(1-\delta\rho_{2T}^2)}}$ by identity (12) in the paper. Now for $i = 1, 2$, the lemma above states that $\rho_{iT} = \rho_{i3}\rho_{3T}$ and because $\psi \equiv \rho_{3T}^2$, we have $\rho_{iT}^2 = \rho_{i3}^2\psi$. It follows that

$$\rho_{12(t)} \equiv corr(\mathbb{X}_1, \mathbb{X}_2 | \mathbb{T} > \theta_0) = \frac{\rho_{12} - \delta\psi\rho_{13}\rho_{23}}{\sqrt{(1-\delta\psi\rho_{13}^2)}\sqrt{(1-\delta\psi\rho_{23}^2)}},$$

as was to be shown. We also have for $i = 1, 2$, $\rho_{i3(t)} \equiv corr(\mathbb{X}_i, \mathbb{X}_3 | \mathbb{T} > \theta_0) = corr(\mathbb{Z}_i, \mathbb{Z}_3 | \mathbb{Z}_T > \theta) = \dfrac{\rho_{i3} - \delta\rho_{iT}\rho_{3T}}{\sqrt{(1-\delta\rho_{iT}^2)}\sqrt{(1-\delta\rho_{3T}^2)}}$ by identity (12) in the paper. Again use $\rho_{iT} = \rho_{i3}\rho_{3T}$, $\rho_{iT}^2 = \rho_{i3}^2\psi$, where as always $\psi \equiv \rho_{3T}^2$. After a little algebra we arrive at

$$\rho_{i3(t)} \equiv corr(\mathbb{X}_i, \mathbb{X}_3 | \mathbb{T} > \theta_0) = \rho_{i3}\sqrt{\frac{1-\delta\psi}{1-\delta\psi\rho_{i3}^2}}, \quad i = 1, 2.$$

And this proves Result 3. QED(3)

Proof of Result 4. $\sigma_{11(t)} \equiv var(\mathbb{X}_1 | \mathbb{T} > \theta_0) = \sigma_{11} var(\mathbb{Z}_1 | \mathbb{Z}_T > \theta) = \sigma_{11}(1 - \delta\rho_{1T})$ by identity (8) in the paper. By the lemma above, $\rho_{1T}^2 = \rho_{13}^2\rho_{3T}^2 \equiv \rho_{13}^2\psi$ and we can re-express this variance as

$$\sigma_{11(t)} \equiv var(\mathbb{X}_1 | \mathbb{T} > \theta_0) = \sigma_{11}(1 - \delta\psi\rho_{13}^2),$$

as was to be shown. The same argument leads to the analogous expression for $\sigma_{22(t)}$. QED(4).

Proof of Result 5. $\mu_{1(t)} \equiv E(\mathbb{X}_1 | \mathbb{T} > \theta_0) = \mu_1 + \sigma_1 E(\mathbb{Z}_1 | \mathbb{Z}_T > \theta) = \mu_1 + \sigma_1\lambda\rho_{1T}$, by identity (6) in the paper. Now using $\rho_{1T} = \rho_{13}\rho_{3T}$ we arrive at

$$\mu_{1(t)} \equiv E(\mathbb{X}_1 | \mathbb{T} > \theta_0) = \mu_1 + \sigma_1\lambda\rho_{13}\rho_{3T},$$

as was to be shown. The same argument leads to the analogous expression for $\mu_{2(t)}$. QED(5).

In the proofs of Results 6-10 below, we continue to use the notation $\theta \equiv (\theta_0 - \mu_T)/\sigma_T$, $\lambda \equiv \phi(\theta)/\Phi(-\theta)$, $\delta \equiv \lambda(\lambda - \theta)$, and $\psi \equiv \rho_{3T}^2$.

Proof of Result 6. By Result 1, $\pi \equiv P(\mathbb{T} > \theta_0) = \Phi(-\theta)$. It follows that $\theta = -\Phi^{-1}(\pi)$. QED(6).

Proof of Result 7. Again, define $\rho \equiv \rho_{3T}$, so $\psi = \rho^2$. We have from Result 2 that $skew_{3(t)} = \text{sgn}(\rho)\lambda D \left(\frac{\rho^2}{1-\delta\rho^2}\right)^{3/2}$, where $D = \theta^2 - 1 - 3\lambda\theta + 2\lambda^2$. Now let $R \equiv |skew_{3(t)}|/(\lambda D)$. It follows that $R^{2/3} = \rho^2/(1-\delta\rho^2)$ and hence $\rho^2 = R^{2/3}/(1+\delta R^{2/3})$. Thus, $\rho = \text{sgn}(\rho)\sqrt{\frac{R^{2/3}}{1+\delta R^{2/3}}}$. Noting that $\rho \equiv \rho_{3T}$ and $\text{sgn}(\rho_{3T}) = \text{sgn}(skew_{3(t)})$ gives the desired result,

$$\rho_{3T} = \text{sgn}(skew_{3(t)})\sqrt{\frac{R^{2/3}}{1+\delta R^{2/3}}}.$$

QED(7).

Proof of Result 8. By Result 3, we have $\rho_{13(t)} = \rho_{13}\frac{\sqrt{1-\delta\psi}}{\sqrt{1-\delta\psi\rho_{13}^2}}$. For simplicity, write this as $A = B\frac{\sqrt{1-\delta\psi}}{\sqrt{1-\delta\psi B^2}}$ and solve for B. Use the fact that $A^2 = B^2\frac{(1-\delta\psi)}{(1-\delta\psi B^2)}$. Algebra leads to $B^2 = \frac{A^2}{1-\delta\psi(1-A^2)}$, and because $\text{sgn}(A) = \text{sgn}(B)$, we arrive at $B = \frac{A}{\sqrt{1-\delta\psi(1-A^2)}}$. That is, we have

$$\rho_{13} = \frac{\rho_{13(t)}}{\sqrt{1-\delta\psi(1-\rho_{13(t)}^2)}},$$

as was to be shown. The same argument leads to the analogous expression for ρ_{23}.

We are left to find the expression for ρ_{12}. By Result 3, we have $\rho_{12(t)} = (\rho_{12} - \delta\psi\rho_{13}\rho_{23})/D$, where $D \equiv \sqrt{(1-\delta\psi\rho_{13}^2)}\sqrt{(1-\delta\psi\rho_{23}^2)}$. Now using the expressions for ρ_{13} and ρ_{23} just derived, we have that

$$\rho_{13}\rho_{23} = \frac{\rho_{13(t)}\rho_{23(t)}}{\sqrt{1-\delta\psi(1-\rho_{13(t)}^2)}\sqrt{1-\delta\psi(1-\rho_{23(t)}^2)}}.$$

After a little algebra, we also have that $(1-\delta\psi\rho_{i3}^2) = \frac{1-\delta\psi}{1-\delta\psi(1-\rho_{i3(t)}^2)}$, for $i = 1, 2$, so that D can be expressed as $D = \frac{1-\delta\psi}{\sqrt{1-\delta\psi(1-\rho_{13(t)}^2)}\sqrt{1-\delta\psi(1-\rho_{23(t)}^2)}}$. With these identities, we arrive at

$$\rho_{12} = \rho_{12(t)}D + \delta\psi\rho_{13}\rho_{23} = \frac{\rho_{12(t)} - \delta\psi(\rho_{12(t)} - \rho_{13(t)}\rho_{23(t)})}{\sqrt{1-\delta\psi(1-\rho_{13(t)}^2)}\sqrt{1-\delta\psi(1-\rho_{23(t)}^2)}},$$

as was to be shown. This proves Result 8. QED(8).

Proof of Result 9. By Result 4, we have $\sigma_{11(t)} = \sigma_{11}(1 - \delta\psi\rho_{13}^2)$. In the proof of Result 8, we also argued that $(1-\delta\psi\rho_{13}^2) = \dfrac{1-\delta\psi}{1-\delta\psi(1-\rho_{13(t)}^2)}$. By these two identities, we can solve for σ_{11} and arrive at

$$\sigma_{11} = \sigma_{11(t)} \left(\frac{1-\delta\psi(1-\rho_{13(t)}^2)}{1-\delta\psi} \right),$$

as was to be shown. The same argument leads to the analogous expression for σ_{22}. QED(9).

Proof of Result 10. By Result 5, we have $\mu_{1(t)} = \mu_1 + \sigma_1\lambda\rho_{13}\rho_{3T}$. By Results 8 and 9, we have that $\sigma_1\rho_{13} = \dfrac{\sigma_{1(t)}\rho_{13(t)}}{\sqrt{1-\delta\psi}}$. It follows that

$$\mu_1 = \mu_{1(t)} - \sigma_1\lambda\rho_{13}\rho_{3T} = \mu_{1(t)} - \sigma_{1(t)}\lambda\rho_{13(t)}\frac{\rho_{3T}}{\sqrt{1-\delta\psi}},$$

as was to be shown. The same argument leads to the analogous expression for μ_2. QED(10).