Appendix for "Quantifying Observed Prior Impact"

David E. Jones[1] **Robert N. Trangucci**[2] **and Yang Chen**[2,*,†]

[1] *Texas A&M University, College Station, TX, USA*

[2] *University of Michigan, Ann Arbor, MI, USA*

Appendices

Appendix A: Proof of Proposition 4.1

In part (a), $\gamma = \mu_n$ and from (4.1) we have

$$W_2(m) = \left(\frac{n}{m}\right)^2 (\mu_n - \bar{y}_n)^2 + \left(\frac{\sigma}{\sqrt{n+z}} - \frac{\sigma}{\sqrt{m}}\right)^2. \tag{A.1}$$

Let m^* denote the minimizer of $W_2(m)$ and suppose that $m^* < n + z$. For $m = n + z$ the second term of (A.1) is zero and the first term is smaller than for m^*, thus $W_2(n + z) < W_2(m^*)$, a contradiction. Therefore, $m^* \geq n + z$. (Aside: note that $m^* = n + z$ holds if $\mu_n = \bar{y}_n$, i.e., if $\mu_0 = \bar{y}_n$.) Furthermore, referring to (4.2), for $m > n$, we have

$$\widetilde{W}_2(m) = (\bar{y}_n - \mu_n)^2 + \left(\frac{\sigma}{\sqrt{n}} - \frac{\sigma}{\sqrt{m+z}}\right)^2 > W_2(m). \tag{A.2}$$

To see this note the following. If $m = n$ then the second term on the right-hand side of (A.1) is equal to the corresponding standard deviation term in (A.2). For $n < m \leq n + z$, the standard deviation term in (A.1) decreases whereas that in (A.2) increases. Regarding the case $m > n + z$, $\sigma/\sqrt{n}$ is larger than $\sigma/\sqrt{n+z}$ and the value being subtracted from these terms is smaller in (A.2) than in (A.1) ($\sigma/\sqrt{m+z}$ versus $\sigma/\sqrt{m}$), and thus again the standard deviation term is larger in (A.2) . Thus, for all $m > n$, the standard deviation term in (A.2) is larger than that in (A.1). This verifies the inequality in (A.2) because the first term (A.2) is necessarily larger than that of (A.1) for $m > n$. Thus, $M_n \geq z$.

In part (b), $\gamma = \bar{y}_n$ and from (4.1) and (4.2) we have

$$W_2(m) = (\mu_n - \bar{y}_n)^2 + \left(\frac{\sigma}{\sqrt{m}} - \frac{\sigma}{\sqrt{n+z}}\right)^2, \tag{A.3}$$

$$\widetilde{W}_2(m) = \left(\frac{n+z}{m+z}\right)^2 (\bar{y}_n - \mu_n)^2 + \left(\frac{\sigma}{\sqrt{n}} - \frac{\sigma}{\sqrt{m+z}}\right)^2. \tag{A.4}$$

*Correspondence to: Yang Chen, Department of Statistics and Michigan Institute for Data Science (MIDAS), University of Michigan, Ann Arbor, e-mail: ychenang@umich.edu

†This work is supported by NSF DMS-1811083 (PI: Yang Chen, 2018 - 2021).

Clearly (A.3) is minimized at $m = n+z$, because in that case the second term on the right-hand side is zero (and the first does not depend on m). The second term on the right-hand side of (A.4) is monotonically increasing for $m \geq n$. Thus, setting ϵ_s to be any value such that $\epsilon_s^2 < \left((\sigma/\sqrt{n}) - (\sigma/\sqrt{n+z})\right)^2$ yields the first part of result (b). The first term on the right-hand side of (A.4) converges to zero as m increases, and the second term is bounded above by σ^2/n. Thus, choosing $\epsilon_l^2 > \sigma^2/n$, the second part of result (b) follows.

In part (c), $\gamma = \mu_0$ and we have

$$W_2(m) = (\mu_n - \mu_{n,m})^2 + \left(\frac{\sigma}{\sqrt{m}} - \frac{\sigma}{\sqrt{n+z}}\right)^2, \tag{A.5}$$

$$\widetilde{W}_2(m) = (\bar{y} - \mu_{n,m+z})^2 + \left(\frac{\sigma}{\sqrt{n}} - \frac{\sigma}{\sqrt{m+z}}\right)^2, \tag{A.6}$$

where $\mu_{n,m} = w_{n,m}\bar{y}_n + (1 - w_{n,m})\mu_0$, and $w_{n,m} = n/m$. Since $m = n + z$ gives $W_2(m) = 0$, and $\widetilde{W}_2(m) > 0$ for all $m \geq n$, we have $M_n = z$.

Appendix B: Proof of Lemma 4.1

We denote the additional samples collected by $s_1, \ldots, s_r$, i.e., $\{x_1, \ldots, x_m\} = \{y_1, \ldots,$
$y_n, s_1, \ldots, s_r\}$, and write $\bar{s}_r$ to denote $\frac{1}{r}\sum_{i=1}^r s_i$. Thus, we have

$$\bar{s}_r | \bar{y}_n, \mu \sim N\left(\mu, \frac{\sigma^2}{r}\right). \tag{B.1}$$

From (4.1) we have

$$\begin{aligned}
W_2(m) &= \left(\mu_n - \frac{n}{m}\bar{y}_n - \frac{r}{m}\bar{s}_r\right)^2 + c_m^2 && \text{(B.2)}\\
&= \left(w_n\bar{y}_n + (1-w_n)\mu_0 - \frac{n}{m}\bar{y}_n - \frac{r}{m}\bar{s}_r\right)^2 + c_m^2 && \text{(B.3)}\\
&= \left(\left(\frac{n}{n+z} - 1 + 1 - \frac{n}{m}\right)\bar{y}_n + (1-w_n)\mu_0 - \frac{r}{m}\bar{s}_r\right)^2 + c_m^2 && \text{(B.4)}\\
&= \left(\left(\frac{r}{m} - \frac{z}{n+z}\right)\bar{y}_n + (1-w_n)\mu_0 - \frac{r}{m}\bar{s}_r\right)^2 + c_m^2 && \text{(B.5)}
\end{aligned}$$

The conditional distribution of $W_2(m)$ given μ and $\bar{y}_n$ stated in Lemma 4.1 then follows from the distribution of $\bar{s}_r$ (the only random quantity in (B.5)). The proof for the conditional distribution of $\widetilde{W}_2(m)$ is similar and is omitted. The proof of Theorem D.1 in Appendix D, which gives the distance distributions conditional on only $\bar{y}_n$, relies on analogous arguments and is also omitted.

Appendix C: Proof of Theorem 4.1

Firstly, suppose that $v > 0$. Let $W = \text{argmin}_{m \geq n} W_2(m)$ and $\widetilde{W} = \text{argmin}_{m \geq n} \widetilde{W}_2(m)$. Then $P(M_n = v|\bar{y}_n) = P(W \geq W_2(v+n), \widetilde{W} \geq W_2(v+n)|\bar{y}_n)$ can be expressed as

$$\int_{\mathbb{R}} \int_{T_v} b(t) \prod_{\substack{m=n \\ m \neq v+n}}^{\infty} (1 - F_{m,\mu}(t_m)) \prod_{m=n}^{\infty} (1 - \widetilde{F}_{m,\mu}(\tilde{t}_m)) h_{v+n,\mu}(t) dt \pi(\mu|\bar{y}) d\mu \tag{C.1}$$

$$= \int_{\mathbb{R}} \int_{T_v} b(t) \prod_{\substack{m=n \\ m \neq v+n}}^{M(t)} (1 - F_{m,\mu}(t_m)) \prod_{m=n}^{\widetilde{M}(t)} (1 - \widetilde{F}_{m,\mu}(\tilde{t}_m)) h_{v+n,\mu}(t) dt \pi(\mu|\bar{y}) d\mu, \tag{C.2}$$

where $b(t) = 1_{\{\sigma^2/(n+z) \geq t\}}$, and for $\sigma^2/(n+z) \geq t$, $M(t)$ denotes the minimum value of $m \in \mathbb{Z}_{\geq n+z}$ such that

$$\left(\frac{\sigma}{\sqrt{m}} - \frac{\sigma}{\sqrt{n+z}} \right)^2 > t, \tag{C.3}$$

whereas $\widetilde{M}(t)$ denotes the minimum value of $m \in \mathbb{Z}_{\geq n+1}$ such that

$$\left(\frac{\sigma}{\sqrt{m+z}} - \frac{\sigma}{\sqrt{n}} \right)^2 > t. \tag{C.4}$$

In particular, for all $t \leq \sigma^2/(n+z)$, $M(t)$ and $\widetilde{M}(t)$ are both finite, and for $m > M(t)$ we have $P(W_2(m) < t|\bar{y}, \mu) = 0$, and similarly for $m > \widetilde{M}(t)$ we have $P(\widetilde{W}_2(m) < t|\bar{y}, \mu) = 0$. This demonstrates the equality of (C.1) and (C.2). The indicator $1_{\{\sigma^2/(n+z) \geq t\}}$ is needed because if $\sigma^2/(n+z) < t$, then the above arguments do not hold and there exists $m^* > v+n$ such that $W_2(m^*) < t$ or $\widetilde{W}_2(m^*) < t$ with probability 1, meaning that $P(W \geq t, \widetilde{W} \geq t|\bar{y}_n, \mu) = 0$. For completeness we set $M(t) = \widetilde{M}(t) = \infty$ if $\sigma^2/(n+z) < t$. Next, $P(W_2(n) > t|\bar{y}, \mu), P(\widetilde{W}_2(n) > t|\bar{y}, \mu) \in \{0, 1\}$, meaning that (C.2) can be written as

$$\int_{\mathbb{R}} \int_{T_v} 1_{\{W_2(n), \widetilde{W}_2(n), \sigma^2/(n+z) \geq t\}} g(t, \mu, v, M(t), \widetilde{M}(t)) dt \pi(\mu|\bar{y}) d\mu \tag{C.5}$$

as in Theorem 4.1. The proof in the case $v < 0$ is analogous.

Lastly, if $v = 0$, then essentially the same derivation holds except that the integral over t is no longer required and $P(M_n = 0|\bar{y}_n)$ simplifies to

$$b(W_2(n)) \int_{\mathbb{R}} \prod_{m=n+1}^{M(W_2(n))} (1 - F_{m,\mu}(s_m)) \prod_{m=n+1}^{\widetilde{M}(W_2(n))} (1 - \widetilde{F}_{m,\mu}(\tilde{s}_m)) \pi(\mu|\bar{y}) d\mu. \tag{C.6}$$

Appendix D: Conditional Distribution of Distances

We report a result similar to Lemma 4.1 for the case in which we only condition on $\boldsymbol{y}$, and not a particular draw of μ. Note that in this case the two distances are not independent. The proof is similar to that for Lemma 4.1 and is omitted.

Lemma D.1. *Distribution of Distances. Using the same notation as in Lemma 4.1, we have*

$$\left[W_2\left(q_n, q_m^b\right)\middle|\boldsymbol{y}\right] \sim \tau_m \chi_1^2\left(\frac{\lambda}{m^2\tau_m}\right) + \left(\frac{\sigma}{\sqrt{n+z}} - \frac{\sigma}{\sqrt{m}}\right)^2,$$

where

$$\tau_m = \frac{r^2}{m^2}\left(\frac{w_n}{n} + \frac{1}{r}\right)\sigma^2, \quad \lambda = \left(n(1-w_n)(\bar{y}_n - \mu_0)\right)^2;$$

and

$$\left[W_2\left(q_m, q_n^b\right)\middle|\boldsymbol{y}\right] \sim \kappa_m \chi_1^2\left(\frac{\delta}{\kappa_m}\right) + \left(\frac{\sigma}{\sqrt{m+z}} - \frac{\sigma}{\sqrt{n}}\right)^2,$$

where

$$\kappa_m = w_m^2 \tau_m, \quad \delta = \frac{\lambda}{n^2}.$$

Appendix E: Details for Regression Model in Section 5.3

Here we give the expressions for the posterior distributions $\{\pi_u^b(\boldsymbol{\beta}), \pi_u(\boldsymbol{\beta})\}$ and the Wasserstein distances $\{W_2(q_n, q_m^b), W_2(q_m), q_n^b\}$ discussed in the example of Section 5.3.

For ease of notation, we define the following quantities, for $u \geq n$:

$$\begin{aligned}
\overline{\omega}_u &= \frac{1}{u}\sum_{i=1}^u \omega_i^{(u)}, \quad \overline{y}_u = \frac{1}{u}\sum_{i=1}^u y_i^{(u)}, \\
\overline{\omega^2}_u &= \frac{1}{u}\sum_{i=1}^u ((\omega_i')^{(u)})^2, \quad \overline{\omega y}_u = \frac{1}{u}\sum_{i=1}^u (\omega_i')^{(u)} y_i^{(u)}, \\
\hat{\beta}_u &= \begin{bmatrix} \overline{y}_u \\ \overline{\omega y}_u / \overline{\omega^2}_u \end{bmatrix}.
\end{aligned}$$

Define $\boldsymbol{\eta}_u$ and Σ_u as

$$\begin{aligned}
\boldsymbol{\eta}_u &= \begin{bmatrix} u + \frac{1}{\tau_1^2} & u\overline{\omega}_u \\ u\overline{\omega}_u & u\overline{\omega^2}_u + \frac{1}{\tau_2^2} \end{bmatrix}^{-1} \begin{bmatrix} u\overline{y}_u + \frac{\mu_0}{\tau_1^2} \\ u\overline{\omega^2}_u \frac{\overline{\omega y}_u}{\overline{\omega^2}_u} + \frac{\gamma_0}{\tau_2^2} \end{bmatrix}, \\
\Sigma_u &= \sigma^2\left(\begin{bmatrix} u + \frac{1}{\tau_1^2} & u\overline{\omega}_u \\ u\overline{\omega}_u & u\overline{\omega^2}_u + \frac{1}{\tau_2^2} \end{bmatrix}\right)^{-1}.
\end{aligned}$$

Then $\pi_u(\boldsymbol{\beta}) = \mathcal{N}\left(\boldsymbol{\eta}_u, \Sigma_u\right)$ while $\pi_u^b(\boldsymbol{\beta}) = \mathcal{N}\left(\hat{\boldsymbol{\beta}}_u, \hat{\Sigma}_u\right)$, where

$$\hat{\boldsymbol{\beta}}_u = \begin{bmatrix} \overline{y}_u \\ \frac{\overline{\omega y}_u}{\overline{\omega^2}_u} \end{bmatrix}, \quad \hat{\Sigma}_u = \sigma^2 \begin{bmatrix} u^{-1} & u\overline{\omega}_u \\ u\overline{\omega}_u & \frac{1}{u\overline{\omega^2}_u} \end{bmatrix}.$$

Since for multivariate Gaussians ν_A, ν_B with mean vectors $\boldsymbol{\mu}_A, \boldsymbol{\mu}_B \in \mathbb{R}^d$ and covariance matrices $\Sigma_A, \Sigma_B \in \mathbb{R}^{d\times d}$, $W_2(\nu_A, \nu_B)$ is

$$\left(||\boldsymbol{\mu}_A - \boldsymbol{\mu}_B||_2^2 + \operatorname{tr}\left(\Sigma_A + \Sigma_B - 2(\Sigma_B^{\frac{1}{2}}\Sigma_A\Sigma_B^{\frac{1}{2}})^{\frac{1}{2}}\right)\right)^{\frac{1}{2}},$$

we can write $W_2(q_n, q_m^b)$ and $W_2(q_m, q_n^b)$ in terms of posterior quantities $\hat{\boldsymbol{\beta}}_u, \boldsymbol{\eta}_u, \hat{\Sigma}_u$, and Σ_u as follows:

$$W_2(q_n, q_m^b) = \left(\left|\left|\hat{\boldsymbol{\beta}}_m - \boldsymbol{\eta}_n\right|\right|_2^2 + \operatorname{tr}\left(\hat{\Sigma}_m + \Sigma_n - 2(\Sigma_n^{\frac{1}{2}}\hat{\Sigma}_m\Sigma_n^{\frac{1}{2}})^{\frac{1}{2}}\right)\right)^{\frac{1}{2}},$$

$$W_2(q_m, q_n^b) = \left(\left|\left|\boldsymbol{\eta}_m - \hat{\boldsymbol{\beta}}_n\right|\right|_2^2 + \operatorname{tr}\left(\Sigma_m + \hat{\Sigma}_n - 2(\hat{\Sigma}_n^{\frac{1}{2}}\Sigma_m\hat{\Sigma}_n^{\frac{1}{2}})^{\frac{1}{2}}\right)\right)^{\frac{1}{2}}.$$