

Irrational Exuberance: Correcting Bias in Probability Estimates.

Supplementary Material

By: Gareth M. James¹, Peter Radchenko² and Bradley Rava^{1,3}

A Proof of Theorem 1

We begin by computing the derivative of the loss function,

$$L(x) = \begin{cases} \frac{1}{x^2} E(p_i^2 | \tilde{p}_i) - \frac{2}{x} E(p_i | \tilde{p}_i) + 1 & x \leq 0.5 \\ \frac{1}{(1-x)^2} E(p_i^2 | \tilde{p}_i) - \frac{2x}{(1-x)^2} E(p_i | \tilde{p}_i) + \left(\frac{x}{1-x}\right)^2 & x > 0.5. \end{cases}$$

We have

$$\begin{aligned} \frac{\partial L}{\partial x} &= \begin{cases} -\frac{2}{x^3} E(p_i^2 | \tilde{p}_i) + \frac{2}{x^2} E(p_i | \tilde{p}_i) & x < 0.5 \\ \frac{2}{(1-x)^3} E(p_i^2 | \tilde{p}_i) - \frac{2(1+x)}{(1-x)^3} E(p_i | \tilde{p}_i) + \frac{2x}{(1-x)^3} & x > 0.5 \end{cases} \\ &\propto \begin{cases} -E(p_i^2 | \tilde{p}_i) + x E(p_i | \tilde{p}_i) & x < 0.5 \\ E(p_i^2 | \tilde{p}_i) - E(p_i | \tilde{p}_i) + x(1 - E(p_i | \tilde{p}_i)) & x > 0.5. \end{cases} \end{aligned}$$

Note that L is a continuous function. If $E(p_i | \tilde{p}_i) \leq 0.5$ and $x^* = E(p_i^2 | \tilde{p}_i) / E(p_i | \tilde{p}_i) \leq 0.5$ then algebraic manipulations show that $\partial L / \partial x$ is negative for all $x < x^*$ and positive for $x > x^*$. Hence, $p_{i0} = x^* = E(p_i | \tilde{p}_i) + \text{Var}(p_i | \tilde{p}_i) / E(p_i | \tilde{p}_i)$ minimizes L . Alternatively, if $E(p_i | \tilde{p}_i) \leq 0.5$ and $x^* = E(p_i^2 | \tilde{p}_i) / E(p_i | \tilde{p}_i) \geq 0.5$ then $\partial L / \partial x$ is negative for all $x < 0.5$ and positive for all $x > 0.5$, so L is minimized by $p_{i0} = 0.5$.

Analogous arguments show that if $E(p_i | \tilde{p}_i) > 0.5$ and $x^* = E(p_i^2 | \tilde{p}_i) / (1 - E(p_i | \tilde{p}_i)) > 0.5$, then $\partial L / \partial x$ is negative for all $x < x^*$, zero at $x = x^*$ and positive for $x > x^*$. Hence, $p_{i0} = x^* = E(p_i | \tilde{p}_i) + \text{Var}(p_i | \tilde{p}_i) / (1 - E(p_i | \tilde{p}_i))$ will minimize L . Alternatively, if $E(p_i | \tilde{p}_i) > 0.5$ and $x^* = E(p_i^2 | \tilde{p}_i) / (1 - E(p_i | \tilde{p}_i)) < 0.5$ then $\partial L / \partial x$ is negative for all $x < 0.5$ and positive for all $x > 0.5$, so L is minimized by $p_{i0} = 0.5$.

To prove the second result, first suppose $E(p_i | \tilde{p}_i) \leq 0.5$ and $p_{i0} < 0.5$, in which case $L(p_{i0}) = 1 - E(p_i^2 | \tilde{p}_i) / p_{i0}^2$. Now let $\tilde{L}(p'_i) = E\left(\left(\frac{p_i - p'_i}{p'_i}\right)^2 | \tilde{p}_i\right) = \frac{1}{p'^2_i} E(p_i^2 | \tilde{p}_i) - \frac{2}{p'_i} E(p_i | \tilde{p}_i) + 1$. Note that $\tilde{L}(p'_i) \leq L(p'_i)$ with equality for $p'_i \leq 0.5$. Hence,

$$L(p'_i) - L(p_{i0}) \geq \tilde{L}(p'_i) - L(p_{i0}) = E(p_i^2 | \tilde{p}_i) \left(\frac{1}{p'^2_i} + \frac{1}{p_{i0}^2} \right) - \frac{2}{p'_i} E(p_i | \tilde{p}_i) = E(p_i^2 | \tilde{p}_i) \left(\frac{1}{p'_i} - \frac{1}{p_{i0}} \right)^2.$$

Now consider the case $E(p_i | \tilde{p}_i) \leq 0.5$ and $p_{i0} = 0.5$. Note that this implies $2E(p_i^2 | \tilde{p}_i) > E(p_i | \tilde{p}_i)$.

¹Department of Data Sciences and Operations, University of Southern California.

²University of Sydney.

³Research is generously supported by the NSF GRFP in Mathematical Statistics.

If $p'_i \leq 0.5$, then

$$L(p'_i) - L(p_{i0}) = E(p_i^2|\tilde{p}_i) \left(\frac{1}{p_i'^2} - 4 \right) - 2E(p_i|\tilde{p}_i) \left(\frac{1}{p'_i} - 2 \right) \geq E(p_i^2|\tilde{p}_i) \left(\frac{1}{p'_i} - \frac{1}{0.5} \right)^2.$$

Also note that

$$\left(\frac{1}{p'_i} - \frac{1}{0.5} \right)^2 \geq \left(\frac{1}{1-p'_i} - \frac{1}{0.5} \right)^2.$$

Alternatively, if $p'_i > 0.5$, then

$$L(p'_i) - L(p_{i0}) \geq \tilde{L}(1-p'_i) - L(p_{i0}) \geq E(p_i^2|\tilde{p}_i) \left(\frac{1}{1-p'_i} - \frac{1}{0.5} \right)^2.$$

Observe that

$$\left(\frac{1}{1-p'_i} - \frac{1}{0.5} \right)^2 \geq \left(\frac{1}{p'_i} - \frac{1}{0.5} \right)^2.$$

Consequently, we have shown that

$$L(p'_i) - L(p_{i0}) \geq E(p_i^2|\tilde{p}_i) \max \left(\left[\frac{1}{p'_i} - \frac{1}{0.5} \right]^2, \left[\frac{1}{1-p'_i} - \frac{1}{0.5} \right]^2 \right)$$

when $E(p_i|\tilde{p}_i) \leq 0.5$ and $p_{i0} = 0.5$.

Thus, we have established the result for the case $E(p_i|\tilde{p}_i) \leq 0.5$. Finally, consider the case $E(p_i|\tilde{p}_i) \geq 0.5$. The result follows by repeating the argument from the case $E(p_i|\tilde{p}_i) < 0.5$ while replacing all of the probabilities with their complements, i.e., by replacing p_i , p_{i0} and p'_i with $1-p_i$, $1-p_{i0}$ and $1-p'_i$, respectively.

B Proof of Theorem 2 and Corollary 1

Throughout the proof, we omit the subscript i , for the simplicity of notation. We let $f_{\tilde{p}}(\tilde{p}|p)$ denote the conditional density of $\tilde{p}$ given that the corresponding true probability equals p , and define $f_X(x|p)$ by analogy for the random variable $X = \log(\tilde{p}/[1-\tilde{p}])$. We will slightly abuse the notation and not distinguish between the random variable and its value in the case of $\tilde{p}$, p and η .

According to model (1), we have $f_{\tilde{p}}(\tilde{p}|p) = B(p/\gamma^*, (1-p)/\gamma^*)^{-1} \tilde{p}^{p/\gamma^*-1} (1-\tilde{p})^{(1-p)/\gamma^*-1}$ where $B(\cdot)$ denotes the beta function. Hence, writing B for $B(p/\gamma^*, (1-p)/\gamma^*)$, we derive

$$\begin{aligned} \log(f_{\tilde{p}}(\tilde{p}|p)) &= -\log B + \left(\frac{p}{\gamma^*} - 1 \right) [\log \tilde{p} - \log(1-\tilde{p})] + (1/\gamma^* - 2) \log(1-\tilde{p}) \\ &= -\log B + \eta x + (1/\gamma^* - 2) \log(1-\tilde{p}) \end{aligned} \quad (34)$$

where $\eta = \frac{p}{\gamma^*} - 1$ and $x = \log \frac{\tilde{p}}{1-\tilde{p}}$. Standard calculations show that

$$f_X(x|p) = f_{\tilde{p}}(\tilde{p}|p) \frac{e^x}{(1+e^x)^2} = f_{\tilde{p}}(\tilde{p}|p) \tilde{p}(1-\tilde{p}). \quad (35)$$

Note that $\log(1 - \tilde{p}) = -\log(1 + e^x)$, and hence

$$\begin{aligned}\log(f_X(x|p)) &= -\log B + \eta x - (1/\gamma^* - 2)\log(1 + e^x) + x - 2\log(1 + e^x) \\ &= -\log B + \eta x + x - 1/\gamma^* \log(1 + e^x) \\ &= -\log B + \eta x + l_h(x),\end{aligned}$$

where $l_h(x) = x - 1/\gamma^* \log(1 + e^x)$.

Consequently, we can apply Tweedie's formula (Efron, 2011) to derive

$$E(p/\gamma^* - 1|\tilde{p}) = E(\eta|x) = v_X(x) - l'_h(x) = v_X(x) - 1 + \frac{1}{\gamma^*} \frac{e^x}{1 + e^x} = v_X(x) + \frac{\tilde{p}}{\gamma^*} - 1,$$

where $v_X(x) = (df_X(x)/dx)/f_X(x)$ and f_X is the density of X . This implies

$$E(p|\tilde{p}) = \tilde{p} + \gamma^* v_X(x).$$

In addition, we have

$$\begin{aligned}\frac{df_X(x)}{dx} &= \frac{df_{\tilde{p}}(\tilde{p})}{d\tilde{p}} \frac{d\tilde{p}}{dx} \frac{e^x}{(1 + e^x)^2} + f_{\tilde{p}}(\tilde{p}) \frac{e^x(1 + e^x)^2 - 2e^{2x}(1 + e^x)}{(1 + e^x)^4} \\ &= \frac{df_{\tilde{p}}(\tilde{p})}{d\tilde{p}} \left(\frac{e^x}{(1 + e^x)^2} \right)^2 + f_{\tilde{p}}(\tilde{p}) \frac{e^x}{(1 + e^x)^2} \frac{1 - e^x}{1 + e^x}.\end{aligned}$$

Using the unconditional analog of formula (35), we derive

$$\begin{aligned}v_X(x) &= \frac{df_X(x)/dx}{f_X(x)} \\ &= \frac{df_{\tilde{p}}(\tilde{p})/d\tilde{p}}{f_{\tilde{p}}(\tilde{p})} \frac{e^x}{(1 + e^x)^2} + \frac{1 - e^x}{1 + e^x} \\ &= v_{\tilde{p}}(\tilde{p})\tilde{p}(1 - \tilde{p}) + 1 - 2\tilde{p},\end{aligned}$$

where $v_{\tilde{p}}(\tilde{p}) = (df_{\tilde{p}}(\tilde{p})/d\tilde{p})/f_{\tilde{p}}(\tilde{p})$. Thus,

$$E(p|\tilde{p}) = \tilde{p} + \gamma^*(\tilde{p}(1 - \tilde{p})v_{\tilde{p}}(\tilde{p}) + 1 - 2\tilde{p}).$$

Similarly, again by Tweedie's formula,

$$Var(p/\gamma^* - 1|\tilde{p}) = Var(\eta|x) = v'_X(x) - l''_h(x) = v'_X(x) + \frac{1}{\gamma^*} \frac{e^x}{(1 + e^x)^2} = v'_X(x) + \frac{\tilde{p}(1 - \tilde{p})}{\gamma^*},$$

which implies

$$Var(p|\tilde{p}) = \gamma^* \tilde{p}(1 - \tilde{p}) + \gamma^{*2} v'_X(x).$$

Noting that

$$v'_X(x) = \tilde{p}(1 - \tilde{p})[v'_{\tilde{p}}(\tilde{p})\tilde{p}(1 - \tilde{p}) + v_{\tilde{p}}(\tilde{p})(1 - 2\tilde{p}) - 2],$$

we derive

$$Var(p|\tilde{p}) = \gamma^{*2}\tilde{p}(1-\tilde{p})[v'_{\tilde{p}}(\tilde{p})\tilde{p}(1-\tilde{p}) + v_{\tilde{p}}(\tilde{p})(1-2\tilde{p}) - 2] + \gamma^*\tilde{p}(1-\tilde{p}).$$

If we define $g^*(\tilde{p}) = \tilde{p}(1-\tilde{p})v_{\tilde{p}}(\tilde{p})$, then

$$\begin{aligned} E(p|\tilde{p}) &= \tilde{p} + \gamma^*(g(\tilde{p}) + 1 - 2\tilde{p}) \\ Var(p|\tilde{p}) &= \gamma^*\tilde{p}(1-\tilde{p}) + \gamma^{*2}\tilde{p}(1-\tilde{p})[g'(\tilde{p}) - 2]. \end{aligned}$$

This completes the proof of Theorem 2.

Finally, we establish some properties of $g^*(\tilde{p})$ and prove Corollary 1. We denote the marginal density of $\tilde{p}$ by f . First note that $g(1-\tilde{p}) = -\tilde{p}(1-\tilde{p})f'(1-\tilde{p}|p)/f(1-\tilde{p}|p)$. If $h(p)$ represents the prior density for p , then

$$f(\tilde{p}) = \int_0^1 B(\alpha, \beta)^{-1} \tilde{p}^{\alpha-1} (1-\tilde{p})^{\beta-1} h(p) dp. \quad (36)$$

Because function h is bounded, differentiation under the integral sign is justified, and hence

$$f'(\tilde{p}) = \int_0^1 B(\alpha, \beta)^{-1} \left\{ (\alpha-1)\tilde{p}^{\alpha-2}(1-\tilde{p})^{\beta-1} - (\beta-1)\tilde{p}^{\alpha-1}(1-\tilde{p})^{\beta-2} \right\} h(p) dp, \quad (37)$$

where $\alpha = p/\gamma^*$ and $\beta = (1-p)/\gamma^*$. Substituting $p^* = 1-p$ we get

$$f(1-\tilde{p}) = \int_0^1 B(\beta, \alpha)^{-1} \tilde{p}^{\alpha-1} (1-\tilde{p})^{\beta-1} h(1-p^*) dp^* = f(\tilde{p})$$

and

$$f'(1-\tilde{p}) = \int_0^1 B(\beta, \alpha)^{-1} \left\{ (\alpha-1)\tilde{p}^{\alpha-2}(1-\tilde{p})^{\beta-1} - (\beta-1)\tilde{p}^{\alpha-1}(1-\tilde{p})^{\beta-2} \right\} h(1-p^*) dp^* = f'(\tilde{p}),$$

provided $h(p) = h(1-p)$. Hence, $g^*(1-\tilde{p}) = -\tilde{p}(1-\tilde{p})f'(\tilde{p})/f(\tilde{p}) = -g^*(\tilde{p})$. By continuity of $g^*(\tilde{p})$ this result also implies $g^*(0.5) = 0$.

To complete the proof of Corollary 1, we note that under the assumption that the distribution of p_i is symmetric, the conditional expected value $E(p_i|\tilde{p}_i)$ lies on the same side of 0.5 as $\tilde{p}_i$.

C Proof of Theorem 3

As before, we denote the marginal density of $\tilde{p}$ by f . First, we derive a bound for g^* . Note that $-1 \leq \alpha - 1 \leq \frac{1}{\gamma^*}$ and, similarly, $-1 \leq \beta - 1 \leq \frac{1}{\gamma^*}$. Hence, by (36) and (37),

$$-\left(1 - \tilde{p} + \frac{\tilde{p}}{\gamma^*}\right) f(\tilde{p}) \leq \tilde{p}(1-\tilde{p})f'(\tilde{p}) \leq \left((1-\tilde{p})\frac{1}{\gamma^*} + \tilde{p}\right) f(\tilde{p}),$$

which implies

$$|g^*(\tilde{p})| \leq \frac{1}{\gamma^*}. \quad (38)$$

Next, note that

$$\lim_{\tilde{p} \rightarrow 0} \tilde{p}(1 - \tilde{p})f(\tilde{p}) = 0 \quad \text{and} \quad (39)$$

$$\lim_{\tilde{p} \rightarrow 1} \tilde{p}(1 - \tilde{p})f(\tilde{p}) = 0. \quad (40)$$

Observe that

$$\begin{aligned} R(g(\tilde{p})) &= E((g(\tilde{p}) - g^*(\tilde{p}))^2) \\ &= Eg(\tilde{p})^2 - 2E\{g(\tilde{p})g^*(\tilde{p})\} + C \\ &= Eg(\tilde{p})^2 - 2 \int_0^1 \left\{ g(\tilde{p})\tilde{p}(1 - \tilde{p}) \frac{f'(\tilde{p})}{f(\tilde{p})} \right\} f(\tilde{p})d\tilde{p} + C \\ &= Eg(\tilde{p})^2 - 2[g(\tilde{p})\tilde{p}(1 - \tilde{p})f(\tilde{p})]_0^1 + 2 \int_0^1 [g(\tilde{p})(1 - 2\tilde{p}) + \tilde{p}(1 - \tilde{p})g'(\tilde{p})] f(\tilde{p})d\tilde{p} + C \\ &= Eg(\tilde{p})^2 + 2 \int_0^1 [g(\tilde{p})(1 - 2\tilde{p}) + \tilde{p}(1 - \tilde{p})g'(\tilde{p})] f(\tilde{p})d\tilde{p} + C \end{aligned} \quad (41)$$

where C is a constant that does not depend on g , and the second to last line follows via integration by parts. Note the last line holds when g is bounded, because by (39),

$$\lim_{\tilde{p} \rightarrow 0} g(\tilde{p})\tilde{p}(1 - \tilde{p})f(\tilde{p}) = 0,$$

and by (40),

$$\lim_{\tilde{p} \rightarrow 1} g(\tilde{p})\tilde{p}(1 - \tilde{p})f(\tilde{p}) = 0.$$

In particular, due to the inequality (38), the relationship (41) holds when g is the true function g^* .

D Proof of Theorem 4

We write $\mathcal{G}_N$ for the class of all natural cubic spline functions g on $[0, 1]$ that correspond to the sequence of n knots located at the observed $\tilde{p}_i$. Given a function g , we define $s_g(\tilde{p}) = 2[g(\tilde{p})(1 - 2\tilde{p}) + \tilde{p}(1 - \tilde{p})g'(\tilde{p})]$ and $I^2(g) = \int_0^1 [g''(\tilde{p})]^2 d\tilde{p}$. We also denote $(1/n) \sum_{i=1}^n g^2(\tilde{p}_i)$ and $\int_0^1 g(\tilde{p})f^*(\tilde{p})d\tilde{p}$ by $\|g\|_n^2$ and $\|g\|^2$, respectively.

By Lemma 1 in Appendix I, there exists $g_N^* \in \mathcal{G}_N$, such that $\|g_N^* - g^*\|^2 = O_p(\lambda_n^2)$ and

$$\|\hat{g} - g_N^*\|^2 + \lambda_n^2 I^2(\hat{g}) \leq O_p(n^{-2/7} \|\hat{g} - g_N^*\|) + O_p(n^{-4/7} I(\hat{g})) + O_p(n^{-4/7} + \lambda_n^2).$$

We consider two possible cases (a) $n^{-4/7} I(\hat{g}) \leq n^{-2/7} \|\hat{g} - g_N^*\| + n^{-4/7} + \lambda_n^2$ and (b) $n^{-4/7} I(\hat{g}) > n^{-2/7} \|\hat{g} - g_N^*\| + n^{-4/7} + \lambda_n^2$.

Under (a) we have

$$\|\hat{g} - g_N^*\|^2 + \lambda_n^2 I^2(\hat{g}) \leq O_p(n^{-2/7} \|\hat{g} - g_N^*\|) + O_p(n^{-4/7} + \lambda_n^2). \quad (42)$$

It follows that $\|\hat{g} - g_N^*\| = O_p(n^{-2/7} + \lambda_n)$ and $I^2(\hat{g}) = O_p(n^{-4/7} \lambda_n^{-2} + 1)$. However, taking into account the case (a) condition, we also have $I^2(\hat{g}) = O_p(n^{4/7} \lambda_n^2 + 1)$, thus leading to $I(\hat{g}) = O_p(1)$.

Under (b) we have

$$\|\hat{g} - g_N^*\|^2 + \lambda_n^2 I^2(\hat{g}) \leq O_p\left(n^{-4/7} I(\hat{g})\right). \quad (43)$$

It follows that $I(\hat{g}) = O_p(n^{-4/7} \lambda_n^{-2})$ and $\|\hat{g} - g_N^*\| = O_p(n^{-4/7} \lambda_n^{-1})$.

Collecting all the stochastic bounds we derived, and using the fact that f^* is bounded away from zero, we deduce

$$\|\hat{g} - g_N^*\| = O_p(n^{-4/7} \lambda_n^{-1} + n^{-2/7} + \lambda_n) \quad \text{and} \quad I(\hat{g}) = O_p(1 + n^{-4/7} \lambda_n^{-2})$$

Using the bound $\|g_N^* - g^*\|^2 = O_p(\lambda_n^2)$, together with the definitions of r_n and s_n , we derive

$$\|\hat{g} - g^*\| = O_p(r_n) \quad \text{and} \quad I(\hat{g} - g^*) = O_p(1 + n^{-4/7} \lambda_n^{-2}). \quad (44)$$

Applying Lemma 10.9 in Van de Geer (2000), which builds on the interpolation inequality of Agmon (1965), we derive $\|\hat{g}' - g^{*'}\| = O_p(\sqrt{r_n s_n})$. This establishes the error bounds for $\hat{g}$ and $\hat{g}'$ with respect to the $\|\cdot\|$ norm.

To derive the corresponding results with respect to the $\|\cdot\|_n$ norm, we first apply bound (50), in which we replace g_N^* with g^* . It follows that

$$\|\hat{g} - g^*\|_n^2 - \|\hat{g} - g^*\|^2 = (\tilde{P}_n - \tilde{P})[\hat{g} - g^*]^2 = o_p\left(\|\hat{g} - g^*\|^2\right) + O_p\left(n^{-1} I^2(\hat{g} - g^*)\right),$$

where we use the notation from the proof of Lemma 1. Because bounds (44) together with the assumption $\lambda_n \gg n^{-8/21}$ imply

$$I(\hat{g} - g^*) = O\left(n^{-4/7} n^{16/21}\right) = O\left(n^{4/21}\right),$$

we can then derive

$$\|\hat{g} - g^*\|_n^2 = O\left(\|\hat{g} - g^*\|^2\right) + O_p\left(n^{-13/21}\right).$$

Because $r_n \geq n^{-2/7}$, we have $r_n^2 \geq n^{-13/21}$. Consequently, $\|\hat{g} - g^*\|_n^2 = O(r_n^2)$, which establishes the analog of the first bound in (44) for the $\|\cdot\|_n$ norm.

It is only left to derive $\|\hat{g}' - g^{*'}\|_n = O_p(\sqrt{r_n s_n})$. Applying Lemma 17 in Meier et al. (2009), in conjunction with Corollary 5 from the same paper, in which we take $\gamma^* = 2/3$ and $\lambda = n^{-3/14}$, we derive

$$(\tilde{P}_n - \tilde{P})[\hat{g}' - g^{*'}]^2 = O_p\left(n^{-5/14} \left[\|\hat{g}' - g^{*'}\|^2 + n^{-2/7} I^2(\hat{g}' - g^{*'})\right]\right).$$

Consequently,

$$\|\hat{g}' - g^{*'}\|_n^2 = O_p\left(\|\hat{g}' - g^{*'}\|^2\right) + O_p\left(n^{-9/14} I^2(\hat{g}' - g^{*'})\right).$$

Taking into account bound (44), the definition of s_n , the assumption $\lambda_n \gg n^{-8/21}$ and the inequality $r_n \geq n^{-2/7}$, we derive

$$n^{-9/14} I^2(\hat{g}' - g^{*'}) = O_p\left(n^{-9/14} s_n^2\right) = O_p\left(n^{-19/42} s_n\right) = O_p\left(r_n s_n\right).$$

Thus, $\|\hat{g}' - g^{*'}\|_n^2 = O_p(\|\hat{g}' - g^{*'}\|^2 + r_n s_n) = O_p(r_n s_n)$, which completes the proof.

E Proof of Theorem 5

We will take advantage of the results in Theorem 6, which are established independently from Theorem 5. We will focus on proving the results involving integrals, because the results for the averages follow by an analogous argument with minimal modifications.

We start by establishing consistency of $\hat{p}$. Fixing an arbitrary positive $\tilde{\epsilon}$, identifying a positive ϵ for which $\tilde{P}(0, \epsilon) + \tilde{P}(1 - \epsilon, 1) \leq \tilde{\epsilon}/2$, and noting that $\hat{p}$ and p_0 fall in $[0, 1]$ for every $\tilde{p}$, we derive

$$\|\hat{p} - p_0\|^2 \leq \tilde{\epsilon}/2 + \int_{\epsilon}^{1-\epsilon} |\hat{p}(\tilde{p}) - p_0(\tilde{p})|^2 f^*(\tilde{p}) d\tilde{p}.$$

By Theorem 6, the second term on the right-hand side of the above display is $o_p(1)$. Consequently,

$$P\left(\|\hat{p} - p_0\|^2 > \tilde{\epsilon}\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

As the above statement holds for every fixed positive $\tilde{\epsilon}$, we have established that $\|\hat{p} - p_0\| = o_p(1)$.

We now focus on showing consistency for $\hat{W}$. Note that

$$[\hat{\mu}^2(\tilde{p}) + \hat{\sigma}^2(\tilde{p})]^2 / \hat{\mu}^2(\tilde{p}) \geq [2\hat{\mu}(\tilde{p})\hat{\sigma}(\tilde{p})]^2 / [\hat{\mu}^2(\tilde{p})] = 4\hat{\sigma}^2(\tilde{p}).$$

Thus, the definition of $\hat{p}$ implies $\hat{p}^2(\tilde{p}) \geq \hat{\sigma}^2(\tilde{p}) \wedge 0.25$, and also $\hat{p}(\tilde{p}) \geq \hat{\mu}(\tilde{p}) \wedge 0.5$, for every $\tilde{p} \in (0, 1)$. Writing p for the true probability corresponding to the observed $\tilde{p}$, we then derive

$$\begin{aligned} \widehat{W}(\tilde{p}) &= \frac{E_p[(p - \hat{p}(\tilde{p}))^2 | \tilde{p}]}{\hat{p}^2(\tilde{p})} \leq \frac{\sigma^2(\tilde{p})}{\hat{p}^2(\tilde{p})} + \frac{[\hat{p}(\tilde{p}) - \mu(\tilde{p})]^2}{\hat{p}^2(\tilde{p})} \\ &\leq \frac{|\hat{\sigma}^2(\tilde{p}) - \sigma^2(\tilde{p})|}{4\hat{\sigma}^2(\tilde{p})} + \frac{[\hat{\mu}(\tilde{p}) - \mu(\tilde{p})]^2}{\hat{\sigma}^2(\tilde{p})} + 7. \end{aligned} \quad (45)$$

By Theorem 6, we have $\|\hat{\sigma}^2 - \sigma^2\| = O_p(\sqrt{r_n s_n}) = o_p(1)$ and $\|\hat{\mu} - \mu\| = O_p(\sqrt{r_n s_n}) = o_p(1)$. Fix an arbitrary positive ϵ and define $A_\epsilon = (0, \epsilon) \cup (1 - \epsilon, 1)$. Applying the Cauchy-Schwarz inequality, and using the imposed technical modification of the ECAP approach to bound $\hat{\sigma}^2$ below, we derive

$$\int_{A_\epsilon} \frac{|\hat{\sigma}^2(\tilde{p}) - \sigma^2(\tilde{p})|}{\hat{\sigma}^2(\tilde{p})} f^*(\tilde{p}) d\tilde{p} \leq [\tilde{P}A_\epsilon]^{1/2} \frac{\|\hat{\sigma}^2 - \sigma^2\|}{c\sqrt{r_n s_n}} = [\tilde{P}A_\epsilon]^{1/2} O_p(1) = O_p(\epsilon^{1/2}).$$

Similarly, we derive

$$\int_{A_\epsilon} \frac{|\hat{\mu}(\tilde{p}) - \mu(\tilde{p})|^2}{\hat{\sigma}^2(\tilde{p})} f^*(\tilde{p}) d\tilde{p} \leq \int_{A_\epsilon} \frac{|\hat{\mu}(\tilde{p}) - \mu(\tilde{p})|}{\hat{\sigma}^2(\tilde{p})} f^*(\tilde{p}) d\tilde{p} = O_p(\epsilon^{1/2}).$$

Note that $|W_0(\tilde{p})| \leq 1$ for every $\tilde{p}$. Thus, combining the bounds for the terms in (45) with the corresponding bound for $|\hat{W} - W_0|$ in Theorem 6, we derive

$$\int_0^1 |\widehat{W}(\tilde{p}) - W_0(\tilde{p})| f^*(\tilde{p}) d\tilde{p} = O_p(\epsilon^{1/2}) + o_p(1).$$

As this bound holds for every positive ϵ , we deduce that $\int_0^1 |\widehat{W}(\tilde{p}) - W_0(\tilde{p})| f^*(\tilde{p}) d\tilde{p} = o_p(1)$.

F Proof of Theorem 6

We build on the results of Theorem 4 to derive the rate of convergence for $\hat{\mu}$ and $\widehat{W}$ for a fixed positive ϵ . Continuity and positivity of $\mu(\tilde{p})$ and $p_0(\tilde{p})$ imply that both functions are bounded away from zero on the interval $[\epsilon, 1 - \epsilon]$. Applying Lemma 10.9 in Van de Geer (2000), we derive $\|\hat{g} - g^*\|_\infty = O_p(r_n^{3/4} s_n^{1/4})$. Because $n^{-8/21} \ll \lambda_n \ll 1$, we have $\|\hat{g} - g^*\|_\infty = o_p(1)$, which implies $\sup_{[\epsilon, 1 - \epsilon]} |\hat{\mu}(\tilde{p}) - \mu(\tilde{p})| = o_p(1)$. Also note that $\hat{p}(\tilde{p}) \geq \hat{\mu}(\tilde{p})$ for all $\tilde{p}$. Consequently, there exists an event with probability tending to one, on which random functions $\hat{p}(\tilde{p})$ and $\hat{\mu}(\tilde{p})$ are bounded away from zero on the interval $[\epsilon, 1 - \epsilon]$. The stated error bounds for $\hat{p}$ then follow directly from this observation and the error bounds for $\hat{g}$ and $\hat{g}'$ in Theorem 4.

For the remainder of the proof we restrict our attention to the event A (whose probability tends to one), on which functions p_0 and $\hat{p}$ are both bounded away from zero on $[\epsilon, 1 - \epsilon]$. We write p for the true probability corresponding to the observed $\tilde{p}$, define $G(q) = E[(p - q)^2 / q^2 | \tilde{p}]$ and note that $G'(q) = 2(q - p^*)E[p|\tilde{p}] / q^3$. Let p^* be the minimizer of G , given by $p^* = E[p|\tilde{p}] + \text{Var}[p|\tilde{p}] / E[p|\tilde{p}]$. Denote by $\hat{p}^*$ our estimator of p^* , which is obtained by replacing the conditional expected value and variance in the above formula by their ECAP estimators. While p^* and $\hat{p}^*$ depend on $\tilde{p}$, we will generally suppress this dependence in the notation for simplicity. Note that for $\tilde{p} \in [\epsilon, 1 - \epsilon]$, functions p^* and $\hat{p}^*$ are both bounded away from zero on the set A.

Fix an arbitrary $\tilde{p} \leq 0.5$. Define events $A_1 = A \cap \{p^* \leq 0.5, \hat{p}^* \leq 0.5\}$, $A_2 = A \cap \{p^* > 0.5, \hat{p}^* \leq 0.5\}$, $A_3 = A \cap \{p^* \leq 0.5, \hat{p}^* > 0.5\}$ and $A_4 = A \cap \{p^* > 0.5, \hat{p}^* > 0.5\}$. Note that A_4 implies $\hat{p} = p_0 = 0.5$. Writing Taylor expansions for function G near p^* and 0.5, we derive the following bounds, which hold for some universal constant c that depends only on ϵ :

$$\begin{aligned} |W_0(\tilde{p}) - \widehat{W}(\tilde{p})| 1_{\{A\}} &= |G(p^*) - G(\hat{p}^*)| 1_{\{A_1\}} + |G(0.5) - G(\hat{p}^*)| 1_{\{A_2\}} + |G(p^*) - G(0.5)| 1_{\{A_3\}} \\ &\leq c(p^* - \hat{p}^*)^2 1_{\{A_1\}} + c(0.5 - \hat{p}^*)^2 1_{\{A_2\}} + c(p^* - 0.5)^2 1_{\{A_3\}} \\ &\leq c(p^* - \hat{p}^*)^2. \end{aligned}$$

Analogous arguments derive the above bound for $\tilde{p} > 0.5$. The rate of convergence for $\widehat{W}$ then follows directly from the error bounds for $\hat{g}$ and $\hat{g}'$ in Theorem 4.

G Proof of Theorem 7

Throughout the proof we drop the subscript i for the simplicity of notation. First note that the derivations in the proof of Theorem 2 also give $E(\gamma^* \alpha | \tilde{p}) = \mu$ and $\text{Var}(\gamma^* \alpha | \tilde{p}) = \sigma^2$, where μ and σ^2 are respectively defined in (8) and (9). These identities hold for both the unbiased and biased versions of the model. The only difference is in how $\gamma^* \alpha$ relates to p . Note that

$$\begin{aligned} E(p|\tilde{p}) &= E(h(\gamma^* \alpha) | \tilde{p}) = (1 - 0.5\theta)E(\gamma^* \alpha | \tilde{p}) - \theta[E(\gamma^{*3} \alpha^3 | \tilde{p}) - 1.5E(\gamma^{*2} \alpha^2 | \tilde{p})] \\ &= (1 - 0.5\theta)\mu - \theta[s_3 + 3\mu\sigma^2 + \mu^3 - 1.5\sigma^2 - 1.5\mu^2], \end{aligned} \tag{46}$$

where we use s_k to denote the k -th conditional central moment of $\gamma^* \alpha$ given $\tilde{p}$. By Lemma 2 in Appendix I, the s_3 term in (46) is $O(\gamma^{*3/2})$, which leads to the stated approximation for $E(p|\tilde{p})$. We also have

$$\text{Var}(p|\tilde{p}) = \text{Var}(h(\gamma^* \alpha) | \tilde{p}) = (1 - 0.5\theta)^2 \sigma^2 + \theta a,$$

where $a = \theta Var(\gamma^{*3}\alpha^3 - 1.5\gamma^{*2}\alpha^2|\tilde{p}) - (1 - 0.5\theta)Cov(\gamma^*\alpha, \gamma^{*3}\alpha^3 - 1.5\gamma^{*2}\alpha^2|\tilde{p})$. It is only left to show that $a = O(\gamma^{*3/2})$. A routine calculation yields

$$a = \sigma_i^2 [3\mu(1 - \mu)(3\theta\mu(1 - \mu) - 0.5\theta + 1)] + O\left(\sum_{k=3}^6 [\sigma^k + s_k]\right).$$

By Lemma 2, the remainder term is $O(\gamma^{*3/2})$, which completes the proof.

H Proof of Theorem 8

We use the notation from the proof of Theorem 2. In particular, we omit the subscript i throughout most of the proof, for the simplicity of the exposition. We represent $\tilde{p}$ as $\sum_{k=1}^K I_{\{\mathcal{I}=k\}}\xi_k$, where $\xi_k|p \sim Beta(\alpha_k, \beta_k)$, $\alpha_k = c_k p / \gamma^*$, $\beta_k = (1 - c_k p) / \gamma^*$, and $\mathcal{I}$ is a discrete random variable independent of p and ξ_k , whose probability distribution is given by $P(\mathcal{I} = k) = w_k$ for $k = 1, \dots, K$.

Note that

$$f_{\xi_k}(\tilde{p}|p) = B\left(\frac{c_k p}{\gamma^*}, \frac{1 - c_k p}{\gamma^*}\right)^{-1} \tilde{p}^{\frac{c_k p}{\gamma^*} - 1} (1 - \tilde{p})^{\frac{1 - c_k p}{\gamma^*} - 1}.$$

Hence, writing B for $B(\frac{c_k p}{\gamma^*}, \frac{1 - c_k p}{\gamma^*})$, we derive

$$\begin{aligned} \log(f_{\xi_k}(\tilde{p}|p)) &= -\log B + \left(\frac{c_k p}{\gamma^*} - 1\right) \log \tilde{p} + \left(\frac{1 - c_k p}{\gamma^*} - 1\right) \log(1 - \tilde{p}) \\ &= -\log B + \frac{c_k p}{\gamma^*} \log \tilde{p} - \log \tilde{p} + \frac{1 - c_k p}{\gamma^*} \log(1 - \tilde{p}) - \log(1 - \tilde{p}) \\ &= -\log B + p \frac{c_k}{\gamma^*} \log\left(\frac{\tilde{p}}{1 - \tilde{p}}\right) - \log \tilde{p} + \frac{1}{\gamma^*} \log(1 - \tilde{p}) - \log(1 - \tilde{p}) \\ &= -\log B + \eta x - \log \tilde{p} + \left(\frac{1 - \gamma^*}{\gamma^*}\right) \log(1 - \tilde{p}), \end{aligned}$$

where we've defined $\eta = p \frac{c_1}{\gamma^*}$ and $x = \log(\tilde{p}/[1 - \tilde{p}])$. Repeating the derivations in the proof of Theorem 2 directly below display (34), we derive

$$\begin{aligned} E(p|\xi_k = \tilde{p}) &= \frac{1}{c_k} \left[\gamma^* (g^*(\tilde{p}) + 1 - 2\tilde{p}) + \tilde{p} \right] \\ Var(p|\xi_k = \tilde{p}) &= \frac{1}{c_k^2} \left[\gamma^{*2} (\tilde{p}(1 - \tilde{p})(g^{*'}(\tilde{p}) - 2) + \gamma^* \tilde{p}(1 - \tilde{p})) \right]. \end{aligned}$$

Consequently,

$$\begin{aligned} E(p|\tilde{p}, \mathcal{I} = k) &= E(p|\xi_k = \tilde{p}) = \frac{1}{c_k} \left[\gamma^* (g^*(\tilde{p}) + 1 - 2\tilde{p}) + \tilde{p} \right] \\ Var(p|\tilde{p}, \mathcal{I} = k) &= Var(p|\xi_k = \tilde{p}) = \frac{1}{c_k^2} \left[\gamma^{*2} (\tilde{p}(1 - \tilde{p})(g^{*'}(\tilde{p}) - 2) + \gamma^* \tilde{p}(1 - \tilde{p})) \right] \end{aligned} \tag{47}$$

Applying the law of total probability and using the fact that $\mathcal{I}$ and $\tilde{p}$ are independent, we derive

$$E(p|\tilde{p}) = \sum_{k=1}^K w_k E(p|\tilde{p}, \mathcal{I} = k) = \sum_{k=1}^K \frac{w_k}{c_k} \left[\gamma \left(g^*(\tilde{p}) + 1 - 2\tilde{p} \right) + \tilde{p} \right].$$

By the law of total variance, we also have

$$\begin{aligned} \text{Var}(p|\tilde{p}) &= \sum_{k=1}^K w_k \text{Var}(p|\tilde{p}, \mathcal{I} = k) + \sum_{k=1}^K w_k E^2(p|\tilde{p}, \mathcal{I} = k) - \left[\sum_{k=1}^K w_k E(p|\tilde{p}, \mathcal{I} = k) \right]^2 \\ &= \sum_{k=1}^K \frac{w_k}{c_k^2} \left[\gamma^{*2} \tilde{p}(1 - \tilde{p})(g^{*'}(\tilde{p}) - 2) + \gamma^* \tilde{p}(1 - \tilde{p}) \right] + \left[\gamma^*(g^*(\tilde{p}) + 1 - 2\tilde{p}) + \tilde{p} \right]^2 \left[\sum_{k=1}^K \frac{w_k}{c_k^2} - \left(\sum_{k=1}^K \frac{w_k}{c_k} \right)^2 \right]. \end{aligned}$$

To complete the proof, we use formulas

$$\mu_i = \tilde{p}_i + \gamma^* [g^*(\tilde{p}_i) + 1 - 2\tilde{p}_i] \quad \text{and} \quad \sigma_i^2 = \gamma^* \tilde{p}_i(1 - \tilde{p}_i) + \gamma^{*2} \tilde{p}_i(1 - \tilde{p}_i) [g^{*'}(\tilde{p}_i) - 2]$$

to rewrite the above expressions as $E(p_i|\tilde{p}_i) = \mu_i \sum_{k=1}^K w_k/c_k$ and

$$\text{Var}(p_i|\tilde{p}_i) = \sum_{k=1}^K \frac{w_k}{c_k^2} \sigma_i^2 + \mu_i^2 \sum_{k=1}^K \frac{w_k}{c_k^2} - \mu_i^2 \left(\sum_{k=1}^K \frac{w_k}{c_k} \right)^2 = (\sigma_i^2 + \mu_i^2) \sum_{k=1}^K \frac{w_k}{c_k^2} - \mu_i^2 \left(\sum_{k=1}^K \frac{w_k}{c_k} \right)^2.$$

I Supplementary Results

Lemma 1 *Under the conditions of Theorem 4, there exists a function $g_N^* \in \mathcal{G}_N$, such that $\|g_N^* - g^*\|^2 = O_p(\lambda_n^2)$ and*

$$\|\hat{g} - g_N^*\|^2 + \lambda_n^2 I^2(\hat{g}) \leq O_p\left(n^{-2/7} \|\hat{g} - g_N^*\|\right) + O_p\left(n^{-4/7} I(\hat{g})\right) + O_p\left(n^{-4/7} + \lambda_n^2\right).$$

Proof of Lemma 1. We will use the empirical process theory notation and write $\tilde{P}_n g$ and $\tilde{P} g$ for $(1/n) \sum_{i=1}^n g(\tilde{p}_i)$ and $\int_0^1 g(\tilde{p}) f^*(\tilde{p}) d\tilde{p}$, respectively. Using the new notation, criterion (14) can be written as follows:

$$Q_n(g) = \tilde{P}_n g^2 + \tilde{P}_n s_g + \lambda_n^2 I^2(g).$$

As we showed in the proof of Theorem 3, equality $\tilde{P} g^2 + \tilde{P} s_g = \|g - g^*\|^2$ holds for every candidate function $g \in \mathcal{G}_N$. Consequently,

$$Q_n(g) = \|g - g^*\|^2 + (\tilde{P}_n - \tilde{P})g^2 + (\tilde{P}_n - \tilde{P})s_g + \lambda_n^2 I^2(g).$$

Let g_N^* be a function in $\mathcal{G}_N$ that interpolates g^* at points $\{0, \tilde{p}_1, \dots, \tilde{p}_n, 1\}$, with two additional constraints: $g_N^{*'}(0) = g^{*'}(0)$ and $g_N^{*'}(1) = g^{*'}(1)$. A standard partial integration argument (similar to that in Green and Silverman, 1993, for example) shows that $I(g_N^*) \leq I(g^*)$, which also implies that $g_N^{*'}$ is uniformly bounded. Furthermore, we have $\|g_N^* - g^*\|_\infty = O_p(\log(n)/n)$ by the maximum spacing results for the uniform distribution (Shorack and Wellner, 2009, for example), the boundedness away from zero assumption on f^* and the boundedness of $g_N^{*'}$. Consequently, $\|g_N^* - g^*\|^2 = O_p(\lambda_n^2)$.

Because $Q_n(\hat{g}) \leq Q_n(g_N^*)$, we then have

$$\|\hat{g} - g_N^*\|^2 + \lambda_n^2 I^2(\hat{g}) \leq (\tilde{P}_n - \tilde{P})[g_N^{*2} - \hat{g}^2] + (\tilde{P}_n - \tilde{P})[s_{g_N^*} - s_{\hat{g}}] + \lambda_n^2 [I^2(g_N^*) + 1]. \quad (48)$$

Note that

$$(\tilde{P}_n - \tilde{P})[g_N^{*2} - \hat{g}^2] = -(\tilde{P}_n - \tilde{P})[\hat{g} - g_N^*]^2 - (\tilde{P}_n - \tilde{P})g_N^*[\hat{g} - g_N^*]. \quad (49)$$

Applying Lemma 17 in Meier et al. (2009), in conjunction with Corollary 5 from the same paper, in which we take $\gamma = 2/5$ and $\lambda = n^{-1/2}$, we derive

$$(\tilde{P}_n - \tilde{P})[g_N^* - \hat{g}]^2 = O_p\left(n^{-1/5}\|\hat{g} - g_N^*\|^2\right) + O_p\left(n^{-1}I^2(\hat{g} - g_N^*)\right). \quad (50)$$

Applying Corollary 5 in Meier et al. (2009) with the same γ and λ yields

$$(\tilde{P}_n - \tilde{P})g_N^*[\hat{g} - g_N^*] = O_p\left(n^{-2/5}\sqrt{\|g_N^*[\hat{g} - g_N^*]\|^2 + n^{-4/5}I^2(g_N^*[\hat{g} - g_N^*])}\right).$$

Using Lemma 10.9 in Van de Geer (2000) to express the L_2 norm of the first derivative in terms of the norms of the second derivative and the original function, we derive

$$(\tilde{P}_n - \tilde{P})g_N^*[\hat{g} - g_N^*] = O_p\left(n^{-2/5}\|\hat{g} - g_N^*\|\right) + O_p\left(n^{-4/5}I(\hat{g} - g_N^*)\right). \quad (51)$$

Applying Corollary 5 in Meier et al. (2009) with $\gamma = 2/3, \lambda = n^{-3/14}$ and using Lemma 10.9 in Van de Geer (2000) again, we derive

$$(\tilde{P}_n - \tilde{P})[s_{g_N^*} - s_{\hat{g}}] = O_p\left(n^{-3/7}\|s_{g_N^*} - s_{\hat{g}}\|\right) + O_p\left(n^{-4/7}I(\hat{g} - g_N^*)\right).$$

Hence, by Lemma 10.9 in Van de Geer (2000),

$$(\tilde{P}_n - \tilde{P})[s_{g_N^*} - s_{\hat{g}}] = O_p\left(n^{-3/7}\|\hat{g} - g_N^*\|^{1/2}I^{1/2}(\hat{g} - g_N^*)\right) + O_p\left(n^{-4/7}I(\hat{g} - g_N^*)\right),$$

which leads to

$$(\tilde{P}_n - \tilde{P})[s_{g_N^*} - s_{\hat{g}}] = O_p\left(n^{-2/7}\|\hat{g} - g_N^*\|\right) + O_p\left(n^{-4/7}I(\hat{g})\right) + O_p\left(n^{-4/7}I(g_N^*)\right) \quad (52)$$

Combining (48)-(52), and noting the imposed assumptions on λ_n , we arrive at

$$\|\hat{g} - g_N^*\|^2 + \lambda_n^2 I^2(\hat{g}) \leq O_p\left(n^{-2/7}\|\hat{g} - g_N^*\|\right) + O_p\left(n^{-4/7}I(\hat{g})\right) + O_p\left(n^{-4/7} + \lambda_n^2\right),$$

which completes the proof of Lemma 1.

Lemma 2 *Under the conditions of Theorem 7,*

$$\sigma^2 = O(\gamma^*) \quad \text{and} \quad s_k = O(\gamma^{*3/2}), \quad \text{for } k \geq 3.$$

Proof of Lemma 2. We first show that $E(t - \tilde{p}|\tilde{p}) = O(\sqrt{\gamma^*})$ as γ^* tends to zero, where we write t for the quantity $\alpha\gamma^* = h_\theta^{-1}(p)$. This result will be useful for establishing the stated bound for s_k .

Throughout the proof we use expression $\gtrsim$ to denote inequality $\geq$ up to a multiplicative factor equal to a positive constant that does not depend on γ^* . We use an analogous agreement for the $\lesssim$ expression. We write $f_c(\tilde{p})$ for the conditional density of $\tilde{p}$ given $t = c$, write $f(\tilde{p})$ for the marginal density of $\tilde{p}$, and write $m_\theta(t)$ for the marginal density of t . In the new notation, we have

$$E(|t - \tilde{p}| | \tilde{p}) = \int_0^1 |t - \tilde{p}| f_t(\tilde{p}) m_\theta(t) [f(\tilde{p})]^{-1} dt.$$

Using Stirling's approximation for the Gamma function, $\Gamma(x) = e^{-x} x^{x-1/2} (2\pi)^{1/2} [1 + O(1/x)]$, and applying the bound $x\Gamma(x) = O(1)$ when t is close to zero or one, we derive the following bounds as τ tends to infinity:

$$\begin{aligned} \sqrt{\tau} E(|t - \tilde{p}| | \tilde{p}) &= \int_0^1 \sqrt{\tau} |t - \tilde{p}| \frac{\Gamma(\tau)}{\Gamma(t\tau)\Gamma(q\tau)} \tilde{p}^{t\tau-1} \tilde{q}^{q\tau-1} m_\theta(t) [f(\tilde{p})]^{-1} dt \\ &\lesssim \int_0^1 \sqrt{\tau} |t - \tilde{p}| \frac{1}{\sqrt{2\pi}} [\tilde{p}/t]^{t\tau} (\tilde{q}/q)^{q\tau} \sqrt{tq\tau} m_\theta(t) [f(\tilde{p})]^{-1} [\tilde{p}\tilde{q}]^{-1} dt \\ &\lesssim \int_0^1 \sqrt{\tau} |t - \tilde{p}| e^{-\frac{\tau(t-\tilde{p})^2}{18}} \sqrt{\tau} dt. \end{aligned}$$

Implementing a change of variable, $v = \sqrt{\tau}(t - \tilde{p})$, we derive

$$\sqrt{\tau} |E(t - \tilde{p} | \tilde{p})| \lesssim \int_{\mathbb{R}} |v| e^{-v^2/18} dv = O(1).$$

Consequently, $E(t - \tilde{p} | \tilde{p}) = O(1/\sqrt{\tau}) = O(\sqrt{\gamma^*})$.

We now bound $E([t - \tilde{p}]^2 | \tilde{p})$ using a similar argument. Following the arguments in the derivations above, we arrive at

$$\tau E([t - \tilde{p}]^2 | \tilde{p}) \lesssim \int_0^1 \tau (t - \tilde{p})^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau(t-\tilde{p})^2}{18}} \sqrt{\tau} m_\theta(t) dt.$$

Implementing a change of variable, $v = \sqrt{\tau}(t - \tilde{p})$, we conclude that

$$\tau E([t - \tilde{p}]^2 | \tilde{p}) \lesssim \int_{\mathbb{R}} v^2 e^{-v^2/18} dv = O(1).$$

Thus, we have established

$$E([t - \tilde{p}]^k | \tilde{p}) = O(\gamma^{*k/2}), \quad \text{for } k \in \{1, 2\}. \quad (53)$$

Analogous arguments lead to bounds $E([t - \tilde{p}]^k | \tilde{p}) = O(\gamma^{*3/2})$ for $k \geq 3$. We complete the proof

of the lemma by noting that

$$s_k = O\left(E\left([t - \tilde{p}]^k \middle| \tilde{p}\right)\right) + O(\gamma^{*3/2}), \quad \text{for } k \geq 2.$$

When $k = 2$ the above approximation follows from $\sigma^2 \leq E([t - \tilde{p}]^2 | \tilde{p})$, and when $k = 3$ it follows from (53) and

$$\begin{aligned} s_3 &= E\left([t - \tilde{p}]^3 \middle| \tilde{p}\right) + 3E\left([t - \tilde{p}]^2 \middle| \tilde{p}\right)E\left(\tilde{p} - t \middle| \tilde{p}\right) + 3E\left(t - \tilde{p} \middle| \tilde{p}\right)E^2\left(\tilde{p} - t \middle| \tilde{p}\right) + E^3\left(\tilde{p} - t \middle| \tilde{p}\right) \\ &= E\left([t - \tilde{p}]^3 \middle| \tilde{p}\right) + O(\gamma^{*3/2}). \end{aligned}$$

The derivations for $k \geq 4$ are analogous.

References

- Agmon, S. (1965). Lectures on elliptic boundary value problems. Van Nostrand, Princeton, NJ.
- Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association* 106(496), 1602–1614.
- Green, P. J. and B. W. Silverman (1993). Nonparametric regression and generalized linear models: a roughness penalty approach. Chapman and Hall/CRC.
- Meier, L., S. Van de Geer, and P. Bühlmann (2009). High-dimensional additive modeling. *The Annals of Statistics* 37(6B), 3779–3821.
- Shorack, G. R. and J. A. Wellner (2009). Empirical processes with applications to statistics. SIAM.
- Van de Geer, S. (2000). Empirical Processes in M-Estimation. Cambridge University Press, Cambridge.