Online Supplement to "Risk Estimation via Regression"

Mark Broadie
Graduate School of Business
Columbia University
email: mnb2@columbia.edu

Yiping Du
Industrial Engineering and Operations Research
Columbia University
email: yd2166@columbia.edu

Ciamac C. Moallemi
Graduate School of Business
Columbia University
email: ciamac@gsb.columbia.edu

July 1, 2015

A. The Weighted Regression Method: Practical Implementation

In this section, we describe the practically implementable variation of the weighted regression method that is demonstrated in the numerical results of Section 6.3. In this two-pass procedure, we will first approximate $L(\omega)$ with an *unweighted* regression and obtain $\hat{r}$; from $\hat{r}$, we will construct an approximation to the weight function $h_{\text{opt}}(\cdot)$ of (37) to be used subsequently in a weighted regression.

Following Remark 6, let $m = 1$ and $n = k$. As $k \to \infty$, Lemma 1 and (11) imply that, for a fixed scenario ω,

$$\sqrt{k}\left(\Phi(\omega)\hat{r} + M(\omega) - L(\omega)\right) \xrightarrow{\mathsf{d}} N\big(\mathbf{0}, \Phi(\omega)\left(\Sigma_M + \Sigma_v\right)\Phi(\omega)^\top\big). \tag{A.1}$$

Equation (A.1) suggests that when k is large, it is reasonable to approximate the distribution of the portfolio loss $L(\omega)$ given $\vec{\omega}$ and $\vec{\zeta}$ by a normal distribution with mean $\Phi(\omega)\hat{r} + M(\omega)$ and variance $\Phi(\omega)\left(\Sigma_M + \Sigma_v\right)\Phi(\omega)^\top/k$. Given this approximation, we can approximate $h_{\text{opt}}(\cdot)$ with its posterior mean:

$$\mathsf{E}\left[h_{\text{opt}}(\omega)\middle|\vec{\omega}, \vec{\zeta}\right] \approx N\left(\frac{\sqrt{k}\left(\Phi(\omega)\hat{r} + M(\omega) - c\right)}{\sqrt{\Phi(\omega)\left(\Sigma_M + \Sigma_v\right)\Phi(\omega)^\top}}\right). \tag{A.2}$$

(Notice that $\hat{r}$ depends on $\vec{\omega}$ and $\vec{\zeta}$.) If the basis functions are well-chosen, the model error $M(\omega)$ is small in magnitude relative to the unweighted regression approximation $\Phi(\omega)\hat{r}$. Further, in practice, we observe that the denominator of the argument in the right-hand side of (A.2) does not vary by more than one order of magnitude. These observations suggest an overall approximation for the globally optimal weight function $h_{\text{opt}}(\cdot)$:

$$h(\omega) = N\left(\frac{\sqrt{k}\left(\Phi(\omega)\hat{r} - c\right)}{\Gamma}\right), \tag{A.3}$$

for some constant $\Gamma > 0$. Using the weight function (A.3), we have a two-pass weighted regression procedure:

- the first pass is an unweighted regression and provides $\Phi(\omega)\hat{r}$, an approximation of $L(\omega)$ used to determine weights;
- based on $\Phi(\omega)\hat{r}$, the second pass is a weighted regression with weights defined by (A.3).

Notice that this two-pass weighted regression procedure assigns weights with a weight function inspired by $h_{\text{opt}}(\cdot)$, and it does not depend on information of $L(\omega)$ or any quantities that are unknown in practice. The two-pass weighted regression method is compared with the unweighted regression method in Section 6.3.

B. Proofs for Section 4

This section presents the proofs of the results in Section 4, where the asymptotic analysis of the regression estimator has been established. First, we show the following lemma, which establishes the existence and uniqueness of the optimal solutions to (10) and (13).

Lemma 3. *Given Assumptions A1 and A2, the function $g(\cdot)$ is strictly convex over $\mathbb{R}^d$, and the function $(1/n)\sum_{i=1}^n G(\cdot,\omega^{(i)},\zeta^{(i)})$ is strictly convex over $\mathbb{R}^d$ almost surely.*

Proof. As a function of r, the Hessian matrix of $(1/n)\sum_{i=1}^n G(\cdot,\omega^{(i)},\zeta^{(i)})$ is

$$\begin{aligned}\nabla^2\left(\frac{1}{n}\sum_{i=1}^n G\left(r,\omega^{(i)},\zeta^{(i)}\right)\right) &= \nabla^2\left(\frac{1}{n}\sum_{i=1}^n\left(\hat{L}(\omega^{(i)},\zeta^{(i)})-\Phi(\omega^{(i)})r\right)^2\right)\\ &= \frac{2}{n}\sum_{i=1}^n \Phi(\omega^{(i)})^\top\Phi(\omega^{(i)}),\end{aligned}$$

which is positive semidefinite almost surely under Assumption A2. Therefore, $\frac{1}{n}\sum_{i=1}^n G(\cdot,\omega^{(i)},\zeta^{(i)})$ is strictly convex almost surely.

The Hessian matrix of $g(r,\omega)$ is

$$\nabla^2\left(g(r,\omega)\right) = \nabla^2\left(\mathsf{E}\left[\left(\hat{L}(\omega,\zeta)-\Phi(\omega)r\right)^2\right]\right) = \nabla^2\left(r\mathsf{E}\left[\Phi(\omega)^\top\Phi(\omega)\right]r\right) = 2I_d.$$

Therefore, $g(r)$ is strictly convex.

■

Given Assumptions A1 and A2, according to Lemma 3, the optimal solutions to (10) and (13) exist and are unique almost surely.

We use the following notation: $\vec{\omega}$ denotes the n outer stage scenarios,

$$\vec{\omega} \triangleq \left(\omega^{(1)},\ldots,\omega^{(n)}\right)^\top,$$

$\vec{\zeta}$ denotes the inner stage uncertainty,

$$\vec{\zeta} \triangleq \left(\zeta^{(1)}, \ldots, \zeta^{(n)}\right)^\top,$$

$\Phi(\vec{\omega})$ is an n-by-d matrix,

$$\Phi(\vec{\omega}) \triangleq \begin{pmatrix} \phi_1(\omega^{(1)}) & \cdots & \phi_d(\omega^{(1)}) \\ \vdots & \ddots & \vdots \\ \phi_1(\omega^{(n)}) & \cdots & \phi_d(\omega^{(n)}) \end{pmatrix},$$

and $\hat{L}\left(\vec{\omega}, \vec{\zeta}\right)$ is an n-by-1 column vector,

$$\hat{L}\left(\vec{\omega}, \vec{\zeta}\right) \triangleq \left(\hat{L}(\omega^{(1)}, \zeta^{(1)}), \ldots, \hat{L}(\omega^{(n)}, \zeta^{(n)})\right)^\top.$$

From regression theory, the unique optimal solution to (13) takes the form

$$\hat{r} = \left(\Phi(\vec{\omega})^\top \Phi(\vec{\omega})\right)^{-1} \Phi(\vec{\omega})^\top \hat{L}\left(\vec{\omega}, \vec{\zeta}\right). \tag{B.1}$$

Then we estimate the risk measure by

$$\hat{\alpha}_{\mathsf{REG}(m,n)} \triangleq \mathsf{E}\left[f\left(\Phi(\omega)\hat{r}\right) \middle| \vec{\omega}, \vec{\zeta}\right].$$

From Assumption A1, the disturbance term $\varepsilon(\omega, \zeta)$ of (12) satisfies

$$\mathsf{E}\left[\varepsilon(\omega, \zeta) \middle| \omega\right] = 0, \tag{B.2}$$

and

$$\mathrm{Var}\left(\varepsilon(\omega, \zeta) \middle| \omega\right) = \frac{v(\omega)}{m}. \tag{B.3}$$

We define

$$\varepsilon\left(\vec{\omega}, \vec{\zeta}\right) \triangleq \left(\varepsilon(\omega^{(1)}, \zeta^{(1)}), \ldots, \varepsilon(\omega^{(n)}, \zeta^{(n)})\right)^\top.$$

From the definition of the model error $M(\cdot)$ in Section 4 and the projection theorem, the basis functions $\phi_1(\cdot), \ldots, \phi_d(\cdot)$ are orthogonal to $M(\cdot)$, i.e.,

$$\mathsf{E}\left[\phi_\ell(\omega) M(\omega)\right] = 0, \tag{B.4}$$

for $\ell = 1, \ldots, d$. We define

$$M(\vec{\omega}) \triangleq \left(M(\omega^{(1)}), \ldots, M(\omega^{(n)})\right)^\top.$$

B.1. Differentiable Case

Lemma 1. *Suppose Assumptions A1, A2, and A3 hold. As the number of scenarios* $n \to \infty$,

$$\sqrt{n}\left(\hat{r} - r^*\right) \xrightarrow{\mathsf{d}} N\left(\mathbf{0}, \Sigma_M + \frac{\Sigma_v}{m}\right),$$

where

$$\Sigma_M \triangleq \mathsf{E}\left[M^2\left(\omega\right)\Phi\left(\omega\right)^\top \Phi\left(\omega\right)\right], \qquad \Sigma_v \triangleq \mathsf{E}\left[v\left(\omega\right)\Phi\left(\omega\right)^\top \Phi\left(\omega\right)\right].$$

Therefore, as $n \to \infty$,

$$\left\|\hat{r} - r^*\right\|_2 = \frac{O_{\mathsf{P}}\left(1\right)}{\sqrt{n}}.$$

Proof. From Theorem 5.3 in White (2001), we have

$$\left(\mathrm{Cov}\left(n^{-\frac{1}{2}}\Phi^\top(\vec{\omega})(M(\vec{\omega}) + \varepsilon(\vec{\omega}, \vec{\zeta}))\right)\right)^{-\frac{1}{2}}\sqrt{n}(\hat{r} - r^*) \xrightarrow{d} N\left(\mathbf{0}, I_d\right). \tag{B.5}$$

Note that using (B.2) and (B.4),

$$\begin{aligned}
&\mathrm{Cov}\left(n^{-\frac{1}{2}}\Phi^\top(\vec{\omega})(M(\vec{\omega}) + \varepsilon(\vec{\omega}, \vec{\zeta}))\right) \\
&= \frac{1}{n}\mathsf{E}\left[\left(\Phi^\top(\vec{\omega})(M(\vec{\omega}) + \varepsilon(\vec{\omega}, \vec{\zeta}))\right)\left(\Phi^\top(\vec{\omega})(M(\vec{\omega}) + \varepsilon(\vec{\omega}, \vec{\zeta}))\right)^\top\right] \\
&= \frac{1}{n}\mathsf{E}\left[\Phi^\top(\vec{\omega})M(\vec{\omega})M^\top(\vec{\omega})\Phi(\vec{\omega})\right] + \frac{1}{n}\mathsf{E}\left[\Phi^\top(\vec{\omega})\mathsf{E}\left[\varepsilon(\vec{\omega}, \vec{\zeta})\varepsilon^\top(\vec{\omega}, \vec{\zeta})\middle|\vec{\omega}, \vec{\zeta}\right]\Phi(\vec{\omega})\right] \\
&+ \frac{2}{n}\mathsf{E}\left[\Phi^\top(\vec{\omega})M(\vec{\omega})\mathsf{E}\left[\varepsilon^\top(\vec{\omega}, \vec{\zeta})\middle|\vec{\omega}, \vec{\zeta}\right]\Phi(\vec{\omega})\right].
\end{aligned}$$

From (B.3),

$$\mathsf{E}\left[\varepsilon(\vec{\omega}, \vec{\zeta})\varepsilon^\top(\vec{\omega}, \vec{\zeta})\middle|\vec{\omega}, \vec{\zeta}\right] = \begin{pmatrix} \frac{v(\omega^{(1)})}{m} & 0 & \cdots & 0 \\ 0 & \frac{v(\omega^{(2)})}{m} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{v(\omega^{(n)})}{m} \end{pmatrix}.$$

Further,

$$
\begin{aligned}
&\frac{1}{n}\mathsf{E}\left[\Phi^\top(\vec{\omega})M(\vec{\omega})M^\top(\vec{\omega})\Phi(\vec{\omega})\right] \\
&= \frac{1}{n}\begin{pmatrix} \mathsf{E}\left[\sum_{i=1}^n \phi_1^2(\omega^{(i)})M^2(\omega^{(i)})\right] & \cdots & \mathsf{E}\left[\sum_{i=1}^n \phi_1(\omega^{(i)})\phi_d(\omega^{(i)})M^2(\omega^{(i)})\right] \\ \vdots & \ddots & \vdots \\ \mathsf{E}\left[\sum_{i=1}^n \phi_d(\omega^{(i)})\phi_1(\omega^{(i)})M^2(\omega^{(i)})\right] & \cdots & \mathsf{E}\left[\sum_{i=1}^n \phi_d^2(\omega^{(i)})M^2(\omega^{(i)})\right] \end{pmatrix} \\
&= \begin{pmatrix} \mathsf{E}\left[\phi_1^2(\omega)M^2(\omega)\right] & \cdots & \mathsf{E}\left[\phi_1(\omega)\phi_d(\omega)M^2(\omega)\right] \\ \vdots & \ddots & \vdots \\ \mathsf{E}\left[\phi_d(\omega)\phi_1(\omega)M^2(\omega)\right] & \cdots & \mathsf{E}\left[\phi_d^2(\omega)M^2(\omega)\right] \end{pmatrix} \\
&= \mathsf{E}\left[M^2(\omega)\Phi(\omega)^\top\Phi(\omega)\right].
\end{aligned}
$$

Therefore,

$$
\text{(B.6)} \qquad \operatorname{Cov}\left(n^{-\frac{1}{2}}\Phi^\top(\vec{\omega})(M(\vec{\omega})+\varepsilon(\vec{\omega},\vec{\zeta}))\right) = \mathsf{E}\left[M^2(\omega)\Phi(\omega)^\top\Phi(\omega)\right] + \frac{1}{m}\mathsf{E}\left[v(\omega)\Phi(\omega)^\top\Phi(\omega)\right].
$$

From (B.5) and (B.6), we have that

$$
\left(\Sigma_M + \frac{1}{m}\Sigma_v\right)^{-\frac{1}{2}}\sqrt{n}\left(\hat{r}-r^*\right) \xrightarrow{d} N\left(\mathbf{0}, I_d\right),
$$

where Σ_M and Σ_v are defined by (17). Based on the result above, we have that

$$
n\left\|\hat{r}-r^*\right\|_2^2 = \left(\sqrt{n}\left(\hat{r}-r^*\right)\right)^\top\left(\sqrt{n}\left(\hat{r}-r^*\right)\right) = O_{\mathsf{P}}(1),
$$

as $n \to \infty$, and the result follows.

■

Theorem 1. *Suppose that Assumptions F1, A1, A2, and A3 hold. Then there exists a sequence of random variables $\{B_{M,n}\}$, for $n = 1, 2, \ldots,$ satisfying*

$$
B_{M,n} \xrightarrow{\mathsf{P}} B_M^* \triangleq \mathsf{E}\left[f\left(\Phi(\omega)r^*\right)\right] - \alpha,
$$

so that

$$
\begin{aligned}
&\sqrt{n}\left(\hat{\alpha}_{\mathsf{REG}(m,n)} - \alpha - B_{M,n}\right) \\
&\qquad \xrightarrow{\mathsf{d}} N\left(0, \mathsf{E}\left[f'\left(L\left(\omega\right)\right)\Phi\left(\omega\right)\right]\left(\Sigma_M + \frac{\Sigma_v}{m}\right)\left(\mathsf{E}\left[f'\left(L\left(\omega\right)\right)\Phi\left(\omega\right)\right]\right)^\top\right),
\end{aligned}
$$

where Σ_M and Σ_v are defined by (17). Further, the asymptotic bias B_M^* satisfies

$$
\left|B_M^* - \mathsf{E}\left[f'\left(L\left(\omega\right)\right)M\left(\omega\right)\right]\right| \le \frac{U_{\mathit{diff}}}{2}\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right].
$$

Proof. By Taylor's theorem,

$$
\begin{aligned}
f\left(\Phi\left(\omega\right)\hat{r}\right)-f\left(L\left(\omega\right)\right) &= f'\left(L\left(\omega\right)\right)\left(\Phi\left(\omega\right)\hat{r}-L\left(\omega\right)\right) \\
&\quad +\frac{1}{2}f''\left(L\left(\omega\right)+\theta\cdot\left(\Phi\left(\omega\right)\hat{r}-L\left(\omega\right)\right)\right)\left(\Phi\left(\omega\right)\hat{r}-L\left(\omega\right)\right)^{2},
\end{aligned}
$$

where $\theta \in (0,1)$ is a random variable. Then

$$
\begin{aligned}
\hat{\alpha}_{\mathsf{REG}(m,n)}-\alpha &= \mathsf{E}\left[f\left(\Phi\left(\omega\right)\hat{r}\right)\middle|\vec{\omega},\vec{\zeta}\right]-\mathsf{E}\left[f\left(L\left(\omega\right)\right)\right] \\
&= \mathsf{E}\left[f'\left(L\left(\omega\right)\right)\left(\Phi\left(\omega\right)\hat{r}-L\left(\omega\right)\right)\middle|\vec{\omega},\vec{\zeta}\right] \\
&\quad + \mathsf{E}\left[\frac{1}{2}f''\left(L\left(\omega\right)+\theta\cdot\left(\Phi\left(\omega\right)\hat{r}-L\left(\omega\right)\right)\right)\left(\Phi\left(\omega\right)\hat{r}-L\left(\omega\right)\right)^{2}\middle|\vec{\omega},\vec{\zeta}\right].
\end{aligned}
\tag{B.7}
$$

For the first term in (B.7),

$$
\mathsf{E}\left[f'\left(L\left(\omega\right)\right)\left(\Phi\left(\omega\right)\hat{r}-L\left(\omega\right)\right)\middle|\vec{\omega},\vec{\zeta}\right]=\mathsf{E}\left[f'\left(L\left(\omega\right)\right)\Phi\left(\omega\right)\right]\left(\hat{r}-r^{*}\right)-\mathsf{E}\left[f'\left(L\left(\omega\right)\right)M\left(\omega\right)\right].
$$

Further, from Lemma 1, we have that

$$
\begin{aligned}
&\sqrt{n}\left(\mathsf{E}\left[f'\left(L\left(\omega\right)\right)\Phi\left(\omega\right)\right]\right)\left(\hat{r}-r^{*}\right) \\
&\quad \xrightarrow{d} N\left(\mathbf{0},\mathsf{E}\left[f'\left(L\left(\omega\right)\right)\Phi\left(\omega\right)\right]\left(\Sigma_{M}+\frac{\Sigma_{v}}{m}\right)\left(\mathsf{E}\left[f'\left(L\left(\omega\right)\right)\Phi\left(\omega\right)\right]\right)^{\top}\right),
\end{aligned}
\tag{B.8}
$$

as $n \to \infty$.

Combining (B.7) and (B.8), and letting

$$
B_{M,n}\triangleq\mathsf{E}\left[f'\left(L\left(\omega\right)\right)M\left(\omega\right)\right]-\mathsf{E}\left[\frac{1}{2}f''\left(L\left(\omega\right)+\theta\cdot\left(\Phi\left(\omega\right)\hat{r}-L\left(\omega\right)\right)\right)\left(\Phi\left(\omega\right)\hat{r}-L\left(\omega\right)\right)^{2}\middle|\vec{\omega},\vec{\zeta}\right],
\tag{B.9}
$$

equation (18) follows. From Lemma 1,

$$
\hat{r}\xrightarrow{\mathsf{P}}r^{*},
$$

and then by the continuous mapping theorem,

$$
\hat{\alpha}_{\mathsf{REG}(m,n)}=\mathsf{E}\left[f\left(\Phi(\omega)\hat{r}\right)\;\middle|\;\vec{\omega},\vec{\zeta}\;\right]\xrightarrow{\mathsf{P}}\mathsf{E}\left[f\left(\Phi(\omega)r^{*}\right)\right].
\tag{B.10}
$$

Also notice that equation (18) implies

$$
\hat{\alpha}_{\mathsf{REG}(m,n)}-\alpha-B_{M,n}\xrightarrow{\mathsf{P}}0.
\tag{B.11}
$$

Combining (B.10) and (B.11),

$$
B_{M,n}\xrightarrow{\mathsf{P}}B_{M}^{*}=\mathsf{E}\left[f\left(\Phi(\omega)r^{*}\right)\right]-\alpha.
\tag{B.12}
$$

Given (16) and (B.9),

$$\left|B_{M,n} - \mathsf{E}\left[f'\left(L\left(\omega\right)\right) M\left(\omega\right)\right]\right| \leq \left|\mathsf{E}\left[\frac{1}{2} f''\left(L\left(\omega\right) + \theta \cdot \left(\Phi\left(\omega\right)\hat{r} - L\left(\omega\right)\right)\right)\left(\Phi\left(\omega\right)\hat{r} - L\left(\omega\right)\right)^2 \middle| \vec{\omega}, \vec{\zeta}\right]\right|$$

$$\leq \frac{U_{\text{diff}}}{2}\mathsf{E}\left[\left(\Phi\left(\omega\right)\hat{r} - L\left(\omega\right)\right)^2 \middle| \vec{\omega}, \vec{\zeta}\right]$$

$$= \frac{U_{\text{diff}}}{2}\mathsf{E}\left[\left(\Phi\left(\omega\right)\hat{r} - \Phi\left(\omega\right)r^*\right)^2 \middle| \vec{\omega}, \vec{\zeta}\right] + \frac{U_{\text{diff}}}{2}\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right], \tag{B.13}$$

where we have used (B.4) in equation (B.13). From Lemma 1, we have

$$\frac{U_{\text{diff}}}{2}\mathsf{E}\left[\left(\Phi\left(\omega\right)\hat{r} - \Phi\left(\omega\right)r^*\right)^2 \middle| \vec{\omega}, \vec{\zeta}\right] = \frac{U_{\text{diff}}}{2}\left(\hat{r} - r^*\right)^\top \mathsf{E}\left[\Phi\left(\omega\right)^\top \Phi\left(\omega\right)\right]\left(\hat{r} - r^*\right)$$

$$= \frac{U_{\text{diff}}}{2}\left\|\hat{r} - r^*\right\|_2^2 = \frac{O_{\mathsf{P}}\left(1\right)}{n}. \tag{B.14}$$

From (B.12), (B.13), and (B.14), we have

$$\left|B_M^* - \mathsf{E}\left[f'\left(L\left(\omega\right)\right) M\left(\omega\right)\right]\right| \leq \frac{U_{\text{diff}}}{2}\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right].$$

■

B.2. Lipschitz Continuous Case

Theorem 2. *Suppose that Assumptions F2, A1, A2, and A3 hold. Then as the number of scenarios $n \to \infty$,*

$$\left(\hat{\alpha}_{\mathsf{REG}(m,n)} - \alpha\right)^2 \leq U_{Lip}^2 \mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right] + O_{\mathsf{P}}\left(\frac{1}{n}\right).$$

Proof. Note that

$$\hat{\alpha}_{\mathsf{REG}(m,n)} - \alpha = \mathsf{E}\left[f\left(\Phi\left(\omega\right)\hat{r}\right) \middle| \vec{\omega}, \vec{\zeta}\right] - \mathsf{E}\left[f\left(L\left(\omega\right)\right)\right] = \mathsf{E}\left[f\left(\Phi\left(\omega\right)\hat{r}\right) - f\left(L\left(\omega\right)\right) \middle| \vec{\omega}, \vec{\zeta}\right].$$

From the Lipschitz continuity condition (21) and Jensen's inequality,

$$\left(\hat{\alpha}_{\mathsf{REG}(m,n)} - \alpha\right)^2 = \left(\mathsf{E}\left[f\left(\Phi\left(\omega\right)\hat{r}\right) - f\left(L\left(\omega\right)\right) \middle| \vec{\omega}, \vec{\zeta}\right]\right)^2$$

$$\leq \mathsf{E}\left[\left(f\left(\Phi\left(\omega\right)\hat{r}\right) - f\left(L\left(\omega\right)\right)\right)^2 \middle| \vec{\omega}, \vec{\zeta}\right]$$

$$\leq U_{\text{Lip}}^2 \mathsf{E}\left[\left(\Phi\left(\omega\right)\left(\hat{r} - r^*\right) - M\left(\omega\right)\right)^2 \middle| \vec{\omega}, \vec{\zeta}\right]$$

$$= U_{\text{Lip}}^2 \mathsf{E}\left[\left(\Phi\left(\omega\right)\left(\hat{r} - r^*\right)\right)^2 \middle| \vec{\omega}, \vec{\zeta}\right] + U_{\text{Lip}}^2 \mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right],$$

where we have used (B.4). Then, by the orthonormality of $\Phi(\cdot)$,

$$\left(\hat{\alpha}_{\mathsf{REG}(m,n)} - \alpha\right)^2 \leq U_{\text{Lip}}^2 \mathsf{E}\left[\left(\Phi\left(\omega\right)\left(\hat{r} - r^*\right)\right)^2 \middle| \vec{\omega}, \vec{\zeta}\right] + U_{\text{Lip}}^2 \mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right]$$

$$= U_{\text{Lip}}^2 \left\|\hat{r} - r^*\right\|_2^2 + U_{\text{Lip}}^2 \mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right]. \tag{B.15}$$

From Lemma 1, as $n \to \infty$,

$$\sqrt{n}\left(\hat{r} - r^*\right) \xrightarrow{d} N\left(\mathbf{0}, \Sigma_M + \frac{\Sigma_v}{m}\right).$$

From the continuous mapping theorem, as $n \to \infty$, $nU_{\text{Lip}}^2 \left\|\hat{r} - r^*\right\|_2^2$ converges to a generalized chi-square distribution. Therefore, for any $\epsilon > 0$, there exist $\Delta_\epsilon > 0$ and $N_\epsilon > 0$, such that for any $n > N_\epsilon$,

$$\mathsf{P}\left(nU_{\text{Lip}}^2 \left\|\hat{r} - r^*\right\|_2^2 > \Delta_\epsilon\right) < \epsilon,$$

which implies that

$$U_{\text{Lip}}^2 \left\|\hat{r} - r^*\right\|_2^2 = O_{\mathsf{P}}\left(\frac{1}{n}\right).$$

With (B.15), the result follows.

■

In order to prove Lemma 2, Theorem 3, and Corollary 2, we need the following lemmas.

Lemma 4. *For any $r \in \mathbb{R}^d$,*

$$g\left(r\right) - g\left(r^*\right) = \left\|r - r^*\right\|_2^2.$$

Therefore, for any $r \in \mathbb{R}^d$ and $\mathcal{R}_\rho$ defined by (22)*, we have*

$$\mathcal{R}_\rho = \left\{r \in \mathbb{R}^d \ : \ g\left(r\right) \le g\left(r^*\right) + \rho\right\}.$$

Proof. By the projection theorem and the fact that Φ is orthonormal,

$$\begin{aligned}
g\left(r\right) &= \mathsf{E}\left[\left(\hat{L}\left(\omega, \zeta\right) - \Phi\left(\omega\right) r\right)^2\right] \\
&= \mathsf{E}\left[\left(\hat{L}\left(\omega, \zeta\right) - \Phi\left(\omega\right) r^* + \Phi\left(\omega\right) r^* - \Phi\left(\omega\right) r\right)^2\right] \\
&= \mathsf{E}\left[\left(\Phi\left(\omega\right)\left(r - r^*\right)\right)^2\right] + \mathsf{E}\left[\left(\hat{L}\left(\omega, \zeta\right) - \Phi\left(\omega\right) r^*\right)^2\right] \\
&\quad - 2\mathsf{E}\left[\left(\Phi\left(\omega\right)\left(r - r^*\right)\right)\left(\hat{L}\left(\omega, \zeta\right) - \Phi\left(\omega\right) r^*\right)\right] \\
&= \left\|r - r^*\right\|_2^2 + g\left(r^*\right) - 2\left(r - r^*\right)^\top \mathsf{E}\left[\Phi\left(\omega\right)^\top\left(\hat{L}\left(\omega, \zeta\right) - \Phi\left(\omega\right) r^*\right)\right] \\
&= \left\|r - r^*\right\|_2^2 + g\left(r^*\right).
\end{aligned}$$

■

Given $\mathcal{R}_\rho$, define

$$\hat{r}_\rho \in \underset{r \in \mathcal{R}_\rho}{\operatorname{argmin}} \ \frac{1}{n}\sum_{i=1}^n G\left(r, \omega^{(i)}, \zeta^{(i)}\right), \tag{B.16}$$

which is the sample optimal solution of r over $\mathcal{R}_\rho$. Under Assumptions A1 and A2, according to Lemma 3, the optimal solution $\hat{r}_\rho$ exists and is unique almost surely.

Lemma 5. *For $\rho > 0$, $\hat{r}_{2\rho} \in \mathcal{R}_\rho$ if and only if $\hat{r} \in \mathcal{R}_\rho$.*

Proof. Notice that $\hat{r} \in \mathcal{R}_\rho$ implies $\hat{r}_{2\rho} = \hat{r}$, and thus $\hat{r} \in \mathcal{R}_\rho$ implies $\hat{r}_{2\rho} \in \mathcal{R}_\rho$.

On the other hand, if $\hat{r}_{2\rho} \in \mathcal{R}_\rho$ and $\hat{r} \notin \mathcal{R}_\rho$, we must have $\hat{r} \notin \mathcal{R}_{2\rho}$. Therefore, we can have

$$\frac{1}{n}\sum_{i=1}^{n} G\left(\hat{r}, \omega^{(i)}, \zeta^{(i)}\right) < \frac{1}{n}\sum_{i=1}^{n} G\left(\hat{r}_{2\rho}, \omega^{(i)}, \zeta^{(i)}\right).$$

Then for any $\varphi \in (0,1)$, by the convexity of $(1/n)\sum_{i=1}^{n} G(\cdot, \omega^{(i)}, \zeta^{(i)})$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} G\left(\varphi\hat{r}_{2\rho} + (1-\varphi)\,\hat{r}, \omega^{(i)}, \zeta^{(i)}\right) &\le \varphi\frac{1}{n}\sum_{i=1}^{n} G\left(\hat{r}_{2\rho}, \omega^{(i)}, \zeta^{(i)}\right) + (1-\varphi)\,\frac{1}{n}\sum_{i=1}^{n} G\left(\hat{r}, \omega^{(i)}, \zeta^{(i)}\right) \\ &< \frac{1}{n}\sum_{i=1}^{n} G\left(\hat{r}_{2\rho}, \omega^{(i)}, \zeta^{(i)}\right),\end{aligned}$$

and thus $\varphi\hat{r}_{2\rho} + (1-\varphi)\,\hat{r} \notin \mathcal{R}_{2\rho}$, i.e.,

$$g\left(\varphi\hat{r}_{2\rho} + (1-\varphi)\,\hat{r}\right) > g\left(r^*\right) + 2\rho.$$

However, we know that $g\left(\cdot\right)$ is convex and thus continuous, and then

$$\rho + g\left(r^*\right) \ge g\left(\hat{r}_{2\rho}\right) = \lim_{\varphi\to 1} g\left(\varphi\hat{r}_{2\rho} + (1-\varphi)\,\hat{r}\right) \ge g\left(r^*\right) + 2\rho,$$

which is a contradiction.

■

Lemma 6. *Suppose that Assumptions F2, A1, A2, A4, and A5 hold. Let $\rho > 0$ be an arbitrary constant. Then consider any $\theta \in (0,\infty)$ and suppose that*

$$n \ge \frac{C'\lambda^2}{\rho}\left(d\ln\left(\frac{2\sqrt{2}C''\Lambda_{2\rho}}{\sqrt{\rho}}\right) + \ln\left(\frac{1}{\theta}\right)\right), \tag{B.17}$$

where λ is defined in Assumption A5, C' and C'' are universal constants, i.e., they do not depend on the problem, and

$$\Lambda_\rho \triangleq \left(2\sqrt{\rho}+1\right)d + 2\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right] + 2\mathsf{E}\left[\left(\varepsilon\left(\omega,\zeta\right)\right)^2\right].$$

Then,

$$\mathsf{P}\left(\hat{r} \notin \mathcal{R}_\rho\right) \le \theta.$$

Proof. When $\theta \ge 1$, the result is trivial. When $\theta \in (0,1)$, the result follows from Corollary 5.20 in Shapiro et al. (2009) and Lemma 5 above. In the setting here, we let $r = 2\rho$, $\varepsilon = \rho$, $\delta = 0$, and then $a = 2\rho$. Also notice that in our problem, $\gamma = 2$, $c = 1$, and $D_a^* = D_{2\rho}^* = 2\sqrt{2\rho}$. Further, compared to the notation of Shapiro et al. (2009), we use $G(\cdot)$ as $F(\cdot)$, $g(\cdot)$ as $f(\cdot)$, $\Psi_{r',r''}(\cdot)$ as $M_{x',x}(\cdot)$, n as N, d as n, and $\Lambda_{2\rho}$ as L.

Assumption (M5) in Shapiro et al. (2009) is from our Assumption A4, i.e., the moment generating functions of $\|\Phi(\omega)\|_2^2$, $(M(\omega))^2$, and $(\varepsilon(\omega,\zeta))^2$ are finite-valued in a neighborhood of zero. In particular,

$$\begin{aligned}
&\left|G(r',\omega,\zeta) - G(r'',\omega,\zeta)\right| \\
&= \left|\left(\hat{L}(\omega,\zeta) - \Phi(\omega)r'\right)^2 - \left(\hat{L}(\omega,\zeta) - \Phi(\omega)r''\right)^2\right| \\
&= \left|\left(\Phi(\omega)r' + \Phi(\omega)r'' - 2\hat{L}(\omega)\right)\Phi(\omega)(r'-r'')\right| \\
&\leq \left|\Phi(\omega)(r'-r^*) + \Phi(\omega)(r''-r^*) + 2\Phi(\omega)r^* - 2\hat{L}(\omega)\right| \|\Phi(\omega)\|_2 \|r'-r''\|_2 \\
&\leq \left(\|\Phi(\omega)\|_2 \|r'-r^*\|_2 + \|\Phi(\omega)\|_2 \|r''-r^*\|_2 + 2\left|\hat{L}(\omega,\zeta) - \Phi(\omega)r^*\right|\right) \|\Phi(\omega)\|_2 \|r'-r''\|_2 \\
&\leq \left(\|\Phi(\omega)\|_2 \sqrt{2\rho} + \|\Phi(\omega)\|_2 \sqrt{2\rho} + 2\left|M(\omega) + \varepsilon(\omega,\zeta)\right|\right) \|\Phi(\omega)\|_2 \|r'-r''\|_2 \\
&= \left(2\sqrt{2\rho}\|\Phi(\omega)\|_2^2 + 2\left|M(\omega) + \varepsilon(\omega,\zeta)\right| \|\Phi(\omega)\|_2\right) \|r'-r''\|_2 \\
&\leq \left((2\sqrt{2\rho}+1)\|\Phi(\omega)\|_2^2 + \left|M(\omega) + \varepsilon(\omega,\zeta)\right|^2\right) \|r'-r''\|_2 \\
&\leq \left((2\sqrt{2\rho}+1)\|\Phi(\omega)\|_2^2 + 2(M(\omega))^2 + 2(\varepsilon(\omega,\zeta))^2\right) \|r'-r''\|_2 .
\end{aligned}$$

Since in a neighborhood of zero, the finiteness of the moment generating functions of $\|\Phi(\omega)\|_2^2$, $(M(\omega))^2$, and $(\varepsilon(\omega,\zeta))^2$ implies the finiteness of the moment generating function of

$$\left(2\sqrt{2\rho}+1\right)\|\Phi(\omega)\|_2^2 + 2(M(\omega))^2 + 2(\varepsilon(\omega,\zeta))^2 ,$$

Assumption (M5) in Shapiro et al. (2009) is satisfied.

Assumption (M6) in Shapiro et al. (2009) is from the Assumption A5. Notice that Assumption A5 is weaker than Assumption (M6) in Shapiro et al. (2009), but according to the discussion after Assumption (M6) in Shapiro et al. (2009), Assumption A5 here is sufficient.

■

Lemma 2. *Suppose that Assumptions F2, A1, A2, A4, and A5 hold. Let $\rho > 0$ be an arbitrary constant. Then for any positive integer n,*

$$\mathsf{P}(\hat{r} \notin \mathcal{R}_\rho) \leq \left(\frac{2\sqrt{2}C''\Lambda_{2\rho}}{\sqrt{\rho}}\right)^d \exp\left(-\frac{\rho n}{C'\lambda^2}\right),$$

where λ is defined in Assumptions A5, C' and C'' are universal constants (i.e., constants that do not depend on the problem), and

$$\Lambda_\rho \triangleq (2\sqrt{\rho}+1)\,d + 2\mathsf{E}\left[(M(\omega))^2\right] + 2\mathsf{E}\left[(\varepsilon(\omega,\zeta))^2\right].$$

Proof. Define

$$\theta \triangleq \left(\frac{2\sqrt{2}C''\Lambda_{2\rho}}{\sqrt{\rho}}\right)^d \exp\left(-\frac{\rho n}{C'\lambda^2}\right).$$

Then,

$$n = \frac{C'\lambda^2}{\rho}\left(d\ln\left(\frac{2\sqrt{2}C''\Lambda_{2\rho}}{\sqrt{\rho}}\right) + \ln\left(\frac{1}{\theta}\right)\right),$$

which satisfies (B.17), and thus, by Lemma 6,

$$\mathsf{P}\left(\hat{r}_{2\rho} \notin \mathcal{R}_\rho\right) \leq \left(\frac{2\sqrt{2}C''\Lambda_{2\rho}}{\sqrt{\rho}}\right)^d \exp\left(-\frac{\rho n}{C'\lambda^2}\right).$$

From Lemma 5,

$$\mathsf{P}\left(\hat{r} \notin \mathcal{R}_\rho\right) = \mathsf{P}\left(\hat{r}_{2\rho} \notin \mathcal{R}_\rho\right).$$

■

Lemma 7. *For any $\rho \geq 2$,*

$$\Lambda_\rho \leq \frac{\sqrt{\rho}}{\sqrt{2}}\Lambda_2.$$

Proof. Notice that for any $\rho \geq 2$,

$$\begin{aligned}
\Lambda_\rho &= \left(2\sqrt{\rho}+1\right)d + 2\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right] + 2\mathsf{E}\left[\left(\varepsilon\left(\omega,\zeta\right)\right)^2\right] \\
&= \sqrt{\rho}\left(\left(2+\frac{1}{\sqrt{\rho}}\right)d + \frac{2\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right]}{\sqrt{\rho}} + \frac{2\mathsf{E}\left[\left(\varepsilon\left(\omega,\zeta\right)\right)^2\right]}{\sqrt{\rho}}\right) \\
&\leq \sqrt{\rho}\left(\left(2+\frac{1}{\sqrt{2}}\right)d + \frac{2\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right]}{\sqrt{2}} + \frac{2\mathsf{E}\left[\left(\varepsilon\left(\omega,\zeta\right)\right)^2\right]}{\sqrt{2}}\right) \\
&= \frac{\sqrt{\rho}}{\sqrt{2}}\left(\left(2\sqrt{2}+1\right)d + 2\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right] + 2\mathsf{E}\left[\left(\varepsilon\left(\omega,\zeta\right)\right)^2\right]\right) \\
&= \frac{\sqrt{\rho}}{\sqrt{2}}\Lambda_2.
\end{aligned}$$

■

Theorem 3. *Suppose that Assumptions F2, A1, A2, A4, and A5 hold, and let $\delta > 0$ be an arbitrary positive constant. Then for any positive integer n,*

$$\begin{aligned}
&\mathsf{E}\left[\left(\Phi\left(\omega\right)\left(\hat{r}-r^*\right)\right)^2\right] = \mathsf{E}\left[\left\|\hat{r}-r^*\right\|_2^2\right] \\
&\leq \frac{1}{n^{1-\delta}} + 2^{\frac{3d}{2}}C'\left(C''\right)^d\left(\Lambda_2\right)^d\lambda^2 n^{\frac{(1-\delta)d}{2}-1}\exp\left(-\frac{n^\delta}{C'\lambda^2}\right) + \frac{2^d C'\left(C''\right)^d\left(\Lambda_2\right)^d\lambda^2}{n}\exp\left(-\frac{n}{C'\lambda^2}\right) \\
&= O\left(n^{-1+\delta}\right).
\end{aligned}$$

Proof. Notice that

$$\begin{aligned}
\mathsf{E}\left[\left(\Phi(\omega)(\hat{r}-r^*)\right)^2\right] &= \mathsf{E}\left[\mathsf{E}\left[\left(\Phi(\omega)(\hat{r}-r^*)\right)^2\middle|\vec{\omega},\vec{\zeta}\right]\right] \\
&= \mathsf{E}\left[(\hat{r}-r^*)^\top \mathsf{E}\left[\Phi(\omega)^\top \Phi(\omega)\right](\hat{r}-r^*)\right] \\
&= \mathsf{E}\left[\|\hat{r}-r^*\|_2^2\right] \\
&= \int_0^\infty \mathsf{P}\left(\|\hat{r}-r^*\|_2^2 > \rho\right) d\rho. \\
&= \int_0^\infty \mathsf{P}\left(\hat{r}\notin \mathcal{R}_\rho\right) d\rho, \qquad \text{(B.18)}
\end{aligned}$$

where we have used Lemma 4.

Without loss of generality, we consider an arbitrary positive constant $\delta \in (0,1)$,

$$\mathsf{E}\left[\|\hat{r}-r^*\|_2^2\right] = \int_0^{\frac{1}{n^{1-\delta}}} \mathsf{P}\left(\hat{r}\notin \mathcal{R}_\rho\right) d\rho + \int_{\frac{1}{n^{1-\delta}}}^1 \mathsf{P}\left(\hat{r}\notin \mathcal{R}_\rho\right) d\rho + \int_1^\infty \mathsf{P}\left(\hat{r}\notin \mathcal{R}_\rho\right) d\rho. \qquad \text{(B.19)}$$

In order to bound (B.19), we bound each term separately. For the first term in (B.19),

$$\int_0^{\frac{1}{n^{1-\delta}}} \mathsf{P}\left(\hat{r}\notin \mathcal{R}_\rho\right) d\rho \le \frac{1}{n^{1-\delta}}.$$

For the second term in (B.19), from Lemma 2 and Lemma 7,

$$\begin{aligned}
\int_{\frac{1}{n^{1-\delta}}}^1 \mathsf{P}\left(\hat{r}\notin \mathcal{R}_\rho\right) d\rho &\le \int_{\frac{1}{n^{1-\delta}}}^1 \left(\frac{2\sqrt{2}C''\Lambda_2}{\sqrt{\rho}}\right)^d \exp\left(-\frac{\rho n}{C'\lambda^2}\right) d\rho \\
&= \left(2\sqrt{2}C''\Lambda_2\right)^d \int_{\frac{1}{n^{1-\delta}}}^1 \rho^{-\frac{d}{2}} \exp\left(-\frac{\rho n}{C'\lambda^2}\right) d\rho.
\end{aligned}$$

Define $\rho' \triangleq n^{1-\delta}\rho$, i.e., $\rho = \rho'/\left(n^{1-\delta}\right)$. Then,

$$\begin{aligned}
\int_{\frac{1}{n^{1-\delta}}}^1 \mathsf{P}\left(\hat{r}\notin \mathcal{R}_\rho\right) d\rho &\le \left(2\sqrt{2}C''\Lambda_2\right)^d \frac{1}{n^{1-\delta}} \int_1^\infty \left(\frac{\rho'}{n^{1-\delta}}\right)^{-\frac{d}{2}} \exp\left(-\frac{\rho' n^\delta}{C'\lambda^2}\right) d\rho' \\
&= \left(2\sqrt{2}C''\Lambda_2\right)^d \frac{1}{\left(n^{1-\delta}\right)^{1-\frac{d}{2}}} \int_1^\infty \left(\rho'\right)^{-\frac{d}{2}} \exp\left(-\frac{\rho' n^\delta}{C'\lambda^2}\right) d\rho' \\
&\le \left(2\sqrt{2}C''\Lambda_2\right)^d \frac{1}{\left(n^{1-\delta}\right)^{1-\frac{d}{2}}} \int_1^\infty \exp\left(-\frac{\rho' n^\delta}{C'\lambda^2}\right) d\rho' \\
&\le \left(2\sqrt{2}C''\Lambda_2\right)^d \frac{1}{\left(n^{1-\delta}\right)^{1-\frac{d}{2}}} \frac{C'\lambda^2}{n^\delta} \int_1^\infty \exp\left(-\frac{\rho' n^\delta}{C'\lambda^2}\right) d\left(\frac{\rho' n^\delta}{C'\lambda^2}\right) \\
&= \left(2\sqrt{2}C''\Lambda_2\right)^d \frac{1}{\left(n^{1-\delta}\right)^{1-\frac{d}{2}}} \frac{C'\lambda^2}{n^\delta} \exp\left(-\frac{n^\delta}{C'\lambda^2}\right).
\end{aligned}$$

For the third term in (B.19), with Lemmas 2 and 7,

$$\begin{aligned}\int_1^\infty \mathsf{P}\left(\hat{r} \notin \mathcal{R}_\rho\right) d\rho &\le \int_1^\infty \left(\frac{2\sqrt{2}C''\Lambda_{2\rho}}{\sqrt{\rho}}\right)^d \exp\left(-\frac{\rho n}{C'\lambda^2}\right) d\rho \\ &\le \int_1^\infty \left(2C''\Lambda_2\right)^d \exp\left(-\frac{\rho n}{C'\lambda^2}\right) d\rho \\ &= \left(2C''\Lambda_2\right)^d \int_1^\infty \exp\left(-\frac{\rho n}{C'\lambda^2}\right) d\rho \\ &= \left(2C''\Lambda_2\right)^d \frac{C'\lambda^2}{n} \exp\left(-\frac{n}{C'\lambda^2}\right).\end{aligned}$$

Therefore, (B.19) becomes

$$\begin{aligned}&\mathsf{E}\left[\|\hat{r}-r^*\|_2^2\right] \\ &\le \frac{1}{n^{1-\delta}} + \left(2\sqrt{2}C''\Lambda_2\right)^d \frac{1}{\left(n^{1-\delta}\right)^{1-\frac{d}{2}}} \frac{C'\lambda^2}{n^\delta} \exp\left(-\frac{n^\delta}{C'\lambda^2}\right) + \left(2C''\Lambda_2\right)^d \frac{C'\lambda^2}{n} \exp\left(-\frac{n}{C'\lambda^2}\right) \\ &= \frac{1}{n^{1-\delta}} + 2^{\frac{3d}{2}} C' \left(C''\right)^d \left(\Lambda_2\right)^d \lambda^2 n^{\frac{(1-\delta)d}{2}-1} \exp\left(-\frac{n^\delta}{C'\lambda^2}\right) + \frac{2^d C' \left(C''\right)^d \left(\Lambda_2\right)^d \lambda^2}{n} \exp\left(-\frac{n}{C'\lambda^2}\right).\end{aligned}$$

■

Corollary 2. *Suppose that Assumptions F2, A1, A2, A4, and A5 hold, and let $\delta > 0$ be an arbitrary positive constant. Then, for any positive integer n,*

$$\begin{aligned}&\mathsf{E}\left[\left(\hat{\alpha}_{\mathsf{REG}(m,n)} - \alpha\right)^2\right] \\ &\le U_{Lip}^2 \left(2^{\frac{3d}{2}} C' \left(C''\right)^d \left(\Lambda_2\right)^d \lambda^2 n^{\frac{(1-\delta)d}{2}-1} \exp\left(-\frac{n^\delta}{C'\lambda^2}\right) + \frac{2^d C' \left(C''\right)^d \left(\Lambda_2\right)^d \lambda^2}{n} \exp\left(-\frac{n}{C'\lambda^2}\right)\right) \\ &+ U_{Lip}^2 \left(\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right] + n^{-1+\delta}\right) \\ &= U_{Lip}^2 \mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right] + O\left(n^{-1+\delta}\right).\end{aligned}$$

Proof. From (B.15) and Theorem 3, we have that

$$\begin{aligned}&\mathsf{E}\left[\left(\hat{\alpha}_{\mathsf{REG}(m,n)} - \alpha\right)^2\right] \\ &\le U_{\mathrm{Lip}}^2 \mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right] + U_{\mathrm{Lip}}^2 \mathsf{E}\left[\|\hat{r}-r^*\|_2^2\right] \\ &\le U_{\mathrm{Lip}}^2 \Bigg(\mathsf{E}\left[\left(M\left(\omega\right)\right)^2\right] \\ &+ \frac{1}{n^{1-\delta}} + 2^{\frac{3d}{2}} C' \left(C''\right)^d \left(\Lambda_2\right)^d \lambda^2 n^{\frac{(1-\delta)d}{2}-1} \exp\left(-\frac{n^\delta}{C'\lambda^2}\right) + \frac{2^d C' \left(C''\right)^d \left(\Lambda_2\right)^d \lambda^2}{n} \exp\left(-\frac{n}{C'\lambda^2}\right)\Bigg).\end{aligned}$$

■

C. Proofs for Section 6

This section presents the proof of Theorem 4 in Section 6.2. In addition to the notation defined in Section B, we define the following: $L(\vec{\omega})$ is an $n \times 1$ column vector,

$$L(\vec{\omega}) \triangleq \left(L(\omega^{(1)}), \dots, L(\omega^{(n)})\right)^\top,$$

and $\vec{H}$ is an $n \times n$ diagonal matrix,

$$\vec{H} \triangleq \begin{pmatrix} h(\omega^{(1)}) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & h(\omega^{(n)}) \end{pmatrix}.$$

We need the following lemma to prove Theorem 4.

Lemma 8. *Given a weight function $h(\cdot)$, if Assumptions F2, A6, A7, A8, and A9 hold, then as $n \to \infty$,*

$$\mathsf{E}\left[(\Phi(\omega)(\hat{r}(\vec{h}) - r^*(h)))^2\right] = O\left(n^{-1+\delta}\right),$$

where $\delta > 0$ is an arbitrary positive constant.

Proof. Following the proof of Theorem 3 for unweighted regression, under appropriate technical assumptions, as $n \to \infty$,

$$\mathsf{E}\left[\|\hat{r} - r^*\|_2^2\right] = O\left(n^{-1+\delta}\right), \tag{C.1}$$

where $\delta > 0$ is an arbitrary positive constant.

Substituting $L(\omega)$ with $\sqrt{h(\omega)}L(\omega)$, $\hat{L}(\omega, \zeta)$ with $\sqrt{h(\omega)}\hat{L}(\omega, \zeta)$, and $\Phi(\omega)$ with $\sqrt{h(\omega)}\Phi(\omega)$, the assumptions in Theorem 3 become Assumptions F2, A6, A7, A8, and A9 here, and the regression coefficients r^* and $\hat{r}$ become $r^*(h)$ and $\hat{r}(\vec{h})$. Then we can directly apply (C.1) and derive that, as $n \to \infty$,

$$\mathsf{E}\left[\|\hat{r}(\vec{h}) - r^*(h)\|_2^2\right] = O\left(n^{-1+\delta}\right).$$

Moreover,

$$\begin{aligned} \mathsf{E}\left[(\Phi(\omega)(\hat{r}(\vec{h}) - r^*(h)))^2\right] &= \mathsf{E}\left[\mathsf{E}\left[(\Phi(\omega)(\hat{r}(\vec{h}) - r^*(h)))^2 \middle| \vec{\omega}, \vec{\zeta}\right]\right] \\ &= \mathsf{E}\left[(\hat{r}(\vec{h}) - r^*(h))^\top \mathsf{E}\left[\Phi(\omega)^\top \Phi(\omega)\right] (\hat{r}(\vec{h}) - r^*(h))\right] \\ &= \mathsf{E}\left[\|\hat{r}(\vec{h}) - r^*(h)\|_2^2\right] \\ &= O\left(n^{-1+\delta}\right), \end{aligned}$$

where $\mathsf{E}\left[\Phi(\omega)^\top \Phi(\omega)\right] = 1$ is from the orthonormality assumed in Assumption A7.

■

Lemma 8 establishes that the mean squared error between our approximation and the best approximation decays at the rate $n^{-1+\delta}$ for any $\delta > 0$. With this lemma, we can establish the following theorem:

Theorem 4. *Given a weight function $h(\cdot)$, if Assumptions F2, A6, A7, A8, and A9 hold, then*

$$\lim_{n\to\infty} \mathsf{E}\left[(\hat{\alpha}_{REG(m,n,h)} - \alpha)^2\right] = \left(\mathsf{E}\left[f(\Phi(\omega) r^*(h))\right] - \mathsf{E}\left[f(L(\omega))\right]\right)^2 .$$

Proof. Decomposing the MSE of $\hat{\alpha}_{\mathsf{REG}(m,n,h)}$, we have

$$\begin{aligned}
&\mathsf{E}\left[\left(\hat{\alpha}_{\mathsf{REG}(m,n,h)} - \alpha\right)^2\right] \\
&= \mathsf{E}\left[\left(\mathsf{E}\left[f(\Phi(\omega)\hat{r}(\vec{h}))\middle|\vec{\omega},\vec{\zeta}\right] - \mathsf{E}\left[f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\right] + \mathsf{E}\left[f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\right] - \mathsf{E}\left[f\left(L\left(\omega\right)\right)\right]\right)^2\right] \\
\text{(C.2)} &= \mathsf{E}\left[\left(\mathsf{E}\left[f(\Phi(\omega)\hat{r}(\vec{h}))\middle|\vec{\omega},\vec{\zeta}\right] - \mathsf{E}\left[f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\right]\right)^2\right] \\
&\quad + \left(\mathsf{E}\left[f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\right] - \mathsf{E}\left[f\left(L\left(\omega\right)\right)\right]\right)^2 \\
&\quad + 2\mathsf{E}\left[\mathsf{E}\left[f\left(\Phi(\omega)\hat{r}\left(h\right)\right) - f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\middle|\vec{\omega},\vec{\zeta}\right]\right]\left(\mathsf{E}\left[f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\right] - \mathsf{E}\left[f\left(L\left(\omega\right)\right)\right]\right).
\end{aligned}$$

We analyze the three terms in (C.2) separately. The first term in (C.2) satisfies

$$\begin{aligned}
\mathsf{E}\left[\left(\mathsf{E}\left[f(\Phi(\omega)\hat{r}(\vec{h}))\middle|\vec{\omega},\vec{\zeta}\right] - \mathsf{E}\left[f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\right]\right)^2\right] &\leq U_{\text{Lip}}^2 \mathsf{E}\left[\left(\mathsf{E}\left[\Phi(\omega)\left(\hat{r}(\vec{h}) - r^*\left(h\right)\right)\middle|\vec{\omega},\vec{\zeta}\right]\right)^2\right] \\
&\leq U_{\text{Lip}}^2 \mathsf{E}\left[\left(\Phi(\omega)(\hat{r}(\vec{h}) - r^*(h))\right)^2\right].
\end{aligned}$$

The third term in (C.2) satisfies

$$\begin{aligned}
&2\mathsf{E}\left[\mathsf{E}\left[f(\Phi(\omega)\hat{r}(\vec{h})) - f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\middle|\vec{\omega},\vec{\zeta}\right]\right]\left(\mathsf{E}\left[f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\right] - \mathsf{E}\left[f\left(L\left(\omega\right)\right)\right]\right) \\
&\quad\leq 2U_{\text{Lip}} \mathsf{E}\left[\mathsf{E}\left[\left|\Phi(\omega)(\hat{r}(\vec{h}) - r^*(h))\right|\middle|\vec{\omega},\vec{\zeta}\right]\right]\left|\mathsf{E}\left[f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\right] - \mathsf{E}\left[f\left(L\left(\omega\right)\right)\right]\right| \\
&\quad\leq 2U_{\text{Lip}}\sqrt{\mathsf{E}\left[\left(\Phi(\omega)(\hat{r}(\vec{h}) - r^*(h))\right)^2\right]}\left|\mathsf{E}\left[f\left(\Phi\left(\omega\right)r^*\left(h\right)\right)\right] - \mathsf{E}\left[f\left(L\left(\omega\right)\right)\right]\right|.
\end{aligned}$$

Combining these inequalities with Lemma 8, the result follows. ■