D Proofs of auxiliary Lemmas and additional details

D.1 Proof of Lemma A.2

We will show that the constraint \( f^*(x_i, 1) \leq f^*(x_j, 1) + \|x_i - x_j\|_X \) holds for all \( i, j \in \{1, \ldots, n\} \).

The argument that \( f^*(x_i, 0) \leq f^*(x_j, 0) + \|x_i - x_j\|_X \) holds for all \( i, j \in \{1, \ldots, n\} \) is similar and omitted. We assume, without loss of generality, that the observations are ordered so that \( d_j = 0 \) for \( j = 1, \ldots, n_0 \) and \( d_i = 1 \) for \( i = n_0 + 1, \ldots, n \). Observe that the bias can be written as

\[
\sum_{i=n_0+1}^{n} (k(x_i, 1) - w(1)) f(x_i, 1) - \sum_{j=1}^{n_0} w(0) f(x_j, 1) \\
+ \sum_{j=1}^{n_0} (k(x_j, 0) + w(0)) f(x_j, 0) + \sum_{i=n_0+1}^{n} w(1) f(x_i, 0).
\]

If \( k(x_i, 1) = w(1) \) for \( i \in \{n_0+1, \ldots, n\} \), we can set \( f^*(x_i, 1) = \min_{j \in \{1, \ldots, n_0\}} \{ f^*(x_j, 1) + \|x_i - x_j\|_X \} \) without affecting the bias, so that we can without loss of generality assume that (24) holds for all \( i \in \{n_0+1, \ldots, n\} \) and all \( j \in \{1, \ldots, n_0\} \).

If \( w(0) = 0 \), then the assumptions on \( k \) imply \( k(x_i, 1) = w(1) \) for \( i > n_0 \), and the value of \( f(\cdot, 1) \) doesn’t affect the bias. If \( w(0) > 0 \), then for each \( j \in \{1, \ldots, n_0\} \), at least one of the constraints \( f^*(x_i, 1) \leq f^*(x_j, 1) + \|x_i - x_j\|_X \), \( i \in \{n_0 + 1, \ldots, n\} \), must bind, otherwise we could decrease \( f^*(x_j, 1) \) and increase the value of the objective function. Let \( i(j) \) denote the index of one of the binding constraints (picked arbitrarily), so that \( f^*(x_{i(j)}, 1) = f^*(x_j, 1) + \|x_{i(j)} - x_j\|_X \). We need to
show that the constraints
\[
\begin{align*}
  f^*(x_i, 1) &\leq f^*(x_{i'}, 1) + \|x_i - x_{i'}\|_X, & i, i' &\in \{n_0 + 1, \ldots, n\}, \quad (S1) \\
  f^*(x_j, 1) &\leq f^*(x_{j'}, 1) + \|x_j - x_{j'}\|_X, & j, j' &\in \{1, \ldots, n_0\}, \quad (S2) \\
  f^*(x_j, 1) &\leq f^*(x_i, 1) + \|x_i - x_j\|_X, & j &\in \{1, \ldots, n_0\}, \; i &\in \{n_0 + 1, \ldots, n\}. \quad (S3)
\end{align*}
\]
are all satisfied. If (S1) doesn’t hold for some \((i, i')\), then by triangle inequality, for all \(j \in \{1, \ldots, n_0\},\)
\[
f^*(x_{i'}, 1) + \|x_i - x_{i'}\|_X < f(x_i, 1) \leq f^*(x_j, 1) + \|x_i - x_j\|_X \leq f^*(x_{j'}, 1) + \|x_i - x_{j'}\|_X + \|x_{i'} - x_j\|_X,
\]
so that \(f^*(x_{i'}, 1) < f^*(x_{j'}, 1) + \|x_{i'} - x_j\|_X\). But then it is possible to increase the bias by increasing \(f^*(x_{i'}, 1)\), which cannot be the case at the optimum. If (S2) doesn’t hold for some \((j, j')\), then by triangle inequality, for all \(i,\)
\[
f^*(x_j, 1) + \|x_i - x_j\|_X > f^*(x_{j'}, 1) + \|x_i - x_{j'}\|_X + \|x_j - x_{j'}\|_X \\
\geq f^*(x_{j'}, 1) + \|x_i - x_{j'}\|_X \geq f^*(x_i, 1).
\]
But this contradicts the assertion that for each \(j\), at least one of the constraints \(f(x_i, 1) \leq f(x_j, 1) + \|x_i - x_j\|_X\) binds. Finally, suppose that (S3) doesn’t hold for some \((i, j)\). Then by triangle inequality,
\[
f^*(x_i, 1) + \|x_i - x_{i(j)}\|_X \leq f^*(x_i, 1) + \|x_i - x_j\|_X + \|x_{i(j)} - x_j\|_X \\
< f^*(x_j, 1) + \|x_{i(j)} - x_j\|_X = f^*(x_{i(j)}, 1),
\]
which violates (S1).

### D.2 Proof of Lemma A.4

We will show that Equations (28), (29) and (30) hold at the optimum for \(d_i, d_{i'} = 1\) and \(d_j, d_{j'} = 0\). The argument that they hold for \(d_i, d_{i'} = 0\) and \(d_j, d_{j'} = 1\) is similar and omitted. The first-order conditions associated with the Lagrangian (31) are
\[
\begin{align*}
  m_j/\sigma^2(0) &= \mu w(0) + \sum_{i=1}^{n_1} \Lambda_{ij}^0, & \mu w(0) &= \sum_{i=1}^{n_1} \Lambda_{ij}^1, & j &= 1, \ldots, n_0, \quad (S4) \\
  m_{i+n_0}/\sigma^2(1) &= \mu w(1) + \sum_{j=1}^{n_0} \Lambda_{ij}^1, & \mu w(1) &= \sum_{j=1}^{n_0} \Lambda_{ij}^0, & i &= 1, \ldots, n_1. \quad (S5)
\end{align*}
\]
If \(w(0) = 0\), the first-order conditions together with the dual feasibility condition \(\Lambda_{ij}^1 \geq 0\) implies that \(m_{i+n_0} = \mu w(1)\sigma^2(1)\), and the assertion of the lemma holds trivially, since \(r_j = \mu w(1)\sigma^2(1)\).
for \( j = 1, \ldots, n \) achieves the optimum. Suppose, therefore, that \( w(0) > 0 \). Then \( \sum_{i=1}^{n_0} \Lambda^1_{ij} > 0 \), so that at least one of the constraints associated with \( \Lambda^1_{ij} \) must bind for each \( j \). Let \( i(j) \) denote the index of one of the binding constraints (picked arbitrarily if it is not unique), so that
\[
\begin{align*}
r_j = m_{i(j)+n_0} + \|x_{i(j)+n_0} - x_j\|_\chi.
\end{align*}
\]
Suppose (28) didn’t hold, so that for some \( j, j’ \in \{1, \ldots, n_0\} \),
\[
r_j > r_{j’} + \|x_j - x_{j’}\|_\chi.
\]
Then by triangle inequality
\[
r_j > r_{j’} + \|x_j - x_{j’}\|_\chi = m_{i(j’)+n_0} + \|x_{i(j’)+n_0} - x_j\|_\chi + \|x_j - x_{j’}\|_\chi \geq m_{i(j’)+n_0} + \|x_{i(j’)+n_0} - x_j\|_\chi,
\]
which violates the constraint associated with \( \Lambda^1_{i(j’)/j} \). Next, if (29) didn’t hold, so that for some \( i, i’ \in \{1, \ldots, n_1\} \),
\[
m_{i+n_0} > m_{i’+n_0} + \|x_{i+n_0} - x_{i’+n_0}\|_\chi,
\]
then for all \( j \in \{1, \ldots, n_0\} \),
\[
r_j \leq m_{i’+n_0} + \|x_{i’+n_0} - x_j\|_\chi \leq m_{i’+n_0} + \|x_{i+n_0} - x_{i’+n_0}\|_\chi + \|x_{i+n_0} - x_j\|_\chi < m_{i+n_0} + \|x_{i+n_0} - x_j\|_\chi,
\]
The complementary slackness condition \( \Lambda^1_{ij}(r_j - m_{i+n_0} - \|x_{i+n_0} - x_j\|_\chi) = 0 \) then implies that
\[
\sum_j \Lambda^1_{ij} = 0,
\]
and it follows from the first-order condition that \( m_{i+n_0} / \sigma^2(1) = \mu w(1) \leq m_{i’+n_0} / \sigma^2(1) \), which contradicts the assertion that \( m_{i+n_0} > m_{i’+n_0} + \|x_{i+n_0} - x_{i’+n_0}\|_\chi \). Finally, if (30) didn’t hold, so that
\[
m_{i+n_0} > r_j + \|x_{i+n_0} - x_j\|_\chi \text{ for some } i \in \{1, \ldots, n_1\} \text{ and } j \in \{1, \ldots, n_0\},
\]
then by triangle inequality
\[
m_{i+n_0} > r_j + \|x_{i+n_0} - x_j\|_\chi = m_{i(j)} + \|x_{i(j)+n_0} - x_j\|_\chi + \|x_{i+n_0} - x_j\|_\chi \geq m_{i(j)} + \|x_{i(j)+n_0} - x_{i+n_0}\|_\chi,
\]
which contradicts (29).

D.3 Derivation of algorithm for solution path

Observe that \( \Lambda^0_{ij} = 0 \) unless for some \( k, i \in R^0_k \) and \( j \in M^0_k \), and similarly \( \Lambda^1_{ij} = 0 \) unless for some \( k, j \in R^1_k \) and \( i \in M^1_k \). Therefore, the first-order conditions (S4) and (S5) can equivalently be written as
\[
\begin{align*}
m_j / \sigma^2(0) &= \mu w(0) + \sum_{i \in R^0_k} \Lambda^0_{ij} \quad j \in M^0_k, \quad \mu w(1) = \sum_{j \in M^0_k} \Lambda^0_{ij} \quad i \in R^0_k, \quad (S6) \\
m_{i+n_0} / \sigma^2(1) &= \mu w(1) + \sum_{j \in R^1_k} \Lambda^1_{ij} \quad i \in M^1_k, \quad \mu w(0) = \sum_{i \in M^1_k} \Lambda^1_{ij} \quad j \in R^1_k. \quad (S7)
\end{align*}
\]
Summing up these conditions then yields
\[
\begin{align*}
\sum_{j \in M^0_k} m_j / \sigma^2(0) &= \mu w(0) \cdot \#M^0_k + \sum_{j \in M^0_k} \sum_{i \in R^0_k} \Lambda^0_{ij} = \#M^0_k \cdot \mu w(0) + \#R^0_k \cdot \mu w(1), \\
\sum_{i \in M^1_k} m_{i+n_0} / \sigma^2(1) &= \mu w(1) \cdot \#M^1_k + \sum_{i \in M^1_k} \sum_{j \in R^1_k} \Lambda^1_{ij} = \#M^1_k \cdot \mu w(1) + \#R^1_k \cdot \mu w(0).
\end{align*}
\]
Following the argument in Osborne et al. (2000, Section 4), by continuity of the solution path, for a small enough perturbation $s$, $N^d(\mu + s) = N^d(\mu)$, so long as the elements of $\Lambda^d(\mu)$ associated with the active constraints are strictly positive. In other words, the set of active constraints doesn’t change for small enough changes in $\mu$. Hence, the partition $\mathcal{M}^d_k$ remains the same for small enough changes in $\mu$ and the solution path is differentiable. Differentiating the preceding display yields

$$\frac{1}{\sigma^2(0)} \sum_{j \in \mathcal{M}^0_k} \frac{\partial m_j(\mu)}{\partial \mu} = \# \mathcal{M}^0_k \cdot w(0) + \# \mathcal{R}^0_k \cdot w(1),$$

$$\frac{1}{\sigma^2(1)} \sum_{i \in \mathcal{M}^1_k} \frac{\partial m_{i+n_0}(\mu)}{\partial \mu} = \# \mathcal{M}^1_k \cdot w(1) + \# \mathcal{R}^1_k \cdot w(0).$$

If $j \in \mathcal{M}^0_k$, then there exists a $j'$ and $i$ such that the constraints associated with $\Lambda^0_{ij}$ and $\Lambda^0_{ij'}$ are both active, so that $m_j + \|x_{i+n_0} - x_j\|_X = r_{i+n_0} = m_{j'} + \|x_{i+n_0} - x_{j'}\|_X$, which implies that $\partial m_j(\mu)/\partial \mu = \partial m_{j'}(\mu)/\partial \mu$. Since all elements in $\mathcal{M}^0_k$ are connected, it follows that the derivative $\partial m_j(\mu)/\partial \mu$ is the same for all $j$ in $\mathcal{M}^0_k$. Similarly, $\partial m_j(\mu)/\partial \mu$ is the same for all $j$ in $\mathcal{M}^1_k$. Combining these observations with the preceding display implies

$$\frac{1}{\sigma^2(0)} \frac{\partial m_j(\mu)}{\partial \mu} = w(0) + \frac{\# \mathcal{R}^0_k(j)}{\# \mathcal{M}^0_k(j)} w(1), \quad \frac{1}{\sigma^2(1)} \frac{\partial m_{i+n_0}(\mu)}{\partial \mu} = w(1) + \frac{\# \mathcal{R}^1_k(i)}{\# \mathcal{M}^1_k(i)} w(0),$$

where $k(i)$ and $k(j)$ are the partitions that $i$ and $j$ belong to. Differentiating the first-order conditions (S6) and (S7) and combining them with the restriction that $\partial \Lambda^d_{ij}(\mu)/\partial \mu = 0$ if $N^d_{ij}(\mu) = 0$ then yields the following set of linear equations for $\partial \Lambda^d(\mu)/\partial \mu$:

$$\frac{\# \mathcal{R}^0_k}{\# \mathcal{M}^0_k} w(1) = \sum_{j \in \mathcal{R}^0_k} \frac{\partial \Lambda^0_{ij}(\mu)}{\partial \mu}, \quad w(1) = \sum_{j \in \mathcal{M}^0_k} \frac{\partial \Lambda^0_{ij}(\mu)}{\partial \mu},$$

$$\frac{\# \mathcal{R}^1_k}{\# \mathcal{M}^1_k} w(0) = \sum_{i \in \mathcal{R}^1_k} \frac{\partial \Lambda^1_{ij}(\mu)}{\partial \mu}, \quad w(0) = \sum_{i \in \mathcal{M}^1_k} \frac{\partial \Lambda^1_{ij}(\mu)}{\partial \mu}, \quad \frac{\partial \Lambda^d_{ij}(\mu)}{\partial \mu} = 0 \quad \text{if} \quad N^d_{ij}(\mu) = 0.$$

Therefore, $m(\mu)$, $\Lambda^0(\mu)$, and $\Lambda^1(\mu)$ are all piecewise linear in $\mu$. Furthermore, since for $i \in \mathcal{R}^0_k$, $r_{i+n_0}(\mu) = m_j(\mu) + \|x_{i+n_0} - x_j\|_X$ where $j \in \mathcal{M}^0_k$, it follows that

$$\frac{\partial r_{i+n_0}(\mu)}{\partial \mu} = \frac{\partial m_j(\mu)}{\partial \mu} = \sigma^2(0) \left[ w(0) + \frac{\# \mathcal{R}^0_k}{\# \mathcal{M}^0_k} w(1) \right].$$

Similarly, since for $j \in \mathcal{R}^1_k$, and $i \in \mathcal{M}^1_k$, $r_j(\mu) = m_{i+n_0}(\mu) + \|x_{i+n_0} - x_j\|_X$, where $j \in \mathcal{M}^0_k$, we have

$$\frac{\partial r_j(\mu)}{\partial \mu} = \frac{\partial m_{i+n_0}(\mu)}{\partial \mu} = \sigma^2(1) \left[ w(1) + \frac{\# \mathcal{R}^1_k}{\# \mathcal{M}^1_k} w(0) \right].$$

S4
Thus, $r(\mu)$ is also piecewise linear in $\mu$.

Differentiability of $m$ and $\Lambda^d$ is violated if the condition that the elements of $\Lambda^d$ associated with the active constraints are all strictly positive is violated. This happens if one of the non-zero elements of $\Lambda^d(\mu)$ decreases to zero, or else if a non-active constraint becomes active, so that for some $i$ and $j$ with $N^0_{ij}(\mu) = 0$, $r_{i+n_0}(\mu) = m_j(\mu) + \|x_{i+n_0} - x_j\|_\chi$, or for some $i$ and $j$ with $N^1_{ij}(\mu) = 0$, $r_j(\mu) = m_{i+n_0}(\mu) + \|x_{i+n_0} - x_j\|_\chi$. This determines the step size $s$ in the algorithm.

### D.4 Proof of Lemma B.2

For ease of notation, let $f_i = f(x_i, d_i)$, $\sigma_i^2 = \sigma^2(x_i, d_i)$, and let $\overline{f}_i = J^{-1} \sum_{j=1}^J f_{\ell_j(i)}$ and $\overline{u}_i = J^{-1} \sum_{j=1}^J u_{\ell_j(i)}$. Then we can decompose

$$
\frac{J+1}{J}(\overline{u}_i^2 - u_i^2) = \frac{J+1}{J}(f_i - \overline{f}_i + u_i - \overline{u}_i)^2 - \frac{J+1}{J}u_i^2
$$

$$
= [(f_i - \overline{f}_i)^2 + 2(u_i - \overline{u}_i)(f_i - \overline{f}_i)] - 2\overline{u}_i u_i + \frac{2}{J^2} \sum_{j=1}^J \sum_{k=1}^{j-1} u_{\ell_j(i)} u_{\ell_k(i)} + \frac{1}{J^2} \sum_{j=1}^J (u_{\ell_j(i)} - u_i^2)
$$

$$
= T_{1i} + 2T_{2i} + 2T_{3i} + T_{4i} + T_{5i} + \frac{1}{J^2} \sum_{j=1}^J (\sigma_{\ell_j(i)}^2 - \sigma_i^2),
$$

where

$$
T_{1i} = [(f_i - \overline{f}_i)^2 + 2(u_i - \overline{u}_i)(f_i - \overline{f}_i)], \quad T_{2i} = \overline{u}_i u_i
$$

$$
T_{3i} = \frac{1}{J^2} \sum_{j=1}^J \sum_{k=1}^{j-1} u_{\ell_j(i)} u_{\ell_k(i)}, \quad T_{4i} = \frac{1}{J^2} \sum_{j=1}^J (u_{\ell_j(i)}^2 - \sigma_{\ell_j(i)}^2), \quad T_{5i} = \sigma_i^2 - u_i^2.
$$

Since $\max_i \|x_{\ell_j(i)} - x_i\| \to 0$ and since $\sigma^2(\cdot, d)$ is uniformly continuous, it follows that

$$
\max_i \max_{1 \leq j \leq J} |\sigma_{\ell_j(i)}^2 - \sigma_i^2| \to 0,
$$

and hence that $|\sum_{i=1}^n a_i J^{-1} \sum_{j=1}^J (\sigma_{\ell_j(i)}^2 - \sigma_i^2)| \leq \max_i \max_{j=1, \ldots, J} (\sigma_{\ell_j(i)}^2 - \sigma_i^2) \sum_{i=1}^n a_i \to 0$. To prove the lemma, it therefore suffices to show that the sums $\sum_{i=1}^n a_i T_{qi}$ all converge to zero.

To that end,

$$
E[\sum_i a_i T_{1i}] \leq \max_i (f_i - \overline{f}_i)^2 \sum_i a_i + 2 \max_i |f_i - \overline{f}_i| \sum_i a_i E|u_i - \overline{u}_i|,
$$

which converges to zero since $\max_i |f_i - \overline{f}_i| \leq \max_{j=1, \ldots, J} (f_i - f_{\ell_j(i)}) \leq C_n \max_i \|x_i - x_{\ell_j(i)}\|_\chi \to \ldots
$$
0. Next, by the von Bahr-Esseen inequality,
\[
E\left|\sum_{i=1}^{n} a_{ni} T_{5i}\right|^{1+1/2K} \leq 2 \sum_{i=1}^{n} a_{ni}^{1+1/2K} E\left|T_{5i}\right|^{1+1/2K} \leq 2 \max_{j} \left\{ a_{nj}^{1/2K} \right\} \max_{j} E\left|T_{5j}\right|^{1+1/2K} \sum_{k=1}^{n} a_{nk} \to 0.
\]

Let \(\mathcal{I}_j\) denote the set of observations for which an observation \(j\) is used as a match. To show that the remaining terms converge to zero, let we use the fact \(#\mathcal{I}_j\) is bounded by \(J\mathcal{L}\), where \(\mathcal{L}\) is the kissing number, defined as the maximum number of non-overlapping unit balls that can be arranged such that they each touch a common unit ball (Miller et al., 1997, Lemma 3.2.1; see also Abadie and Imbens, 2008). \(\mathcal{L}\) is a finite constant that depends only on the dimension of the covariates (for example, \(\mathcal{L} = 2\) if \(\text{dim}(x_i) = 1\)). Now,
\[
\sum_{i} a_{ni} T_{4i} = \frac{1}{J^2} \sum_{j=1}^{n} \left(u_j - \sigma_j^2\right) \sum_{i \in \mathcal{I}_j} a_{ni},
\]
and so by the von Bahr-Esseen inequality,
\[
E\left|\sum_{i} a_{ni} T_{4i}\right|^{1+1/2K} \leq \frac{2}{J^2+1/K} \sum_{j=1}^{n} E\left|u_j - \sigma_j^2\right|^{1+1/2K} \left(\sum_{i \in \mathcal{I}_j} a_{ni}\right)^{1+1/2K} \leq \frac{(J\mathcal{L})^{1/2K}}{J^2+1/K} \max_{k} E\left|u_k - \sigma_k^2\right|^{1+1/2K} \max_{i} a_{ni}^{1+1/2K} \sum_{j=1}^{n} \sum_{i \in \mathcal{I}_j} a_{ni},
\]
which is bounded by a constant times \(\max_{i} a_{ni}^{1+1/2K} \sum_{j=1}^{n} \sum_{i \in \mathcal{I}_j} a_{ni} = \max_{i} a_{ni}^{1+1/2K} J \sum_{i} a_{ni} \to 0\).

Next, since \(E[u_i u_{i'} u_{\ell_j(i)} u_{\ell_k(i')}\] is non-zero only if either \(i = i'\) and \(\ell_j(i) = \ell_k(i')\), or else if \(i = \ell_k(i')\) and \(i' = \ell_j(i)\), we have \(\sum_{i'=1}^{n} a_{ni'} E[u_i u_{i'} u_{\ell_k(i)} u_{\ell_k(i')}\] \(\leq \max_{i'} a_{ni'} \left(\sigma^2_{\ell_j(i)} \sigma^2_{\ell_k(i')} + \sigma^2_{\ell_j(i')} \sigma^2_{\ell_k(i)}\right)\), so that
\[
\text{var} \left(\sum_{i} a_{ni} T_{2i}\right) = \frac{1}{J^2} \sum_{i,j,k,i'} a_{ni} a_{ni'} E[u_i u_{i'} u_{\ell_k(i)} u_{\ell_k(i')}\] \(\leq 2K^2 \max_{i'} a_{ni'} \sum_{i} a_{ni} \to 0\).
\]

Similarly for \(j \neq k\) and \(j' \neq k\), \(\sum_{i'=1}^{n} a_{ni'} E[u_{\ell_j(i)} u_{\ell_k(i)} u_{\ell_j(i')} u_{\ell_k(i')}\] \(\leq \max_{i'} 2\sigma^2_{\ell_j(i)} \sigma^2_{\ell_k(i)}\), so that
\[
\text{var} \left(\sum_{i} a_{ni} T_{3i}\right) = \frac{1}{J^4} \sum_{i,i',j,j',k,k'=1}^{j,j'+1} a_{ni} a_{ni'} E[u_{\ell_j(i)} u_{\ell_k(i)} u_{\ell_j(i')} u_{\ell_k(i')}\] \(\leq 2K^2 \max_{i'} a_{ni'} \sum_{i} a_{ni} \to 0\).
References

