Online Appendix

Supplement to "Identification of Nonparametric Monotonic Regression Models with Continuous Nonclassical Measurement Errors"

Yingyao Hu, Susanne Schennach, and Ji-Liang Shiu

(Not for publication)

A. Alternative Proof of Theorem 2.1

We provide an alternative proof of the main nonparametric identification result in Theorem 2.1.

We first derive the basic integral equation that needs to be solved. Combining Assumptions 2.2(i) and 2.3(i), we can obtain the relationship between the observed density and the unobserved ones:

$$
\begin{aligned}
(1) \qquad f_{Y,X}(y,x) &= \int_{\mathcal{X}^*} f_{Y,X,X^*}(y,x,x^*)dx^* \\
&= \int_{\mathcal{X}^*} \frac{f_{Y,X,X^*}(y,x,x^*)}{f_{X,X^*}(x,x^*)} f_{X,X^*}(x,x^*)dx^* \\
&= \int_{\mathcal{X}^*} f_{Y|X^*}(y|x^*) f_{X,X^*}(x,x^*)dx^* \\
&= \int_{\mathcal{X}^*} f_\eta(y-m_0(x^*)) f_{X|X^*}(x|x^*) f_{X^*}(x^*)dx^*.
\end{aligned}
$$

Since a characteristic function of any random variable completely determines its probability distribution, the above equation is equivalent to

$$
\begin{aligned}
(2) \qquad \phi_{f_{Y,X=x}}(t) &\equiv \int_{\mathcal{Y}} e^{ity} f_{Y,X}(y,x)dy \\
&= \int_{\mathcal{Y}} e^{ity} \int_{\mathcal{X}^*} f_\eta(y-m_0(x^*)) f_{X|X^*}(x|x^*) f_{X^*}(x^*)dx^* dy \\
&= \int_{\mathcal{X}^*} \int_{\mathcal{Y}} e^{ity} f_\eta(y-m_0(x^*))dy f_{X|X^*}(x|x^*) f_{X^*}(x^*)dx^* \\
&= \phi_\eta(t) \int_{\mathcal{X}^*} e^{itm_0(x^*)} f_{X|X^*}(x|x^*) f_{X^*}(x^*)dx^*, \\
&= |\phi_\eta(t)| \int_{\mathcal{X}^*} e^{i(tm_0(x^*)+e(t))} f_{X|X^*}(x|x^*) f_{X^*}(x^*)dx^*,
\end{aligned}
$$

for all real-valued t, where $\phi_\eta(t) = \int_\eta e^{it\eta} f_\eta(\eta)d\eta$ and we define $e(t)$ such that the following holds $\phi_\eta(t) \equiv |\phi_\eta(t)|e^{ie(t)}$ and $e(t)$ is the phase of the function. Then Eq. (2) can expressed in terms of two real equations:

$$(3) \qquad Re\phi_{f_{Y,X=x}}(t) = |\phi_\eta(t)| \int_{\mathcal{X}^*} \cos(tm_0(x^*) + e(t)) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*,$$

$$(4) \qquad Im\phi_{f_{Y,X=x}}(t) = |\phi_\eta(t)| \int_{\mathcal{X}^*} \sin(tm_0(x^*) + e(t)) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*.$$

Without loss of generality, we can make the following assumption:

Assumption A.1. *(Locally symmetric range) The range of the regression function $\{m_0(x^*) : x^* \in \mathcal{X}^*\}$ has an open subset containing zero.*

Assumption A.1 is not restrictive because one may always shift the mean of the dependent variable Y and redefine the regression function accordingly. Also, the range of the regression is never reduced to a point, by the strict monotonicity imposed by Assumption 2.4.

Using Assumptions A.1 and 2.4 we can rescale the range of the regression function such that the range is equal to the interval $[-c, d]$ for positive constants c, d and $c+d < \pi$. Because $|\phi_\eta(t)|$ is continuous at 0 (a property of any characteristic function) and $|\phi(0)| = 1$, we can find a $\bar{t} \leq \pi$ such that $0 < |\phi_\eta(t)| < b_1$ for all t in $[0, \bar{t}]$ and a constant b_1. Denote the variance of the regression error as σ_η^2. Choose a constant t_u such that

$$0 < t_u < \min\left\{\bar{t}, \sqrt{\frac{2}{\sigma_\eta^2}}\right\}.$$

Use Eq. (3) to derive an operator equivalence relationship as following: for an arbitrary $h \in \mathcal{L}^2([0, t_u])$

$$\begin{aligned}
(5) \qquad (L_{Re\phi_{f_{Y,X}}} h)(x) &= \int Re\phi_{f_{Y,X=x}}(t) h(t) dt, \\
&= \int |\phi_\eta(t)| \int_{\mathcal{X}^*} \cos(tm_0(X^*) + e(t)) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^* h(t) dt \\
&= \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*) f_{X^*}(x^*) \left(\int \cos(tm_0(x^*) + e(t)) |\phi_\eta(t)| h(t) dt \right) dx^* \\
&= \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*) f_{X^*}(x^*) \left(\int \cos(tm_0(x^*) + e(t)) (\Delta_{|\phi_\eta|} h)(t) dt \right) dx^* \\
&= \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*) f_{X^*}(x^*) \left(L_{\cos_{m_0,e}} \Delta_{|\phi_\eta|} h)(x^*) \right) dx^* \\
&= \left(L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\cos_{m_0,e}} \Delta_{|\phi_\eta|} h \right)(x),
\end{aligned}$$

where we have used (i) Eq. (3), (ii) an interchange of the order of integration (justified by

Fubini's theorem), (iii) the definition of $\Delta_{|\phi_\eta|}$, (iv) the definition of $L_{\cos_{m_0,e}}$ operating on the function $\Delta_{|\phi_\eta|}h$, and (v) the definition of $L_{f_{X|X^*}}\Delta_{f_{X^*}}$ operating on the function $L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|}h$. Thus, we obtain

$$(6) \qquad L_{Re\phi_{f_{Y,X}}} = L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|} \equiv L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{Re\phi_{f_{Y|X^*}}},$$

We can also express Eq. (4) as the following operator equivalence relationships:

$$(7) \qquad L_{Im\phi_{f_{Y,X}}} = L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\sin_{m_0,e}}\Delta_{|\phi_\eta|} \equiv L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{Im\phi_{f_{Y|X^*}}}.$$

Both $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$ are bounded linear operators from $\mathcal{L}^2([0,t_u])$ to $\mathcal{L}^2(\mathcal{X}^*)$ because operators in the right hand side are all bounded by Assumption 2.1 and continuity of characteristic functions.

Our identification technique is to derive a spectral decomposition of an observed integral operator and show the uniqueness of the decomposition under our assumptions. We can derive some primitive conditions for the invertibility of the operators $L_{Re\phi_{f_{Y,X}}}$, and $L_{Im\phi_{f_{Y,X}}}$ which are related to the invertibility of the operator $L_{f_{X|X^*}}$ and the invertibility of the operators $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$.

Lemma A.1. *Assumptions 2.1 and 2.3(ii), $L^{-1}_{f_{X|X^*}}$ exists and is densely defined over $\mathcal{L}^2(\mathcal{X})$.*

The proof of Lemma A.1.1. Because the orthogonal complement of the null space of $L_{f_{X|X^*}}$, denoted $\mathcal{N}(L_{f_{X|X^*}})^\perp$, is the closure of the range of $L^*_{f_{X|X^*}}$ [1] and $L_{f_{X|X^*}}$ is one-to-one over $\mathcal{N}(L_{f_{X|X^*}})^\perp$, $L^{-1}_{f_{X|X^*}}$ exists.

Next, we show that the adjoint of $L_{f_{X|X^*}}$ also admits an integral representation. Because $L_{f_{X|X^*}}$ is a bounded linear operator, there exists a unique adjoint operator $L^*_{f_{X|X^*}}$ such that

$$\langle L_{f_{X|X^*}}(f_1), f_2\rangle = \langle f_1, L^*_{f_{X|X^*}}(f_2)\rangle.$$

We have

$$\langle L_{f_{X|X^*}}(f_1), f_2\rangle = \int_{\mathcal{X}}\int_{\mathcal{X}^*} f_{X|X^*}(x|x^*)f_1(x^*)f_2(x)dx^*dx.$$

[1] Lemma VI. 2.8 in Dunford and Schwartz (1971).

A direct verification gives the integral representation for $L^*_{f_{X|X^*}}$,

$$(L^*_{f_{X|X^*}} f_2)(x^*) = \int f_{X|X^*}(x|x^*) f_2(x) dx. \tag{8}$$

Assumption 2.3(ii) implies that the adjoint operator $L^*_{f_{X|X^*}}$ is one-to-one. By the part (b) of Corollaries of Theorem 4.12 in Rudin (1991), the range of $L_{f_{X|X^*}}$ is dense in $\mathcal{L}^2(\mathcal{X})$. Therefore, $L^{-1}_{f_{X|X^*}}$ can be densely defined over $\mathcal{L}^2(\mathcal{X})$. QED.

The above result show $L_{f_{X|X^*}}$ is onto and the injectivity of the operators $L_{f_{X|X^*}}$ is directly assumed from the first part of Assumption 2.3(ii). Therefore, $L^{-1}_{f_{X|X^*}}$ exists and $L^{-1}_{f_{X|X^*}} L_{f_{X|X^*}} = L_{f_{X|X^*}} L^{-1}_{f_{X|X^*}} = I$ where I is the identity map from $\mathcal{L}^2(\mathcal{X}^*)$ to itself. The discussion hereafter focuses on the conditions for the completeness of $L_{\cos_{m_0,e}}$, and $L_{\sin_{m_0,e}}$. Define $ns(t) = 1 - \cos(e(t)) = \frac{|\phi_\eta| - Re(\phi_\eta)}{|\phi_\eta|}$ as a measure of degree of non-symmetry. If the distribution of the error term η is symmetric then $\phi_\eta(t)$ is real-valued and $ns(t) = 0$ for $t \in [0, t_u]$. Continuity of characteristic functions and Assumption A.1 are sufficient conditions for the invertibility of the operators $L_{\cos_{m_0,e}}$, and $L_{\sin_{m_0,e}}$. We have

Lemma A.2. *If Assumption A.1 holds, then each of systems,* $\{\cos(tm_0(x^*) + e(t))|\phi_\eta(t)| : x^* \in \mathcal{X}^*\}$ *and* $\{\sin(tm_0(x^*) + e(t))|\phi_\eta(t)| : x^* \in \mathcal{X}^*\}$*, is complete over* $\mathcal{L}^2([0, t_u])$*. This implies the operators* $L_{Re\phi_{f_{Y|X^*}}}$ *and* $L_{Im\phi_{f_{Y|X^*}}}$ *are both injective from* $\mathcal{L}^2([0, t_u])$ *to* $\mathcal{L}^2(\mathcal{X}^*)$*.*

The injectivity implies the inverses of $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$ exist and can defined over the range of the operators. To show this primitive conditions for the invertibility, we utilize results from Fourier analysis. We provide the following result of the trigonometric system.[2]

Lemma A.3. *If* $1 < p < \infty$ *and* λ_k *is a sequence of distinct real or complex numbers for which*

$$|\lambda_k| \leq k + \frac{1}{2p}, \qquad k = 1, 2, 3, ..., \tag{9}$$

then the sequence $\{e^{it\lambda_k}\}_{k=1}^{\infty}$ *is complete in* $\mathcal{L}^p([-\pi, \pi])$*.*

We can directly use this neat result to establish the following completeness.

[2]See Theorem 4 of page 119 in Young (1980).

Lemma A.4. *If the range of the regression function $\{m_0(x^*) : x^* \in \mathcal{X}^*\}$ contains a sequence of distinct numbers $\{\lambda_1, \lambda_2, \lambda_3, ...\}$ such that*

$$|\lambda_k| \leq k + \frac{1}{4}, \qquad k = 1, 2, 3, ..., \tag{10}$$

then the family of the functions $\{e^{itm_0(x^*)} : x^* \in \mathcal{X}^*\}$ is complete in $\mathcal{L}^2([-\pi, \pi])$.

Next, we establish the completeness of the two systems: $\{\cos(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ and $\{\sin(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ over $\mathcal{L}^2([0, t_u])$.

Lemma A.5. *If the range of the regression function $\{m_0(x^*) : x^* \in \mathcal{X}^*\}$ contains a sequence of distinct numbers $\{\lambda_1, \lambda_2, \lambda_3, ...\}$ such that*

$$|\lambda_k| \leq k + \frac{1}{4}, \qquad k = 1, 2, 3, ..., \tag{11}$$

then the families of the functions $\{\cos(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ and $\{\sin(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ are complete in $\mathcal{L}^2([0, t_u])$ for any $t_u \leq \pi$.

The proof of Lemma A.5. This proof is established by the claim that the two families $\{\cos(t\lambda_k) : k = 1, 2, 3, ...\}$ and $\{\sin(t\lambda_k) : k = 1, 2, 3, ...\}$ are complete over $\mathcal{L}^2([0, \pi])$ with $|\lambda_k| \leq k + \frac{1}{4}, \quad k = 1, 2, 3, ...,$. Suppose that $\int_{[0,\pi]} \cos(t\lambda_k) h_2(t) dt = 0$ for some $h_2 \in \mathcal{L}^2([0, \pi])$. Then we can extend the function h_2 to an even function in $\mathcal{L}^2([-\pi, \pi])$ by $h_2(-x) = h_2(x)$. It follows that $\int_{[-\pi,0]} \cos(t\lambda_k) h_2(t) dt = \int_{[0,\pi]} \cos(t\lambda_k) h_2(t) dt = 0$ and $\int_{[-\pi,0]} \sin(t\lambda_k) h_2(t) dt = -\int_{[0,\pi]} \sin(t\lambda_k) h_2(t) dt$. Thus, $\int_{[-\pi,\pi]} \cos(t\lambda_k) h_2(t) dt = 0$ and $\int_{[-\pi,\pi]} \sin(t\lambda_k) h_2(t) dt = 0$. This implies $\int_{[-\pi,\pi]} [\cos(t\lambda_k) + i \sin(t\lambda_k)] h_2(t) dt = \int_{[-\pi,\pi]} e^{it\lambda_k} h_2(t) dt = 0$. The cosine part of the claim can be ensured by the completeness of $\{e^{it\lambda_k} : k = 1, 2, 3, ...\}$ over $\mathcal{L}^2([-\pi, \pi])$ with $|\lambda_k| \leq k + \frac{1}{4}, \quad k = 1, 2, 3, ...,$ in Lemma A.4. The sine part of the claim can be reached by the observation that we can extend a function h_3 over $[0, \pi]$ to an odd function over $[-\pi, \pi]$ by $h_3(-x) = -h_3(x)$. Therefore, we have reached the families of the functions $\{\cos(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ and $\{\sin(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ are complete in $\mathcal{L}^2([0, t_b])$ because if a family is complete for functions over $[0, \pi]$, then it is complete over $[0, t_b]$ with $t_b < \pi$. QED.

The above gives the invertibility of the operators $L_{\cos_{m_0,e}}$, and $L_{\sin_{m_0,e}}$ under the symmetric distribution of the regression error η, i.e., $e(t) = 0 \ \forall t$. Next, we try to generalize the invertibility or completeness of the symmetric case to a non-symmetric case. Comparing the function in the symmetric case $\cos(tm_0(x^*))$ with the function in the non-symmetric case

$\cos(tm_0(x^*) + e(t))$ suggests that we can look into a situation where $e(t)$ is under "small" perturbations around zero (symmetry; $e(t) = 0\ \forall t$) and investigate what restrictions on the range of $e(t)$ leads to invertibility of operators. In this way, questions about "small" perturbations can be regarded as questions about the stability of completeness because we have already provided a sufficient condition for the symmetric case in Lemma A.5. We will adopt a stability criterion to study completeness. The following result can be found in Young (1980).[3]

Lemma A.6. *Let $\{b_k\}$ be a complete sequence for a Hilbert space $(H, \|\cdot\|)$, and suppose that $\{f_k\}$ is sequence of elements of H such that*

$$\|\sum_{k=1}^{n} c_k(b_k - f_k)\| \leq \lambda \|\sum_{k=1}^{n} c_k b_k\|$$

for some constant $0 \leq \lambda < 1$, and all choices of the scalar $\{c_1, c_2, c_3, ..., c_n\}$, Then $\{f_k\}$ is complete for H.

Lemma A.6 is based on the fact that a bounded linear operator T on a Banach space is invertible whenever $\|I - T\| < 1$ because the inverse operator of T can exist by the formula $T^{-1} = \sum_{k=0}^{\infty}(I-T)^k$.[4] Define $ns(t) = 1 - \cos(e(t)) = \frac{|\phi_\eta| - Re(\phi_\eta)}{|\phi_\eta|}$ as a measure of degree of non-symmetry. If the distribution of the error term η is symmetric then $\phi_\eta(t)$ is real-valued and $ns(t) = 0$. The following result provides an upper bound on the absolute values of $ns(t)$ and it will be used to prove Lemma A.2.

Lemma A.7. *For $t \in [0, t_u]$, $ns(t)$ is nonnegative and its maximum is less than 1 .*

The proof of Lemma A.7. Suppose $\phi_s(t)$ is the characteristic function of a random variable s. If s has zero expectation and finite variance σ_s^2, then the real part of the characteristic function satisfies the following inequality,[5]

$$Re\phi_s(t) \geq 1 - \frac{\sigma_s^2 t^2}{2}. \tag{12}$$

Because η has zero expectation, we can apply the result. By the definition of t_u, we obtain

$$Re\phi_\eta(t) \geq 1 - \frac{\sigma_\eta^2 t^2}{2} > 1 - \frac{\sigma_\eta^2 t_u^2}{2} > 0 \text{ for } t \in [0, t_u]. \tag{13}$$

[3]See Problem 2 in page 41. The result is stated for a Banach space and the dense property. Here we adopt Hilbert space version by an important consequence of the Hahn-Banach theorem and the Riesz representation theorem that the dense property is equivalent to the completeness in a Hilbert space.

[4]The result is like ordinary numbers: if $|1-t| < 1$, then t^{-1} exists. More discussions can be found in Young (1980).

[5]See Theorem 2.3.2 of page 89 in Ushakov (1999).

This implies $0 < c_\eta < \frac{Re(\phi_\eta)}{|\phi_\eta|} \leq 1$ for some positive constant c_η over $[0, t_u]$ and then $0 \leq ns(t) = 1 - \frac{Re(\phi_\eta)}{|\phi_\eta|} < 1 - c_\eta$ for $t \in [0, t_u]$. QED.

Applying the stability criterion and Lemma A.7 to Lemma A.5 under Assumptions A.1 and 2.4, we can prove Lemma A.2.

The proof of Lemma A.2. First, if the range contains any open set with zero, we can always pick a sequence of distinct numbers $\{\pm\lambda_1, \pm\lambda_2, \pm\lambda_3, ...\}$ such that $|\lambda_k| \leq k + \frac{1}{4}, \quad k = 1, 2, 3, ...$ We use subsequence of this sequence of distinct numbers to show that $\{\cos(t\lambda_k)\cos(e(t)) : k = 1, 2, 3, ...\}$ is complete for $\mathcal{L}^2([0, t_u])$ by applying Lemma A.6 to $b_k(t) = \cos(t\lambda_k)$ and $f_k(t) = \cos(t\lambda_k)\cos(e(t))$ with $|\lambda_k| \leq k + \frac{1}{4}, \quad k = 1, 2, 3, ...$ Write $c_k(b_k(t) - f_k(t)) = ns(t)c_k\cos(t\lambda_k)$ for some constant c_k. By Lemma A.7, $|ns(t)| < \lambda < 1$ for $t \in [0, t_u]$ for some constant λ. For all choices of the scalars $\{c_1, c_2, c_3, ..., c_n\}$,

$$\| \sum_{k=1}^{n} c_k(b_k(\cdot) - f_k(\cdot))\| = \|ns(\cdot)\sum_{k=1}^{n} c_k b_k(\cdot)\| < \lambda \|\sum_{k=1}^{n} c_k b_k(\cdot)\|, \tag{14}$$

with $\lambda < 1$. Because $\{\cos(t\lambda_k) : k = 1, 2, 3, ...\}$ is complete by Lemma A.5, $\{\cos(t\lambda_k)\cos(e(t)) : k = 1, 2, 3, ...\}$ is complete by the stability criterion in Lemma A.6.

The next step is to show $\{\cos(t\lambda + e(t)) : \lambda = \pm\lambda_1, \pm\lambda_2, \pm\lambda_3, ...\}$ is complete over $\mathcal{L}^2([0, t_u])$. Suppose that there exists h_2 such that $\int \cos(t\lambda_k + e(t))h_2(t)dt = 0$, $\forall \lambda = \pm\lambda_1, \pm\lambda_2, \pm\lambda_3, ...$ Recall that $\cos(t\lambda + e(t)) = \cos(t\lambda)\cos(e(t)) - \sin(t\lambda)\sin(e(t))$. Plugging $\lambda = \lambda_1$ and $\lambda = -\lambda_1$ back to $\int \cos(t\lambda_k + e(t))h_2(t)dt = 0$ and then summing up those two identities, we obtain $\int \cos(t\lambda_1)\cos(e(t))h_2(t)dt = 0$. Applying the same derivation to $\pm\lambda_2, \pm\lambda_3, ...,$ yields $\int \cos(t\lambda_k)\cos(e(t))h_2(t)dt = 0$, $k = 1, 2, 3, ...$ Since we have shown $\{\cos(t\lambda_k)\cos(e(t)) : k = 1, 2, 3, ...\}$ is complete, it follows that $h_2 = 0$ which proves that completeness of $\{\cos(tm_0(x^*) + e(t)) : x^* \in \mathcal{X}^*\}$. Since $|\phi_\eta(t)| \neq 0$ for $[0, t_u]$ by continuity , $\{\cos(tm_0(x^*) + e(t))|\phi_\eta(t)| : x^* \in \mathcal{X}^*\}$ is also complete. As for the sine part, in a similar manner, we first show $\{\sin(t\lambda_k)\cos(e(t)) : k = 1, 2, 3, ...\}$ is complete by the stability criterion in Lemma A.6 and then utilize $|\phi_\eta(t)| \neq 0$ for $[0, t_u]$ and the identities $\sin(t\lambda_k + e(t)) = \sin(t\lambda_k)\cos(e(t)) + \cos(t\lambda_k)\sin(e(t))$ and $\sin(-t\lambda_k + e(t)) = -\sin(t\lambda_k)\cos(e(t)) + \cos(t\lambda_k)\sin(e(t))$ to show the completeness of $\{\sin(tm_0(x^*) + e(t))|\phi_\eta(t)| : x^* \in \mathcal{X}^*\}$. QED.

In order to provide the onto property of the operators $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$, we need a variant of the stability result as in Lemma A.6. To provide the result, we introduce the following notations and statements. Any function f in a Hilbert space can be expressed as a linear combination of the basis function with a unique sequence of scalars $\{c_1, c_2, c_3, ...\}$. Therefore, we can consider c_n as a function of f. In fact, $c_n(\cdot)$ is the so-called coefficient functional.[6]

Definition A.1. *If $\{f_1, f_2, f_3, ...\}$ is a basis in a Hilbert space $\mathcal{H}$, then every function f in $\mathcal{H}$ has a unique series $\{c_1, c_2, c_3, ...\}$ such that*

$$f = \sum_{n=1}^{\infty} c_n(f) f_n.$$

Each c_n is a function of f. The functionals c_n $(n = 1, 2, 3, ...)$ are called the coefficient functionals associated with the basis $\{f_1, f_2, f_3, ...\}$. Because c_n is a coefficient functional from $\mathcal{H}$ to $\mathbb{R}$. Define its norm by

$$\|c_n\| = \sup\{|c_n(f)| : f \in \mathcal{H}, \quad \|f\| \le 1\}.$$

The following results regarding the coefficient functionals are from Theorem 3 in section 6 in Young (1980).

Lemma A.8. *If $\{f_1, f_2, f_3, ...\}$ is a basis in a Hilbert space $\mathcal{H}$. Define c_n as coefficient functionals associated with the basis. Then, there exists a constant M such that*

$$1 \le \|f_n\| \cdot \|c_n\| \le M, \tag{15}$$

for all n.

Lemma A.9. *Denote H as a Hilbert space. Suppose that*

i) the sequence $\{e_k(\cdot) : k = 1, 2, ...\}$ is a basis in a Hilbert space $\mathcal{H}$;

ii) the sequence $\{f_k(\cdot) : k = 1, 2, ...\}$ in $\mathcal{H}$ is $\omega-$independent;

iii) $\sum_{n=1}^{\infty} \frac{\|f_k(\cdot) - e_k(\cdot)\|}{\|e_k(\cdot)\|} < \infty$.

Then, the sequence $\{f_k(\cdot) : k = 1, 2, ...\}$ is a basis in $\mathcal{H}$.

Proof of Lemma A.9: Consider for any function $f \in \mathcal{H}$

$$f = \sum_{n=1}^{\infty} c_n(f) e_n,$$

[6]The introduction of coefficient functional can be found in the page 22 of Young (1980).

where $c_n(f)$ is the coefficient functional corresponding to the basis $\{e_n\}$. It is clear that $c_n(f)$ is a linear function of f. We mimic the proof of Theorem 12 in Young (1980). Consider

$$\begin{aligned}
\left\|\sum_{n=1}^{\infty} c_n(f)\,(e_n - f_n)\right\| &\leq \sum_{n=1}^{\infty} \|c_n(f)\,(e_n - f_n)\| \\
&\leq \left(\sum_{n=1}^{\infty} \|e_n - f_n\|\,\|c_n\|\right)\|f\| \\
&\leq \left(\sum_{n=1}^{\infty} \frac{\|e_n - f_n\|}{\|e_n\|}\,\|e_n\|\,\|c_n\|\right)\|f\| \\
&\leq M\left(\sum_{n=1}^{\infty} \frac{\|e_n - f_n\|}{\|e_n\|}\right)\|f\|,
\end{aligned}$$

where we have used (i) the triangle inequality, (ii) the definition of functional, and (iii) Lemma A.8, $1 \leq \|e_n\|\,\|c_n\| \leq M$. Applying the condition (iii) $\sum_{n=1}^{\infty} \frac{\|e_n - f_n\|}{\|e_n\|} < \infty$ to the above inequality suggests that the infinite sum $\sum_{n=1}^{\infty} c_n(f)\,(e_n - f_n)$ is absolutely converge. Hence, define an operator $T : \mathcal{H} \to \mathcal{H}$ as

$$Tf = \sum_{n=1}^{\infty} c_n(f)\,(e_n - f_n)\,.$$

It is clear that T is linear. Since $c_n(e_n) = 1$ and $c_k(e_n) = 0$ for $k \neq n$, we have

$$Te_n = \sum_{n=1}^{\infty} c_n(e_n)\,(e_n - f_n) = e_n - f_n.$$

The relationship above implies that the linear operator T is bounded if

$$\sum_{n=1}^{\infty} \frac{\|e_n - f_n\|}{\|e_n\|} < \infty,$$

which is the condition (iii). We then show that T is a compact operator. Set

$$T_N f = \sum_{n=1}^{N} c_n(f)\,(e_n - f_n)\,.$$

Start with

$$\begin{aligned}
\|(T-T_n)f\| &= \left\|\sum_{n=N+1}^{\infty} c_n(f)\,(e_n-f_n)\right\| \\
&\le \sum_{n=N+1}^{\infty} \|c_n(f)\,(e_n-f_n)\| \\
&\le \left(\sum_{n=N+1}^{\infty} \|e_n-f_n\|\,\|c_n\|\right)\|f\|\,.
\end{aligned}$$

Follow the previous derivation of $\|\sum_{n=1}^{\infty} c_n(f)\,(e_n-f_n)\|$, we can obtain

$$\|(T-T_n)f\| \le M\left(\sum_{n=N+1}^{\infty} \frac{\|e_n-f_n\|}{\|e_n\|}\right)\|f\|\,.$$

This implies that $\|(T-T_n)\| \le M\left(\sum_{n=N+1}^{\infty} \frac{\|e_n-f_n\|}{\|e_n\|}\right)$. Assumption iii) of Lemma 10 suggests that $\|T-T_n\| \to 0$. Since each T_N has finite dimensional range and $\|T-T_N\| \to 0$ as $N \to \infty$, T is an compact operator.[7]

Next, we show that $Ker(I-T) = \{0\}$, i.e., $(I-T)$ is invertible. Consider

$$\begin{aligned}
0 &= (I-T)\,f \\
&= f - \sum_{n=1}^{\infty} c_n(f)\,(e_n-f_n) \\
&= \sum_{n=1}^{\infty} c_n(f)e_n - \sum_{n=1}^{\infty} c_n(f)e_n + \sum_{n=1}^{\infty} c_n(f)f_n \\
&= \sum_{n=1}^{\infty} c_n(f)f_n
\end{aligned}$$

Since $\{f_n(\cdot)\}$ is an $\omega-$independent sequence, we have $c_n(f) = 0$ for all n, and therefore, $0 = (I-T)\,f$ implies $f = 0$.

Therefore, T is a compact operator defined in a Hilbert space $\mathcal{H}$ with $Ker(I-T) = \{0\}$. Since T is bounded, $(I-T)$ is also bounded. By the Fredholm alternative, this shows that $(I-T)$ is a bounded invertible operator.[8] Clearly, we have $(I-T)\,e_n = f_n$. Consider any $h \in \mathcal{H}$. Then, $(I-T)^{-1}h$ has an unique series expression $(I-T)^{-1}h = \sum_{n=1}^{\infty} c_n e_n$ since $\{e_n(\cdot)\}$ is a basis. Since $(I-T)$ is bounded, applying $(I-T)$ to the expression above results

[7] If an bounded linear operator T is the limit of operators of finite rank, then T is compact. See Exercise 13 on page 112 in Rudin (1991).

[8] See the Fredholm alternative in Rudin (1991), Exercise 13 on page 112.

in $h = \sum_{n=1}^{\infty} c_n f_n$. This series expansion is unique because $\{f_n(\cdot)\}$ is ω-independent. The argument above shows that every element $h \in \mathcal{H}$ has a unique series expansion in terms of f_n. Thus, $\{f_n(\cdot)\}$ is also a basis for $\mathcal{H}$. QED

Lemma A.10. *If Assumptions A.1 and 2.4 hold, then each of systems,* $\{\cos(tm_0(x^*)+e(t)) : t \in [0, t_u]\}$ *and* $\{\sin(tm_0(x^*)+e(t)) : t \in [0, t_u]\}$*, is complete over* $\mathcal{L}^2(\mathcal{X}^*)$*. This implies that the inverse operators* $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ *and* $L^{-1}_{Im\phi_{f_{Y|X^*}}}$ *exist and are densely defined over* $\mathcal{L}^2(\mathcal{X}^*)$*.*

The proof of Lemma A.10. Set $e_k(w) = \cos(t_k w)$ and $f_k(w) = \cos(t_k w - t_k c + e(t_k))$. Define the relative deviation as $\frac{\|e_k(\cdot)-f_k(\cdot)\|}{\|e_k(\cdot)\|}$. By Lemma A.5, we can have the completeness of the system, $\{\cos(t_k w) : k = 1, 2, 3, ...\}$ over $\mathcal{L}^2([0, c+d])$ for a sequence of distinct numbers $\{t_1, t_2, t_3, ...\} \subset [0, t_u]$ converging to 0 such that $\frac{\|e_k(\cdot)-1\|}{\|e_k(\cdot)\|} < 1/2^k$ and $\frac{\|1-f_k(\cdot)\|}{\|e_k(\cdot)\|} < 1/2^k$. The existence of such subsequence comes from $t_k \to 0$ as $k \to \infty$ and the relative deviation is continuous and has zero value at 0. Because a complete sequence contain a subsequence as a basis, we can extract a subsequence $\{t_{s1}, t_{s2}, t_{s3}, ...\}$ such that $\{\cos(t_{sk} w) : k = 1, 2, 3, ...\}$ is a basis over $\mathcal{L}^2([0, c+d])$. According to the second Theorem in Erdös and Straus (1953), any linearly independent sequence in a normed space contains an $\omega-$ independent subsequence. Because $\{\cos(t_k w - t_k c + e(t_k)) : k = 1, 2, 3, ...\}$ is linear independent for any sequence of distinct numbers $\{t_1, t_2, t_3, ...\}$[9], we can extract a subsequence $\{t_{l1}, t_{l2}, t_{l3}, ...\}$ such that $\{\cos(t_{lk} w - t_{lk} c + e(t_{lk})) : k = 1, 2, 3, ...\}$ is $\omega-$ independent. Next, we try to apply Lemma A.9 to $b_{sk}(w) = \cos(t_{sk} w)$ and $f_{lk}(w) = \cos(t_{lk} w - t_{lk} c + e(t_{lk}))$ with the total deviation $\sum_{k=1}^{\infty} \frac{\|b_{sk}(\cdot)-f_{lk}(\cdot)\|}{\|b_{sk}(\cdot)\|} = \sum_{k=1}^{\infty} \frac{\|b_{sk}(\cdot)-1\|}{\|b_{sk}(\cdot)\|} + \frac{\|1-f_{lk}(\cdot)\|}{\|b_{sk}(\cdot)\|} < \infty$. Therefore, the sequence $\{\cos(t_{lk} w - t_{lk} c + e(t_{lk})) : k = 1, 2, 3, ...\}$ contain a basis and then the sequence is complete over $\mathcal{L}^2([0, c+d])$. Similarly, we can use the completeness of $\{\sin(t_k w) : k = 1, 2, 3, ...\}$ to show the system $\{\sin(t_k w - t_k c + e(t_k)) : k = 1, 2, 3, ...\}$ is complete over $\mathcal{L}^2([0, c+d])$.

Suppose there exists $h \in \mathcal{L}^2(\mathcal{X}^*)$ such that $\int_{\mathcal{X}^*} \cos(t_k m_0(x^*) + e(t_k)) h(x^*) dx^* = 0$ for $k = 1, 2, 3, ...$ By the monotonicity in Assumption 2.4, we can do the change of the variables between x^* and w using $w = m_0(x^*)+c$. This yields $\int_{[0,c+d]} \cos(t_k w - t_k c + e(t_k)) \frac{h(m_0^{-1}(w-c))}{m_0'(m_0^{-1}(w-c))} dw = 0$ for $k = 1, 2, 3, ...$ The completeness of $\{\cos(t_k w - t_k c + e(t_k)) : k = 1, 2, 3, ...\}$ and the monotonicity of m implies $h = 0$ and then we achieve the completeness of $\{\cos(tm_0(x^*)+e(t)) : t \in [0, t_u]\}$ over $\mathcal{L}^2(\mathcal{X}^*)$. Similarly, we have the completeness of $\{\sin(tm_0(x^*) + e(t)) : t \in [0, t_u]\}$ over $\mathcal{L}^2(\mathcal{X}^*)$.

[9] The 2, 4,..., 2(K-1) times differentiation of these functions can be expressed as a Vandermonde matrix whose determinant is non-zero and this leads to linear independence of the system.

Following the derivation in Lemma A.1, we know the adjoint operators of $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$ are as the following:

$$L^*_{Re\phi_{f_{Y|X^*}}} : \mathcal{L}^2(\mathcal{X}^*) \to \mathcal{L}^2([0, t_u]) \text{ with } (L^*_{\cos_{m_0,e}} h)(t) = \int \cos(tm_0(x^*) + e(t))|\phi_\eta(t)|h(x^*)dx^*,$$

$$L^*_{Im\phi_{f_{Y|X^*}}} : \mathcal{L}^2(\mathcal{X}^*) \to \mathcal{L}^2([0, t_u]) \text{ with } (L^*_{\sin_{m_0,e}} h)(t) = \int \sin(tm_0(x^*) + e(t))|\phi_\eta(t)|h(x^*)dx^*.$$

The completeness of the systems $\{\cos(tm_0(x^*) + e(t)) : t \in [0, t_u]\}$ and $\{\sin(tm_0(x^*) + e(t)) : t \in [0, t_u]\}$ and $0 < |\phi_\eta(t)| < b_1$ for all t in $[0, t_u]$ implies these adjoint operators $L^*_{Re\phi_{f_{Y|X^*}}}$ and $L^*_{Im\phi_{f_{Y|X^*}}}$ are one-to-one from $\mathcal{L}^2(\mathcal{X}^*)$ to $\mathcal{L}^2([0, t_u])$. By the part (b) of Corollaries of Theorem 4.12 in Rudin (1991), the ranges of $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$ are dense in $\mathcal{L}^2(\mathcal{X}^*)$.[10] This implies that the inverse operators $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ and $L^{-1}_{Im\phi_{f_{Y|X^*}}}$ exist and are densely defined over $\mathcal{L}^2(\mathcal{X}^*)$. QED.

The completeness results in Lemma A.2 imply the injectivity of $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$ while Lemma A.10 gives the onto property of these operators. Therefore, the operators invertible with $L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{Re\phi_{f_{Y|X^*}}} = L_{Re\phi_{f_{Y|X^*}}} L^{-1}_{Re\phi_{f_{Y|X^*}}} = I$ and $L^{-1}_{Im\phi_{f_{Y|X^*}}} L_{Im\phi_{f_{Y|X^*}}} = L_{Im\phi_{f_{Y|X^*}}} L^{-1}_{Im\phi_{f_{Y|X^*}}} = I$, where I is the identity map from $\mathcal{L}^2([0, t_u])$ to itself.

Define L_{K_1} as

$$L_{K_1} = L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{Im\phi_{f_{Y|X^*}}}$$

by the existence of $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ over $\mathcal{L}^2(\mathcal{X}^*)$ by Lemma A.10. We can elicit simpler representations of the operator L_{K_1} under Assumption A.1. Furthermore, this simpler representation of L_{K_1} implies the angle function $e(t)$ is identified.

Lemma A.11. *If Assumption A.1 holds, then L_{K_1} is a multiplier operator such that $(L_{K_1}h)(t) = \tan(e(t))h(t)$ or $(L_{K_1}h)(t) = \frac{Im\phi_\eta(t)}{Re\phi_\eta(t)}h(t)$ for $t \in [0, t_u]$.*

The proof of Lemma A.11. By the definition of L_{K_1}, we have $L_{Re\phi_{f_{Y|X^*}}} L_{K_1} = L_{Im\phi_{f_{Y|X^*}}}$, i.e. $\int \cos(tm_0(x^*) + e(t))|\phi_\eta(t)|(L_{K_1}h)(t)dt = \int \sin(tm_0(x^*) + e(t))|\phi_\eta(t)|h(t)dt$ for all $h(t) \in \mathcal{L}^2([0, t_u])$. Since $\{m_0(x^*) : x^* \in \mathcal{X}^*\}$ contains an open set with zero, there exists a sequence of distinct numbers $\{\pm\lambda_1, \pm\lambda_2, \pm\lambda_3, ...\}$ in the range. We can plug $m_0(x^*) = \lambda_k$ and $m_0(x^*) = -\lambda_k$ for all $k \in \{1, 2, 3, ...\}$ into the equation and then sum up those plugged equations. Using two identities, $\cos(t\lambda_k + e(t)) + \cos(-t\lambda_k + e(t)) = 2\cos(t\lambda_k)\cos(e(t))$ and

[10] The statement of the part (b) of Corollaries of Theorem 4.12 in Rudin (1991) is the following: Suppose X and Y are Banach spaces, and T is a bounded linear operator from X to Y. Then the range of T is dense in Y if and only if its adjoint T^* is one-to-one.

$\sin(t\lambda_k + e(t)) + \sin(-t\lambda_k + e(t)) = 2\cos(t\lambda_k)\sin(e(t))$, we obtain

$$(16)\qquad 2\int \cos(t\lambda_k)\cos(e(t))|\phi_\eta(t)|(L_{K_1}h)(t)dt = 2\int \cos(t\lambda_k)\sin(e(t))|\phi_\eta(t)|h(t)dt.$$

Rearranging the term, we have

$$(17)\qquad \int \cos(t\lambda_k)|\phi_\eta(t)|\left[\cos(e(t))(L_{K_1}h)(t) - \sin(e(t))h(t)\right]dt = 0.$$

By the completeness of the system $\{\cos(t\lambda_k) : k = 1, 2, 3, ...\}$ and $|\phi_\eta(t)| \neq 0$ over $[0, t_u]$, we have $(L_{K_1}h)(t) = \tan(e(t))h(t) = \frac{Im\phi_\eta(t)}{Re\phi_\eta(t)}h(t)$. QED.

We now are ready to prove the main theorem.

Alternate proof of Theorem 2.1. We start with the operator equivalence relationships in Eqs. (6) and (7):

$$\begin{aligned} L_{Re\phi_{f_{Y,X}}} &= L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|} \equiv L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{Re\phi_{f_{Y|X^*}}}, \\ L_{Im\phi_{f_{Y,X}}} &= L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\sin_{m_0,e}}\Delta_{|\phi_\eta|} \equiv L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{Im\phi_{f_{Y|X^*}}}, \end{aligned}$$

Those operator equivalence relationships may not provide enough information to derive the spectral decomposition of the operator of interest. In order to solicit more useful operator equivalence relationships, we take derivative with respect to t in Eq. (2). It gives that

$$\begin{aligned} (18)\qquad \frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t) = &\left(\frac{\partial}{\partial t}|\phi_\eta(t)|\right)\int_{\mathcal{X}^*} e^{i(tm_0(x^*)+e(t))}f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\ &+ i\left(\frac{\partial}{\partial t}e(t)\right)|\phi_\eta(t)|\int_{\mathcal{X}^*} e^{i(tm_0(x^*)+e(t))}f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\ &+ i|\phi_\eta(t)|\int_{\mathcal{X}^*} e^{i(tm_0(x^*)+e(t))}m_0(x^*)f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^*. \end{aligned}$$

We split Eq. (18) into a real part and an imaginary part:

$$
\begin{aligned}
(19)\quad Re\frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t) = & \left(\frac{\partial}{\partial t}|\phi_\eta(t)|\right)\int_{\mathcal{X}^*}\cos(tm_0(x^*)+e(t))f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\
& -\left(\frac{\partial}{\partial t}e(t)\right)|\phi_\eta(t)|\int_{\mathcal{X}^*}\sin(tm_0(x^*)+e(t))f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\
& -|\phi_\eta(t)|\int_{\mathcal{X}^*}\sin(tm_0(x^*)+e(t))m_0(x^*)f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^*,
\end{aligned}
$$

(20)

$$
\begin{aligned}
Im\frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t) = & \left(\frac{\partial}{\partial t}|\phi_\eta(t)|\right)\int_{\mathcal{X}^*}\sin(tm_0(x^*)+e(t))f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\
& +\left(\frac{\partial}{\partial t}e(t)\right)|\phi_\eta(t)|\int_{\mathcal{X}^*}\cos(tm_0(x^*)+e(t))f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\
& +|\phi_\eta(t)|\int_{\mathcal{X}^*}\cos(tm_0(x^*)+e(t))m_0(x^*)f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^*.
\end{aligned}
$$

We define operators as follows:

$$(21)\quad L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}} : \mathcal{L}^2([0,t_u]) \to \mathcal{L}^2(\mathcal{X}) \text{ with } (L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}}h)(x) = \int Re\frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t)h(t)dt,$$

$$(22)\quad L_{Im\frac{\partial}{\partial t}\phi_{f_{Y,X}}} : \mathcal{L}^2([0,t_u]) \to \mathcal{L}^2(\mathcal{X}) \text{ with } (L_{Im\frac{\partial}{\partial t}\phi_{f_{Y,X}}}h)(x) = \int Im\frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t)h(t)dt,$$

$$(23)\quad \Delta_{\partial|\phi_\eta|} : \mathcal{L}^2([0,t_u]) \to \mathcal{L}^2([0,t_u]) \text{ with } (\Delta_{\partial|\phi_\eta|}h)(t) = \left(\frac{\partial}{\partial t}|\phi_\eta(t)|\right)h(t),$$

$$(24)\quad \Delta_{\partial e} : \mathcal{L}^2([0,t_u]) \to \mathcal{L}^2([0,t_u]) \text{ with } (\Delta_{\partial e}h)(t) = \left(\frac{\partial}{\partial t}e(t)\right)h(t),$$

$$(25)\quad \Delta_{m_0} : \mathcal{L}^2(\mathcal{X}^*) \to \mathcal{L}^2(\mathcal{X}^*) \text{ with } (\Delta_{m_0}h)(x^*) = m_0(x^*)h(x^*).$$

Similarly to the derivation in Eq. (3), we can obtain operator equivalence relationships to Eqs. (19) and (20) as the following:

$$
\begin{aligned}
(26)\qquad L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}} = & L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\cos_{m_0,e}}\Delta_{\partial|\phi_\eta|} - L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\sin_{m_0,e}}\Delta_{|\phi_\eta|}\Delta_{\partial e} \\
& - L_{f_{X|X^*}}\Delta_{f_{X^*}}\Delta_{m_0}L_{\sin_{m_0,e}}\Delta_{|\phi_\eta|}, \\
(27)\qquad L_{Im\frac{\partial}{\partial t}\phi_{f_{Y,X}}} = & L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\sin_{m_0,e}}\Delta_{\partial|\phi_\eta|} + L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|}\Delta_{\partial e} \\
& + L_{f_{X|X^*}}\Delta_{f_{X^*}}\Delta_{m_0}L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|}.
\end{aligned}
$$

Define $\Delta_{\partial \ln|\phi_\eta|} : \mathcal{L}^2([0,t_u]) \to \mathcal{L}^2([0,t_u])$ with $(\Delta_{\partial \ln|\phi_\eta|}h)(t) = \left(\frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|}\right)h(t)$. The following derivation is dedicated to the identification of

$$L_A = L^{-1}_{Re\phi_{f_{Y|X^*}}} \Delta_{m_0} L_{Re\phi_{f_{Y|X^*}}},$$

where $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ exists and is densely defined over $\mathcal{L}^2(\mathcal{X})$ by Lemma A.10. We will show L_A is identified and use it to construct a spectral decomposition. Note that the invertibility of the operators $L_{Re\phi_{f_{Y,X}}}$ and $L_{Im\phi_{f_{Y,X}}}$ is equivalent to the invertibility of operators, $L_{f_{X|X^*}}$, $L_{Re\phi_{f_{Y|X^*}}}$, and $L_{Im\phi_{f_{Y|X^*}}}$ and the boundedness of f_{X^*}. While Assumption 2.3(ii) and Lemma A.1.1 permits the invertibility of $L_{f_{X|X^*}}$, Lemma A.2, and Lemma A.10 guarantee the invertibility of $L_{Re\phi_{f_{Y|X^*}}}$, and $L_{Im\phi_{f_{Y|X^*}}}$. The boundedness is ensured by Assumption 2.1. Post-multiplying $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ to Eq. (6) yields

$$L_{Re\phi_{f_{Y,X}}} L^{-1}_{Re\phi_{f_{Y|X^*}}} = L_{f_{X|X^*}} \Delta_{f_{X^*}},$$

which is justified by Lemma A.10. Use this relation to rewrite Eq. (26) as

$$\begin{aligned}
L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}} &= L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\cos_{m_0,e}} \Delta_{\partial|\phi_\eta|} - L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \Delta_{\partial e} \\
&\quad - L_{f_{X|X^*}} \Delta_{f_{X^*}} \Delta_{m_0} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|}, \\
&= \left[L_{Re\phi_{f_{Y,X}}} L^{-1}_{Re\phi_{f_{Y|X^*}}} \right] L_{\cos_{m_0,e}} \Delta_{\partial|\phi_\eta|} - \left[L_{Re\phi_{f_{Y,X}}} L^{-1}_{Re\phi_{f_{Y|X^*}}} \right] L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \Delta_{\partial e} \\
&\quad - \left[L_{Re\phi_{f_{Y,X}}} L^{-1}_{Re\phi_{f_{Y|X^*}}} \right] \Delta_{m_0} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|}, \\
&= L_{Re\phi_{f_{Y,X}}} \left[L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{\cos_{m_0,e}} \Delta_{\partial|\phi_\eta|} - L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \Delta_{\partial e} \right. \\
&\quad \left. - L^{-1}_{Re\phi_{f_{Y|X^*}}} \Delta_{m_0} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \right]
\end{aligned}$$

Because $L_{Re\phi_{f_{Y,X}}}$ is injective by the injectivity of operators, $L_{f_{X|X^*}}$, $L_{Re\phi_{f_{Y|X^*}}}$, and f_{X^*},

$L^{-1}_{Re\phi_{f_{Y,X}}} L_{Re\phi_{f_{Y,X}}} = I$. This implies

$$
\begin{aligned}
(28) \qquad L_{B_1} &\equiv L^{-1}_{Re\phi_{f_{Y,X}}} L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}} \\
&= \left(L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|}\right)^{-1} L_{\cos_{m_0,e}}\Delta_{\partial|\phi_\eta|} - \left(L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{\sin_{m_0,e}}\Delta_{|\phi_\eta|}\right)\Delta_{\partial e} \\
&\quad - \left(L^{-1}_{Re\phi_{f_{Y|X^*}}}\Delta_{m_0} L_{Re\phi_{f_{Y|X^*}}}\right)\left(L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{\sin_{m_0,e}}\Delta_{|\phi_\eta|}\right) \\
&= \Delta_{\partial \ln|\phi_\eta|} - L_{K_1}\Delta_{\partial e} - L_A L_{K_1},
\end{aligned}
$$

where we have used $L_{Re\phi_{f_{Y|X^*}}} L^{-1}_{Re\phi_{f_{Y|X^*}}} = I$. Similar, using Eqs. (7) and (27), we obtain

$$
\begin{aligned}
(29) \qquad L_{B_2} &\equiv L^{-1}_{Im\phi_{f_{Y,X}}} L_{Im\frac{\partial}{\partial t}\phi_{f_{Y,X}}} \\
&= \left(L_{\sin_{m_0,e}}\Delta_{|\phi_\eta|}\right)^{-1} L_{\sin_{m_0,e}}\Delta_{\partial|\phi_\eta|} + \left(L^{-1}_{Im\phi_{f_{Y,X}}} L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|}\right)\Delta_{\partial e} \\
&\quad + \left(L^{-1}_{Im\phi_{f_{Y|X^*}}} L_{Re\phi_{f_{Y|X^*}}}\right)\left(L^{-1}_{Re\phi_{f_{Y|X^*}}}\Delta_{m_0} L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|}\right) \\
&= \Delta_{\partial \ln|\phi_\eta|} + L^{-1}_{K_1}\Delta_{\partial e} + L^{-1}_{K_1} L_A.
\end{aligned}
$$

We eliminate the operator L_A in Eqs. (28) and (29) by applying L_{K_1} to the left and right sides of Eq. (29) and then adding with Eq. (28):

$$
\begin{aligned}
L_C &= L_{B_1} + L_{K_1} L_{B_2} L_{K_1} \\
&= \Delta_{\partial \ln|\phi_\eta|} - L_{K_1}\Delta_{\partial e} + L_{K_1}\Delta_{\partial \ln|\phi_\eta|} L_{K_1} + \Delta_{\partial e} L_{K_1} \\
(30) \qquad &= \Delta_{\partial \ln|\phi_\eta|} + L_{K_1}\Delta_{\partial \ln|\phi_\eta|} L_{K_1},
\end{aligned}
$$

where we have used $L_{K_1}\Delta_{\partial e} = \Delta_{\partial e} L_{K_1}$ which is justified by Lemma A.11. Note that LHS are observable and $\Delta_{\partial \ln|\phi_\eta|}$ is the unobservable operators in RHS. Applying the observed operator L_C in Eq. (30) to the constant function $\mathbf{1}$ and using Lemma A.11 yields

$$
\begin{aligned}
(31) \qquad (L_C\mathbf{1})(t) &= \frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|} + \tan(e(t))^2 \frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|} \\
&= (1 + \tan(e(t))^2)\frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|}.
\end{aligned}
$$

Because L_{K_1}, and therefore $e(t)$, are identified, this implies that both $\frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|}$ is identified. It follows that L_A is identified from Eq. (29) as follows:

$$L_A = L_{K_1}\left(L_{B_2} - \Delta_{\partial \ln|\phi_\eta|}\right) - \Delta_{\partial e}.$$

Pre-multiplying the operator $L_{f_{X|X^*}}\Delta_{f_{X^*}}$ to the both sides of the equation $L_{Re\phi_{f_{Y|X^*}}} L_A = \Delta_{m_0} L_{Re\phi_{f_{Y|X^*}}}$, we have

$$L_{Re\phi_{f_{Y,X}}} L_A \;=\; L_{f_{X|X^*}}\Delta_{f_{X^*}}\Delta_{m_0} L_{Re\phi_{f_{Y|X^*}}}. \tag{32}$$

Post-multiplying the operator $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ to the both sides of Eq. (32) (justified by Lemma A.10) yields

$$L_{Re\phi_{f_{Y,X}}} L_A L^{-1}_{Re\phi_{f_{Y|X^*}}} \;=\; L_{f_{X|X^*}}\Delta_{f_{X^*}}\Delta_{m_0}. \tag{33}$$

Because $\Delta^{-1}_{f_{X^*}}$ and $L^{-1}_{f_{X|X^*}}$ both defined over a dense subset of their domain spaces (Assumption 2.1 and Lemma A.1.1), we post-multiply these operators to Eq. (33) to obtain

$$\begin{aligned}
\underbrace{L_{Re\phi_{f_{Y,X}}} L_A L^{-1}_{Re\phi_{f_{Y,X}}}}_{\text{Identified}} &= \left(L_{Re\phi_{f_{Y,X}}} L_A L^{-1}_{Re\phi_{f_{Y|X^*}}}\right)\Delta^{-1}_{f_{X^*}} L^{-1}_{f_{X|X^*}} \\
&= L_{f_{X|X^*}}\Delta_{f_{X^*}}\Delta_{m_0}\Delta^{-1}_{f_{X^*}} L^{-1}_{f_{X|X^*}} \\
&= L_{f_{X|X^*}}\Delta_{m_0} L^{-1}_{f_{X|X^*}}.
\end{aligned} \tag{34}$$

The above operator to be diagonalized is defined in terms of observable operators, while the resulting eigenvalues $m_0(x^*)$ and eigenfunctions $f_{X|X^*}(\cdot|x^*)$ (both indexed by x^*) provide the unobserved function of interest including the regression function and the joint distribution of the joint distribution of the unobserved regressor x^* and the observed regressor x. Assumptions 2.3(iii) and 2.4 ensure the uniqueness of the spectral decomposition of the observed operator Eq. (32). Similar to Eq. (1), we have $f_{Y,X}(y,x) = \int_{\mathcal{X}^*} f_{Y,X^*}(y,x^*) f_{X|X^*}(x|x^*)dx^*$ and it implies that for any $y \in \mathcal{Y}$, $(L_{f_{X|X^*}} f_{Y,X^*})(x) = f_{Y,X}(y,x)$. Thus the identification of $f_{X|X^*}$ induces the identification of f_{Y,X^*} as follow, for any $y \in \mathcal{Y}$,

$$f_{Y,X^*}(y,x^*) = (L^{-1}_{f_{X|X^*}} f_{Y,X})(x^*),$$

where the inverse is justified by the first part of 2.3(ii). Therefore, the densities $f_{Y|X^*}$ and f_{X^*} are identified and so is the regression error distribution f_η. We have reached our main result. QED.

B. A Sieve Maximum Likelihood Estimator

Our asymptotic analysis relies on regularity restrictions on the unknown functions to be estimated. We thus introduce a typical space of smooth functions, i.e., the Hölder space. Given a $d \times 1$ vector of nonnegative integers, $a = (a_1, ..., a_d)^T$ and denote $[a] = a_1 + ... + a_d$ and let D^a denote the differential operator defined by $D^a = \frac{\partial^{[a]}}{\partial \xi_1^{a_1} ... \partial \xi_d^{a_d}}$. Let $\underline{\gamma}$ denote the largest integer satisfying $\gamma > \underline{\gamma}$ and set $\gamma = \underline{\gamma} + p$. The Hölder space $\Lambda^\gamma(\nu)$ of order $\gamma > 0$ is a collection of functions which are $\underline{\gamma}$ times continuously differentiable on ν and the $\underline{\gamma}-$th derivative are Hölder continuous with the exponent p. The Hölder space becomes a Banach space with the Hölder norm, i.e., $\forall g \in \Lambda^\gamma(\nu)$

$$\|g\|_{\Lambda^\gamma} = \sup_{\xi \in \nu} |g(\xi)| + \max_{a_1 + ... + a_d = \underline{\gamma}} \sup_{\xi \neq \xi' \in \nu} \frac{|D^a g(\xi) - D^a g(\xi')|}{\|\xi - \xi'\|_E^p}. \tag{35}$$

The weighted Hölder norm is defined as $\|g\|_{\Lambda^{\gamma,\omega}} \equiv \|\widetilde{g}\|_{\Lambda^\gamma}$ for $\widetilde{g}(\xi) \equiv g(\xi)\omega(\xi)$ and the corresponding weighted Hölder space is $\Lambda^{\gamma,\omega}(\nu)$. Define a weighted Hölder ball as $\Lambda_c^{\gamma,\omega}(\nu) \equiv \{g \in \Lambda^{\gamma,\omega}(\nu) : \|g\|_{\Lambda^{\gamma,\omega}} \leq c < \infty\}$. Let $\eta \in \mathbb{R}$, $\gamma_1 > 1$, and $W \in \mathcal{W}$ with $\mathcal{W}$ a compact convex subset in $\mathbb{R}^{d_w}$. Without loss of generality, we consider strictly increasing m (decreasing m can be handled simiarly). Define the following sets:

$$\begin{aligned}
\tilde{\mathcal{F}}_1 &= \{\sqrt{f_1(\cdot)} \in \Lambda_c^{\gamma_1,\omega}(\mathbb{R}) : f_1(\cdot) \geq 0 \text{ and } \int_{\mathbb{R}} f_1(\eta) d\eta = 1\}, \\
\tilde{\mathcal{F}}_2 &= \{f_2(\cdot|\cdot) \in \Lambda_c^{\gamma_1,\omega}(\mathcal{X} \times \mathcal{X}^*) : f_2(\cdot|\cdot) \geq 0 \text{ and } f_2 \text{ satisfies Assumption} \\
&\qquad 2.3(ii)\&(iii) \text{ and } \int_{\mathcal{X}} f_2(x|x^*) dx = 1, \text{ for } x^* \in \mathcal{X}^*\}, \\
\tilde{\mathcal{F}}_3 &= \{\sqrt{f_3(\cdot)} \in \Lambda_c^{\gamma_1,\omega}(\mathcal{X}^*) : f_3(\cdot) \geq 0 \text{ and } \int_{\mathcal{X}^*} f_3(x^*) dx^* = 1\}, \\
\tilde{\mathcal{F}}_4 &= \{\sqrt{m_4'(\cdot;\theta,h_4)} \in \Lambda_c^{\gamma_1,\omega}(\mathcal{X}^*) : m_4'(\cdot;\theta,h_4) > 0\},
\end{aligned}$$

where $m_4'(\cdot;\theta,h_4)$ is the derivative of $m_4(\cdot;\theta,h_4)$ with respect to x^*. Working with square roots is a convenient device to enforce positiveness. For maximum generality, we phrase our estimation result for subsets of those maximal sets $\mathcal{F}_i \subseteq \tilde{\mathcal{F}}_i$, $i = 1,2,3,4$, thus allowing practitioners to impose other constraints that may be known to hold in the population. This is helpful in cases where the sets $\tilde{\mathcal{F}}_i$ are too "rich" to allow simple primitive verifications of the assumptions of our asymptotic theory while such verification could be possible on suitably defined subsets $\mathcal{F}_i$. With $\gamma_1 > 1$, assume that the square roots of the nonparametric components f_η, f_{X^*} and m_0' belong to the spaces, $\mathcal{F}_1$, $\mathcal{F}_3$, and $\mathcal{F}_4$ respectively, and $f_{X^*|X}$ belongs to $\mathcal{F}_2$, and that the parametric vector in the regression function belongs to the space Θ.

Assumption B.1. *(i) all the assumptions of Theorem 2.1 hold; (ii)* $\sqrt{f_1(\cdot)} \in \mathcal{F}_1$*; (iii)* $f_2(\cdot) \in \mathcal{F}_2$*; (iv)* $\sqrt{f_3(\cdot)} \in \mathcal{F}_3$*; (v)* $\sqrt{h_4(\cdot)} \in \mathcal{F}_4$.

Set $\mathcal{A} = \Theta \times \mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3 \times \mathcal{F}_4$ and $\alpha \equiv (\theta, f_{1s}, f_2, f_{3s}, h_{4s})$, where the lower subscript s indicates the square roots. Let $\alpha_0 \equiv (\theta_0, \sqrt{f_\eta}, f_{X|X^*}, \sqrt{f_{X^*}}, \sqrt{h_0})$ denote the true parameter. Our sieve MLE $\widehat{\alpha}_n$ is obtained by maximizing

$$\widehat{\alpha}_n = \arg\max_{\alpha \in \mathcal{A}^n} \widehat{Q}_n(\alpha).$$

where $\mathcal{A}^n \equiv \Theta \times \mathcal{F}_1^n \times \mathcal{F}_2^n \times \mathcal{F}_3^n \times \mathcal{F}_4^n$ is a sequence of approximation spaces to $\mathcal{A}$, where

$$\widehat{Q}_n(\alpha) = \frac{1}{n} \sum\nolimits_{i=1}^{n} \ln \left(\int_{\mathcal{X}^*} f_1(y_i - m_0(x^*; \widetilde{\theta})) f_2(x_i|x^*) f_3(x^*) dx^* \right).$$

and where $\{y_i, x_i\}_{i=1}^n$ denotes the observed sample.

This estimator is a direct application of the general semi-parametric sieve MLE presented by Shen (1997), Chen and Shen (1998), and Ai and Chen (2003). Ai and Chen (2003) shows that $\widehat{\alpha}_n$ is a consistent estimator, and the parametric component of α has an asymptotically normal distribution. For completeness, we present all the standard assumptions for consistency of all unknown parameters and root-n normality of the parametric part in the remainder of this appendix.

B.1. Consistency and Convergence Rates

In this subsection, we first obtain consistency of the sieve MLE $\widehat{\alpha}_n$ for α_0 under a strong norm $\|\cdot\|_s$, as in Newey and Powell (2003) and Ai and Chen (2003). Using the consistency as a starting point, we then establish that $\widehat{\alpha}_n$ converges to α_0 at a rate faster than $n^{-1/4}$ under a suitably constructed weaker Fisher norm, $\|\cdot\|$. Define

$$\|\alpha\|_s = \|\theta\|_E + \|f_{1s}\|_{s,\omega} + \|f_2\|_{s,\omega} + \|f_{3s}\|_{s,\omega} + \|h_{4s}\|_{s,\omega} \tag{36}$$

where $\|\theta\|_E$ is the Euclidean norm and $\|g\|_{s,\omega} \equiv \sup_\xi |g(\xi)\omega(\xi)|$ with $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}$, for some $\varsigma > 0$. The weighting function ω is introduced to deal with unbounded support. Define

$$\ell_D(d_i; \alpha) = \ln f_D(d_i; \alpha) = \ln \left(\int_{\mathcal{X}^*} f_1(y_i - m_0(x^*; \widetilde{\theta})) f_2(x_i|x^*) f_3(x^*) dx^* \right),$$

where d_i is a realization of a random variable $D \equiv (Y, X)$ in the sample.

Assumption B.2. *(i) The data $\{(D_i)_{i=1}^n\}$ are i.i.d.; (ii) The density function of D, f_D, satisfies $\int \omega(\zeta)^{-2} f_D(\zeta) d\zeta < \infty$.*

Assumption B.3. *(i) $\theta_0 \in \mathcal{B}$, a compact subset of $\mathbb{R}^{d_\theta}$; (ii) Assumption B.1 holds for a neighborhood of α_0 under the norm $\|\cdot\|_s$.*

Assumption B.4. *(i) For any $\alpha \in \mathcal{A}$, there exists $\Pi_n \alpha \in \mathcal{A}^n$ such that $\|\Pi_n \alpha - \alpha\|_s = o(1)$; (ii) $k_{i,n} \to +\infty$ and $k_{i,n}/n \to 0$ for $i = 1, 2, 3, 4$ as $n \to +\infty$.*

Definition B.1. *$\ell_D(d_i; \alpha)$ is Hölder continuous with respect to $\alpha \in \mathcal{A}$ if there exists a measurable function $c(D)$ with $E\{c(D)^2\} < \infty$ such that, for all $\alpha_1, \alpha_2 \in \mathcal{A}$, and D, we have*

$$|\ell_D(d_i; \alpha_1) - \ell_D(d_i; \alpha_2)| \leq c(D) \|\alpha_1 - \alpha_2\|_s. \tag{37}$$

The next assumption ensures $\ell_D(d_i; \alpha)$ is Hölder continuous with respect to $\alpha \in \mathcal{A}$.

Assumption B.5. *(i) $E\{|\ell_D(d_i; \alpha)|^2\}$ is bounded; (ii) There exits a positive measurable function $\widetilde{h}(D)$ with $E\{\widetilde{h}(D)^2\} < \infty$ such that, for any $\bar{\alpha}_{12} = (\bar{\theta}, \bar{f_{1s}}, \bar{f_2}, \bar{f_{3s}}, \bar{h}_{4s})$ and $\bar{\omega}(\varepsilon, \eta, x, x^*) \equiv \left[1, \omega^{-1}(\varepsilon), \omega^{-1}(\eta), \omega^{-1}(x, x^*), \omega^{-1}(x^*)\right]^T$, we have $|h_1(d_i, \bar{\alpha}, \bar{\omega})| < \widetilde{h}(D)$, where the function $h_1(d_i, \bar{\alpha}, \bar{\omega})$ is constructed by the path derivatives of $\ln f_D(d_i; \alpha)$ and the explicit expression of $h_1(d_i, \bar{\alpha}, \bar{\omega})$ can be found in the proof of Lemma 3.1.*

Assumption B.5(ii) implies that $\ell_D(d_i; \alpha)$ is Hölder continuous in α.

Lemma B.1. *Under Assumptions B.1-B.5, we obtain $\|\widehat{\alpha}_n - \alpha_0\|_s = o_p(1)$.*

See the online appendix for the proof.

For simplicity, Assumption B.2(i) rules out serially dependent observations and could easily be relaxed. Assumptions B.2(ii) and B.3(i) and (ii) are standard conditions imposed for series approximation. The series approximation can approximate any function with a small mean-square error, which holds for power series, wavelets, Fourier series, and splines. Assumption B.4(i) states that there is a finite dimensional approximation $\mathcal{A}^n$ to $\mathcal{A}$ and Assumption B.4(ii) imposes that the number of terms in the sieve grows slower than the sample size to control the variance. Assumption B.5 ensures the Hölder continuity of the log likelihood function.

Next, we consider $n^{-1/4}$ convergence rates of $\widehat{\alpha}_n$ under weaker metrics, which are sufficient to establish the asymptotic normality and $\sqrt{n}$-consistency results. We first define the weaker Fisher metric $\|\cdot\|_F$ introduced by Ai and Chen (2003). Assume the function space $\mathcal{A}$ is convex. For any $v \in \bar{V}$, define the pathwise derivative as:

$$\frac{d\ell_D(d_i;\alpha)}{d\alpha}[v] \equiv \left.\frac{d\ell_D(d_i;\alpha+\tau v)}{d\tau}\right|_{\tau=0} \qquad \text{a.s. } D.$$

For any $\alpha_1, \alpha_2 \in \mathcal{A}$, the Fisher norm is defined as:

$$\|\alpha_1-\alpha_2\|_F^2 \equiv \mathrm{E}\left\{\left(\frac{d\ell_D(d_i;\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\right)^2\right\}. \tag{38}$$

We make the following assumptions to obtain a rate faster than $n^{-1/4}$.

Assumption B.6. *Let k_n be the total number of sieve coefficients in the sieve estimator $\widehat{\alpha}_n$, i.e., $k_n = k_{1,n} + k_{2,n}^2 + k_{3,n} + k_{4,n}$. Then, $(k_n n^{-1/2} \ln n) \times \sup_{\xi \in (\mathbb{R}\cup\mathcal{X}\times\mathcal{X}^*\cup\mathcal{X}^*)} \|p^{k_n}(\xi)\|_E^2 = o(1)$.*

Assumption B.7. (i) *There exist a measurable function $c(D)$ with $E\{c(D)^4\} < \infty$ such that $|\ell_D(d_i;\alpha)| \le c(D)$ for all D and $\alpha \in \mathcal{A}^n$; (ii) $\ell_D(d_i;\alpha) \in \Lambda_c^{\tau,\omega}(\mathcal{X}\times\mathcal{Y})$ with $\tau > \dim D/2$, for all $\alpha \in \mathcal{A}^n$, where $\dim D$ is the dimension of D.*

Assumption B.8. *$\mathcal{A}$ is convex in α_0, and $m_0(x^*;\widetilde{\theta})$ is pathwise differentiable at (θ_0, h_0).*

Assumption B.9. $\ln N(\delta, \mathcal{A}^n) = O(k_n \ln(k_n/\delta))$ *where $N(\delta, \mathcal{A}^n)$ is the minimum number of balls with radius δ under the $\|\cdot\|_s$ norm covering $\mathcal{A}^n$.*

Assumption B.10. *There exists c_1, $c_2 > 0$,*

$$c_1 E\left(\ln \frac{f_D(d_i;\alpha_0)}{f_D(d_i;\alpha)}\right) \le \|\alpha-\alpha_0\|_F^2 \le c_2 E\left(\ln \frac{f_D(d_i;\alpha_0)}{f_D(d_i;\alpha)}\right)$$

holds for all $\alpha \in \mathcal{A}^n$ with $\|\alpha-\alpha_0\|_s = o(1)$.

Assumption B.11. *For any $\alpha \in \mathcal{A}$, there exists $\Pi_n\alpha \in \mathcal{A}^n$ such that $\|\Pi_n\alpha - \alpha\|_F = o(k_n^{-\mu 1})$ and $k_n^{-\mu 1} = o(n^{-1/4})$.*

Theorem B.1. *If Assumptions B.1-B.11 hold, then* $\|\widehat{\alpha}_n - \alpha_0\|_F = o_p(n^{-1/4})$.

See the online appendix for the proof.

Assumption B.7(i) and (ii) impose a dominance condition and smoothness condition on $\ell_D(d_i;\alpha)$. Envelope conditions are assumed to restrict a change of the objective function as the parameters change and secure stochastic equi-continuity. Assumption B.8 implies that the Fisher norm in Eq. (38) is well defined. Assumption B.9 requires that the size of the sieve space $\mathcal{A}^n$ does not grow too fast in terms of the covering number. Commonly used sieve spaces, such as power series, Fourier series, splines, and wavelet linear sieves, satisfy this assumption. Assumption B.10 assumes that the population criterion function is locally equivalent to the Fisher norm. This condition helps derive the convergence of the parameter and $\sqrt{n}$-normality of the parametric component. Assumption B.11 controls the approximation error of $\Pi_n\alpha$ to α and the selection of k_n such that the error goes to zero uniformly at the rate $o_p(n^{-1/4})$ over $\alpha \in \mathcal{A}$.

B.2. Asymptotic Normality

In this section, we consider the asymptotic normality of the parametric component θ which contains the parameter of interest in the regression function. Let $\bar{V}$ be completion of the linear space spanned by $\mathcal{A} - \alpha_0$ under the Fisher norm $\|\cdot\|_F$. Then, $\left(\bar{V}, \|\cdot\|_F\right)$ is a Hilbert space with the inner product

$$\langle v_1, v_2\rangle \equiv E\left\{\left(\frac{d}{d\alpha}\ell_D(d_i;\alpha_0)[v_1]\right)\left(\frac{d}{d\alpha}\ell_D(d_i;\alpha_0)[v_2]\right)\right\},$$

and $\langle v, v\rangle = \|v\|_F$. For any fixed and nonzero $\lambda \in \mathbb{R}^{d_\theta}$, $f_\lambda(\alpha - \alpha_0) \equiv \lambda^T(\theta - \theta_0)$ is linear in $\alpha - \alpha_0$ and $f_\lambda(\alpha - \alpha_0)$ is a linear functional on $\left(\overline{V}, \|\cdot\|_F\right)$. Shen (1997) and van der Vaart (1991) show that $f(\alpha) \equiv \lambda^T\theta$ is a bounded linear functional on $\bar{V}$ under the operator norm. That is:

$$|||f_\lambda||| \equiv \sup_{\{\alpha\in\mathcal{A}:\|\alpha-\alpha_0\|>0\}} \frac{|f_\lambda(\alpha-\alpha_0)|}{\|\alpha-\alpha_0\|_F} < \infty. \tag{39}$$

By the Riesz representation theorem, there exists $v^* \in \bar{V}$ such that for any $\alpha \in \mathcal{A}$, we have $f_\lambda(\alpha - \alpha_0) = \langle \alpha - \alpha_0, v^*\rangle$. and $\|f_\lambda\|_F = \|v^*\|_F$. Denote $\bar{V} = \mathbb{R}^{d_\theta} \times \overline{W}$ and $\overline{W} \equiv \overline{\mathcal{F}_1^n \times \mathcal{F}_2^n \times \mathcal{F}_3^n \times \mathcal{F}_4^n} - (\sqrt{f_\eta}, f_{X|X^*}, \sqrt{f_{X^*}}, \sqrt{h_0})$.

For each component θ_j of the parametric component θ, $j = 1, 2, ..., d_\theta$, define $w_j^* \in \overline{W}$ to be the solution to the following minimization problem associated with the denominator of the operator norm,

$$\begin{aligned}
w_j^* &\equiv (f_{1sj}^*, f_{2j}^*, f_{3sj}^*, h_{4sj}^*)^T \\
&= \arg \min_{w_j=(f_{1s},f_2,f_{3s},h_{4s})^T \in \overline{W}} E\bigg(\frac{d\ell_D(d_i;\alpha_0)}{d\theta_j} - \frac{d\ell_D(d_i;\alpha_0)}{df_{1s}}[f_{1s}] - \frac{d\ell_D(d_i;\alpha_0)}{df_2}[f_2] \\
&\qquad\qquad - \frac{d\ell_D(d_i;\alpha_0)}{df_{3s}}[f_{3s}] - \frac{d\ell_D(d_i;\alpha_0)}{dh_{4s}}[h_{4s}] \bigg)^2.
\end{aligned}$$

Define $w^* = (w_1^*, ..., w_{d_\theta}^*)$ and let $\frac{d\ell_D(d_i;\alpha_0)}{df}[w^*]$ be the vector with elements (indexed by $j = 1, \ldots, d_\theta$):

$$\begin{aligned}
\frac{d\ell_D(d_i;\alpha_0)}{df}[w_j^*] &= \frac{d\ell_D(d_i;\alpha_0)}{df_{1s}}[f_{1sj}^*] + \frac{d\ell_D(d_i;\alpha_0)}{df_2}[f_{2j}^*] \\
&\quad + \frac{d\ell_D(d_i;\alpha_0)}{df_{3s}}[f_{3sj}^*] + \frac{d\ell_D(d_i;\alpha_0)}{dh_{4s}}[h_{4sj}^*],
\end{aligned}$$

and

$$H_{w^*}(d_i) \equiv \frac{d\ell_D(d_i;\alpha_0)}{d_\theta^T} - \frac{d\ell_D(d_i;\alpha_0)}{df}[w^*].$$

With these notation,

$$\|f_\lambda\|^2 = \sup_{\{\alpha\in\mathcal{A}:\|\alpha-\alpha_0\|>0\}} \frac{|f_\lambda(\alpha-\alpha_0)|^2}{\|\alpha-\alpha_0\|_F} = \lambda^T \left(E\{H_{w^*}(D)^T H_{w^*}(D)\}\right)^{-1} \lambda,$$

$v^* \equiv (v_\theta^*, v_h^*) \in \overline{V}$ with $v_\theta^* = \left(E\{H_{w^*}(D)^T H_{w^*}(D)\right)^{-1} \lambda$ and $v_h^* = -w^* \times v_\theta^*$. In addition, $f_\lambda(\alpha-\alpha_0) = \lambda'(\theta-\theta_0) = \langle \alpha-\alpha_0, v^*\rangle$ by the Riesz representation theorem and $\frac{d\ell_D(d_i;\alpha_0)}{d\alpha}[v^*] = H_{w^*}(d_i)v_\theta^*$. This implies that the asymptotic distribution of parametric component $\widehat{\theta}_n$ reduces to when the linear functional f_λ is bounded and what is the asymptotic distribution of $\langle \widehat{\alpha}_n - \alpha_0, v^*\rangle$. That is:

$$\begin{aligned}
\lambda^T \left(\widehat{\theta}_n - \theta_0\right) &= \langle \widehat{\alpha}_n - \alpha_0, v^*\rangle = \frac{1}{n}\sum_{i=1}^n \frac{d\ell_D(d_i;\alpha_0)}{d\alpha}[v^*] + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n \lambda^T \left(E\{H_{w^*}(D)^T H_{w^*}(D)\right)^{-1} H_{w^*}(d_i)^T + o_p(n^{-1/2}),
\end{aligned}$$

and $\sqrt{n}(\widehat{\theta}_n - \theta_0) \to N(0, \left(E\{H_{w^*}(D)^T H_{w^*}(D)\}\right)^{-1})$.

We make the following conditions for the $\sqrt{n}-$normality of $\widehat{\theta}_n$ which are also conditions in Ai and Chen (2003) and Hu and Schennach (2008):

Assumption B.12. *(i) $E\{H_{w^*}(D)^T H_{w^*}(D)\}$ is positive-definite and bounded; (ii) $\theta_0 \in int(\mathcal{B})$.*

Assumption B.13. *There is a $v_n^* = (v_\theta^*, -\Pi_n w^* \times v_\theta^*) \in \mathcal{A}^n - \alpha_0$ such that $\|v_n^* - v^*\|_F = o_p(n^{-1/4})$.*

We use the $\sqrt{n}$ consistency results in the previous section to focus on a smaller neighborhood of α_0. Define $\mathcal{N}_{0n} \equiv \{\alpha \in \mathcal{A}^n : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\|_F = o(n^{-1/4})\}$ and $\mathcal{N}_0 \equiv \{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\|_F = o(n^{-1/4})\}$.

Assumption B.14. *There exits a measurable function $h(D)$ with $E\{h(D)^2\} < \infty$ such that, for any $\bar{\alpha}_1 = (\bar{\theta}, \bar{f}_{1s}, \bar{f}_2, \bar{f}_{3s}, \bar{h}_{4s})$, we have*

$$|h_1(d_i, \bar{\alpha}, \bar{\omega})| + |h_2(d_i, \bar{\alpha}, \bar{\omega})| \leq h(D),$$

where the function $h_1(d_i, \bar{\alpha}, \bar{\omega})$ is the first path derivatives of $\ln f_D(d_i; \alpha)$ in Lemma 3.1 and $h_2(d_i, \bar{\alpha}, \bar{\omega})$ is the term controlling the second path derivatives of $\ln f_D(d_i; \alpha)$ in the proof of Theorem 3.2.

Assumption B.14 implies that there exits a non-negative measurable function s with $\lim_{\delta \to 0} s(\delta) = 0$ such that for all $\alpha \in \mathcal{N}_{0n}$,

$$\sup_{\bar{\alpha} \in \mathcal{N}_0} \left| \frac{d^2 \ell_D(D; \bar{\alpha})}{d\alpha d\alpha^T} [\alpha - \alpha_0, v^*] \right|_{t=0} \leq h(D) \cdot s(\|\alpha - \alpha_0\|_s)$$

Assumption B.15. *Uniformly over $\bar{\alpha} \in \mathcal{N}_0$ and $\alpha \in \mathcal{N}_{0n}$,*

$$E\left[\frac{d^2 \ell_D(D; \bar{\alpha})}{d\alpha d\alpha^T} [\alpha - \alpha_0, v^*] - \frac{d^2 \ell_D(D; \alpha_0)}{d\alpha d\alpha^T} [\alpha - \alpha_0, v^*] \right] = o(n^{-1/2}). \tag{40}$$

Because our estimator takes the form of a single-step semiparametric sieve MLE, the general treatment of Shen (1997) and Chen and Shen (1998) can be used to establish asymptotic normality, root-n consistency, and efficiency under these assumptions.

Theorem B.2. *Suppose that α_0 is identified and Assumptions B.6-B.11 and B.12-B.15 hold, then $\sqrt{n}(\widehat{\theta}_n - \theta_0) \Rightarrow N(0, V^{-1})$ where $V = E\{H_{w^*}(D)^T H_{w^*}(D)\}$ and the matrix V is the efficient information matrix.*

See the online appendix for the proof.

Assumption B.12 ensures that the asymptotic variance exists and that the population parametric component is an interior solution. This assumption also ensures that the estimation problem for θ_0 is well-posed, which is made possible by a semiparametric treatment. Assumption B.13 is a "no asymptotic bias" which means the representor v^* can be approximated by the sieve v_n^* with an asymptotically negligible error. Assumption B.14 represents the boundedness and smoothness restrictions of the second pathwise derivative of the log likelihood function. Assumption B.15 controls the higher order terms in terms of asymptotic expansion. These assumptions control the local quadratic behavior of the criterion difference are common in the literature on the method of sieve.[11]

C. Proofs of Consistency and Asymptotic Normality

The proof of Lemma B.1. The consistency result is a direct application of Lemma 3.1 of Ai and Chen (2003). and the proof will be provided by checking the conditions in the lemma. Most conditions are assumed directly in our assumptions and the only thing we have to show is that $\ell_D(d_i;\alpha)$ is Hölder continuous in α. The difference of $\ell_D(d_i;\cdot)$ at α_1 and α_2 is given by

$$
\begin{aligned}
&\ell_D(d_i;\alpha_1) - \ell_D(d_i;\alpha_2) \\
&= \frac{d}{d\alpha}\ell_D(d_i;\bar{\alpha}_{12})[\alpha_1 - \alpha_2] \\
&= \frac{d}{dt}\ell_D(d_i;\bar{\alpha}_{12} + t(\alpha_1 - \alpha_2))\Big|_{t=0},
\end{aligned}
$$

where $\bar{\alpha}_{12} = (\bar{\theta}, \bar{f_{1s}}, \bar{f_2}, \bar{f_{3s}}, \bar{h}_{4s})$, a mean value between α_1 and α_2, and $\bar{\alpha}_{12} + t(\alpha_1 - \alpha_2) = (\bar{\theta} + t(\theta_1 - \theta_2), \bar{f_{1s}} + t(f_{1s1} - f_{1s2}), \bar{f_2} + t(f_{21} - f_{22}), \bar{f_{3s}} + t(f_{3s1} - f_{3s2}), \bar{h}_{4s} + t(h_{4s1} - h_{4s2}))$.

[11] A detailed discussion of these can be found in Newey (1997), Shen (1997), and Ai and Chen (2003).

Consider

$$
\begin{aligned}
&\frac{d}{dt}\ell_D(d_i;\bar{\alpha}_{12}+t(\alpha_1-\alpha_2))\Big|_{t=0}\\
&=\frac{1}{f_D(d_i;\bar{\alpha}_{12})}\bigg(\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}(\theta_1-\theta_2)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^*\\
&\quad+\int_{\mathcal{X}^*}2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\left(f_{1s1}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))-f_{1s2}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\right)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^*\\
&\quad+\int_{\mathcal{X}^*}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\left(f_{21}(x_i|x^*)-f_{22}(x_i|x^*)\right)\bar{f}_3(x^*)dx^*\\
&\quad+\int_{\mathcal{X}^*}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\left(f_{3s1}(x^*)-f_{3s2}(x^*)\right)dx^*\\
&\quad+\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}(h_{4s1}-h_{4s2})\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^*\bigg).
\end{aligned}
$$

Then we can obtain the bounds for Hölder continuous as follows:

$$
\begin{aligned}
&\Big|\frac{d}{dt}\ell_D(d_i;\bar{\alpha}_{12}+t(\alpha_1-\alpha_2))\Big|_{t=0}\\
&\leq\frac{1}{|f_D(d_i;\bar{\alpha}_{12})|}\bigg(\int_{\mathcal{X}^*}\Big|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}\omega^{-1}(\varepsilon)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\Big|dx^*\\
&\qquad\times\|\theta_1-\theta_2\|_s\\
&\quad+\int_{\mathcal{X}^*}\Big|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega^{-1}(\eta)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\Big|dx^*\|f_{1s1}-f_{1s2}\|_{s,\omega}\\
&\quad+\int_{\mathcal{X}^*}\Big|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega^{-1}(x_i,x^*)\bar{f}_3(x^*)\Big|dx^*\|f_{21}-f_{22}\|_{s,\omega}\\
&\quad+\int_{\mathcal{X}^*}\Big|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega^{-1}(x^*)\Big|dx^*\|f_{3s1}-f_{3s2}\|_{s,\omega}\\
&\quad+\int_{\mathcal{X}^*}\Big|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega^{-1}(x^*)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\Big|dx^*\|h_{4s1}-h_{4s2}\|_{s,\omega}\bigg)\\
&\leq\frac{1}{|f_D(d_i;\bar{\alpha}_{12})|}\bigg(\int_{\mathcal{X}^*}\Big|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}\omega^{-1}(\varepsilon)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\Big|dx^*\\
&\quad+\int_{\mathcal{X}^*}\Big|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega^{-1}(\eta)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\Big|dx^*\\
&\quad+\int_{\mathcal{X}^*}\Big|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega^{-1}(x_i,x^*)\bar{f}_3(x^*)\Big|dx^*\\
&\quad+\int_{\mathcal{X}^*}\Big|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega^{-1}(x^*)\Big|dx^*\\
&\quad+\int_{\mathcal{X}^*}\Big|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega^{-1}(x^*)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\Big|dx^*\bigg)\|\alpha_1-\alpha_2\|_s
\end{aligned}
$$

(41)

$$
\equiv h_1(d_i,\bar{\alpha},\bar{\omega})\|\alpha_1-\alpha_2\|_s,
$$

where $\bar{\omega}(\varepsilon, \eta, x, x^*) \equiv \left[1, \omega^{-1}(\varepsilon), \omega^{-1}(\eta), \omega^{-1}(x, x^*), \omega^{-1}(x^*)\right]^T$. Therefore, Assumption B.5(ii) implies

$$|\ell_D(D_i; \alpha_1) - \ell_D(D_i; \alpha_2)| \leq \widetilde{h}(D_i)\|\alpha_1 - \alpha_2\|_s,$$

and $\ell_D(d_i; \alpha)$ is Hölder continuous in α.

The proof of Theorem B.1. We prove the results by checking the conditions in Theorem 3.1 in Ai and Chen (2003). As discussed in Hu and Schennach (2008), our single-step semiparametric sieve MLE is simpler than the setup in Ai and Chen (2003) because we do not need to estimate the conditional mean as a function of the unknown parameter. The assumptions in Theorem 3.1 in Ai and Chen (2003) are directly being assumed, we obtain the consistency result.

The proof of Theorem B.2. The likelihood function $f_D(d_i; \alpha)$ has a similar expression as the likelihood function in Hu and Schennach (2008) which applied Theorem 1 of Shen (1997) to establish an asymptotic normality result. The proof there can directly apply to our case. We prove the results by showing an envelope condition on the second derivative of the likelihood function (Assumption B.14). Set $\bar{\alpha}_1 = (\bar{\theta}, \bar{f_{1s}}, \bar{f_2}, \bar{f_{3s}}, \bar{h}_{4s})$ and $v_n = \Pi_n(\alpha - \alpha_0 - v^*) = ([v_n]_\theta, [v_n]_{f_{1s}}, [v_n]_{f_2}, [v_n]_{f_{3s}}, [v_n]_{h_{4s}})$. Consider the second derivative of pathwise derivative as follows

$$\begin{aligned}
&\left|\sup_{\alpha \in \mathcal{N}_{0n}} \frac{d^2 \ell_D(d_i; \bar{\alpha}_1)}{d\alpha d\alpha^T}[v_n, \alpha - \alpha_0]\right| \\
&\leq \sup_{\alpha \in \mathcal{N}_{0n}} \left| \frac{1}{f_D(d_i; \bar{\alpha}_1)} \frac{d^2 f_D(d_i; \bar{\alpha}_1)}{d\alpha d\alpha^T}[v_n, \alpha - \alpha_0] \right. \\
&\qquad \left. - \frac{d\ell_D(d_i; \bar{\alpha}_1)}{d\alpha}[v_n] \frac{d\ell_D(d_i; \bar{\alpha}_1)}{d\alpha}[\alpha - \alpha_0] \right| \\
&\leq \sup_{\alpha \in \mathcal{N}_{0n}} \left(\left| \frac{1}{f_D(d_i; \bar{\alpha}_1)} \frac{d^2 f_D(d_i; \bar{\alpha}_1)}{d\alpha d\alpha^T}[v_n, \alpha - \alpha_0] \right| \right. \\
&\qquad \left. + \left| \frac{d\ell_D(d_i; \bar{\alpha}_1)}{d\alpha}[v_n] \right| \left| \frac{d\ell_D(d_i; \bar{\alpha}_1)}{d\alpha}[\alpha - \alpha_0] \right| \right).
\end{aligned}$$

We divide this into three terms and find the bounds for them. Consider

$$
\begin{aligned}
&\frac{d\ell_D(d_i;\bar{\alpha}_1)}{d\alpha}[\alpha-\alpha_0] \\
&= \frac{1}{f_D(d_i;\bar{\alpha}_1)}\Bigg(\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}(\theta-\theta_0)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^* \\
&\quad+\int_{\mathcal{X}^*}2\bar{f}_{1s}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\left(f_{1s}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))-\sqrt{f_\eta}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\right)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^* \\
&\quad+\int_{\mathcal{X}^*}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\left(f_2(x_i|x^*)-f_{X|X^*}(x_i|x^*)\right)\bar{f}_3(x^*)dx^* \\
&\quad+\int_{\mathcal{X}^*}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\bar{f}_2(x_i|x^*)2\bar{f}_{3s}(x^*)\left(f_{3s}(x^*)-\sqrt{f_{X^*}}(x^*)\right)dx^* \\
&\quad+\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}(h_{4s}-\sqrt{h_0})\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^*\Bigg).
\end{aligned}
$$

Therefore, similar to the derivation of the Hölder continuity in the proof of Lemma B.1, we obtain

$$
\left|\frac{d\ell_D(d_i;\bar{\alpha}_1)}{d\alpha}[\alpha-\alpha_0]\right| \leq h_1(d_i,\bar{\alpha},\bar{\omega})\|\alpha_1-\alpha_2\|_s,
$$

and

$$
\left|\frac{d\ell_D(d_i;\bar{\alpha}_1)}{d\alpha}[v_n]\right| \leq h_1(d_i,\bar{\alpha},\bar{\omega})\|v_n\|_s,
$$

where $h_1(d_i,\bar{\alpha},\bar{\omega})$ is defined in Equation (41).

Expanding out the term $\frac{1}{f_D(d_i;\bar{\alpha}_1)}\frac{d^2 f_D(d_i;\bar{\alpha}_1)}{d\alpha d\alpha^T}[v_n,\alpha-\alpha_0]$:

$$
\begin{aligned}
&\frac{1}{f_D(d_i;\bar{\alpha}_1)}\frac{d^2 f_D(d_i;\bar{\alpha}_1)}{d\alpha d\alpha^T}[v_n,\alpha-\alpha_0] \\
&= \frac{1}{f_D(d_i;\bar{\alpha}_1)}\Bigg(\int_{\mathcal{X}^*}\frac{d^2}{d^2\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2 m_0(x^*;\bar{\theta},\bar{h}_4)}{d^2\theta}[v_n]_\theta(\theta-\theta_0)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}2\frac{d}{d\eta}\left(\bar{f_{1s}}\cdot\left(f_{1s}-\sqrt{f_\eta}\right)\right)(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}[v_n]_\theta\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}[v_n]_\theta\left(f_2(x_i|x^*)-f_{X|X^*}(x_i|x^*)\right)\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}[v_n]_\theta\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\left(f_{3s}(x^*)-\sqrt{f_{X^*}}(x^*)\right)dx^* \\
&+\int_{\mathcal{X}^*}\frac{d^2}{d^2\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2 m_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta dh}[v_n]_\theta 2\bar{h}_{4s}(h_{4s}-\sqrt{h_0})\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}2\left(f_{1s}-\sqrt{f_\eta}\right)(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))[v_n]_{f_{1s}}f_2(x_i|x^*)\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))[v_n]_{f_{1s}}\left(f_2(x_i|x^*)-f_{X|X^*}(x_i|x^*)\right)\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))[v_n]_{f_{1s}}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\left(f_{3s}(x^*)-\sqrt{f_{X^*}}(x^*)\right)dx^* \\
&+\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}(\theta-\theta_0)[v_n]_{f_2}\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}2\left(\bar{f_{1s}}\cdot\left(f_{1s}-\sqrt{f_\eta}\right)\right)(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))[v_n]_{f_2}\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))[v_n]_{f_2}2\bar{f_{3s}}(x^*)\left(f_{3s}(x^*)-\sqrt{f_{X^*}}(x^*)\right)dx^* \\
&+\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}(h_{4s}-\sqrt{h_0})[v_n]_{f_2}\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}(\theta-\theta_0)\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)[v_n]_{f_{3s}}dx^* \\
&+\int_{\mathcal{X}^*}2\left(\bar{f_{1s}}\cdot\left(f_{1s}-\sqrt{f_\eta}\right)\right)(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)[v_n]_{f_{3s}}dx^* \\
&+\int_{\mathcal{X}^*}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\left(f_2(x_i|x^*)-f_{X|X^*}(x_i|x^*)\right)2\bar{f_{3s}}(x^*)[v_n]_{f_{3s}}dx^* \\
&+\int_{\mathcal{X}^*}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\bar{f}_2(x_i|x^*)2\left(f_{3s}(x^*)-\sqrt{f_{X^*}}(x^*)\right)[v_n]_{f_{3s}}dx^* \\
&+\int_{\mathcal{X}^*}\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}(h_{4s}-\sqrt{h_0})\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)[v_n]_{f_{3s}}dx^* \\
&+\int_{\mathcal{X}^*}\frac{d^2}{d^2\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2 m_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta dh}2\bar{h}_{4s}[v_n]_{h_{4s}}(\theta-\theta_0)\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}2\frac{d}{d\eta}\left(\bar{f_{1s}}\cdot\left(f_{1s}-\sqrt{f_\eta}\right)\right)(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}[v_n]_{h_{4s}}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^*
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{\mathcal{X}^*} \frac{d}{d\eta}\bar{f}_1(y_i - m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh} 2\bar{h}_{4s}[v_n]_{h_{4s}} \left(f_2(x_i|x^*) - f_{X|X^*}(x_i|x^*)\right)\bar{f}_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*} \frac{d}{d\eta}\bar{f}_1(y_i - m_0(x^*;\bar{\theta},\bar{h}_4))\frac{-dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh} 2\bar{h}_{4s}[v_n]_{h_{4s}} \bar{f}_2(x_i|x^*)2\bar{f}_{3s}(x^*)\left(f_{3s}(x^*) - \sqrt{f_{X^*}}(x^*)\right)dx^* \\
&+\int_{\mathcal{X}^*} \frac{d^2}{d^2\eta}\bar{f}_1(y_i - m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2m_0(x^*;\bar{\theta},\bar{h}_4)}{d^2h} 2\bar{h}_{4s}[v_n]_{h_{4s}}(h_{4s} - \sqrt{h_0})\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^*\Bigg).
\end{aligned}
$$

Letting $\left[\omega^{-1}(\varepsilon), \omega^{-1}(\eta), \omega^{-1}(x,x^*), \omega^{-1}(x^*)\right] \equiv \left[\omega_\theta^{-1}, \omega_1^{-1}, \omega_2^{-1}, \omega_{34}^{-1}\right]$, this term can be bounded:

$$
\begin{aligned}
&\left|\frac{1}{f_D(d_i;\bar{\alpha}_1)}\frac{d^2 f_D(d_i;\bar{\alpha}_1)}{d\alpha d\alpha^T}[v_n,\alpha-\alpha_0]\right| \\
\leq &\frac{1}{f_D(d_i;\bar{\alpha}_1)}\bigg(\int_{\mathcal{X}^*}\left|\frac{d^2}{d^2\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2 m_0(x^*;\bar{\theta},\bar{h}_4)}{d^2\theta}\omega_\theta^{-1}\omega_\theta^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\|[v_n]_\theta\|_s\|\theta-\theta_0\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|2\frac{d}{d\eta}\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_1^{-1}\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}\omega_\theta^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^*\|f_{1s}-\sqrt{f_\eta}\|_s\|[v_n]_\theta\|_s\right. \\
&\quad+\int_{\mathcal{X}^*}\left|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_1^{-1}\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}\omega_\theta^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)dx^*\|\frac{d}{d\eta}f_{1s}-\frac{d}{d\eta}\sqrt{f_\eta}\|_s\|[v_n]_\theta\|_s\right. \\
&\quad+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}\omega_\theta^{-1}\omega_2^{-1}\bar{f}_3(x^*)\right|dx^*\|[v_n]_\theta\|_s\|f_2(x_i|x^*)-f_{X|X^*}(x_i|x^*)\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}\omega_\theta^{-1}[v_n]_\theta\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^* \\
&\quad\times\|[v_n]_\theta\|_s\|f_{3s}(x^*)-\sqrt{f_{X^*}}(x^*)\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|\frac{d^2}{d^2\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2 m_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta dh}\omega_\theta^{-1}2\bar{h}_{4s}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\|[v_n]_\theta\|_s\|h_{4s}-\sqrt{h_0}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|2\omega_1^{-1}\omega_1^{-1}f_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\|f_{1s}-\sqrt{f_\eta}\|_s\|[v_n]_{f_{1s}}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_1^{-1}\omega_2^{-1}\bar{f}_3(x^*)\right|dx^*\|[v_n]_{f_{1s}}\|_s\|f_2-f_{X|X^*}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_1^{-1}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\|[v_n]_{f_{1s}}\|_s\|f_{3s}-\sqrt{f_{X^*}}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}\omega_\theta^{-1}\omega_2^{-1}\bar{f}_3(x^*)\right|dx^*\|\theta-\theta_0\|_s\|[v_n]_{f_2}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_1^{-1}\omega_2^{-1}\bar{f}_3(x^*)\right|dx^*\|f_{1s}-\sqrt{f_\eta}\|_s\|[v_n]_{f_2}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_2^{-1}2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\|[v_n]_{f_2}\|_s\|f_{3s}-\sqrt{f_{X^*}}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\omega_2^{-1}\bar{f}_3(x^*)\right|dx^*\|h_{4s}-\sqrt{h_0}\|_s\|[v_n]_{f_2}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}\omega_\theta^{-1}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\|\theta-\theta_0\|_s\|[v_n]_{f_{3s}}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_1^{-1}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\|f_{1s}-\sqrt{f_\eta}\|_s\|[v_n]_{f_{3s}}\|_s \\
&\quad+\int_{\mathcal{X}^*}\left|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_2^{-1}2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\|f_2-f_{X|X^*}\|_s\|[v_n]_{f_{3s}}\|_s
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{\mathcal{X}^*}\left|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\bar{f}_2(x_i|x^*)2\omega_{34}^{-1}\omega_{34}^{-1}\right|dx^*\|f_{3s}-\sqrt{f_{X^*}}\|_s\|[v_n]_{f_{3s}}\|_s\\
&+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\|h_{4s}-\sqrt{h_0}\|_s\|[v_n]_{f_{3s}}\|_s\\
&+\int_{\mathcal{X}^*}\left|\frac{d^2}{d^2\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2m_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta dh}2\bar{h}_{4s}\omega_{\theta}^{-1}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\|[v_n]_{h_{4s}}\|_s\|\theta-\theta_0\|_s\\
&+\int_{\mathcal{X}^*}\left|2\frac{d}{d\eta}\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_1^{-1}\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\|f_{1s}-\sqrt{f_\eta}\|_s\|[v_n]_{h_{4s}}\|_s\\
&+\int_{\mathcal{X}^*}\left|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\|\frac{d}{d\eta}f_{1s}-\frac{d}{d\eta}\sqrt{f_\eta}\|_s\|[v_n]_{h_{4s}}\|_s\\
&+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\omega_2^{-1}\bar{f}_3(x^*)\right|dx^*\|[v_n]_{h_{4s}}\|_s\|f_2-f_{X|X^*}\|_s\\
&+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\|[v_n]_{h_{4s}}\|_s\|f_{3s}-\sqrt{f_{X^*}}\|_s\\
&+\int_{\mathcal{X}^*}\left|\frac{d^2}{d^2\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2m_0(x^*;\bar{\theta},\bar{h}_4)}{d^2h}2\bar{h}_{4s}\omega_{34}^{-1}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\|[v_n]_{h_{4s}}\|_s\|h_{4s}-\sqrt{h_0}\|_s\Bigg).
\end{aligned}
$$

The definition of the norm leads to the following bound:

(42)

$$
\begin{aligned}
&\left|\frac{1}{f_D(d_i;\bar\alpha_1)}\frac{d^2 f_D(d_i;\bar\alpha_1)}{d\alpha d\alpha^T}[v_n,\alpha-\alpha_0]\right| \\
\leq&\frac{1}{f_D(d_i;\bar\alpha_1)}\Bigg(\int_{\mathcal{X}^*}\Big|\frac{d^2}{d^2\eta}\bar f_1(y_i-m_0(x^*;\bar\theta,\bar h_4))\frac{d^2 m_0(x^*;\bar\theta,\bar h_4)}{d^2\theta}\omega_\theta^{-1}\omega_\theta^{-1}\bar f_2(x_i|x^*)\bar f_3(x^*)\Big|dx^* \\
&+\int_{\mathcal{X}^*}\Big|2\frac{d}{d\eta}\bar{f_{1s}}(y_i-m_0(x^*;\bar\theta,\bar h_4))\omega_1^{-1}\frac{dm_0(x^*;\bar\theta,\bar h_4)}{d\theta}\omega_\theta^{-1}\bar f_2(x_i|x^*)\bar f_3(x^*)dx^* \\
&+\int_{\mathcal{X}^*}\Big|2\bar{f_{1s}}(y_i-m_0(x^*;\bar\theta,\bar h_4))\omega_1^{-1}\frac{dm_0(x^*;\bar\theta,\bar h_4)}{d\theta}\omega_\theta^{-1}\bar f_2(x_i|x^*)\bar f_3(x^*)dx^*\frac{\|\frac{d}{d\eta}f_{1s}-\frac{d}{d\eta}\sqrt{f_\eta}\|_s}{\|f_{1s}-\sqrt{f_\eta}\|_s} \\
&+\int_{\mathcal{X}^*}\Big|\frac{d}{d\eta}\bar f_1(y_i-m_0(x^*;\bar\theta,\bar h_4))\frac{dm_0(x^*;\bar\theta,\bar h_4)}{d\theta}\omega_\theta^{-1}\omega_2^{-1}\bar f_3(x^*)\Big|dx^* \\
&+\int_{\mathcal{X}^*}\Big|\frac{d}{d\eta}\bar f_1(y_i-m_0(x^*;\bar\theta,\bar h_4))\frac{dm_0(x^*;\bar\theta,\bar h_4)}{d\theta}\omega_\theta^{-1}[v_n]_\theta\bar f_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\Big|dx^* \\
&+\int_{\mathcal{X}^*}\Big|\frac{d^2}{d^2\eta}\bar f_1(y_i-m_0(x^*;\bar\theta,\bar h_4))\frac{d^2 m_0(x^*;\bar\theta,\bar h_4)}{d\theta dh}\omega_\theta^{-1}2\bar h_{4s}\omega_{34}^{-1}\bar f_2(x_i|x^*)\bar f_3(x^*)\Big|dx^* \\
&+\int_{\mathcal{X}^*}\Big|2\omega_1^{-1}\omega_1^{-1}f_2(x_i|x^*)\bar f_3(x^*)\Big|dx^* \\
&+\int_{\mathcal{X}^*}\Big|2\bar{f_{1s}}(y_i-m_0(x^*;\bar\theta,\bar h_4))\omega_1^{-1}\omega_2^{-1}\bar f_3(x^*)\Big|dx^* \\
&+\int_{\mathcal{X}^*}\Big|2\bar{f_{1s}}(y_i-m_0(x^*;\bar\theta,\bar h_4))\omega_1^{-1}\bar f_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\Big|dx^* \\
&+\int_{\mathcal{X}^*}\Big|\frac{d}{d\eta}\bar f_1(y_i-m_0(x^*;\bar\theta,\bar h_4))\frac{dm_0(x^*;\bar\theta,\bar h_4)}{d\theta}\omega_\theta^{-1}\omega_2^{-1}\bar f_3(x^*)\Big|dx^* \\
&+\int_{\mathcal{X}^*}\Big|2\bar{f_{1s}}(y_i-m_0(x^*;\bar\theta,\bar h_4))\omega_1^{-1}\omega_2^{-1}\bar f_3(x^*)\Big|dx^*
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{\mathcal{X}^*}\left|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_2^{-1}2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\omega_2^{-1}\bar{f}_3(x^*)\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta}\omega_\theta^{-1}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_1^{-1}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_2^{-1}2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\bar{f}_2(x_i|x^*)2\omega_{34}^{-1}\omega_{34}^{-1}\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|\frac{d^2}{d^2\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2m_0(x^*;\bar{\theta},\bar{h}_4)}{d\theta dh}2\bar{h}_{4s}\omega_\theta^{-1}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|2\frac{d}{d\eta}\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\omega_1^{-1}\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|2\bar{f_{1s}}(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\frac{\|\frac{d}{d\eta}f_{1s}-\frac{d}{d\eta}\sqrt{f_\eta}\|_s}{\|f_{1s}-\sqrt{f_\eta}\|_s}\\
&+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\omega_2^{-1}\bar{f}_3(x^*)\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|\frac{d}{d\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{dm_0(x^*;\bar{\theta},\bar{h}_4)}{dh}2\bar{h}_{4s}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)2\bar{f_{3s}}(x^*)\omega_{34}^{-1}\right|dx^*\\
&+\int_{\mathcal{X}^*}\left|\frac{d^2}{d^2\eta}\bar{f}_1(y_i-m_0(x^*;\bar{\theta},\bar{h}_4))\frac{d^2m_0(x^*;\bar{\theta},\bar{h}_4)}{d^2h}2\bar{h}_{4s}\omega_{34}^{-1}\omega_{34}^{-1}\bar{f}_2(x_i|x^*)\bar{f}_3(x^*)\right|dx^*\Bigg)\|[v_n]\|_s\|\alpha-\alpha_0\|_s\\
&\equiv h_2(d_i,\bar{\alpha},\bar{\omega})\|[v_n]\|_s\|\alpha-\alpha_0\|_s
\end{aligned}
$$

Then Assumption B.14 guarantees the envelope condition and help us to control the linear approximation of the likelihood function near α_0.

D. Proofs related to estimation

Lemma D.1. *Consider the finite-dimensional sieve representation:*

$$
f_2(x|x^*)=\sum_{i=1}^{k_{2,n}}\sum_{j=1}^{k_{2,n}}\beta_{2ij}p_i(x)p_j(x^*). \tag{43}
$$

where the $p_i, i = 1, 2, \ldots$ form an orthonogal basis of $L^2(\mathcal{X})$ while the $p_j, j = 1, 2, \ldots$ form an orthonogal basis of $L^2(\mathcal{X}^*)$. The condition

$$\int f_2(x|x^*)h(x^*)dx^* = 0 \; \forall x \in \mathcal{X} \Rightarrow h(x^*) = 0, \tag{44}$$

for functions h in the space spanned by the p_j, for $j = 1, \ldots, k_{2,n}$, is equivalent to imposing that the square coefficient matrix $[\beta_{2ij}]_{k_{2,n} \times k_{2,n}}$ is invertible.

Proof We also have a sieve expressions for h as

$$h(x^*) = \sum_{k=1}^{k_{2,n}} \gamma_k p_k(x^*).$$

Substituting this sieve expression and Equation (43) into Equation (44) and then applying the orthogonality of $\{p_j(x^*) : j = 1, 2, 3, ...\}$ yields

$$0 = \int \left(\sum_{i=1}^{k_{2,n}} \sum_{j=1}^{k_{2,n}} \beta_{2ij} p_i(x) p_j(x^*) \right) \left(\sum_{k=1}^{k_{2,n}} \gamma_k p_k(x^*) \right) dx^*, \tag{45}$$

$$= \sum_{i=1}^{k_{2,n}} \sum_{j=1}^{k_{2,n}} \beta_{2ij} \gamma_j p_i(x). \tag{46}$$

Because $\{p_i(x) : i = 1, 2, 3, ...\}$ is an orthonormal basis, it follows that

$$0 = \sum_{j=1}^{k_{2,n}} \beta_{2ij} \gamma_j \text{ for all } i.$$

We can express the above relation using a matrix notation as

$$[0]_{k_{2,n} \times 1} = [\beta_{2ij}]_{k_{2,n} \times k_{2,n}} [\gamma_j]_{k_{2,n} \times 1}.$$

If the square coefficient matrix $[\beta_{2ij}]_{k_{2,n} \times k_{2,n}}$ is invertible, then the vector of coefficients $[\gamma_j]_{k_{2,n} \times 1} = 0$. This implies that $h(x^*) = 0$.

References

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

Chen, X., and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 66(2), 289–314.

Dunford, N., and J. T. Schwartz (1971): *Linear Operators*. New-York: Wiley.

Erdös, P., and E. Straus (1953): "On Linear Independence of Sequences in a Banach Space," *Pacific J. Math*, 3, 689–694.

Hu, Y., and S. Schennach (2008): "Supplement to Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1), 195–216.

Newey, W., and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

Newey, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79(1), 147–168.

Rudin, W. (1991): *Functional Analysis*. McGraw-Hill, Inc., New York.

Shen, X. (1997): "On Methods of Sieves and Penalization," *Annals of Statistics*, 25, 2555–2591.

Ushakov, N. G. (1999): *Selected Topics in Characteristic Functions*. Walter de Gruyter.

van der Vaart, A. W. (1991): "On Differentiable Functionals," *Annals of Statistics*, 19, 178–204.

Young, R. (1980): *An Introduction to Nonharmonic Fourier Series*. Academic Press.