Statistica Sinica (2007): Preprint

Nonparametric Identification and Estimation of Nonclassical Errors-in-Variables Models Without Additional Information

Xiaohong Chen, Yingyao Hu, and Arthur Lewbel

Yale University, Johns Hopkins University, and Boston College

Supplementary Material

S1. Appendix

Proof of Theorem 2.1. Notice that $\frac{\partial}{\partial t}\left|\phi_{\eta}(0)\right| = 0$ and $\frac{\partial}{\partial t}a(0) = 0$. we define

$$\frac{\partial}{\partial \mathbf{t}}\Phi_{Y,X}(\mathbf{t}) = \begin{pmatrix} iE\left[Y|X=1\right]f_X(1) & \frac{\partial}{\partial t}\phi_{Y,X=1}(t_2) & ... & \frac{\partial}{\partial t}\phi_{Y,X=1}(t_J) \\ iE\left[Y|X=2\right]f_X(2) & \frac{\partial}{\partial t}\phi_{Y,X=2}(t_2) & ... & \frac{\partial}{\partial t}\phi_{Y,X=2}(t_J) \\ ... & ... & ... & ... \\ iE\left[Y|X=J\right]f_X(J) & \frac{\partial}{\partial t}\phi_{Y,X=J}(t_2) & ... & \frac{\partial}{\partial t}\phi_{Y,X=J}(t_J) \end{pmatrix}.$$

By taking the derivative with respect to scalar t, we have from equation (2.3)

$$\begin{aligned} \frac{\partial}{\partial t}\phi_{Y,X=x}(t) &= \left(\frac{\partial}{\partial t}\left|\phi_{\eta}(t)\right|\right)\sum_{x^*}\exp\left(itm(x^*)+ia(t)\right)f_{X,X^*}(x,x^*) \\ &\quad +i\left(\frac{\partial}{\partial t}a(t)\right)\left|\phi_{\eta}(t)\right|\sum_{x^*}\exp\left(itm(x^*)+ia(t)\right)f_{X,X^*}(x,x^*) \\ &\quad +i\left|\phi_{\eta}(t)\right|\sum_{x^*}\exp\left(itm(x^*)+ia(t)\right)m(x^*)f_{X,X^*}(x,x^*). \end{aligned} \tag{S1.1}$$

Equation (S1.1) is equivalent to

$$\begin{aligned} \frac{\partial}{\partial \mathbf{t}}\Phi_{Y,X}(\mathbf{t}) &= F_{X,X^*}\Phi_{m,a}(\mathbf{t})D_{\partial|\phi|}(\mathbf{t}) \qquad \text{(S1.2)} \\ &\quad +iF_{X,X^*}\Phi_{m,a}(\mathbf{t})D_{|\phi|}(\mathbf{t})D_{\partial a}(\mathbf{t})+iF_{X,X^*}D_m\Phi_{m,a}(\mathbf{t})D_{|\phi|}(\mathbf{t}), \end{aligned}$$

where $D_{\partial|\phi|}(\mathbf{t}) = Diag\{0, \frac{\partial}{\partial t}|\phi_\eta(t_2)|, ..., \frac{\partial}{\partial t}|\phi_\eta(t_J)|\}$, $D_{\partial a}(\mathbf{t}) = Diag\{0, \frac{\partial}{\partial t}a(t_2), ..., \frac{\partial}{\partial t}a(t_J)\}$, $D_m = Diag\{m_1, ..., m_J\}$. Since by definition, $D_{\partial|\phi|}(\mathbf{t})$ and $D_{\partial a}(\mathbf{t})$ are real-valued, we also have from equation (S1.2)

$$\begin{aligned} Re\{\frac{\partial}{\partial \mathbf{t}}\Phi_{Y,X}(\mathbf{t})\} &= F_{X,X^*}\, Re\{\Phi_{m,a}(\mathbf{t})\}D_{\partial|\phi|}(\mathbf{t}) \\ &\quad -F_{X,X^*}\, Im\{\Phi_{m,a}(\mathbf{t})\}D_{|\phi|}(\mathbf{t})D_{\partial a}(\mathbf{t}) \\ &\quad -F_{X,X^*}D_m\, Im\{\Phi_{m,a}(\mathbf{t})\}D_{|\phi|}(\mathbf{t}). \end{aligned}$$

In order to replace the singular matrix $Im\{\Phi_{m,a}(\mathbf{t})\}$ with the invertible $(Im\{\Phi_{m,a}(\mathbf{t})\} + \Upsilon)$, we define

$$\Upsilon_{E[Y|X]} = \begin{pmatrix} E[Y|X=1]\, f_X(1) & 0 & ... & 0 \\ E[Y|X=2]\, f_X(2) & 0 & ... & 0 \\ ... & ... & ... & ... \\ E[Y|X=J]\, f_X(J) & 0 & ... & 0 \end{pmatrix} = F_{X,X^*}D_m\Upsilon.$$

We then have

$$\begin{aligned} & \left(Re\{\frac{\partial}{\partial \mathbf{t}}\Phi_{Y,X}(\mathbf{t})\} - \Upsilon_{E[Y|X]}\right) && \text{(S1.3)} \\ = \;& F_{X,X^*}\, Re\{\Phi_{m,a}(\mathbf{t})\}D_{\partial|\phi|}(\mathbf{t}) && \text{(S1.4)} \\ - \;& F_{X,X^*}\left(Im\{\Phi_{m,a}(\mathbf{t})\} + \Upsilon\right)D_{|\phi|}(\mathbf{t})D_{\partial a}(\mathbf{t}) \\ - \;& F_{X,X^*}D_m\left(Im\{\Phi_{m,a}(\mathbf{t})\} + \Upsilon\right)D_{|\phi|}(\mathbf{t}), \end{aligned}$$

where $\Upsilon D_{|\phi|}(\mathbf{t})D_{\partial a}(\mathbf{t}) = 0$ and $\Upsilon = \Upsilon D_{|\phi|}(\mathbf{t})$. Similarly, we have

$$\begin{aligned} Im\{\frac{\partial}{\partial \mathbf{t}}\Phi_{Y,X}(\mathbf{t})\} &= F_{X,X^*}\, Im\{\Phi_{m,a}(\mathbf{t})\}D_{\partial|\phi|}(\mathbf{t}) \\ &\quad +F_{X,X^*}\, Re\{\Phi_{m,a}(\mathbf{t})\}D_{|\phi|}(\mathbf{t})D_{\partial a}(\mathbf{t}) \\ &\quad +F_{X,X^*}D_m\, Re\{\Phi_{m,a}(\mathbf{t})\}D_{|\phi|}(\mathbf{t}) \\ &= F_{X,X^*}\left(Im\{\Phi_{m,a}(\mathbf{t})\} + \Upsilon\right)D_{\partial|\phi|}(\mathbf{t}) \\ &\quad +F_{X,X^*}\, Re\{\Phi_{m,a}(\mathbf{t})\}D_{|\phi|}(\mathbf{t})D_{\partial a}(\mathbf{t}) \\ &\quad +F_{X,X^*}D_m\, Re\{\Phi_{m,a}(\mathbf{t})\}D_{|\phi|}(\mathbf{t}), \end{aligned}$$

where $\Upsilon D_{\partial|\phi|}(\mathbf{t}) = 0$. Define $\Phi_{Y|X^*}(\mathbf{t}) = \Phi_{m,a}(\mathbf{t}) D_{|\phi|}(\mathbf{t})$, then we have

$$Re\{\Phi_{Y|X^*}(\mathbf{t})\} = Re\{\Phi_{m,a}(\mathbf{t})\} D_{|\phi|}(\mathbf{t}),$$

$$\left(Im\{\Phi_{Y|X^*}(\mathbf{t})\} + \Upsilon\right) = \left(Im\{\Phi_{m,a}(\mathbf{t})\} + \Upsilon\right) D_{|\phi|}(\mathbf{t}).$$

In summary, we have

$$\begin{aligned}
Re\{\Phi_{Y,X}(\mathbf{t})\} &= F_{X,X^*}\, Re\{\Phi_{Y|X^*}(\mathbf{t})\}, && \text{(S1.5)}\\
\left(Im\{\Phi_{Y,X}(\mathbf{t})\} + \Upsilon_X\right) &= F_{X,X^*}\left(Im\{\Phi_{Y|X^*}(\mathbf{t})\} + \Upsilon\right), && \text{(S1.6)}\\
\left(Re\,\frac{\partial}{\partial \mathbf{t}}\Phi_{Y,X}(\mathbf{t}) - \Upsilon_{E[Y|X]}\right) &= F_{X,X^*}\, Re\,\Phi_{m,a}(\mathbf{t}) D_{\partial|\phi|}(\mathbf{t}) && \text{(S1.7)}\\
&\quad - F_{X,X^*}\left(Im\,\Phi_{Y|X^*}(\mathbf{t}) + \Upsilon\right) D_{\partial a}(\mathbf{t})\\
&\quad - F_{X,X^*} D_m \left(Im\,\Phi_{Y|X^*}(\mathbf{t}) + \Upsilon\right),\\
Im\,\frac{\partial}{\partial \mathbf{t}}\Phi_{Y,X}(\mathbf{t}) &= F_{X,X^*}\left(Im\,\Phi_{m,a}(\mathbf{t}) + \Upsilon\right) D_{\partial|\phi|}(\mathbf{t})\\
&\quad + F_{X,X^*}\, Re\,\Phi_{Y|X^*}(\mathbf{t}) D_{\partial a}(\mathbf{t})\\
&\quad + F_{X,X^*} D_m\, Re\,\Phi_{Y|X^*}(\mathbf{t}). && \text{(S1.8)}
\end{aligned}$$

The left-hand sides of these equations are all observed, while the right-hand sides contain all the unknowns. Assumption 2.3(i) also implies that F_{X,X^*}, $Re\{\Phi_{m,a}(\mathbf{t})\}$ and $(Im\{\Phi_{m,a}(\mathbf{t})\} + \Upsilon)$ are invertible in equations (2.5) and (2.7). Recall the definition of the observed matrix $C_{\mathbf{t}}$, which by equations (S1.5) and (S1.6) equals

$$C_{\mathbf{t}} \equiv \left(Re\,\Phi_{Y,X}(\mathbf{t})\right)^{-1}\left(Im\,\Phi_{Y,X}(\mathbf{t}) + \Upsilon_X\right) = \left(Re\,\Phi_{Y|X^*}(\mathbf{t})\right)^{-1}\left(Im\,\Phi_{Y|X^*}(\mathbf{t}) + \Upsilon\right).$$

Denote $A_{\mathbf{t}} \equiv \left(Re\,\Phi_{Y|X^*}(\mathbf{t})\right)^{-1} D_m\, Re\,\Phi_{Y|X^*}(\mathbf{t})$. With equations (S1.5) and (S1.7), we consider

$$
\begin{aligned}
B_R &\equiv (Re\,\Phi_{Y,X}(\mathbf{t}))^{-1}\left(Re\,\frac{\partial}{\partial \mathbf{t}}\Phi_{Y,X}(\mathbf{t}) - \Upsilon_{E[Y|X]}\right) \\
&= \left(Re\,\Phi_{m,a}(\mathbf{t})D_{|\phi|}(\mathbf{t})\right)^{-1} Re\,\Phi_{m,a}(\mathbf{t})D_{\partial|\phi|}(\mathbf{t}) \\
&\quad - \left(Re\,\Phi_{Y|X^*}(\mathbf{t})\right)^{-1}\left(Im\,\Phi_{Y|X^*}(\mathbf{t}) + \Upsilon\right) D_{\partial a}(\mathbf{t}) \qquad \text{(S1.9)} \\
&\quad - \left(Re\,\Phi_{Y|X^*}(\mathbf{t})\right)^{-1} D_m\left(Im\,\Phi_{Y|X^*}(\mathbf{t}) + \Upsilon\right) \\
&= [D_{|\phi|}(\mathbf{t})]^{-1} D_{\partial|\phi|}(\mathbf{t}) - C_{\mathbf{t}} D_{\partial a}(\mathbf{t}) \qquad \text{(S1.10)} \\
&\quad - \left(\left(Re\,\Phi_{Y|X^*}(\mathbf{t})\right)^{-1} D_m\, Re\,\Phi_{Y|X^*}(\mathbf{t})\right) C_{\mathbf{t}} \\
&\equiv D_{\partial \ln|\phi|}(\mathbf{t}) - C_{\mathbf{t}} D_{\partial a}(\mathbf{t}) - A_{\mathbf{t}} C_{\mathbf{t}}. \qquad \text{(S1.11)}
\end{aligned}
$$

Similarly, we have by equations (S1.6) and (S1.8)

$$
\begin{aligned}
B_I &\equiv (Im\,\Phi_{Y,X}(\mathbf{t}) + \Upsilon_X)^{-1}\left(Im\,\frac{\partial}{\partial \mathbf{t}}\Phi_{Y,X}(\mathbf{t})\right) \\
&= \left(\left(Im\,\Phi_{m,a}(\mathbf{t}) + \Upsilon\right) D_{|\phi|}(\mathbf{t})\right)^{-1}\left(Im\,\Phi_{m,a}(\mathbf{t}) + \Upsilon\right) D_{\partial|\phi|}(\mathbf{t}) \\
&\quad + \left(Im\,\Phi_{Y|X^*}(\mathbf{t}) + \Upsilon\right)^{-1} Re\,\Phi_{Y|X^*}(\mathbf{t}) D_{\partial a}(\mathbf{t}) \qquad \text{(S1.12)} \\
&\quad + \left(Im\,\Phi_{Y|X^*}(\mathbf{t}) + \Upsilon\right)^{-1} D_m\, Re\,\Phi_{Y|X^*}(\mathbf{t}), \\
&= D_{\partial \ln|\phi|}(\mathbf{t}) + C_{\mathbf{t}}^{-1} D_{\partial a}(\mathbf{t}) + C_{\mathbf{t}}^{-1} A_{\mathbf{t}} \qquad \text{(S1.13)}
\end{aligned}
$$

We eliminate the matrix $A_{\mathbf{t}}$ in equations (S1.11) and (S1.13) to have

$$
\begin{aligned}
&B_R + C_{\mathbf{t}} B_I C_{\mathbf{t}} \\
= \;& D_{\partial \ln|\phi|}(\mathbf{t}) + C_{\mathbf{t}} D_{\partial \ln|\phi|}(\mathbf{t}) C_{\mathbf{t}} + D_{\partial a}(\mathbf{t}) C_{\mathbf{t}} - C_{\mathbf{t}} D_{\partial a}(\mathbf{t}).
\end{aligned} \qquad \text{(S1.14)}
$$

Notice that both $D_{\partial \ln|\phi|}(\mathbf{t})$ and $D_{\partial a}(\mathbf{t})$ are diagonal, Assumption 2.3(ii) implies that $D_{\partial \ln|\phi|}(\mathbf{t})$ and $D_{\partial a}(\mathbf{t})$ are uniquely identified from equation (S1.14).

Since the diagonal terms of $(D_{\partial a}(\mathbf{t})C_{\mathbf{t}} - C_{\mathbf{t}}D_{\partial a}(\mathbf{t}))$ are zeros, we have

$$\begin{aligned} diag\left(B_R + C_{\mathbf{t}}B_I C_{\mathbf{t}}\right) &= diag\left(D_{\partial \ln|\phi|}(\mathbf{t})\right) + \left(C_{\mathbf{t}} \circ C_{\mathbf{t}}^T\right) diag\left(D_{\partial \ln|\phi|}(\mathbf{t})\right) \\ &\quad + D_{\partial a}(\mathbf{t}) diag\left(C_{\mathbf{t}}\right) - D_{\partial a}(\mathbf{t}) diag\left(C_{\mathbf{t}}\right) \\ &= \left[\left(C_{\mathbf{t}} \circ C_{\mathbf{t}}^T\right) + I\right] diag\left(D_{\partial \ln|\phi|}(\mathbf{t})\right), \end{aligned}$$

where the function $diag(\cdot)$ generates a vector of the diagonal entries of its argument and the notation "$\circ$" stands for the Hadamard product or the element-wise product. By assumption 2.5(i), we may solve $D_{\partial \ln|\phi|}(\mathbf{t})$ as follows:

$$diag\left(D_{\partial \ln|\phi|}(\mathbf{t})\right) = \left\{\left(C_{\mathbf{t}} \circ C_{\mathbf{t}}^T\right) + I\right\}^{-1} diag\left(B_R + C_{\mathbf{t}}B_I C_{\mathbf{t}}\right). \quad \text{(S1.15)}$$

Furthermore, equation (S1.14) implies that

$$\begin{aligned} U &\equiv B_R + C_{\mathbf{t}}B_I C_{\mathbf{t}} - D_{\partial \ln|\phi|}(\mathbf{t}) - C_{\mathbf{t}}D_{\partial \ln|\phi|}(\mathbf{t})C_{\mathbf{t}} \quad \text{(S1.16)} \\ &= D_{\partial a}(\mathbf{t})C_{\mathbf{t}} - C_{\mathbf{t}}D_{\partial a}(\mathbf{t}), \end{aligned}$$

Define a J by 1 vector $e_1 = (1, 0, 0, ..., 0)^T$. The definition of $D_{\partial a}(\mathbf{t})$ implies that $e_1^T D_{\partial a}(\mathbf{t}) = 0$. Therefore, equation S1.16 implies $e_1^T U = -e_1^T C_{\mathbf{t}} D_{\partial a}(\mathbf{t})$. Assumption 2.5(ii) implies that all the entries in the row vector $e_1^T C_{\mathbf{t}}$ are nonzero. Let $e_1^T C_{\mathbf{t}} \equiv (c_{11}, c_{12}, ..., c_{1J})$. The vector $Diag\left(D_{\partial a}(\mathbf{t})\right)$ is then uniquely determined as: $Diag\left(D_{\partial a}(\mathbf{t})\right) = -\left(Diag\{c_{11}, ..., c_{1J}\}\right)^{-1} U^T e_1$. We can then identify the $A_{\mathbf{t}}$ which is defined as $A_{\mathbf{t}} \equiv \left(Re\,\Phi_{Y|X^*}(\mathbf{t})\right)^{-1} D_m\, Re\,\Phi_{Y|X^*}(\mathbf{t})$ from equation (S1.13): $A_{\mathbf{t}} = C_{\mathbf{t}}\left(B_I - D_{\partial \ln|\phi|}(\mathbf{t})\right) - D_{\partial a}(\mathbf{t})$. Notice that

$$Re\,\Phi_{Y|X^*}(\mathbf{t}) = \left(F_{X,X^*}\right)^{-1} Re\,\Phi_{Y,X}(\mathbf{t}) = \left(F_{X|X^*}F_{X^*}\right)^{-1} Re\,\Phi_{Y,X}(\mathbf{t})$$

where $F_{X,X^*} = F_{X|X^*}F_{X^*}$, with $F_{X^*} = Diag\{f_{X^*}(1), ..., f_{X^*}(J)\}$. Thus,

$$\begin{aligned} Re\,\Phi_{Y,X}(\mathbf{t})A_{\mathbf{t}}\left(Re\,\Phi_{Y,X}(\mathbf{t})\right)^{-1} &= \left(F_{X|X^*}F_{X^*}\right) D_m \left(F_{X|X^*}F_{X^*}\right)^{-1} \\ &= F_{X|X^*} D_m \left(F_{X|X^*}\right)^{-1}. \quad \text{(S1.17)} \end{aligned}$$

Equation (S1.17) implies that the unknowns m_j in matrix D_m are eigenvalues of a directly estimatable matrix on the left-hand side, and each column in the matrix $F_{X|X^*}$ is an eigenvector. Assumption 2.4 guarantees that all the eigenvalues are distinctive and nonzero in the diagonalization in equation (S1.17). We may then identify m_j as the roots of $\det(A_{\mathbf{t}} - m_j I) = 0$. To be specific, m_j may be identified as the j-th smallest root. Equation (S1.17) also implies that the j-th column in the matrix $F_{X|X^*}$ is the eigenvector corresponding to the eigenvalue m_j. Notice that each eigenvector is already normalized because each column of $F_{X|X^*}$ is a conditional density and the sum of entries in each column equals one. Therefore, each column of $F_{X|X^*}$ is identified as normalized eigenvectors corresponding to each eigenvalue m_j. Finally, we may identify f_{Y,X^*} through equation (2.1) as follows, for any $y \in \mathcal{Y}$.

$$\begin{aligned} & \begin{pmatrix} f_{Y,X^*}(y,1) & f_{Y,X^*}(y,2) & \dots & f_{Y,X^*}(y,J) \end{pmatrix}^T \\ = \; & F_{X|X^*}^{-1} \begin{pmatrix} f_{Y,X}(y,1) & f_{Y,X}(y,2) & \dots & f_{Y,X}(y,J) \end{pmatrix}^T . \end{aligned}$$

The identification of the joint distribution f_{Y,X^*} implies that both the latent model $f_{Y|X^*}$ and the marginal distribution of X^*, i.e., f_{X^*}, are identified. □

Acknowledgment We thank participants at June 2007 North American Summer Meetings of the Econometric Society at Duke for helpful comments. Chen acknowledges support from NSF Grant SES-0631613.

Department of Economics, Yale University, New Haven, CT 06520, USA

E-mail: (xiaohong.chen@yale.edu)

Department of Economics, Johns Hopkins University, Baltimore, MD 21218, USA

E-mail: (yhu@jhu.edu)

Department of Economics, Boston College, Boston, MA 02467, USA

E-mail: (lewbel@bc.edu)

Table S1.1: Example 1 with sample size n=1000

Value of x^*:	1	2	3	4
Regression function $m(x^*)$:				
– true value	1.3500	2.8000	5.9500	11.400
– mean estimate	1.2535	3.0927	6.1445	11.505
– standard error	0.45839	0.59858	0.54063	0.59807
Marginal distribution $\Pr(x^*)$:				
– true value	0.2	0.3	0.3	0.2
– mean estimate	0.24207	0.27678	0.28511	0.19604
– standard error	0.18769	0.20536	0.059821	0.026285
Misclasification Prob. $f_{x\|x^*}(\cdot\|x^*)$:				
– true value	0.6	0.2	0.1	0.1
	0.2	0.6	0.1	0.1
	0.1	0.1	0.7	0.1
	0.1	0.1	0.1	0.7
– mean estimate	0.54112	0.21293	0.096892	0.097281
	0.26198	0.54021	0.10390	0.096685
	0.095379	0.15299	0.69051	0.10076
	0.10152	0.093865	0.10870	0.70527
– standard error	0.10306	0.077743	0.031280	0.026116
	0.095323	0.10473	0.047729	0.032475
	0.051416	0.085318	0.077425	0.048582
	0.032624	0.033011	0.052891	0.054561

Table S1.2: Example 1 with sample size n=500

Value of x^*:	1	2	3	4
Regression function $m(x^*)$:				
– true value	1.3500	2.8000	5.9500	11.400
– mean estimate	1.2438	3.2815	6.3482	11.620
– standard error	0.56137	0.88858	1.0824	0.92184
Marginal distribution $\Pr(x^*)$:				
– true value	0.2	0.3	0.3	0.2
– mean estimate	0.26163	0.25911	0.28491	0.19435
– standard error	0.29562	0.38938	0.17555	0.063135
Misclasification Prob. $f_{x\|x^*}(\cdot\|x^*)$:				
– true value	0.6	0.2	0.1	0.1
	0.2	0.6	0.1	0.1
	0.1	0.1	0.7	0.1
	0.1	0.1	0.1	0.7
– mean estimate	0.51101	0.21140	0.097463	0.095530
	0.28206	0.50973	0.11049	0.095992
	0.10445	0.18175	0.66800	0.10439
	0.10248	0.097114	0.12405	0.70409
– standard error	0.11994	0.087891	0.044424	0.036802
	0.10616	0.13413	0.062208	0.042780
	0.065647	0.11954	0.12420	0.056882
	0.043470	0.042879	0.099284	0.073266

Table S1.3: Example 1 with sample size n=200

Value of x^*:	1	2	3	4
Regression function $m(x^*)$:				
– true value	1.3500	2.8000	5.9500	11.400
– mean estimate	1.2653	3.5975	6.7691	11.927
– standard error	0.69055	1.3787	1.8141	1.4144
Marginal distribution $\Pr(x^*)$:				
– true value	0.2	0.3	0.3	0.2
– mean estimate	0.33554	0.18632	0.28717	0.19097
– standard error	0.42358	0.69672	0.47600	0.087511
Misclasification Prob. $f_{x\|x^*}(\cdot\|x^*)$:				
– true value	0.6	0.2	0.1	0.1
	0.2	0.6	0.1	0.1
	0.1	0.1	0.7	0.1
	0.1	0.1	0.1	0.7
– mean estimate	0.46238	0.21570	0.10419	0.10039
	0.31637	0.44537	0.12597	0.10601
	0.11719	0.23344	0.61520	0.11320
	0.10406	0.10550	0.15464	0.68041
– standard error	0.13274	0.10449	0.063146	0.053476
	0.10870	0.15334	0.076515	0.062934
	0.079978	0.14483	0.16768	0.078462
	0.052014	0.057346	0.14346	0.10967

Table S1.4: Example 1 with sample size n=100

Value of x^*:	1	2	3	4
Regression function $m(x^*)$:				
– true value	1.3500	2.8000	5.9500	11.400
– mean estimate	1.2584	4.0330	7.3602	12.417
– standard error	0.78434	1.7707	2.4457	1.9903
Marginal distribution $\Pr(x^*)$:				
– true value	0.2	0.3	0.3	0.2
– mean estimate	0.42243	0.13077	0.24800	0.19879
– standard error	0.72523	1.1647	0.80987	0.32021
Misclasification Prob. $f_{x\|x^*}(\cdot\|x^*)$:				
– true value	0.6	0.2	0.1	0.1
	0.2	0.6	0.1	0.1
	0.1	0.1	0.7	0.1
	0.1	0.1	0.1	0.7
– mean estimate	0.42054	0.22427	0.12112	0.10406
	0.33386	0.38130	0.14933	0.12054
	0.14139	0.26949	0.53992	0.13939
	0.10421	0.12494	0.18963	0.63601
– standard error	0.14234	0.11361	0.084698	0.074355
	0.10895	0.15907	0.099828	0.087488
	0.095047	0.15521	0.20407	0.11304
	0.058203	0.075012	0.16123	0.16887

Table S1.5: Example 2 with sample size n=1000

Value of x^*:	1	2	3	4
Regression function $m(x^*)$:				
– true value	1.3500	2.8000	5.9500	11.400
– mean estimate	1.2655	3.5624	6.5922	11.810
– standard error	0.69075	1.1590	1.4835	1.3222
Marginal distribution $\Pr(x^*)$:				
– true value	0.2	0.3	0.3	0.2
– mean estimate	0.32699	0.22362	0.25335	0.19604
– standard error	0.59020	0.89173	0.38530	0.068205
Misclasification Prob. $f_{x\|x^*}(\cdot\|x^*)$:				
– true value	0.5220	0.1262	0.2180	0.2994
	0.1881	0.4968	0.1719	0.2489
	0.1829	0.1699	0.4126	0.0381
	0.1070	0.2071	0.1976	0.4137
– mean estimate	0.41602	0.19474	0.22935	0.28958
	0.27239	0.38766	0.17172	0.24737
	0.17525	0.22750	0.38095	0.054362
	0.13634	0.19010	0.21799	0.40869
– standard error	0.11541	0.10058	0.062443	0.040801
	0.088955	0.10607	0.072264	0.045202
	0.054288	0.074511	0.11160	0.047605
	0.041825	0.037886	0.062883	0.044622

Table S1.6: Example 2 with sample size n=500

Value of x^*:	1	2	3	4
Regression function $m(x^*)$:				
– true value	1.3500	2.8000	5.9500	11.400
– mean estimate	1.2967	3.8909	7.0886	12.159
– standard error	0.78357	1.4776	1.9645	1.7987
Marginal distribution $\Pr(x^*)$:				
– true value	0.2	0.3	0.3	0.2
– mean estimate	0.33689	0.23911	0.24200	0.18200
– standard error	0.48300	0.69234	0.44014	0.16836
Misclasification Prob. $f_{x\|x^*}(\cdot\|x^*)$:				
– true value	0.5220	0.1262	0.2180	0.2994
	0.1881	0.4968	0.1719	0.2489
	0.1829	0.1699	0.4126	0.0381
	0.1070	0.2071	0.1976	0.4137
– mean estimate	0.38113	0.20487	0.22888	0.28304
	0.29806	0.36538	0.18042	0.24700
	0.17469	0.23807	0.36040	0.069700
	0.14612	0.19168	0.23031	0.40027
– standard error	0.11423	0.10008	0.076727	0.061401
	0.092344	0.10526	0.093464	0.068755
	0.062018	0.084518	0.13151	0.085666
	0.046722	0.045024	0.076361	0.074851

Table S1.7: Example 2 with sample size n=200

Value of x^*:	1	2	3	4
Regression function $m(x^*)$:				
– true value	1.3500	2.8000	5.9500	11.400
– mean estimate	1.4450	4.5772	8.1681	12.951
– standard error	0.97746	1.8481	2.6165	2.3871
Marginal distribution $\Pr(x^*)$:				
– true value	0.2	0.3	0.3	0.2
– mean estimate	0.38658	0.27126	0.15170	0.19045
– standard error	0.72173	0.97354	0.86185	0.58516
Misclasification Prob. $f_{x\|x^*}(\cdot\|x^*)$:				
– true value	0.5220	0.1262	0.2180	0.2994
	0.1881	0.4968	0.1719	0.2489
	0.1829	0.1699	0.4126	0.0381
	0.1070	0.2071	0.1976	0.4137
– mean estimate	0.34295	0.21459	0.23541	0.27153
	0.31561	0.32675	0.20579	0.24831
	0.18730	0.26303	0.30208	0.093243
	0.15414	0.19563	0.25672	0.38692
– standard error	0.11151	0.098387	0.090719	0.084699
	0.089650	0.10369	0.10193	0.093076
	0.080727	0.095566	0.14814	0.10467
	0.056924	0.062477	0.10268	0.10288

Table S1.8: Example 2 with sample size n=100

Value of x^*:	1	2	3	4
Regression function $m(x^*)$:				
– true value	1.3500	2.8000	5.9500	11.400
– mean estimate	1.5340	5.2229	9.2296	13.794
– standard error	1.0386	2.2521	3.0175	2.8465
Marginal distribution $\Pr(x^*)$:				
– true value	0.2	0.3	0.3	0.2
– mean estimate	0.44919	0.25324	0.14859	0.14898
– standard error	0.72846	0.97166	0.66633	0.32666
Misclasification Prob. $f_{x\|x^*}(\cdot\|x^*)$:				
– true value	0.5220	0.1262	0.2180	0.2994
	0.1881	0.4968	0.1719	0.2489
	0.1829	0.1699	0.4126	0.0381
	0.1070	0.2071	0.1976	0.4137
– mean estimate	0.32472	0.22633	0.24651	0.26078
	0.31834	0.30906	0.22748	0.24987
	0.19551	0.26009	0.25832	0.12802
	0.16143	0.20451	0.26769	0.36132
– standard error	0.11697	0.10124	0.10612	0.10644
	0.098660	0.10838	0.11884	0.11472
	0.087243	0.10084	0.14835	0.13381
	0.065870	0.079621	0.12507	0.14271

Table S1.9: Example 2 (n=1000) with t randomly picked from a standard normal

Value of x^*:	1	2	3	4
Regression function $m(x^*)$:				
– true value	1.3500	2.8000	5.9500	11.400
– mean estimate	1.2460	3.7240	6.9216	12.121
– standard error	0.72486	1.3954	1.8783	1.7830
Marginal distribution $\Pr(x^*)$:				
– true value	0.2	0.3	0.3	0.2
– mean estimate	0.30060	0.27676	0.22457	0.19807
– standard error	0.30619	0.57637	0.65251	0.41582
Misclasification Prob. $f_{x\|x^*}(\cdot\|x^*)$:				
– true value	0.5220	0.1262	0.2180	0.2994
	0.1881	0.4968	0.1719	0.2489
	0.1829	0.1699	0.4126	0.0381
	0.1070	0.2071	0.1976	0.4137
– mean estimate	0.40281	0.19205	0.22770	0.28828
	0.28227	0.39096	0.17600	0.24100
	0.17675	0.22192	0.37572	0.068580
	0.13818	0.19507	0.22058	0.40214
– standard error	0.11480	0.090727	0.062887	0.051100
	0.094142	0.10648	0.083599	0.061313
	0.055361	0.077751	0.12000	0.082745
	0.040389	0.040630	0.067471	0.064290

Table S1.10: Sieve MLE with sample size n=1000

true value of β:	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_3 = 1$
ignoring meas. error:			
– mean estimate	2.275	1.657	0.9371
– standard error	0.1797	0.1775	0.1260
– root mse	1.287	0.6803	0.1408
infeasible MLE:			
– mean estimate	0.9928	1.023	0.9808
– standard error	0.09695	0.1178	0.1047
– root mse	0.09722	0.1201	0.1064
sieve MLE:			
– mean estimate	0.9225	1.025	0.9991
– standard error	0.2630	0.1321	0.3306
– root mse	0.2741	0.1344	0.3306

Table S1.11: Sieve MLE with sample size n=500

true value of β:	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_3 = 1$
ignoring meas. error:			
– mean estimate	2.216	1.627	0.9450
– standard error	0.2683	0.2294	0.2112
– root mse	1.245	0.6674	0.2182
infeasible MLE:			
– mean estimate	0.9810	1.039	0.9754
– standard error	0.1423	0.1146	0.1565
– root mse	0.1436	0.1209	0.1584
sieve MLE:			
– mean estimate	0.8731	1.108	1.005
– standard error	0.2773	0.1972	0.4222
– root mse	0.3050	0.2247	0.4222

Table S1.12: Sieve MLE with sample size n=200

true value of β:	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_3 = 1$
ignoring meas. error:			
– mean estimate	2.192	1.664	0.9435
– standard error	0.4093	0.3692	0.4194
– root mse	1.261	0.7597	0.4232
infeasible MLE:			
– mean estimate	0.9999	1.082	0.9453
– standard error	0.2592	0.1913	0.3343
– root mse	0.2592	0.2083	0.3387
sieve MLE:			
– mean estimate	0.8641	1.289	0.9144
– standard error	0.3186	0.2578	0.4762
– root mse	0.3463	0.3871	0.4839

Table S1.13: Sieve MLE with sample size n=100

true value of β:	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_3 = 1$
ignoring meas. error:			
– mean estimate	2.052	1.665	0.9545
– standard error	0.5395	0.5865	0.5289
– root mse	1.183	0.8865	0.5309
infeasible MLE:			
– mean estimate	0.9396	1.134	0.9401
– standard error	0.3898	0.3065	0.4554
– root mse	0.3944	0.3345	0.4593
sieve MLE:			
– mean estimate	0.8539	1.420	0.8753
– standard error	0.3133	0.3358	0.4122
– root mse	0.3457	0.5378	0.4307