On Goodness of Fit in Non-parametric Measurement Error Model

Chitradipa Chakraborty
Symposium

Statistics and Mathematics Unit
Indian Statistical Institute, Bangalore Centre

August 22, 2019

- Data set : $\{y, x_1, x_2, \ldots, x_p\}$
- y : dependent variable
- $x_1, x_2, \ldots, x_p$: independent variables / regressors
- Multiple linear regression model : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$
- $\beta_0, \beta_1, \ldots, \beta_p$: regression coefficients (unknown)
- ϵ : i.i.d. random error with $E(\epsilon) = 0$ and $V(\epsilon) = \sigma_\epsilon^2$
- Ordinary least square estimation (OLSE) : minimize residual sum of squares $\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip})^2$ w.r.t. $\beta_0, \beta_1, \ldots, \beta_p$.
- $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$: OLS estimators

Goodness of Fit in Multiple Linear Regression Model

- Fitted y : $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_p x_{ip}$
- Coefficient of determination (R^2) : measures how well observed data are replicated by the model.
- $R^2 = \dfrac{\text{sum of squares due to regression (SSR)}}{\text{total sum of squares (TSS)}}$
- $\text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \text{SSE} + \text{SSR}$
- $\bar{y} = \dfrac{1}{n} \sum_{i=1}^{n} y_i$
- SSE and SSR are independently distributed and mutually orthogonal components.

Non-parametric Regression Model

- Data set : $\{y, x_1, x_2, \ldots, x_p\}$
- y : dependent variable
- $x_1, x_2, \ldots, x_p$: independent variables / regressors
- Non-parametric regression model : $y_i = m(x_{i1}, x_{i2}, \ldots, x_{ip}) + \epsilon_i$
- $m(.)$: regression function (unknown)
- ϵ : i.i.d. random error with $E(\epsilon) = 0$ and $V(\epsilon) = \sigma_\epsilon^2$
- Kernel based estimators :
  - Nadarya-Watson estimator (see Nadarya (1964), Watson (1964))
  - Pristley-Chao estimator (see Pristley and Chao (1970))

Non-parametric Measurement Error Model

- $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})' \xrightarrow{\mathbf{x}_i^* = \mathbf{x}_i + \boldsymbol{\eta}_i} \mathbf{x}_i^* = (x_{i1}^\star, x_{i2}^\star, \dots, x_{ip}^\star)'$
  true but unobserved — observed
- $\boldsymbol{\eta}$: measurement error or error-in-variables, the difference between the true and observed values of a variable
- $\boldsymbol{\eta}$: $(p \times 1)$ i.i.d. random vectors with $E(\boldsymbol{\eta}_i) = \mathbf{0}$ and $E(\boldsymbol{\eta}_i \boldsymbol{\eta}_j') = 0 \; \forall i \neq j$
- Non-parametric measurement error model : $y_i = m(x_{i1}^\star, x_{i2}^\star, \dots, x_{ip}^\star) + \epsilon_i$
- $m(.)$: regression function (unknown)
- ϵ : i.i.d. random error with $E(\epsilon) = 0$ and $V(\epsilon) = \sigma_\epsilon^2$
- ϵ is uncorrelated with every component of $\boldsymbol{\eta}$.
- Kernel based estimators :
  - Nadarya-Watson estimator (see Nadarya (1964), Watson (1964))
  - Pristley-Chao estimator (see Pristley and Chao (1970))

- R^2 statistic is based on the partitioning of TSS into two orthogonal components, viz., SSR and SSE
- in case of measurement error models, such partitioning of sum of squares is not possible
- the traditional R^2 cannot be used as goodness of fit statistic

Why Non-parametric Measurement Error Model?

Cheng, Shalabh and Garg (2014, 2016) proposed a measure to judge the goodness of fit statistic in the multiple measurement error model but their measures are dependent on the choice of the additional information :

- covariance matrix of measurement errors associated with the regressors is known, or
- reliability matrix associated with the regressors is known

  $\rightarrow$ Such information may not always be available in practice

- non-parametric procedures are free from such limitations

For a fixed $\mathbf{x}$,

$$\boxed{R_n^2(\mathbf{x}) = \frac{\hat{m}_n^2(\mathbf{x})}{\widehat{\sigma_{\epsilon,n}^2} + \hat{m}_n^2(\mathbf{x})}}$$

- $\hat{m}_n(\mathbf{x})$: consistent estimator of $m(\mathbf{x})$
- $\widehat{\sigma_{\epsilon,n}^2} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{m}_n(\mathbf{x}_i^\star))^2$: consistent estimator of σ_ϵ^2.
- $\rho^2(\mathbf{x}) = \frac{m^2(\mathbf{x})}{\sigma_\epsilon^2 + m^2(\mathbf{x})}$: the population counterpart of $R_n^2(\mathbf{x})$ for a fixed $\mathbf{x}$

Kernel based estimators

- Nadarya-Watson estimator : $\hat{m}_n^{NW}(\mathbf{x}) = \dfrac{\sum_{i=1}^{n} y_i K\left(\dfrac{\mathbf{x}-\mathbf{x}_i^*}{h_n}\right)}{\sum_{i=1}^{n} K\left(\dfrac{\mathbf{x}-\mathbf{x}_i^*}{h_n}\right)}$

- Pristley-Chao estimator : $\hat{m}_n^{PC}(\mathbf{x}) = \dfrac{\sum_{i=1}^{n} y_i K\left(\dfrac{\mathbf{x}-\mathbf{x}_i^*}{h_n}\right)}{n h_n^p}$

$K(.)$ is the multivariate kernel function, and $\{h_n\}$ is a sequence of bandwidth parameters.

Assumptions

- **(A1)** K is centrally symmetric about zero satisfying $\int \mathbf{u}K(\mathbf{u})d\mathbf{u} = \mathbf{0}$ and $\int \mathbf{u}'\mathbf{u}K(\mathbf{u})d\mathbf{u} < \infty$.
- **(A2)** $f_{\boldsymbol{\eta}}$, probability density function of $\boldsymbol{\eta}$, is bounded and twice differentiable. Moreover, each derivative is bounded function.
- **(A3)** Each component of $\mathbf{x}_i$, $x_{ij} \in [a, b], j = 1, 2, \ldots, p$, are fixed, where $a \in \mathbb{R}$ and $b \in \mathbb{R}$.
- **(A4)** The sequence of bandwidth $\{h_n\}$ is such that $h_n \to 0$, $nh_n^{p-4} \to \infty$ and $\lim_{n\to\infty} nh_n^{p+4} = \lambda^2$ with $0 \leq \lambda < \infty$ as $n \to \infty$.
- **(A5)** $m(.)$ is thrice continuously differentiable function.

- **Theorem 3.1** Under (A1)–(A5), for a fixed $\mathbf{x}$,

$$\sqrt{nh_n^p}\left(R_n^{NW,2}(\mathbf{x})-\rho^2(\mathbf{x})\right)\xrightarrow{d}\frac{\sigma_\epsilon^2 Z_1(\mathbf{x})}{(\sigma_\epsilon^2+m^2(\mathbf{x}))^2},$$

where $Z_1(\mathbf{x})$ is a random variable associated with normal distribution with mean $=\dfrac{2\lambda m(\mathbf{x})b(\mathbf{x})}{\int_{[a,b]^p} f_{\boldsymbol{\eta}}(\mathbf{x}-\mathbf{y})d\mathbf{y}}$ and variance $=\dfrac{4\sigma_\epsilon^2 m^2(\mathbf{x})\int K(\mathbf{u})^2 d\mathbf{u}}{\int_{[a,b]^p} f_{\boldsymbol{\eta}}(\mathbf{x}-\mathbf{y})d\mathbf{y}}$. Here, $b(\mathbf{x})=\frac{1}{2}\int\int_{[a,b]^p}\mathbf{u}'\nabla^2 m(\mathbf{x})\mathbf{u}f_{\boldsymbol{\eta}}(\mathbf{x}-\mathbf{y})K(\mathbf{u})d\mathbf{u}d\mathbf{y}+$
$\int\int_{[a,b]^p}\mathbf{u}'\nabla m(\mathbf{x})'\nabla f_{\boldsymbol{\eta}}(\mathbf{x}-\mathbf{y})\mathbf{u}K(\mathbf{u})d\mathbf{u}d\mathbf{y}$, where $a\in\mathbb{R}$ and $b\in\mathbb{R}$.

- **Theorem 3.2** Under (A1)–(A5), for a fixed $\mathbf{x}$,

$$\sqrt{nh_n^p}\left(R_n^{PC,2}(\mathbf{x}) - \rho^2(\mathbf{x})\right) \xrightarrow{d} \frac{\sigma_\epsilon^2 Z_2(\mathbf{x})}{\left(\sigma_\epsilon^2 + \left(m(\mathbf{x})\int_{[a,b]^p} f_{\boldsymbol{\eta}}(\mathbf{x}-\mathbf{y})d\mathbf{y}\right)^2\right)^2},$$

where $Z_2(\mathbf{x})$ is a random variable associated with normal distribution with mean $= 2\lambda m(\mathbf{x})b(\mathbf{x})\int_{[a,b]^p} f_{\boldsymbol{\eta}}(\mathbf{x}-\mathbf{y})d\mathbf{y}$ and variance $= 4\sigma_\epsilon^2 m^2(\mathbf{x})\int K(\mathbf{u})^2 d\mathbf{u}\left[\int_{[a,b]^p} f_{\boldsymbol{\eta}}(\mathbf{x}-\mathbf{y})d\mathbf{y}\right]^2$. Here, $b(\mathbf{x}) = \frac{1}{2}\int\int_{[a,b]^p} \mathbf{u}'\nabla^2 m(\mathbf{x})\mathbf{u} f_{\boldsymbol{\eta}}(\mathbf{x}-\mathbf{y})K(\mathbf{u})d\mathbf{u}d\mathbf{y} + \int\int_{[a,b]^p} \mathbf{u}'\nabla m(\mathbf{x})'\nabla f_{\boldsymbol{\eta}}(\mathbf{x}-\mathbf{y})\mathbf{u}K(\mathbf{u})d\mathbf{u}d\mathbf{y}$, where $a \in \mathbb{R}$ and $b \in \mathbb{R}$.

Finite Sample Study

Monte Carlo simulation experiments with $k = 1000$ replications

- exact values of $R_n^2(\mathbf{x})$
- empirical absolute bias $= EAB\left(R_n^2(\mathbf{x})\right) := \frac{1}{k}\sum_{i=1}^{k}\left|R_{n,i}^2(\mathbf{x}) - \rho^2(\mathbf{x})\right|$
- empirical mean squared error $= EMSE\left(R_n^2(\mathbf{x})\right) := \frac{1}{k}\sum_{i=1}^{k}\left(R_{n,i}^2(\mathbf{x}) - \rho^2(\mathbf{x})\right)^2$

Finite Sample Study

In the numerical study,

- Five regressors, namely, x_1, x_2, x_3, x_4 and x_5 are considered, which are fixed values in $[0, 1]$
- $\boldsymbol{\eta}$ is uniformly distributed on $[-1, 1]^5$
- $\epsilon \overset{\text{i.i.d}}{\sim} N(0, \sigma_\epsilon^2)$

Investigating the performances of $R_n^{NW,2}(\mathbf{x})$ and $R_n^{PC,2}(\mathbf{x})$ of the model without and with intercept term for

- **Example 1:** $m(\mathbf{x}) = \sum_{i=1}^{5} x_i^2$
- **Example 2:** $m(\mathbf{x}) = \sum_{i=1}^{5} \sin x_i$

		$\mathbf{x} = (1.41, 0.92, 1.22, 1.21, 1.01)'$					
n	σ_ϵ^2	$R_n^{NW,2}(\mathbf{x})$	$EAB\left(R_n^{NW,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{NW,2}(\mathbf{x})\right)$	$R_n^{PC,2}(\mathbf{x})$	$EAB\left(R_n^{PC,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{PC,2}(\mathbf{x})\right)$
50	0.1	0.99	0.004	0.00002	0.92	0.08	0.007
	0.5	0.98	0.02	0.0004	0.9	0.09	0.01
	1	0.97	0.04	0.002	0.88	0.1	0.01
	50	0.8	0.6	0.5	0.6	0.4	0.2
	100	0.67	0.8	0.7	0.56	0.5	0.3
100	0.1	0.99	0.002	0.00001	0.97	0.03	0.00008
	0.5	0.99	0.01	0.0003	0.96	0.03	0.0009
	1	0.98	0.03	0.001	0.94	0.05	0.001
	50	0.9	0.6	0.4	0.7	0.32	0.07
	100	0.87	0.8	0.6	0.6	0.34	0.09
		$\mathbf{x} = (1, 1.41, 0.92, 1.22, 1.21, 1.01)'$					
	σ_ϵ^2	$R_n^{NW,2}(\mathbf{x})$	$EAB\left(R_n^{NW,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{NW,2}(\mathbf{x})\right)$	$R_n^{PC,2}(\mathbf{x})$	$EAB\left(R_n^{PC,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{PC,2}(\mathbf{x})\right)$
50	0.1	0.99	0.002	0.000005	0.7	0.3	0.08
	0.5	0.97	0.008	0.00007	0.7	0.3	0.08
	1	0.96	0.02	0.0003	0.67	0.36	0.09
	50	0.83	0.5	0.2	0.45	0.37	0.1
	100	0.7	0.6	0.4	0.43	0.4	0.12
100	0.1	0.99	0.001	0.000003	0.72	0.27	0.07
	0.5	0.97	0.007	0.00006	0.71	0.28	0.07
	1	0.95	0.01	0.0002	0.7	0.32	0.08
	50	0.83	0.44	0.2	0.5	0.35	0.09
	100	0.69	0.6	0.37	0.45	0.37	0.1

Table: The values, absolute bias and mean squared errors for Example 1 with different values of σ_ϵ^2 based on $n = 50$ and 100

$\mathbf{x} = (1.41, 0.92, 1.22, 1.21, 1.01)'$							
n	σ_ϵ^2	$R_n^{NW,2}(\mathbf{x})$	$EAB\left(R_n^{NW,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{NW,2}(\mathbf{x})\right)$	$R_n^{PC,2}(\mathbf{x})$	$EAB\left(R_n^{PC,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{PC,2}(\mathbf{x})\right)$
50	0.1	0.99	0.007	0.00005	0.9	0.08	0.008
	0.5	0.97	0.03	0.001	0.88	0.09	0.01
	1	0.95	0.07	0.005	0.85	0.09	0.02
	50	0.76	0.7	0.6	0.6	0.4	0.3
	100	0.72	0.9	0.75	0.57	0.49	0.32
100	0.1	0.99	0.006	0.00004	0.93	0.05	0.003
	0.5	0.98	0.02	0.0008	0.92	0.06	0.003
	1	0.97	0.04	0.001	0.9	0.08	0.003
	50	0.82	0.65	0.46	0.63	0.39	0.19
	100	0.78	0.8	0.65	0.6	0.45	0.27
$\mathbf{x} = (1, 1.41, 0.92, 1.22, 1.21, 1.01)'$							
	σ_ϵ^2	$R_n^{NW,2}(\mathbf{x})$	$EAB\left(R_n^{NW,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{NW,2}(\mathbf{x})\right)$	$R_n^{PC,2}(\mathbf{x})$	$EAB\left(R_n^{PC,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{PC,2}(\mathbf{x})\right)$
50	0.1	0.97	0.004	0.00003	0.64	0.35	0.1
	0.5	0.94	0.02	0.0006	0.62	0.36	0.12
	1	0.93	0.05	0.002	0.6	0.4	0.13
	50	0.8	0.7	0.5	0.43	0.4	0.14
	100	0.65	0.8	0.7	0.42	0.45	0.18
100	0.1	0.98	0.003	0.00001	0.67	0.32	0.08
	0.5	0.95	0.015	0.0002	0.65	0.35	0.09
	1	0.94	0.04	0.002	0.61	0.38	0.11
	50	0.81	0.69	0.48	0.46	0.39	0.13
	100	0.7	0.75	0.68	0.44	0.4	0.15

Table: The values, absolute bias and mean squared errors for Example 2 with different values of σ_ϵ^2 based on $n = 50$ and 100

- $EAB\left(R_n^2(\mathbf{x})\right) \downarrow$ and $EMSE\left(R_n^2(\mathbf{x})\right) \downarrow$ as $n \uparrow$
- when $\sigma_\epsilon^2 \uparrow$, $R_n^{NW,2}(\mathbf{x}) \downarrow$ and $R_n^{PC,2}(\mathbf{x}) \downarrow$
- for large values of σ_ϵ^2, $EAB(R_n^{NW,2}(\mathbf{x})) > EAB(R_n^{PC,2}(\mathbf{x}))$ and

$$EMSE(R_n^{NW,2}(\mathbf{x})) > EMSE(R_n^{PC,2}(\mathbf{x}))$$

- for small values of σ_ϵ^2, $EAB(R_n^{NW,2}(\mathbf{x})) < EAB(R_n^{PC,2}(\mathbf{x}))$ and

$$EMSE(R_n^{NW,2}(\mathbf{x})) < EMSE(R_n^{PC,2}(\mathbf{x}))$$

Hence, if some prior information about the value of σ_ϵ^2 is known, one can then decide which estimator will be used to estimate $\rho^2(\mathbf{x})$.

With Intercept Model

- in both the cases, the *EAB* and the *EMSE* of $R_n^{NW,2}(\mathbf{x})$ do not differ much.
- when σ_ϵ^2 is small, the *EAB* and the *EMSE* of $R_n^{PC,2}(\mathbf{x})$ for the model with intercept are higher than that of the model without intercept.

For both the model, $R_n^{NW,2}(\mathbf{x})$ performs satisfactorily, but for the model with intercept term and for the small values of σ_ϵ^2, the use of $R_n^{PC,2}(\mathbf{x})$ as goodness of fit statistic may not be advisable.

If Number of Regressors Increases

- three more variables are added to the earlier data set, i.e., $p = 8$.
- we compute the EAB and the $EMSE$ of $R_n^{NW,2}(\mathbf{x})$ and $R_n^{PC,2}(\mathbf{x})$.
- the values of $R_n^{NW,2}(\mathbf{x})$ decrease slightly with the increase in the number of explanatory variables in the model but the values of $R_n^{PC,2}(\mathbf{x})$ for $p = 8$ are lower than that of for $p = 5$.

Therefore from this study, it is not readily evident in case of non-parametric measurement error model that the value of goodness of fit statistic always increases with the increase in the number of explanatory variables.

$\mathbf{x} = (1.41, 0.92, 1.22, 1.21, 1.01, 2.09, 0.73, 0.82)'$							
n	σ_ϵ^2	$R_n^{NW,2}(\mathbf{x})$	$EAB\left(R_n^{NW,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{NW,2}(\mathbf{x})\right)$	$R_n^{PC,2}(\mathbf{x})$	$EAB\left(R_n^{PC,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{PC,2}(\mathbf{x})\right)$
50	0.1	0.95	0.0008	0.000007	0.68	0.37	0.09
	0.5	0.94	0.004	0.0001	0.65	0.38	0.1
	1	0.93	0.007	0.0005	0.64	0.4	0.12
	50	0.77	0.6	0.4	0.42	0.4	0.14
	100	0.75	0.8	0.4	0.34	0.43	0.15
100	0.1	0.97	0.0006	0.000005	0.68	0.35	0.07
	0.5	0.95	0.0009	0.00008	0.67	0.36	0.09
	1	0.94	0.006	0.0005	0.65	0.37	0.1
	50	0.8	0.5	0.32	0.45	0.37	0.12
	100	0.76	0.5	0.34	0.34	0.4	0.13

Table: The values, empirical absolute bias and empirical mean squared errors of $R_n^{NW,2}(\mathbf{x})$ and $R_n^{PC,2}(\mathbf{x})$ for Example 1 with different values of σ_ϵ^2 when sample size = 50 and 100

$\mathbf{x} = (1.41, 0.92, 1.22, 1.21, 1.01, 2.09, 0.73, 0.82)'$							
n	σ_ϵ^2	$R_n^{NW,2}(\mathbf{x})$	$EAB\left(R_n^{NW,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{NW,2}(\mathbf{x})\right)$	$R_n^{PC,2}(\mathbf{x})$	$EAB\left(R_n^{PC,2}(\mathbf{x})\right)$	$EMSE\left(R_n^{PC,2}(\mathbf{x})\right)$
50	0.1	0.96	0.0007	0.000009	0.7	0.35	0.06
	0.5	0.94	0.005	0.0003	0.68	0.37	0.09
	1	0.91	0.009	0.0007	0.66	0.4	0.11
	50	0.76	0.7	0.6	0.45	0.42	0.13
	100	0.7	0.9	0.8	0.4	0.43	0.15
100	0.1	0.95	0.0006	0.000007	0.72	0.3	0.05
	0.5	0.93	0.002	0.0001	0.7	0.34	0.06
	1	0.92	0.006	0.0006	0.69	0.37	0.09
	50	0.78	0.5	0.36	0.5	0.39	0.11
	100	0.75	0.7	0.37	0.42	0.41	0.13

Table: The values, empirical absolute bias and empirical mean squared errors of $R_n^{NW,2}(\mathbf{x})$ and $R_n^{PC,2}(\mathbf{x})$ for Example 2 with different values of σ_ϵ^2 for the sample sizes 50 and 100

- Data set : The Pig data
- Collected by : the Statistical Laboratory of Iowa State University under contract to the Statistical Reporting Service, U.S. Department of Agriculture
- Previously investigated by : Battese, Fuller and Hickman (1976) and Fuller (1987)
- Two variables :
  - Y : the number of sows farrowing
  - X : the number of breeding hogs on hand
- n = 184

Fuller (1987) considered the linear regression model under parametric set-up

- in the presence of measurement errors in the data, $R^2 = 0.36$
- we model this data by non-parametric measurement error model

x	$R_n^{NW,2}(x)$	$R_n^{PC,2}(x)$
-11.29	0.95	0.98
1.31	0.73	0.99
10.04	0.96	0.99
51.74	0.91	1
98.29	0.92	1
162.26	0.72	0.98

Table: The values of $R_n^{NW,2}(x)$ and $R_n^{PC,2}(x)$ for the data

- The proposed measure is not location invariant. If the dependent variable is shifted by a constant, the estimated function *m* changes its value substantially, while the errors stay the same, and thus the proposed measure can change from a very small value to a very large one without any change in the predictive capability of the model.

Suggestion : Presence of intercept term in multiple linear regression model leads to the location invariant R^2 due to correction around the mean. However if $\beta_0 = 0$ the model reduces to no-intercept model and there R^2 is not location invariant. Non-parametric multiple regression model does not have any generic intercept term, which may lead to the fact that the proposed estimator is not location invariant.

Limitation 2

- Regression goodness of fit typically compares variability of signal to variability of noise. At a single point, the signal consists of a single point, the conditional mean, which is unknown to us and it looks like I cannot thus judge its strength. For example, imagine a linear regression of y on x going through the origin. At $x = 0$, the predicted value is 0, which implies the proposed measure of fit is 0 even if the errors are arbitrarily small, but that does not mean there is no good fit. To address this, the traditional goodness of fit measures are typically global, that is they summarize the performance on the whole support of the explanatory variables rather than at a single point as the proposed measure does.

Suggestion : Deriving the process convergence of R_n^2 is equivalent to deriving the process convergence of the Nadarya-Watson and the Priestley-Chao estimators in the presence of the measurement errors in the regressors. However, to the best of my knowledge, the process convergence of the kernel based estimator of the non-parametric regression function even without having any measurement error has not yet been studied in the literature.

References

- Abarin, T. and Wang, L. (2012). Instrumental variable approach to covariate measurement error in generalized linear models. *Annals of the Institute of Statistical Mathematics*, **64**, 475–493.
- Battese, G. E., Fuller, W. A., and Hickman, R. D. (1976). Estimation of response variances from interview-reinterview surveys. *Journal of the Indian Society of Agricultural Statistics*, **28**, 1–14.
- Carroll, R. J., Maca, J. D. and Ruppert, D. (1999). Nonparametric Regression in the Presence of Measurement Error. *Biometrika*, **86(3)**, 541–554.
- Cheng, C.-L., Shalabh, Garg, G. (2014). Coefficient of determination for multiple measurement error models. *Journal of Multivariate Analysis*, **126**, 137–152.
- Cheng, C.-L., Shalabh, Garg, G. (2016). Goodness of fit in restricted measurement error models. *Journal of Multivariate Analysis*, **145**, 101–116.
- Cheng, C.-L. and VanNess, J. W. (1991). On the unreplicated ultrastructural model. *Biometrika*, **78**, 442–445.
- Cheng, C.-L. and VanNess, J. W. (1999). Statistical Regression with Measurement Errors. *Arnold, London*.
- Elandt-Johnson, R. C. and Johnson, N. L. (1998). Survival Models and Data Analysis. *John Wiley & Sons*.
- Fan, J. and Troung, Y. K. (1993). Nonparametric regression with errors in variables. *Annals of Statistics*, **21**, 1900–1925.
- Fang. K. T., Kotz, S. and Ng, K. W. (1989). Symmetric Multivariate and Related Distributions. *New York: Chapman and Hall*.
- Fuller, W. A. (1987). Measurement Error Models. *John Wiley & Sons*.
- Gasser, T. and Müller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, **11**, 171–185.

References

Gleser, L. J. (1992). The importance of assessing measurement reliability in multivariate regression. *Journal of the American Statistical Association*, **87(419)**, 696–707.

Gleser, L. J. (1992). Estimators of slopes in linear errors-in-variables regression models when the predictors have known reliability matrix. *Statistics and Probability Letters*, **17**, 113–121.

Huang, Y-H., Wen, C-C. and Hsu, Y-H. (2015). The Extensively Corrected Score for Measurement Error Models. *Scandinavian Journal of Statistics*, **42**, 911–924.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H. and Lee, T-C. (1988). Introduction to the Theory and Practice of Econometrics. *John Wiley & Sons*.

Kulcsar, E. (2009). Multiple Regression Analysis of Main Economic Indicators in Tourism. *Journal of Tourism - Studies and Research in Tourism*, **8(8)**, 59–64.

Li, T. and Vuong, Q. (1998). Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators. *Journal of Multivariate Analysis*, **65**, 139–165.

Montgomery, D.C., Peck, E. A. and Vining, G. G. (2001). Introduction to Linear Regression Analysis. *Wiley*.

Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**, 141–142.

Priestley, M. E. and Chao, M. T. (1972). Nonparametric function fitting. *Journal of The Royal Statistical Society, Series B*, **34**, 385–392.

Rao, C. R., Toutenburg, H., Shalabh and Heumann, C. (2008). Linear Models and Generalizations - Least Squares and Alternatives. *Springer*.

Schneeweiss, H. (1976). Consistent estimation of a regression with errors in the variables. *Metrika*, **23**, 101–115.

Shalabh (1998). Improved estimation in measurement error models through Stein-rule procedure. *Journal of Multivariate Analysis*, **67**, 35–48.

References

Shalabh (2000). Corrigendum. *Journal of Multivariate Analysis*, **74**, 162.

Shalabh (2003). Consistent estimation of coefficients in measurement error models under non-normality. *Journal of Multivariate Analysis*, **86(2)**, 227–241.

Shalabh, Garg, G. and Misra, N. (2007). Restricted regression estimation in measurement error models. *Computational Statistics and Data Analysis*, **52(2)**, 1149–1166.

Shalabh, Garg, G. and Misra, N. (2009). Use of prior information in the consistent estimation of regression coefficients in measurement error models. *Journal of Multivariate Analysis*, **100(7)**, 1498–1520.

Shalabh, Garg, G. and Misra, N. (2010). Consistent estimation of regression coefficients in measurement error model using stochastic a priori information. *Statistical Papers*, **51**, 717–748.

Shalabh, Garg, G. and Misra, N. (2011). Estimation of regression coefficients in a restricted measurement error model using instrumental variables. *Communications in Statistics - Theory and Methods*, **40**, 3614–3629.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. *Chapman & Hall, London, New York*.

Sørensen, Ø., Frigessi, A. and Thoresen, M. (2015). Measurement Error in Lasso: Impact and Likelihood Bias Correction. *Statistica Sinica*, **25**, 809–829.

Stuart, A and Ord, K. (1994). Kendall's Advanced Theory of Statistics. *Arnold, London*.

Watson, G.S. (1964). Smooth regression analysis. *Sankhy A: The Indian Journal of Statistics, Series A*, **26**, 359–372.

Zhang, J., Zhu, L. and Zhu, L. (2014). Surrogate Dimension Reduction in Measurement Error Regressions. *Statistica Sinica*, **24**, 1341–1363.

Thank You