Investigation of Non-normality in a Simple Errors-in-variables Model

Lee Chun Yin [3035469140]

4-5-2021

Supervisor: Dr. Raymond W.L. Wong

STAT3799 Directed studies in statistics

Introduction

Background

- In a classical linear regression setting, we often assume:
  - that the explanatory variable is nonrandom without any observation error
  - the errors are normally distributed
- These assumptions may not hold in real life datasets.
- We use the computer simulation technique to demonstrate the impacts of non-normality in the *errors-in-variables* model.
- We present numerical results from simulations based on different error distributions, and for different estimators of slope parameter β and residual variance σ_ϵ^2.

Regression setting – Observation error

- Regression through the origin: $\boxed{y = \beta x + \epsilon}$
- Obtain observation data $(\tilde{x}_i, \tilde{y}_i)$
- Under additive measurement error, $\tilde{x}_i = x_i + u_i$, $\tilde{y}_i = y_i + v_i$
  - The observed values are not the true values the model is assumed on
- Assume that measurement error u and v have:
  - mean zero
  - constant variances σ_u^2 and σ_v^2
  - uncorrelated and independent to the true values x and y
- "errors-in-variables" model

Regression setting – Non-normality

- Often assume that $\epsilon \sim N(0, \sigma_\epsilon^2)$
  - Necessary for performing statistical inference of regression parameters
  - Ordinary-least-squares (OLS) is the maximum likelihood estimator (MLE) for normal-distributed ϵ, but this may not hold for other distributions of ϵ

- Normal distribution: Light tailed, symmetric.
  - May not hold in real life settings

- Non-normality may happen in both residual ϵ and measurement error u, v

Outline

- Focus on estimation problem in regression

- 1. Literature review
- 2. Methodology: Computer simulation technique
- 3. Results: Estimation of slope parameter β
- 4. Results: Estimation of residual variance σ_ϵ^2.
- 5. Conclusion

Literature review

Errors-in-variables model (Pischke, 2007)

- Suppose we want to estimate the relationship $y = \beta x + \epsilon$, but we only have data on $\tilde{x} = x + u$.
- Further assume that $\sigma_v^2 = 0$, so there is only measurement error in x.

$$y = \beta(\tilde{x} - u) + \epsilon = \beta\tilde{x} + (\epsilon - \beta u)$$

- Measurement error becomes part of residual error.
- Exogeneity assumption of Gauss-Markov theorem (OLS is BLUE) is violated as $cov(\epsilon - \beta u, \tilde{x}) \neq 0$

OLS for β (Pische, 2007)

- $\boxed{\hat{\beta}_{OLS} = \frac{cov(\tilde{x},y)}{var(\tilde{x})}} = \frac{cov(x+u,\beta x+\epsilon)}{var(x+u)}$

- $plim\ \hat{\beta}_{OLS} = \frac{\beta\sigma_x^2}{\sigma_x^2+\sigma_u^2} = \lambda\beta,$ where $\lambda = \frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2}$

- $\hat{\beta}_{OLS}$ biased towards zero, inconsistent estimator

MME for β (Pische, 2007)

- Need to use method-of-moments estimator (MME) to obtain consistent estimates for β

- An MME can be obtained if we have some prior knowledge on the observation error
  - E.g. we can estimate σ_u^2 via repeated measurements (Gilldard, 2014)

- $$\boxed{\hat{\beta}_{MME} = \frac{1}{\lambda} \cdot \hat{\beta}_{OLS} = \frac{cov(\tilde{x},y)}{var(\tilde{x}) - \sigma_u^2}}$$

OLS for σ_ϵ^2 (Pische, 2007)

- Estimated residual $\hat{\epsilon}$

$$
\begin{aligned}
\hat{\epsilon} &= y - \hat{\beta}_{OLS}\,\tilde{x} \\
&= y - \hat{\beta}_{OLS}(x+u) \\
&= \epsilon - (y - \beta x) + y - \hat{\beta}_{OLS}\,x - \hat{\beta}_{OLS}\,u \\
&= \epsilon + (\beta - \hat{\beta}_{OLS})\,x - \hat{\beta}_{OLS}\,u
\end{aligned}
$$

- $\hat{\epsilon}$ is obfuscated by two additional sources of error
- $\boxed{\widehat{\sigma_\epsilon^2}_{OLS} = \frac{\sum \hat{\epsilon}^2}{n-1}}$
- $plim\ \widehat{\sigma_\epsilon^2}_{OLS} = \sigma_\epsilon^2 + (1-\lambda)^2\beta^2\sigma_x^2 + \lambda^2\beta^2\sigma_u^2$

MME for σ_ϵ^2

- Similarly, need to use MME to obtain consistent estimates for σ_ϵ^2

- After obtaining a consistent estimation $\hat{\beta}_{MME}$, can substitute into the asymptotic bias of $\widehat{\sigma_\epsilon^2}_{OLS}$, rearrange to obtain consistent $\widehat{\sigma_\epsilon^2}_{MME}$ (Gillard, 2014)

- $\boxed{\widehat{\sigma_\epsilon^2}_{MME} = \widehat{\sigma_\epsilon^2}_{OLS} - (1-\lambda)^2\left(\hat{\beta}_{MME}\right)^2\left(var(\tilde{x}) - \sigma_u^2\right) + \lambda^2\left(\hat{\beta}_{MME}\right)^2\sigma_u^2}$

Methodology

Procedures

- Computer simulation method
- Use common $\beta_{truth}, n, x_{lo}, x_{hi}$

- β_{truth}: underlying ground truth β to be estimated
- n: number of observations
- x_{lo}, x_{hi}: lower bound and upper bound of x being sampled respectively

Procedures

- 1. Fix a certain σ_u^2 and σ_ϵ^2
- 2. For observation error u, choose distribution with variance σ_u^2
- 3. For residual error ϵ, choose distribution with variance σ_ϵ^2
- 4. Pick n instances of x uniformly from $[x_{lo}, x_{hi}]$
- 5. Generate observation pair $(\tilde{x}_i, \tilde{y}_i)$
  - $\tilde{x}_i = x_i + u_i$
  - $\tilde{y}_i = y_i = \beta_{truth} \cdot x_i + \epsilon_i$
- 6. Compute OLS and MME for β and σ_ϵ^2, compute error metrics

Iterations and error computation

- Perform the procedures for 10,000 iterations for each pair of $(\sigma_u^2, \sigma_\epsilon^2)$
- Obtain mean bias error (MBE) and mean squared error (MSE) over all iterations

Choice of distribution

- To investigate non-normality, need to use some non-normal distributions for measurement error and residual error
- Requirement: Finite variance, easy to compute parameters from given variance, mean zero

- Normal distribution
- Student's t distribution
- χ^2 distribution, re-centered at mean zero

- For simplicity, choose the same type of distribution for both measurement error and residual error

Experiment parameters used

- $n = 30$
- $\beta_{truth} = 2$
- $x_{lo} = -20$
- $x_{hi} = 20$

- $\sigma_u^2 \in \{0, 1.5, 2, 2.5, 3.0, 4.5, 6.0, 8.0, 10.0\}$
- $\sigma_\epsilon^2 \in \{1.5, 2, 2.5\}$

- Implemented in Python with Jupyter notebook, **`scipy`** package for distributions

Procedure 1: Computer simulation procedure

Initialize β_{truth}, n, x_{lo}, x_{hi};

forall $\sigma_u^2 \in$ {*list of observation variance*} **do**

    forall $\sigma_\epsilon^2 \in$ {*list of residual variance*} **do**

        Initialize $MSE_{\hat{\beta}_{OLS}} = 0$, $MSE_{\hat{\beta}_{MME}} = 0$, $MBE_{\hat{\beta}_{OLS}} = 0$, $MBE_{\hat{\beta}_{MME}} = 0.$;

        Initialize $MSE_{\hat{\sigma^2_\epsilon}_{OLS}} = 0$, $MSE_{\hat{\sigma^2_\epsilon}_{MME}} = 0$, $MBE_{\hat{\sigma^2_\epsilon}_{OLS}} = 0$, $MBE_{\hat{\sigma^2_\epsilon}_{MME}} = 0$;

        for *10000 iterations* **do**

            Pick n instances of explanatory variable x_i uniformly from the interval $[x_{lo}, x_{hi}]$;

            Compute n observations of $y_i = \beta_{truth} \cdot x_i + \epsilon_i$, where ϵ_i is drawn independently from a distribution with variance σ_ϵ^2;

            Compute n instances of observed explanatory variable $\tilde{x}_i = x_i + u_i$, where observation u_i is drawn independently from a distribution with variance σ_u^2;

            Calculate $\hat{\beta}_{OLS}$, $\hat{\beta}_{MME}$, $\hat{\sigma^2_\epsilon}_{OLS}$, $\hat{\sigma^2_\epsilon}_{MME}$ from the observations $\tilde{x}$, y, and prior information on observation error σ_u^2;

            Update $MSE_{\hat{\beta}_{OLS}}$, $MSE_{\hat{\beta}_{MME}}$, $MBE_{\hat{\beta}_{OLS}}$, $MBE_{\hat{\beta}_{MME}}$, $MSE_{\hat{\sigma^2_\epsilon}_{OLS}}$, $MSE_{\hat{\sigma^2_\epsilon}_{MME}}$, $MBE_{\hat{\sigma^2_\epsilon}_{OLS}}$, $MBE_{\hat{\sigma^2_\epsilon}_{MME}}$;

        end

    end

end

Estimation of β

Estimators

- $\hat{\beta}_{OLS} = \frac{cov(\tilde{x},y)}{var(\tilde{x})}$
- $\hat{\beta}_{MME} = \frac{cov(\tilde{x},y)}{var(\tilde{x})-\sigma_u^2}$

Absence of observation error

- Under the absence of observation error ($\sigma_u^2 = 0$), the OLS and MME estimators of β are the same.
- MBE: no clear trend (centered at zero)
- MSE: Increasing with σ_ϵ^2, slightly greater with Student's t errors

Table 2: Estimation MBE of β when $\sigma_u^2 = 0$, under different distributions and variance of ϵ.

	Normal	Student's t	χ^2
$\sigma_\epsilon^2 = 1.5$	-0.000260	0.000252	0.000052
$\sigma_\epsilon^2 = 2.0$	0.000151	0.000114	-0.000356
$\sigma_\epsilon^2 = 2.5$	-0.000176	-0.000105	0.000112

Table 3: Estimation MSE of β when $\sigma_u^2 = 0$, under different distributions and variance of ϵ.

	Normal	Student's t	χ^2
$\sigma_\epsilon^2 = 1.5$	0.000352	0.000351	0.000343
$\sigma_\epsilon^2 = 2.0$	0.000470	0.000478	0.000458
$\sigma_\epsilon^2 = 2.5$	0.000588	0.000594	0.000573

Normal-distributed measurement error and residual error

- OLS:
  - MBE is negative and decreasing with σ_u^2
  - MSE is increasing with σ_u^2
- MME:
  - MBE is positive and increasing with σ_u^2
  - MSE is increasing with σ_u^2
- Absolute value of MBE and MSE for MME are lower than that of OLS

Table 4: Estimation MBE of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{MME}$ under normal-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$
$\sigma_u^2 = 1.5$	-0.018406	0.001800	-0.019667	0.000505	-0.019292	0.000895
$\sigma_u^2 = 2.0$	-0.025018	0.001851	-0.024511	0.002371	-0.025436	0.001416
$\sigma_u^2 = 2.5$	-0.031520	0.001958	-0.029421	0.004139	-0.031215	0.002285
$\sigma_u^2 = 3.0$	-0.036707	0.003436	-0.036961	0.003156	-0.037188	0.002923
$\sigma_u^2 = 4.5$	-0.054875	0.004901	-0.056101	0.003559	-0.055655	0.004057
$\sigma_u^2 = 6.0$	-0.073502	0.005575	-0.074400	0.004531	-0.072289	0.006886
$\sigma_u^2 = 8.0$	-0.097213	0.007125	-0.097049	0.007313	-0.096943	0.007427
$\sigma_u^2 = 10.0$	-0.120748	0.008388	-0.119521	0.009902	-0.119075	0.010326

Table 5: Estimation MSE of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{MME}$ under normal-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$
$\sigma_u^2 = 1.5$	0.002026	0.001781	0.002168	0.001873	0.002304	0.002029
$\sigma_u^2 = 2.0$	0.002732	0.002267	0.002830	0.002396	0.002987	0.002506
$\sigma_u^2 = 2.5$	0.003544	0.002801	0.003559	0.002964	0.003688	0.002967
$\sigma_u^2 = 3.0$	0.004319	0.003334	0.004455	0.003456	0.004653	0.003648
$\sigma_u^2 = 4.5$	0.007088	0.004862	0.007457	0.005112	0.007360	0.005045
$\sigma_u^2 = 6.0$	0.010532	0.006487	0.010886	0.006739	0.010648	0.006833
$\sigma_u^2 = 8.0$	0.016103	0.009102	0.016091	0.009100	0.016123	0.009138
$\sigma_u^2 = 10.0$	0.022169	0.011236	0.022276	0.011799	0.022216	0.011831

Student's t-distributed measurement error and residual error

- OLS:
  - MBE is negative and decreasing with σ_u^2, in general closer to zero than normal case
  - MSE higher than normal case for $\sigma_u^2 \leq 8.0$
- MME:
  - MBE is positive and increasing with σ_u^2, in general farther from zero than normal case
  - MSE higher than normal case
- For low σ_u^2, absolute value of MBE and MSE for MME are lower than OLS

Table 6: Estimation MBE of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{MME}$ under Student's t-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$
$\sigma_u^2 = 1.5$	-0.018811	0.001391	-0.019199	0.000988	-0.018163	0.002050
$\sigma_u^2 = 2.0$	-0.024563	0.002314	-0.024635	0.002238	-0.024720	0.002157
$\sigma_u^2 = 2.5$	-0.031062	0.002428	-0.031050	0.002448	-0.030458	0.003071
$\sigma_u^2 = 3.0$	-0.034826	0.005397	-0.035187	0.005009	-0.035580	0.004603
$\sigma_u^2 = 4.5$	-0.045805	0.014550	-0.048785	0.011340	-0.045181	0.015206
$\sigma_u^2 = 6.0$	-0.054096	0.026632	-0.054759	0.025926	-0.054605	0.026035
$\sigma_u^2 = 8.0$	-0.061836	0.046525	-0.060023	0.048540	-0.061507	0.046848
$\sigma_u^2 = 10.0$	-0.066070	0.070679	-0.067018	0.069615	-0.065713	0.071101

Table 7: Estimation MSE of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{MME}$ under Student's t-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$
$\sigma_u^2 = 1.5$	0.002112	0.001854	0.002179	0.001905	0.002279	0.002051
$\sigma_u^2 = 2.0$	0.002912	0.002481	0.003040	0.002609	0.003132	0.002699
$\sigma_u^2 = 2.5$	0.004094	0.003409	0.004211	0.003528	0.004270	0.003635
$\sigma_u^2 = 3.0$	0.005661	0.004913	0.005470	0.004671	0.005644	0.004827
$\sigma_u^2 = 4.5$	0.009523	0.008639	0.010456	0.009277	0.009140	0.008293
$\sigma_u^2 = 6.0$	0.014130	0.013794	0.013723	0.013223	0.013692	0.013231
$\sigma_u^2 = 8.0$	0.017969	0.019371	0.016112	0.017761	0.017037	0.018482
$\sigma_u^2 = 10.0$	0.019388	0.024422	0.020242	0.025071	0.018737	0.023675

Re-centered χ^2-distributed measurement error and residual error

- OLS:
  - MBE is negative and decreasing with σ_u^2
  - MSE is increasing with σ_u^2, higher than normal case
- MME:
  - MBE is positive and increasing with σ_u^2
  - MSE is increasing with σ_u^2, higher than normal case
- Absolute value of MBE and MSE for MME are lower than OLS

Table 10: Estimation MBE of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{MME}$ under χ^2-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$
$\sigma_u^2 = 1.5$	-0.018624	0.001580	-0.017994	0.002220	-0.018908	0.001285
$\sigma_u^2 = 2.0$	-0.024808	0.002076	-0.025817	0.001020	-0.024370	0.002517
$\sigma_u^2 = 2.5$	-0.030362	0.003170	-0.031010	0.002494	-0.030079	0.003481
$\sigma_u^2 = 3.0$	-0.036793	0.003351	-0.037685	0.002393	-0.037465	0.002621
$\sigma_u^2 = 4.5$	-0.054270	0.005575	-0.055211	0.004503	-0.056487	0.003124
$\sigma_u^2 = 6.0$	-0.072231	0.006964	-0.072339	0.006804	-0.072609	0.006531
$\sigma_u^2 = 8.0$	-0.097065	0.007307	-0.097076	0.007311	-0.096631	0.007895
$\sigma_u^2 = 10.0$	-0.118317	0.011277	-0.116507	0.013404	-0.118055	0.011505

Table 11: Estimation MSE of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{MME}$ under χ^2-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{MME}$
$\sigma_u^2 = 1.5$	0.002205	0.001958	0.002274	0.002053	0.002460	0.002206
$\sigma_u^2 = 2.0$	0.002944	0.002501	0.003080	0.002584	0.003201	0.002792
$\sigma_u^2 = 2.5$	0.003712	0.003059	0.003838	0.003142	0.003863	0.003232
$\sigma_u^2 = 3.0$	0.004493	0.003508	0.004917	0.003890	0.004644	0.003598
$\sigma_u^2 = 4.5$	0.007488	0.005382	0.007612	0.005388	0.007981	0.005624
$\sigma_u^2 = 6.0$	0.010957	0.007206	0.011047	0.007281	0.011212	0.007411
$\sigma_u^2 = 8.0$	0.016616	0.009711	0.016813	0.009960	0.016772	0.010004
$\sigma_u^2 = 10.0$	0.022696	0.012718	0.022178	0.012613	0.022977	0.013127

Estimation of σ_ϵ^2

Estimators

- $\widehat{\sigma_\epsilon^2}_{OLS} = \frac{\sum \hat{\epsilon}^2}{n-1}$

- $\widehat{\sigma_\epsilon^2}_{MME} = \widehat{\sigma_\epsilon^2}_{OLS} - (1-\lambda)^2 \left(\hat{\beta}_{MME}\right)^2 (var(\tilde{x}) - \sigma_u^2) + \lambda^2 \left(\hat{\beta}_{MME}\right)^2 \sigma_u^2$

Absence of measurement error

- Under the absence of observation error ($\sigma_u^2 = 0$), the OLS and MME estimators of σ_ϵ^2 are the same.
- MBE: no clear trend (centered at zero)
- MSE: Increasing with σ_ϵ^2, greater with Student's t and χ^2 errors

Table 14: Estimation MBE of σ_ϵ^2 when $\sigma_u^2 = 0$, under different distributions and variance of ϵ.

	Normal	Student's t	χ^2
$\sigma_\epsilon^2 = 1.5$	0.000038	0.006738	-0.009608
$\sigma_\epsilon^2 = 2.0$	0.000135	-0.010998	-0.005749
$\sigma_\epsilon^2 = 2.5$	-0.001355	0.018754	0.026852

Table 15: Estimation MSE of σ_ϵ^2 when $\sigma_u^2 = 0$, under different distributions and variance of ϵ.

	Normal	Student's t	χ^2
$\sigma_\epsilon^2 = 1.5$	0.154338	0.367452	1.325430
$\sigma_\epsilon^2 = 2.0$	0.278130	1.806585	1.919026
$\sigma_\epsilon^2 = 2.5$	0.437989	10.128111	2.414450

Normal-distributed measurement error and residual error

- OLS:
  - MBE is positive and increasing with σ_u^2
  - MSE is increasing with σ_u^2
- MME:
  - MBE is negative and decreasing with σ_u^2
  - MSE is increasing with σ_u^2
- Absolute value of MBE and MSE for MME are lower than that of OLS

Table 16: Estimation MBE of $\hat{\sigma^2_{\epsilon}}_{OLS}$ and $\hat{\sigma^2_{\epsilon}}_{MME}$ under normal-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\sigma^2_{\epsilon}}_{OLS}$	$\hat{\sigma^2_{\epsilon}}_{MME}$	$\hat{\sigma^2_{\epsilon}}_{OLS}$	$\hat{\sigma^2_{\epsilon}}_{MME}$	$\hat{\sigma^2_{\epsilon}}_{OLS}$	$\hat{\sigma^2_{\epsilon}}_{MME}$
$\sigma_u^2 = 1.5$	5.900404	-0.050233	5.955353	0.012206	5.947017	0.001355
$\sigma_u^2 = 2.0$	7.888963	-0.018928	7.901321	-0.010901	7.879309	-0.025664
$\sigma_u^2 = 2.5$	9.784296	-0.068626	9.802717	-0.071802	9.869316	0.012851
$\sigma_u^2 = 3.0$	11.798199	-0.002881	11.757166	-0.041108	11.793174	-0.002928
$\sigma_u^2 = 4.5$	17.597952	0.047601	17.466602	-0.062097	17.455414	-0.081391
$\sigma_u^2 = 6.0$	23.120196	-0.063780	23.033594	-0.129041	23.132393	-0.083139
$\sigma_u^2 = 8.0$	30.406714	-0.150267	30.219614	-0.342906	30.389711	-0.176612
$\sigma_u^2 = 10.0$	37.647515	-0.097025	37.724051	-0.077960	37.508399	-0.310914

Table 17: Estimation MSE of $\hat{\sigma^2_{\epsilon}}_{OLS}$ and $\hat{\sigma^2_{\epsilon}}_{MME}$ under normal-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\sigma^2_{\epsilon}}_{OLS}$	$\hat{\sigma^2_{\epsilon}}_{MME}$	$\hat{\sigma^2_{\epsilon}}_{OLS}$	$\hat{\sigma^2_{\epsilon}}_{MME}$	$\hat{\sigma^2_{\epsilon}}_{OLS}$	$\hat{\sigma^2_{\epsilon}}_{MME}$
$\sigma_u^2 = 1.5$	38.530776	3.780289	39.882151	4.471836	40.209234	4.935117
$\sigma_u^2 = 2.0$	68.339624	6.241460	69.219747	6.910293	69.534180	7.638777
$\sigma_u^2 = 2.5$	104.624534	9.105856	105.756458	9.905827	108.054738	10.961756
$\sigma_u^2 = 3.0$	151.461698	12.734043	151.520531	13.621373	153.743436	15.163531
$\sigma_u^2 = 4.5$	335.041092	26.578374	331.019140	27.398219	331.821175	28.443274
$\sigma_u^2 = 6.0$	576.830404	45.826542	575.198317	47.750947	581.386469	49.216860
$\sigma_u^2 = 8.0$	995.830908	79.148108	984.510480	79.308313	998.785315	83.400332
$\sigma_u^2 = 10.0$	1520.990888	117.975374	1528.874963	120.145671	1516.974233	125.296896

Student’s t-distributed measurement error and residual error

- OLS:
  - MBE is positive and increasing with σ_u^2, closer to zero than normal case
  - MSE higher than normal case
- MME:
  - MBE is negative and decreasing with σ_u^2, (much) farther from zero than normal case
  - MSE higher than normal case
- Using MME instead of OLS doesn’t provide that much improvement in absolute MBE or MSE
  - MSE of MME worse than OLS for larger σ_u^2

Table 18: Estimation MBE of $\hat{\sigma}^2_{\epsilon OLS}$ and $\hat{\sigma}^2_{\epsilon MME}$ under Student’s t-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\sigma}^2_{\epsilon OLS}$	$\hat{\sigma}^2_{\epsilon MME}$	$\hat{\sigma}^2_{\epsilon OLS}$	$\hat{\sigma}^2_{\epsilon MME}$	$\hat{\sigma}^2_{\epsilon OLS}$	$\hat{\sigma}^2_{\epsilon MME}$
$\sigma_u^2 = 1.5$	5.965444	0.017129	5.949026	0.002994	5.852522	-0.099986
$\sigma_u^2 = 2.0$	7.873037	-0.038917	7.794577	-0.117039	7.837349	-0.073789
$\sigma_u^2 = 2.5$	9.613666	-0.245337	9.697142	-0.162316	9.591657	-0.274076
$\sigma_u^2 = 3.0$	11.240567	-0.587908	11.061106	-0.762236	11.091512	-0.727538
$\sigma_u^2 = 4.5$	14.674276	-3.058067	15.130487	-2.549836	14.349906	-3.392302
$\sigma_u^2 = 6.0$	17.308219	-6.392124	17.300402	-6.380649	17.289769	-6.394398
$\sigma_u^2 = 8.0$	19.391143	-12.403962	18.999681	-12.842478	19.122997	-12.675239
$\sigma_u^2 = 10.0$	20.359688	-19.757988	20.806443	-19.278803	20.337719	-19.788499

Table 19: Estimation MSE of $\hat{\sigma}^2_{\epsilon OLS}$ and $\hat{\sigma}^2_{\epsilon MME}$ under Student’s t-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\sigma}^2_{\epsilon OLS}$	$\hat{\sigma}^2_{\epsilon MME}$	$\hat{\sigma}^2_{\epsilon OLS}$	$\hat{\sigma}^2_{\epsilon MME}$	$\hat{\sigma}^2_{\epsilon OLS}$	$\hat{\sigma}^2_{\epsilon MME}$
$\sigma_u^2 = 1.5$	42.851635	7.429251	44.646925	9.417269	47.303721	13.233213
$\sigma_u^2 = 2.0$	90.739181	29.905215	88.083920	28.419856	96.305849	35.838509
$\sigma_u^2 = 2.5$	167.755054	79.503490	181.578188	92.099801	182.441974	94.903265
$\sigma_u^2 = 3.0$	284.186452	168.747879	278.572350	166.286730	276.831587	163.997266
$\sigma_u^2 = 4.5$	621.524421	453.668034	712.260002	535.305676	588.075696	427.984632
$\sigma_u^2 = 6.0$	1059.949532	894.596872	988.210829	815.597278	1118.336892	940.571666
$\sigma_u^2 = 8.0$	1433.607905	1384.842799	1208.084274	1158.996324	1213.814914	1157.835613
$\sigma_u^2 = 10.0$	1430.156559	1641.357813	1618.590469	1813.400195	1429.762567	1631.738065

Re-centered χ^2-distributed measurement error and residual error

- OLS:
  - > MBE is positive and increasing with σ_u^2
  - > MSE is increasing with σ_u^2, higher than normal case
- MME:
  - > MBE is negative and decreasing with σ_u^2, farther from zero than normal case
  - > MSE is increasing with σ_u^2, higher than normal case
- Absolute value of MBE and MSE for MME are lower than OLS

Table 22: Estimation MBE of $\hat{\sigma^2_\epsilon}_{OLS}$ and $\hat{\sigma^2_\epsilon}_{MME}$ under χ^2-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\sigma^2_\epsilon}_{OLS}$	$\hat{\sigma^2_\epsilon}_{MME}$	$\hat{\sigma^2_\epsilon}_{OLS}$	$\hat{\sigma^2_\epsilon}_{MME}$	$\hat{\sigma^2_\epsilon}_{OLS}$	$\hat{\sigma^2_\epsilon}_{MME}$
$\sigma_u^2 = 1.5$	5.893220	-0.056369	5.925753	-0.027765	5.910017	-0.038214
$\sigma_u^2 = 2.0$	7.941942	0.031871	7.883551	-0.018494	7.897804	-0.016332
$\sigma_u^2 = 2.5$	9.786872	-0.078416	9.747000	-0.111938	9.862407	-0.006246
$\sigma_u^2 = 3.0$	11.764364	-0.036215	11.771377	-0.019330	11.577529	-0.215025
$\sigma_u^2 = 4.5$	17.515282	-0.048596	17.354157	-0.191983	17.339678	-0.183992
$\sigma_u^2 = 6.0$	23.076080	-0.143101	23.001147	-0.215374	22.952380	-0.258514
$\sigma_u^2 = 8.0$	30.188767	-0.377915	30.370164	-0.198146	30.555481	-0.029103
$\sigma_u^2 = 10.0$	37.399951	-0.459851	37.362384	-0.572358	37.111804	-0.761288

Table 23: Estimation MSE of $\hat{\sigma^2_\epsilon}_{OLS}$ and $\hat{\sigma^2_\epsilon}_{MME}$ under χ^2-distributed observation error with different variances σ_u^2 and residual error with different variances σ_ϵ^2.

	$\sigma_\epsilon^2 = 1.5$		$\sigma_\epsilon^2 = 2.0$		$\sigma_\epsilon^2 = 2.5$	
	$\hat{\sigma^2_\epsilon}_{OLS}$	$\hat{\sigma^2_\epsilon}_{MME}$	$\hat{\sigma^2_\epsilon}_{OLS}$	$\hat{\sigma^2_\epsilon}_{MME}$	$\hat{\sigma^2_\epsilon}_{OLS}$	$\hat{\sigma^2_\epsilon}_{MME}$
$\sigma_u^2 = 1.5$	57.156508	23.096732	59.199434	24.756849	59.595066	25.384590
$\sigma_u^2 = 2.0$	94.420321	32.593721	94.155136	33.209791	96.123602	35.016405
$\sigma_u^2 = 2.5$	135.176838	41.299873	134.988694	41.911471	142.479071	47.127930
$\sigma_u^2 = 3.0$	187.194973	51.476957	187.669689	52.236293	182.631566	51.413928
$\sigma_u^2 = 4.5$	385.846689	85.801536	378.435062	83.682260	379.322162	85.748375
$\sigma_u^2 = 6.0$	638.422049	118.133857	636.923293	119.730214	636.076559	122.171158
$\sigma_u^2 = 8.0$	1057.790558	169.895222	1077.712211	178.676540	1093.595726	183.465213
$\sigma_u^2 = 10.0$	1597.776001	237.689528	1599.654260	240.683626	1580.838713	245.632506

Conclusion

Conclusion

- Presented numerical results on the case of non-normality under the errors-in-variables model

- Shown how non-normality in the observation error affects the estimation of the regression parameters β and σ_ϵ^2.

- Compared the OLS and MME estimators under different distributions, by comparing their bias and squared error

Future work

- Inference of statistical parameters under non-normality and measurement errors
- Non-normality on multivariate errors-in-variables models
- Other estimators of β and σ_ϵ^2

Acknowledgments

- Dr. Raymond W.L. Wong, for his support and guidance for this directed studies.

- HKU Department of Statistics and Actuarial Science, for providing me with this opportunity to pursue in this directed studies capstone project.

Q&A