Monte Carlo Study Project Report

1 Introduction

In this report, we will study the following model.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i \quad (1)$$

where $\epsilon_i \sim i.i.d.t(0, \sigma^2, v)$, and $i = 1, 2, \ldots, N$. Note $x_{1i} = 3x_{2i} + u_i$ with $u_i \sim i.i.d.N(0, 1)$. In particular, we will test the robustness of the OLS estimators when some assumptions are not strictly met. We will study the properties of the OLS estimators $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. Our findings are

1. The Gauss Markov assumptions are all satisfied when $v > 2$.
2. The OLS estimators are unbiased and consistent when $v > 1$.
3. $NVar(\hat{\beta}_j)$ for $j = 1, 2\ 3$ will converge asymptotically to the normal distribution as N approaches infinity.
4. The estimates for $\sqrt{NVar(\hat{\beta}_k)}$ for $k = 1, 2, 3$ will be unbiased and consistent when $v > 2$.
5. The t statistics are asymptotically valid when N is large and $v > 1$ despite the error term being non-normally distributed.

The rest of the report is arranged as follows: we present our theoretical findings in Section 2 and give the Monte Carlo evidence to support this in Section 3 before Section 4 concludes.

2 Theoretical Results Related to the Econometrics Model

The error term defined as ϵ_i in model (1) follows a t-distribution with v degrees of freedom. Furthermore, the following conditions are implemented on the error term: $E(\epsilon_i) = 0$ for $v > 1$ and $Var(\epsilon_i) = \frac{v}{v-2}\sigma^2$ for $v > 2$. These conditions indicate $E(\epsilon_i)$ not exist for $0 < v \leq 1$ and $Var(\epsilon_i)$ does not exist for $0 < v \leq 2$ These conditions are crucial for defining the assumptions satisfied by and properties of model.

It can be shown all of the Gauss Markov (GM) assumptions hold for both finite and large sample settings when $v > 2$ (Topic 1: p.7, Topic 4: p.4). Some Gauss Markov assumptions hold regardless of v including the first assumption for both sample settings. Additionally, we have the condition $x_{1i} = 3x_{2i} + u_i$ in model (1). With the existence of $u_i \sim i.i.d.N(0,1)$, a normally distributed error term we cannot prove there is a linear relationship between x_1 and x_2. The second finite sample GM assumption of exogeneity (Topic 1: p.7): $E(\epsilon_i|x_i) = 0$, holds only when $v > 1$, as only then can the mean of the error term exist. The final GM assumption for finite sample holds on the condition $v > 2$. This allows $Var(\epsilon_i|x_i) = E(\epsilon_i^2|x_i) = \frac{v}{v-2}\sigma^2$ so we can define $Cov(\epsilon_i^2, x_i) = \frac{v}{v-2}\sigma^2\, Var(x_i) = 0$. Thus, we have no heteroskedasticity in the model given ϵ_i is not a function of x_i.

The second GM large sample assumption (Topic 4: p.7) is satisfied because ϵ_i and x_i follow independent, identically distributed and random distributions. The third GM condition is satisfied when $v > 1$, as $E(\epsilon_i) = E(\epsilon_i\, x_{ki}) = 0$. Lastly, one can show $E(\epsilon_i^2) = Var(\epsilon_i) = \frac{v}{v-2}\sigma^2$ when the condition $v > 2$ is met. Therefore, the OLS estimators are unbiased and consistent on the condition $v > 2$ as all the Gauss Markov assumptions are met. The results are summarized by the following theories.

Theorem 1 *The OLS estimators for the slopes and intercept are defined respectively below:*

$$\hat{\beta}_1 = \beta_{1,} + \frac{\sum_{i=1}^{N} \hat{r}_{1i}\, \epsilon_i}{\sum_{i=1}^{N} \hat{r}_{1i}^2} \tag{2}$$

$$\hat{\beta}_2 = \beta_2 + \frac{\sum_{i=1}^{N} \hat{r}_{2i}\, \epsilon_i}{\sum_{i=1}^{N} \hat{r}_{2i}^2} \tag{3}$$

$$\hat{\beta}_0 = \overline{y} - \overline{x}_1\hat{\beta}_1 - \overline{x}_2\hat{\beta}_2 \tag{4}$$

$$\hat{\beta}_0 = \beta_0 + \overline{\epsilon} - \overline{x}_1 \frac{\sum_{i=1}^{N} \hat{r}_{1i}\, \epsilon_i}{\sum_{i=1}^{N} \hat{r}_{1i}^2}\; \overline{x}_2 \frac{\sum_{i=1}^{N} \hat{r}_{2i}\, \epsilon_i}{\sum_{i=1}^{N} \hat{r}_{2i}^2} \tag{5}$$

The expression $\hat{r}_{ki}$ is the estimated residual from regressing x_k on an intercept and all other explanatory variables.

The existence of ϵ_i in the OLS estimators for $\hat{\beta}_1$ and $\hat{\beta}_2$ implies that when taking expectations of both sides, $E(\epsilon_i)$ does not exist for $0 < v \leq 1$. This would not allow for the OLS estimators to be accurate, hence $E(\beta_1) \neq \beta_1$ and the model is biased. The following conditions show unbiasedness.

$\overline{x_1,} = \frac{\sum_{i=1}^{N} x_{1i}}{N}$, $\overline{x_2} = \frac{\sum_{i=1}^{N} x_{2i}}{N}$ and $\overline{y} = \frac{\sum_{i=1}^{N} y_i}{N}$. Then $E\big(\hat{\beta}_1\big) = \beta_1$, $E\big(\hat{\beta}_2\big) = \beta_2$ and $E\big(\hat{\beta}_0\big) = \beta_0$.

To test the consistency of the OLS estimators we apply the Law of Large Numbers. Consider the probability limit $plim\overline{x} = \mu$ as a consistent estimator of the population mean. Under this assumption we can derive $plim\hat{\beta}_k = \beta_k + \frac{0}{E(\hat{r}_{ki}^2)} = \beta_k$ (Topic 4: p.5). Given the estimators are unbiased and consistent, we can express the true variance of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$.

Theorem 2 *The true variance of the OLS estimators for the slopes and intercept are defined as:*

$$\widehat{Var}(\hat{\beta}_1) = Var(\hat{\beta}_1 \mid x_{1i}) = \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{1i}^2} \tag{6}$$

$$\widehat{Var}(\hat{\beta}_2) = Var(\hat{\beta}_2 \mid x_{2i}) = \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{2i}^2} \tag{7}$$

$$\widehat{Var}(\hat{\beta}_0) = \frac{\frac{v}{v-2}\sigma^2}{N} + \overline{x}_1^2 \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{1i}^2} \overline{x}_2^2 \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{2i}^2} \tag{8}$$

Using eviews, we can obtain $\sum_{i=1}^{N} \hat{r}_{1i}^2$ and $\sum_{i=1}^{N} \hat{r}_{2i}^2$ to find the true variance. The value for $\hat{r}_1$ can be seen in (*n=500v=3.wf1*). Given OLS estimates for $\sqrt{NVar(\hat{\beta}_k)}$ produced by the eviews equation object, we can compare to the true variance to determine whether the estimates are unbiased and consistent.

Given the error term is non-normally distributed, we cannot state under finite settings that the OLS estimators are normally distributed. However, according to standard econometrics textbooks including Wooldridge (2008 [1]), given Gauss Markov assumptions hold for large sample settings, β_j for $j = 0$, 1 and 2 will follow normal distribution regardless of the fact ϵ_i is not normally distributed. This is supported by the Central Limit Theorem, $\bar{x} \overset{a}{\sim} N\left(E(X_i), \frac{Var(X_i)}{N}\right)$, which states the sample average, $\overline{x}$, asymptotically follows the normal distribution as N tends to infinity.

Theorem 3 *The OLS estimators will asymptotically follow the normal distribution:*

$$\widehat{\beta_1} \mid X \sim N\left(\beta_1, \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{1i}^2}\right) \tag{9}$$

$$\widehat{\beta_2} \mid X \sim N\left(\beta_2, \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{2i}^2}\right) \tag{10}$$

$$\widehat{\beta_0} \mid X \sim N\left(\beta_0, \frac{\frac{v}{v-2}\sigma^2}{N} + \overline{x}_1^2 \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{1i}^2} \overline{x}_2^2 \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{2i}^2}\right) \tag{11}$$

Thus, $NVar(\hat{\beta}_j)$ will converge asymptotically to be normally distributed when v is large. We can estimate $\widehat{\frac{v}{v-2}\sigma^2} = \frac{SSR}{N-v-1}$ in the model allowing us to calculate the true variance. Thus, we can state $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{Var}(\hat{\beta}_j)}} \xrightarrow{d} t_{N-K-1}$ for large sample settings, meaning $\hat{\beta}_j$ is very close to standard normal distribution. In section (3), we observe how the distribution behaves according to N and v, proving whether this theorem holds. This would prove that the t statistics are asymptotically valid despite the error term being non-normally distributed.

3 Monte Carlo Experiments and Results

In this section, we use a large number of Monte Carlo simulations to study the theories discussed in section (2). We set the true values in the model as $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 3$ and $\sigma^2 = 1$. All results are based on 10,000 simulations of the Monte Carlo, which is crucially large to ensure our results are accurate and figures we observe including the true variance are valid. The data is generated according to model (1), and we have saved the OLS estimates and their descriptive statistics to find their true properties. We will vary v and N to test the theories stated in section (2). The values we will use are $N = 5,\ 50,\ 200,\ 500$ and $v = 0.5,\ 1.5,\ 3,\ 300$.

Tables 1, 2 and 3 show the summary statistics for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ respectively. For these tables, we have used $v = 300$, as this satisfies all of the conditions stated in section (1), allowing us to test whether the OLS estimators are asymptotically normally distributed. Within ($N = 500v = 0.5.wf1$, $N = 500v = 1.5.wf1$ and N $= 500v = 3.wf1$), one can select ($beta1hat$, $beta2hat$, and $beta3hat$) to confirm the JB p-value is 0.000 for these small values of v. With increased degrees of freedom, we have more information about the population's standard deviation, so we expect the t-distribution to asymptotically follow the normal distribution.

Tables (1, 2 and 3) show the OLS estimators are unbiased and consistent, given the means of $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are very close to their true values regardless of N. The range between the minimum and maximum values decreases as N increases. Furthermore, the estimated standard deviations tend to 0 as N increases. These are signs of consistency.

Table 1: Summary Statistics for $\hat{\beta}_0$ when $v = 300$ from Simulations

N	mean	median	max.	min.	std. dev	JB p-value
5	0.960	0.917	14.082	-12.897	3.489	0.558
50	0.997	0.998	2.402	-0.270	0.340	0.484
200	1.000	0.998	1.561	0.362	0.148	0.245
500	1.001	1.001	1.326	0.673	0.091	0.628

Table 2: Summary Statistics for $\hat{\beta}_1$ when $v = 300$ from Simulations

N	mean	median	max.	min.	std. dev	JB p-value
5	2.007	2.003	5.216	-0.999	0.736	0.056
50	2.002	2.002	2.557	1.397	0.150	0.444
200	2.001	2.002	2.277	1.702	0.078	0.436
500	2.000	2.000	2.189	1.834	0.046	0.695

Table 3: Summary Statistics for $\hat{\beta}_2$ when $v = 300$ from Simulations

N	mean	median	max.	min.	std. dev	JB p-value
5	3.041	3.102	18.545	-12.645	3.591	0.364
50	3.003	3.007	5.752	-0.212	0.765	0.467
200	2.997	2.994	4.438	1.493	0.362	0.644
500	2.998	2.998	3.817	2.127	0.212	0.548

Using the eviews equation object for when v and N are large shows the true variance converges asymptotically to $NVar(\hat{\beta}_0) = 4\sigma^2$, $NVar(\hat{\beta}_1) = \sigma^2$ and $NVar(\hat{\beta}_2) = \sigma^2$. From the Wikipedia entry on t-distribution in the error term [2], we can see the skewness is 0 for $v > 3$, meaning we can expect the distribution to be symmetrical and the large value for $NVar(\hat{\beta}_2)$ to be due to the heavy tails omitted by the t-distribution. Regardless, in tables (1, 2 and 3), the OLS estimates for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ and for standard deviation, $\sqrt{NVar(\hat{\beta}_k)}$ for $k = 1, 2, 3$, are unbiased and consistent.

When v is large, the Jarque-Bera probability exceeds 10% even with a small sample meaning we cannot reject the normality hypothesis when N is large. Also, the JB p-value generally rises as N increases- take table (2). At $N = 5$ the value is actually below 10% however all other values are above 10% with 69.5% by $N = 500$.

To test whether the estimates are unbiased and consistent at different values of v and N, the following results tables summarize.

Table 4: OLS estimates of $\hat{\beta}_0$ given v and N

$E(\widehat{\beta}_0)$	v = 0.5	v = 1.5	v = 3
N=5	<-100	0.455	1.009
N=50	<-100	1.006	1.019
N=200	<-100	1.019	1.003
N=500	>100	1.026	1.000

Table 5: OLS estimates of $\hat{\beta}_1$ given v and N

$E(\widehat{\beta}_1)$	v = 0.5	v = 1.5	v = 3
N=5	<-100	2.043	1.995
N=50	>100	2.013	2.000
N=200	<-100	2.010	1.999
N=500	<-100	1.993	2.001

Table 6: OLS estimates of $\hat{\beta}_2$ given v and N

$E(\widehat{\boldsymbol{\beta}}_2)$	v = 0.5	v = 1.5	v = 3
N=5	71.905	3.484	2.999
N=50	>100	2.967	2.966
N=200	>100	2.973	2.999
N=500	>100	3.002	2.998

These estimates yield the following standard deviations:

Table 7: OLS estimates of standard deviation of $\hat{\beta}_0$ given v and N

$\sqrt{NVar(\widehat{\boldsymbol{\beta}}_0)}$	v = 0.5	v = 1.5	v = 3
N=5	>100	73.166	6.774
N=50	>100	4.321	0.613
N=200	>100	3.961	0.248
N=500	>100	3.036	0.157

Table 8: OLS estimates of standard deviation of $\hat{\beta}_1$ given v and N

$\sqrt{NVar(\widehat{\boldsymbol{\beta}}_1)}$	v = 0.5	v = 1.5	v = 3
N=5	>100	12.362	1.357
N=50	>100	2.116	0.266
N=200	>100	1.017	0.135
N=500	>100	1.048	0.079

Table 9: OLS estimates of standard deviation of $\hat{\beta}_2$ given v and N

$\sqrt{NVar(\widehat{\boldsymbol{\beta}}_2)}$	v = 0.5	v = 1.5	v = 3
N=5	>100	96.677	7.346
N=50	>100	10.776	1.373
N=200	>100	4.695	0.616
N=500	>100	4.188	0.368

As predicted by section (2), every result for $v = 0.5$ is invalid. The OLS estimates for $v = 3$ and $v = 1.5$ are all accurate and converge to those in tables (1, 2 and 3) as N increases, showing signs of consistency. The estimates for $\sqrt{NVar(\hat{\beta}_k)}$ for $k = 1, 2, 3$ are biased given $\widehat{Var}(\hat{\epsilon}_i)$ does not exist for $0 < v \leq 2$ as alluded to in theorem (2), however these estimates are unbiased and consistent for $v = 3$.

None of the JB p-values were valid for $v = 0.5,\ 1.5,\ 3$ due to the heavy tails of the t-distribution. Because of this, the t statistics constructed for finite samples may not be valid, however the Central Limit Theorem outlined in theorem (2) should displace this issue when N is large. The t statistics are constructed by $t_j = \frac{\hat{\beta}_j - \beta_{j,0}}{SE(\hat{\beta}_j)}$. The standard error estimates are:

Table 10: Standard error estimates for $\hat{\beta}_0$

	v=0.5	v=1.5	v=3	v=300
N=5	>100	10.301	4.701	3.133
N=50	>100	1.778	0.554	0.339
N=200	>100	1.000	0.245	0.146
N=500	>100	0.716	0.154	0.091

Table 11: Standard error estimates for $\hat{\beta}_1$

	v=0.5	v=1.5	v=3	v=300
N=5	>100	2.170	0.991	0.660
N=50	>100	0.785	0.245	0.150
N=200	>100	0.533	0.131	0.078
N=500	>100	0.360	0.077	0.046

Table 12: Standard error estimates for $\hat{\beta}_2$

	v=0.5	v=1.5	v=3	v=300
N=5	>100	10.628	4.851	3.233
N=50	>100	4.005	1.248	0.763
N=200	>100	2.481	0.609	0.362
N=500	>100	1.676	0.360	0.212

Tables (10, 11 and 12) show the t statistics will be asymptotically valid when $v > 1$, as the standard error estimates converge to the true values as N increases. The p curves following section (4) support this hypothesis.

4 Conclusion

In this report, we have studied how an error term following t-distribution effects the OLS estimates at different sample sizes and degrees of freedom. We have found the OLS estimations are more reliable with greater degrees of freedom. The results also reflect consistency when $v > 1$ which should be considered when studying the properties of the OLS estimates.

Table 11 The following key applies for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$

	v
BLUE	0.5
ORANGE	1.5
GREY	3
YELLOW	300

Along the x-axis are the results with the concentration of results shown by the change in y-axis value. With a sudden convergence to the middle, a large number of results are populated at the respective x-axis value.

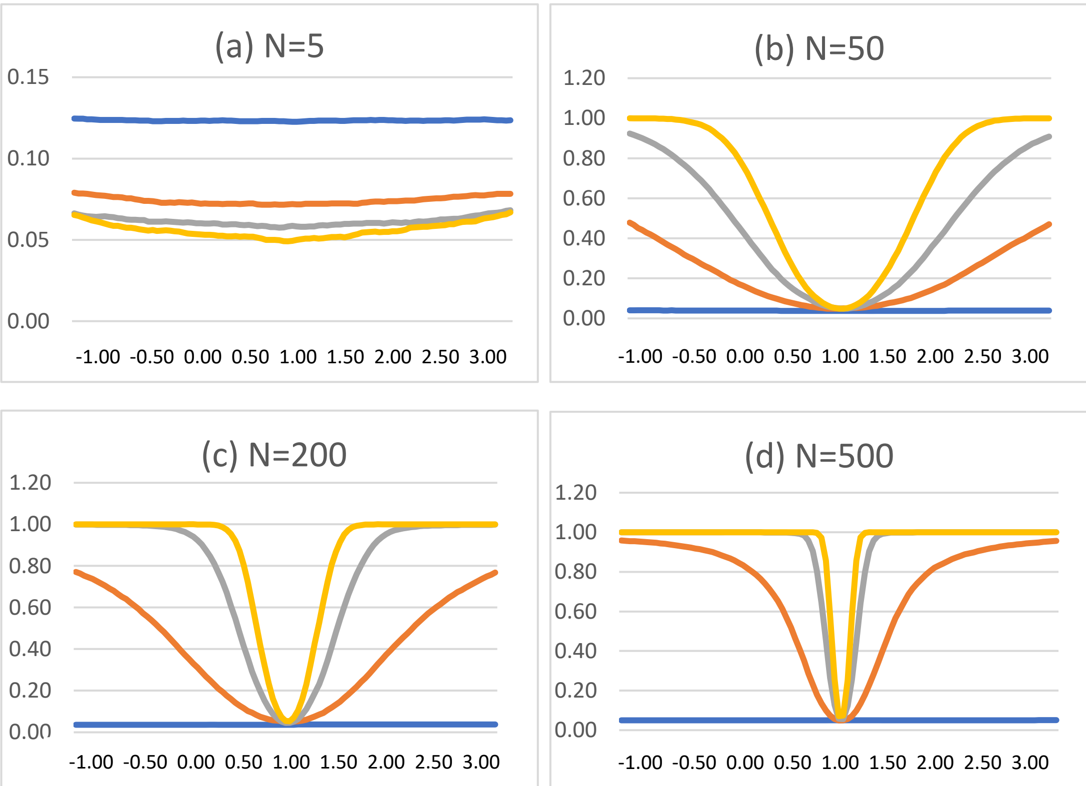

Figure 1: The Power Curves of the T tests for $\hat{\beta}_0$

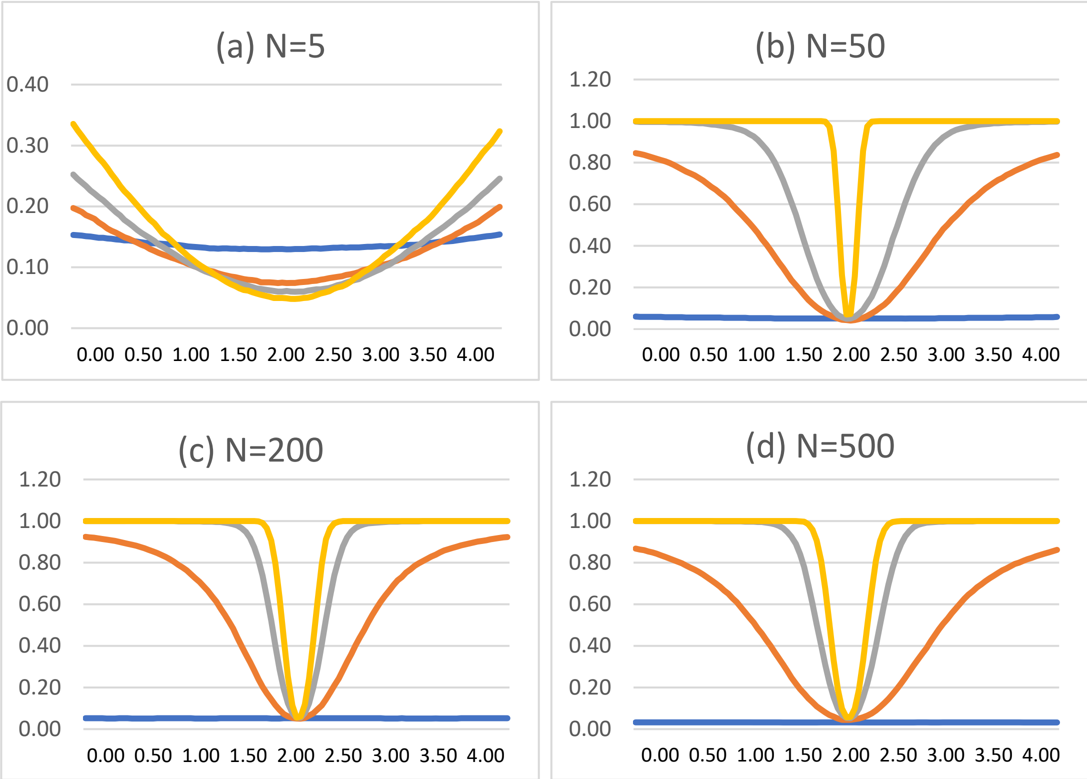

Figure 2: The Power Curves of the T tests for $\hat{\beta}_1$

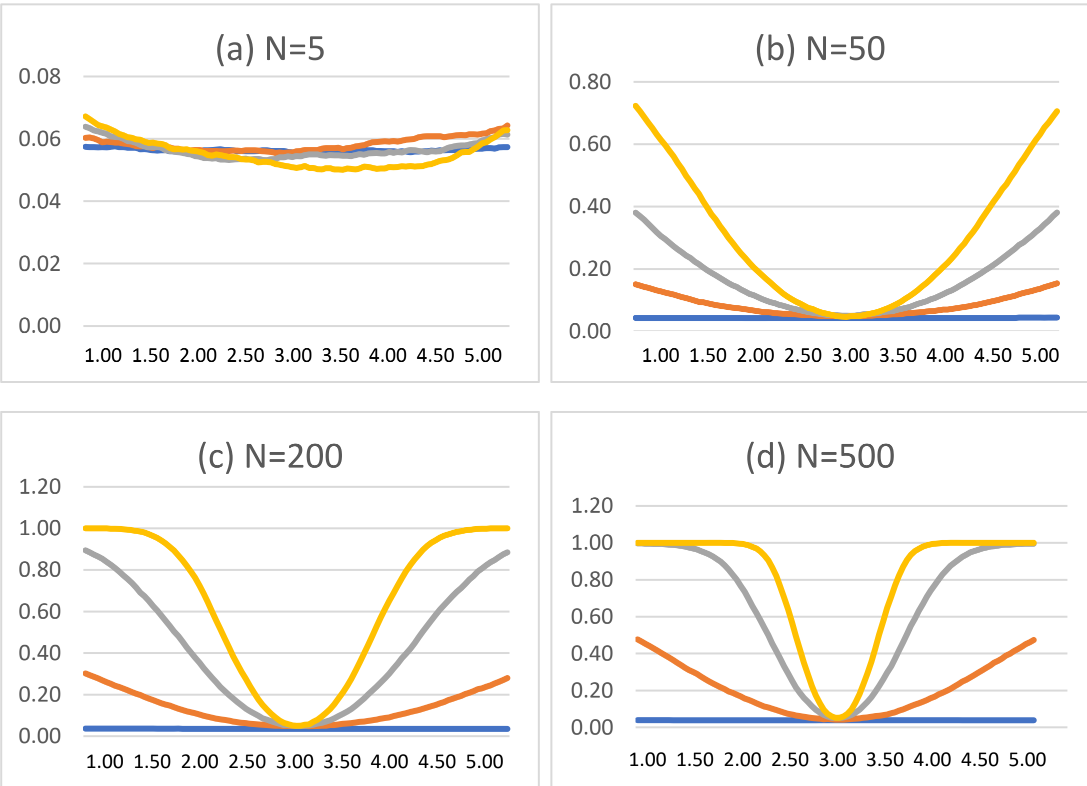

Figure 3: The Power Curves of the T tests for $\hat{\beta}_1$

A Appendix

A.1 Proof for Theorem 1

Given the first Gauss Markov assumption which holds for both finite and large samples of N:

$$\hat{\beta}_0 = \overline{y} - \overline{x}_1\hat{\beta}_1 - \overline{x}_2\hat{\beta}_2$$

$$x_{1i} = \hat{\gamma}_1 + x_{2i}\hat{\gamma}_2 + \hat{r}_{1i} = \hat{x}_{1i} + \hat{r}_{1i}$$

Considering all $\hat{\gamma}$ values as OLS estimators, $\hat{r}$ is orthogonal to $\hat{x}_1$ and $\hat{x}_2$. This gives us the following FOC for $\hat{\beta}_1$:

$$\sum_{i=1}^{N} x_{1i}\hat{\epsilon}_i = \sum_{i=1}^{N} (\hat{x}_{1i} + \hat{r}_{1i})\hat{\epsilon}_i = \sum_{i=1}^{N} \hat{r}_{1i}\hat{\epsilon}_i = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} \hat{r}_{1i} y_i}{\sum_{i=1}^{N} \hat{r}_{1i}^2} = \frac{\sum_{i=1}^{N} \hat{r}_{1i} y_i}{\sum_{i=1}^{N} \hat{r}_{1i}} = \frac{\sum_{i=1}^{N} \hat{r}_{1i}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i)}{\sum_{i=1}^{N} \hat{r}_{1i} x_{1i}} = \beta_1 + \frac{\sum_{i=1}^{N} \hat{r}_{1i}\epsilon_i}{\sum_{i=1}^{N} \hat{r}_{1i}^2}$$

The derivation for $\hat{\beta}_2$ is the same as above for $\hat{\beta}_1$

$\hat{\beta}_0 = \beta_0 + \overline{\epsilon} - \overline{x}_1 \frac{\sum_{i=1}^{N} \hat{r}_{1i}\,\epsilon_i}{\sum_{i=1}^{N} \hat{r}_{1i}^2}\, \overline{x}_2 \frac{\sum_{i=1}^{N} \hat{r}_{2i}\,\epsilon_i}{\sum_{i=1}^{N} \hat{r}_{2i}^2}$ by substitution of $\hat{\beta}_1$ and $\hat{\beta}_2$ into $\hat{\beta}_0 = \overline{y} - \overline{x}_1\hat{\beta}_1 - \overline{x}_2\hat{\beta}_2$.

Given $\hat{r}_{ki}$ is orthogonal to all other regressors, we can show the proof for consistency:

$$\begin{aligned}
\beta_k = \frac{\sum_{i=1}^{N} \hat{r}_{ki} y_i}{\sum_{i=1}^{N} \hat{r}_{ki}^2} &= \frac{\sum_{i=1}^{N} \hat{r}_{ki}(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i} + \epsilon_i)}{\sum_{i=1}^{N} \hat{r}_{ki}^2} = \beta_k \frac{\sum_{i=1}^{N} \hat{r}_{ki} x_{ki}}{\sum_{i=1}^{N} \hat{r}_{ki}^2} + \frac{\sum_{i=1}^{N} \hat{r}_{ki}\epsilon_i}{\sum_{i=1}^{N} \hat{r}_{ki}^2} \\
&= \beta_k \frac{\sum_{i=1}^{N} \hat{r}_{ki}(\hat{x}_{ki} + \hat{r}_{ki})}{\sum_{i=1}^{N} \hat{r}_{ki}^2} + \frac{\frac{\sum_{i=1}^{N} \hat{r}_{ki}\epsilon_i}{N}}{\frac{\sum_{i=1}^{N} \hat{r}_{ki}^2}{N}} = \beta_k + \frac{\frac{\sum_{i=1}^{N} \hat{r}_{ki}\epsilon_i}{N}}{\frac{\sum_{i=1}^{N} \hat{r}_{ki}^2}{N}}.
\end{aligned}$$

$$\begin{aligned}
plim \frac{\sum_{i=1}^{N} \hat{r}_{ki}\epsilon_i}{N} &= plim \frac{\sum_{i=1}^{N} (x_{ki} - \hat{\gamma}_0 - \hat{\gamma}_1 x_{1i} - \hat{\gamma}_2 x_{2i})\epsilon_i}{N} \\
&= plim \frac{\sum_{i=1}^{N} x_{ki}\epsilon_i}{N} - plim\hat{\gamma}_0 plim \frac{\sum_{i=1}^{N} \epsilon_i}{N} - plim\hat{\gamma}_1 plim \frac{\sum_{i=1}^{N} x_{1i}\epsilon_i}{N} \\
&- plim\hat{\gamma}_2 plim \frac{\sum_{i=1}^{N} x_{2i}\epsilon_i}{N} \\
&= E(x_{ki}\epsilon_i) - plim\hat{\gamma}_0\, E(\epsilon_i) - plim\hat{\gamma}_1\, E(x_{1i}\epsilon_1) - plim\hat{\gamma}_2\, E(x_{2i}\epsilon_i) = 0
\end{aligned}$$

Taking probability limit on both sides yields $\hat{\beta}_k = \beta_k + \frac{0}{E(\hat{r}_{ki}^2)} = \beta_k$

A.2 Proof for Theorem 2

$$Var(\hat{\beta}_{1,2} \mid x_{1i,2i}) = Var\left(\frac{\sum_{i=1}^{N} \hat{r}_{1i,2i}\epsilon_i}{\sum_{i=1}^{N} \hat{r}_{1i,2i}^2} \middle| x_{1i,2i}\right) = \frac{Var(\sum_{i=1}^{N} \hat{r}_{1i,2i}\epsilon_i \mid x_{1i,2i})}{\left[\sum_{i=1}^{N} \hat{r}_{1i,2i}^2\right]^2}$$

$$= \frac{\sum_{i=1}^{N} \hat{r}_{1i,2i}^2 \, Var(\epsilon_i \mid x_{1i,2i})}{\left[\sum_{i=1}^{N} \hat{r}_{1i,2i}^2\right]^2} = \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{1i,2i}^2}$$

For $Var(\hat{\beta}_0)$:

$\hat{\beta}_0 = \beta_0 + \overline{\epsilon} - \overline{x}_1 \frac{\sum_{i=1}^{N} \hat{r}_{1i}\,\epsilon_i}{\sum_{i=1}^{N} \hat{r}_{1i}^2}\ \overline{x}_2 \frac{\sum_{i=1}^{N} \hat{r}_{2i}\,\epsilon_i}{\sum_{i=1}^{N} \hat{r}_{2i}^2}$, taking variance of both sides yields $Var(\hat{\beta}_0) = Var(\overline{\epsilon} - \overline{x}_1 \frac{\sum_{i=1}^{N} \hat{r}_{1i}\,\epsilon_i}{\sum_{i=1}^{N} \hat{r}_{1i}^2}\ \overline{x}_2 \frac{\sum_{i=1}^{N} \hat{r}_{2i}\,\epsilon_i}{\sum_{i=1}^{N} \hat{r}_{2i}^2})$ because constant $\beta_0 = 0$. Defining and substituting $Var(\overline{\epsilon}) = \frac{\frac{v}{v-2}\sigma^2}{N}$ gives $\widehat{Var}(\hat{\beta}_0) = \frac{\frac{v}{v-2}\sigma^2}{N} + \overline{x}_1^2 \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{1i}^2}\ \overline{x}_2^2 \frac{\frac{v}{v-2}\sigma^2}{\sum_{i=1}^{N} \hat{r}_{2i}^2}$.