# Problem Set 2 Solutions

Alberto Ramírez

February 20, 2014

## Problem 1

### Part B: Conditional Mean $E(y|x, \theta)$ and Conditional Variance $V(y|x)$

We can compute the conditional mean and variance of a general binary regression as follows. For the conditional mean we have, for a general CDF $F(x'\theta)$:

$$E(y|x, \theta) = 0 \cdot [1 - F(x'\theta)] + 1 \cdot [F(x'\theta)] = F(x'\theta)$$

And for the conditioanl variance we would have:

$$V(y|x) = E(y^2|x) - [E(y|x, \theta)]^2 = [0^2 \cdot (1 - F(x'\theta)) + 1^2 \cdot (F(x'\theta))] - [F(x'\theta)]^2 = F(x'\theta)[1 - F(x'\theta)]$$

Since we are dealing with the logistic distribution, we set the above CDF to the logistic CDF $F(x'\theta) = \Lambda(x'\theta)$ and obtain:

$$E(y|x, \theta) = \Lambda(x'\theta) \quad \text{and} \quad V(y|x) = \Lambda(x'\theta)[1 - \Lambda(x'\theta)]$$

### Part C: Propose at least 2 Moment Conditions

We may propose two sets of moment conditions: one from the score and another from the above conditional mean and variance. Since we have two parameters to estimate both sets will yield two moment conditions (this is less directly seen by the score, but recall from the previous homework set that the score will give a $2 \times 1$ vector of parameters).

#### Moment Conditions based on the Score

This will be less interesting as GMM estimation based on the score will give the same result as MLE estimation (obviously, since the score fully describes the distribution of the logit model and we have assumed that the errors are logisticaly distributed). So we have, given that the expectation of the score is 0:

$$m_2(\theta) \equiv E\left(\frac{\partial L(\theta|y_t)}{\partial \theta}\right) = 0$$

for any consistant estimate of $\theta$, $\widehat{\theta}$, such that $E(\widehat{\theta}) - \theta = 0$. Then the 2 moment conditions based on the score are given, in matrix form by:

$$m_2(\widehat{\theta}) \equiv X'[Y - \Lambda(X\widehat{\theta})]$$

**Moment Conditions based on the Conditional Mean and Conditional Variance**

From the results in part B we have the population moments of the mean and variance. We can thus define the following moment conditions

$$m_\mu(\theta) \equiv E(y|x,\theta) - \widehat{\mu} = \mu_{y|x}(\theta) - \widehat{\mu} = \Lambda(x'\theta) - \widehat{\mu}$$

$$m_{var}(\theta) \equiv V(y|x) - \widehat{V}(y) = \Lambda(x'\theta)[1 - \Lambda(x'\theta)] - \widehat{V}(y)$$

where the sample moments, $\widehat{\mu}$ and $\widehat{V}(y)$, are given by

$$\widehat{\mu} = \frac{1}{T}\sum_{i=1}^{T} y_i \quad \text{and} \quad \widehat{V}(y) = \frac{1}{T}\sum_{i=1}^{T}(y_i - \widehat{\mu})^2$$

since the sample moments are asymptotically consistant estimates of the population moments. Therefore, the estimate of $\theta$, $\widehat{\theta}$, such that the population moment equals the sample moment, that is $m_n(\widehat{\theta}) \equiv 0$, is the estimate we seek in GMM estimation.

The matrix form with the estimate $\widehat{\theta}$ that we will find through GMM is then:

$$m_2(\widehat{\theta}) = \begin{bmatrix} \Lambda(x'\widehat{\theta}) - \frac{1}{T}\sum_{i=1}^{T} y_i \\ \\ \Lambda(x'\widehat{\theta})[1 - \Lambda(x'\widehat{\theta})] - \frac{1}{T-1}\sum_{i=1}^{T}(y_i - \widehat{\mu})^2 \end{bmatrix}$$

# Problem 2

# Problem 3

We have the following two equations describing our process, the distribution of the error terms, and are given that $x_t$ is **strictly** exogenous to both error terms:

$$y_{t1} = \alpha_1 + \alpha_2 y_{t2} + \varepsilon_{t1} \tag{1}$$

$$y_{t2} = \beta_1 + \beta_1 x_t + \varepsilon_{t2} \tag{2}$$

$$\begin{pmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \end{pmatrix} \sim \left(0, \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$$

## Part A: OLS Estimation of $\alpha$ in Eq 1 Consistant?

From (1) it is clear that $y_{t2}$ is endogenous to $\varepsilon_{t1}$. So OLS will not yield consistant estimates of our parameter $\alpha$. This is evident from the distribution of $\varepsilon_{ti}$ since the off diagnol elements are not equal to 0.

We will proceed as usual, deriving the sampling error of the OLS estimator of $\alpha$, then applying a WLLN along with Slutzky's and Cramer's theorems. We begin by defining the concatenated matrix $Y = [1 \; y_{t2}]$, our $n \times 2$ design matrix, and the vector $y_{t1}$ of dim $n \times 1$. Then the OLS estimator of $\alpha$ is the $2 \times 1$ vector given by:

$$\widehat{\alpha}_{OLS} = (Y'Y)^{-1}Y'y_{t1}$$

$$\widehat{\alpha}_{OLS} - \alpha = (Y'Y)^{-1}Y'y_{t1} - \alpha =$$

$$(Y'Y)^{-1}Y'(Y\alpha + \varepsilon_{t1}) - \alpha =$$

$$\cancelto{1}{(Y'Y)^{-1}Y'Y}\alpha + (Y'Y)^{-1}Y'\varepsilon_{t1} - \alpha = (Y'Y)^{-1}Y'\varepsilon_{t1}$$

Vectorizing the above matrices and dividing by T we get:

$$\widehat{\alpha}_{OLS} - \alpha = \left(\frac{\sum_{t=1}^{T} y_t y_t'}{T}\right)^{-1} \left(\frac{\sum_{t=1}^{T} y_t \varepsilon_{t1}}{T}\right)$$

Now we can apply the WLLN to get

$$\underset{T\to\infty}{plim} \frac{\sum_{t=1}^{T} y_t y_t'}{T} = E(y_t y_t')\ ,\ \underset{T\to\infty}{plim} \frac{\sum_{t=1}^{T} y_t \varepsilon_{t1}}{T} = E(y_t \varepsilon_{t1})$$

And by Slutzky's theorem, assuming that we have $E(y_t y_t')$ is square and of full column rank (and so invertible):

$$\underset{T\to\infty}{plim} \left(\frac{\sum_{t=1}^{T} y_t y_t'}{T}\right)^{-1} = E(y_t y_t')^{-1}$$

Finally, Cramer's theorem gives the following result (in matrix form):

$$\underset{T\to\infty}{plim}\ (\widehat{\alpha}_{OLS} - \alpha) = E(Y'Y)^{-1}E(Y'\varepsilon_{t1})$$

Then for $\widehat{\alpha}_{OLS}$ to be a consistant estimate of $\alpha$ we must have that $E(Y'\varepsilon_{t1}) = 0$. We can first decompose $E(Y'\varepsilon_{t1})$ into $E(1 \cdot \varepsilon_{t1})$ and $E(y_{t2}\varepsilon_{t1})$, where obviously $E(1 \cdot \varepsilon_{t1}) = 0$. So now we must investigate if $E(y_{t2}\varepsilon_{t1}) = 0$.

$$E(y_{t2}\varepsilon_{t1}) = E[(\beta_1 + \beta_2 x_t + \varepsilon_{t2}) \cdot \varepsilon_{t1}] =$$

$$\beta_1 \cancelto{0}{E(\varepsilon_{t1})} + \beta_2 \cancelto{0}{E(x_t \varepsilon_{t1})} + E(\varepsilon_{t2}\varepsilon_{t1}) = \rho_{12}\sigma_1\sigma_2 \neq 0$$

This shows that the sampling error of equation (1) is not equal to 0 in expectation, as $E(y_{t2}\varepsilon_{t1}) = \rho_{12}\sigma_1\sigma_2 \neq 0$. So the estimate $\widehat{\alpha}_{OLS}$ is not a consistant estimator of $\alpha$ given the above derived bias in the estimation.

## Part B: OLS Estimation of $\beta$ in Eq 2 Consistant?

It is straightforward to show that $\widehat{\beta}_{OLS}$ is indeed a consistant estimator of $\beta$ since $x_t$ is strictly exogenous to both error terms. The procedure is the same as that done in part A and amounts to showing that in expectation the sampling error $(X'X)^{-1}X'\varepsilon_{t2}$ is indeed equal to 0, where $X = [1\ x_t]$ is our $n \times 2$ design matrix. Therefore $plim_{T\to\infty}\widehat{\beta}_{OLS} = \beta$.

## Part C: Consistant Estimator of $\alpha$

There are two methods to consistantly estimate $\alpha$, each with it's own advantages and disadvantages. However, we will first show why naively attempting to insert (2) into

(1) will not yield us the desired estimates for $\alpha$ even though the resulting estimator is consistant.[1]

$$y_{t1} = \alpha_1 + \alpha_2(\beta_1 + \beta_2 x_t + \varepsilon_{t2}) + \varepsilon_{t1}$$
$$y_{t1} = \alpha_1 + \alpha_2\beta_1 + \alpha_2\beta_2 x_t + \alpha_2\varepsilon_{2t} + \varepsilon_{t1}$$

We can then define the following parameters:

$$\delta_1 = \alpha_1 + \alpha_2\beta_1 \quad , \quad \delta_2 = \alpha_2\beta_2 \quad , \quad \nu_t = \alpha_2\varepsilon_{t2} + \varepsilon_{t1} \tag{3}$$

The reduced form representation of the system is then:

$$y_{t1} = \delta_1 + \delta_2 x_t + \nu_t$$

And since $\nu_t$ is a linear combination of random variables $\varepsilon_{t1}$ and $\varepsilon_{t2}$ and $x_t$ is exogenous to these error terms, it will also be exogenous to $\nu_t$. So it is trivial to show that the expectation of the sampling error, $(X'X)^{-1}X'\nu_t$, is 0. Thus, the estimator $\widehat{\delta}_{OLS}$ is a consistant estimate of the reduced form parameter $\delta$. However, given that we have four structural parameters and only two estimates of the reduced form parameters we would be unable to identify $\alpha$ and $\beta$. This is clear from (3).

### 2SLS Regression of Equation (1)

If we ONLY care about consistently estimating the parameters then we can run a 2SLS on the system. However, this method will yield inconsistant and inefficient estimates of the second stage standard errors as it carries into it the error from the first stage; thus the second stage standard errors would require correction. This can be done by either applying an analytical correction (which is not as easy as it sounds) or bootstrapping the standard errors of the second stage (much easier).

First we would estimate $\widehat{\beta}_{OLS}$ since this is a consistant estimate of $\beta$. We can then generate the estimated $\widehat{y}_{t2} = X\widehat{\beta}_{OLS}$. In the second stage we would regress $y_{t1}$ on the estimated $\widehat{y}_{t2}$ (not the actual $y_{t2}$) and this will give us a consistant estimate of $\alpha$.

This works for the following reason:
Adding and subtracting $\alpha_2\widehat{y}_{t2}$ to (1) gives us:

$$y_{t1} = \alpha_1 + \alpha_2\widehat{y}_{t2} + [\varepsilon_{t1} + \alpha_2(y_{t2} - \widehat{y}_{t2})] \quad , \quad \widehat{\varepsilon}_{t2} = (y_{t2} - \widehat{y}_{t2}) : \ \operatorname*{plim}_{T\to\infty} \widehat{\varepsilon}_{t2} = \varepsilon_{t2}$$

The bracketed term is the new error term, which we can call $\eta_t$ for simplicity. Since $\widehat{y}_{t2}$ is a consistant prediction it should be equal to the least square projection matrix $\widehat{E}^*(y_{t2}|1, x_t)$, which is uncorrelated with $\eta_t$ (by definition of the least square projection matrix): $\widehat{y}_{t2}$ is uncorrelated with $\varepsilon_{t1}$ since $\widehat{y}_{t2}$ is a linear function of $x_t$ - which is strictly exogenous to $\varepsilon_{t1}$ - and $(y_{t2} - \widehat{y}_{t2})$ is uncorrelated with $\widehat{y}_{t2}$ because it is the least squares projection error and thus orthogonal to $\widehat{y}_{t2}$. Thus 2SLS is one method for a consistant estimation of $\alpha$.

[1] The result of plugging (2) in (1) is called the reduced form representation of the system. Thus the parameters of the reduced form will be functional forms of the structural system and as we will see, in this case, we would be unable to identify the structural parameters.

### IV Estimation of Equation (1)

The second method, the preferred method[2], is the IV estimation of (1) using as an instrument $x_t$ and a constant. Thus we would define $Z = [1\ x_t]$ as our set of instruments. This would work because $x_t$ is correlated with $y_{t2}$ (since $E[x_t y'_{t2}] \neq 0$, by equation (2)) and by strict exogeneity our instruments are uncorrelated with $\varepsilon_{t1}$ (implies $E(Z'\varepsilon_{t1}) = 0$). Our instruments therefore satisfy the two properties required. Therefore our IV estimator will be:

$$\widehat{\alpha}_{IV} = (Z'Y)^{-1} Z' y_{t1} \quad \text{where } Y = [1\ y_{t2}]$$

We can show that with these instruments $\widehat{\alpha}_{IV}$ consistantly estimates $\alpha$. A similar exercise was done in PS 1 and will not be reproduced here.

## Part D: Consistant Estimator of $\beta$

Since we already have that $\widehat{\beta}_{OLS}$ is a consistant estimator of $\beta$ we do not need to propose a consistant estimator.

# Problem 4

The poisson density is given by:

$$f_X(x) = \left\{ \begin{array}{c} (e^{-\lambda}\lambda^x)(x!)^{-1}\ ,\ x \in \{0, 1, 2, ...\} \\ 0\ ,\ otherwise \end{array} \right\} \tag{4}$$

where we have an i.i.d. sequence $\{x_i\}_{i=1,2,3,\ldots,n}$ and wish to estimate the parameter $\lambda$.

## Part A: Derive the LLF

Because the sequence $\{x_i\}$ is i.i.d we can take the product of each distribution of $f_X(x)$ in (4). Then taking logs we will have the log-likelihood function:

$$\mathcal{L}(\lambda|x) = lnf_X(x;\lambda) = ln\left[\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right] = \sum_{i=1}^{n}[x_i ln\lambda - lnx_i! - \lambda] \tag{5}$$

## Part B: Compute ML Estimator of $\lambda$

Take the F.O.C. of (5) with respect to $\lambda$:

$$\max_{\lambda} \mathcal{L}(\lambda|x)$$

$$\frac{\partial \mathcal{L}(\lambda|x)}{\partial \lambda} = \sum_{i=1}^{n} \frac{x_i}{\widehat{\lambda}} - \sum_{i=1}^{n} 1 = 0 \ \Leftrightarrow\ \widehat{\lambda}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

This shows that the MLE estimator of $\lambda$ is the sample mean of $x_i$. Since the expectation of each observation is $\lambda$ the sample mean also has this expectation. So $\widehat{\lambda}_{ML}$ is an unbiased estimator of $\lambda$.

[2]It is preferred simply because this method directly gives us consistant estimates of our standard errors if they are desired vs. the extra computation required with 2SLS to obtain them. Recall from PS 1 that IV estimation may not produce efficient estimates of our standard errors, however.

## Part C: 2 Moment Conditions for GMM Estimation of $\lambda$

Since the poisson distribution has the first and second moments equal to the parameter $\lambda$, specifying two moment conditions based on the sample moments yields overidentification. From the mean we can define the first moment condition as:

$$m_\mu(\lambda) = \lambda - \frac{1}{n}\sum_{i=1}^{n} x_i$$

The second moment condition will then be given by the sample variance (define $\bar{x}$ as the sample mean):

$$m_{var}(\lambda) = \lambda - \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Then in matrix form we have our $2 \times 1$ moment vector as:

$$m_2(\lambda) = \begin{bmatrix} \lambda - \frac{1}{n}\sum_{i=1}^{n} x_i \\ \\ \lambda - \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{bmatrix}$$

So our GMM estimator, $\widehat{\lambda}_{GMM}$, is defined as:

$$\widehat{\lambda}_{GMM}(\widehat{W}) = \arg\min_{\lambda} \; n \cdot m_2(\lambda)'\widehat{W} m_2(\lambda)$$

## Part D: Disscussion of GMM and ML Estimator of $\lambda$

The ML estimator $\widehat{\lambda}_{ML}$ maximizes the likelihood of obtaining the observed sample. To derive the estimator we implicitely made an assumption of the underlying distribution of the sample we observed. If we make an incorrect distributional assumption, then our estimator will generally not be consistant.[3]

On the otherhand, the GMM estimator we proposed is based on sample moments. Therefore it should be robust to distributional misspecifications as long as we have that the moments are correctly specified. This comes with a possible loss of efficiency as we would not be specifying all the possible moments in the estimation; whereas the ML estimator exploits the parametric form of the (assumed) density function of the data (which GMM does not do).[4]

However, in our case the GMM estimator reaches the Cramer-Rao lower bound and so both estimators reach the same asymptotic efficiency. We can take the S.O.C. of (5) to obtain:

$$\frac{\partial^2 \mathcal{L}(\lambda|x)}{\partial \lambda^2} = \frac{-\sum_{i=1}^{n} x_i}{\lambda^2}$$

Then the Fischer information matrix is given by:[5]

$$I(\lambda) = -E\left(\frac{-\sum_{i=1}^{n} x_i}{\lambda^2}\right) = \frac{1}{\lambda^2} E\left(\sum_{i=1}^{n} x_i\right) \rightarrow_p \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$

[3] Quasi-Maximum Likelihood Estimation would be consistant to a distributional misspecification, for example.

[4] In general the asymptotic variance of any GMM estimator can be no smaller than that of the ML estimator, as both variances have as their lower bound the inverse of the information matrix (the Cramer-Rao lower bound).

[5] Recall that the Fischer Information Matrix is a square positive-semidefinite matrix. Since we have only one parameter the matrix is actually a scalar.

And the asymptotic variance of the mean is given by:

$$V(\bar{x}) = \frac{1}{n}V(x) \to_p \frac{\lambda}{n} = \frac{1}{I(\lambda)}$$

Therefore, the sample mean is an efficient estimator of the parameter $\lambda$ as its asymptotic variance reaches the Cramer-Rao lower bound. So both the ML estimator and a GMM estimator consisting solely of the sample mean as its moment condition are both asymptotically efficient and equivalent.

Inclusion of the sample variance as a second moment in the GMM estimation does not increase efficiency since the expected variance of the poisson distribution is also $\lambda$. So the sample variance is also expected to asymptotically reach the Cramer-Rao lower bound. Therefore, adding a second fully efficient moment condition to conduct GMM estimation will not add any more efficiency gains. This shows that both $\widehat{\lambda}_{ML}$ and $\widehat{\lambda}_{GMM}$ are both asymptotically efficient.

# Problem 5

## Part B: Discuss the Results from Part A

Here we have conducted a Monte Carlo experiment with 1000 simulations of the below regression

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{6}$$

based on a DGP where $y_t$ and $x_t$ are independent random walk processes with Gausian white noise. The results of one of these simulations, randomly drawn, is shown in the Appendix. Also included in the appendix are the histograms from the experiment showing the distributions of $R^2$, the t-statistic, and of $\beta_1$ and its standard error. (Note that in the code we have set the random number generators to default, so we always obtain the same results.)

We will concentrate on the results from the Monte Carlo Simulation, as this gives us a better understanding of the spurious relationship that we see from the one realization while allowing us to account for things we may otherwise miss.

- **The Slope Coefficient**: The slope coefficient has estimates ranging between -2 to 2 with a mean of -0.066. The Monte Carlo simulation rejects the null ($H_o : \beta_1 = 0$) 93.2% of the time when using a 5% level test. In reality we should see the opposite: that, within simulation error, the null is rejected about 5% of the time. That we know these results should not be shows that the classical model is failing to account for the spurious correlation between $y_t$ and $x_t$.

- **Distribution of $R^2$**: Our simulation shows the range of values that $R^2$ may assume. It suggests that depending on the realization of the particular time series the classical model attributes a large amount of the variation in y to x, even though the two are independent of one another. The median value is 0.168 (reported as such given the skewness of the distribution). We also note that in a substantial proportion of simulations the value of $R^2$ is quite small and near to 0 - despite this, the model still finds the $\beta_1$ associated with these small $R^2$ to be significantly non-zero.

- **The t-Statistic**: The profile of the histogram for our t-statistics does not seem to conform to that of the t distribuition. (Ours has a mean of -0.066 whereas the t distribution should have a mean of 0; and the standard deviation of the t-statistics we have generated in the simulation is 26.36 and not 1.002 as it should be.) Specifically, we know from asymptotic theory that at such a high sample size the distribution should have a similar profile to that of a standard normal, $\mathcal{N}(0,1)$. But our histogram is too "fat". This shape, with its high standard deviation, indicates that the classical model is finding highly significant $\beta_1$s. As previously alluded, this means that in only 68 simulations a t statistic at or below 1.96 was generated; 932 times the value was greater, in some cases much greater. Since we calculated the p values, we can see that on average the size of the test is 0.026; however the median is much smaller and basically 0 (Matlab gives us a result of $0.203 \times 10^{-50}$).

Given these results we conclude that the results are nonsensical. Further, the above analysis did not conduct a HAC correction of the standard errors, which would affect our t statistics and our standard errors. This will be explained in the following section.

## Part C: Variance of $y_t$ and $x_t$

We can recursively substitute $y_t$ and $x_t$ to obtain the following results:

$$\begin{aligned} y_1 &= \upsilon_1 \\ y_2 &= y_1 + \upsilon_2 = \upsilon_1 + \upsilon_2 \\ y_t &= \sum_{i=1}^{t} \upsilon_i \end{aligned}$$

Similarly we obtain for $x_t$:

$$x_t = \sum_{i=1}^{t} \upsilon_i$$

In the above we have automatically used that $y_0 = 0$ and $x_0 = 0$. Then the autocovariance of $y_t$ is given by:

$$\gamma_j = E[(y_t)(y_{t-j})] - \cancelto{0}{E(y_t)E(y_{t-j})} = E\left[\left(\sum_{i=1}^{t} \upsilon_i\right)\left(\sum_{i=1}^{t-j} \upsilon_i\right)\right] = \sum_{i=1}^{t-j} \sigma_\upsilon^2 = (t-j)\cdot\sigma_\upsilon^2 \ \ , \forall j \geq 0$$

where setting $j = 0$ gives us the variance of the process. Obviously we obtain a similar representation for $x_t$. Therefore the variances $y_t$ and $x_t$ depend on time $t$. This implies the process is not covariance-stationary since we have violation of the condition that they not be time dependent.[6] Finally, the process is not ergodic for the mean as the above random processes, by having time dependent autocovariances, fail to satisfy the condition of absolute summability ($\sum_{j=0}^{\infty}|\gamma_j| < \infty$) for any $t$, even $t = 1$, as the sum equals $\infty$.

Given the above analysis we do not conduct a HAC correction of standard errors in the Monte Carlo simulation of part B for two reasons:

[6] This is due to the fact that for covariance-stationarity we must have finite second moments: $E(y_t y_t) < \infty$. As $t \to \infty$ we have that $E(y_t y_t) \to \infty$, clearly a violation.

- Even though the Ljung-Box Q test rejects the null of no autocorrelation consistantly, strictly speaking the use of this test in this present context is invalid because the condition of "own" conditional homoskedasticity is violated (see page 142 in Hayashi).[7] This is a result of the variance of the error term being a function of time, as derived below in the generalized situation of serial covariance:

$$E(\varepsilon_t \varepsilon_{t-j}) = (t-j)\sigma_y^2 + \beta_1^2 (t-j)\sigma_x^2 = (t-j)(1+\beta_1^2) \tag{7}$$

  with the variance is given by $j = 0$. Thus our error term is heteroskedastic.

- We have ex-ante no reason to believe that serial correlations will die out over time, as the DGPs of the two random walks display peristant autocorrelations. This effect is carried into our specified model as was shown in (7). Given the discussion in the previous paragraph, these serial covariances are not abosolutely summable. Therefore, the error term of the regression is also non-stationary.

## Part D: Violations of Assumptions of the Classical Linear Model

The following is the list of the assumptions of the classical linear model, as pulled from Hayashi, page 109, 110:

- Linearity
- Ergodic Stationarity
- Predetermined Regressors (weak exogeneity)
- Rank Condition (nonsingular, and thus invertible, $E(X'X)$ matrix)
- $\varepsilon_t$ is a martingale difference sequence with finite second moments

Given what we have found in part C, we have violation of assumption 2. It is this violation that caused us issues, since the violation of stationarity implies that the autocovariates of the regressors are not going to 0 quickly enough so that we get $\frac{1}{n}(X'X) \not\to \Sigma_{XX}$. This matrix not being stationary affects our estimation of the variance-covariance matrix, and thus our spurious results.

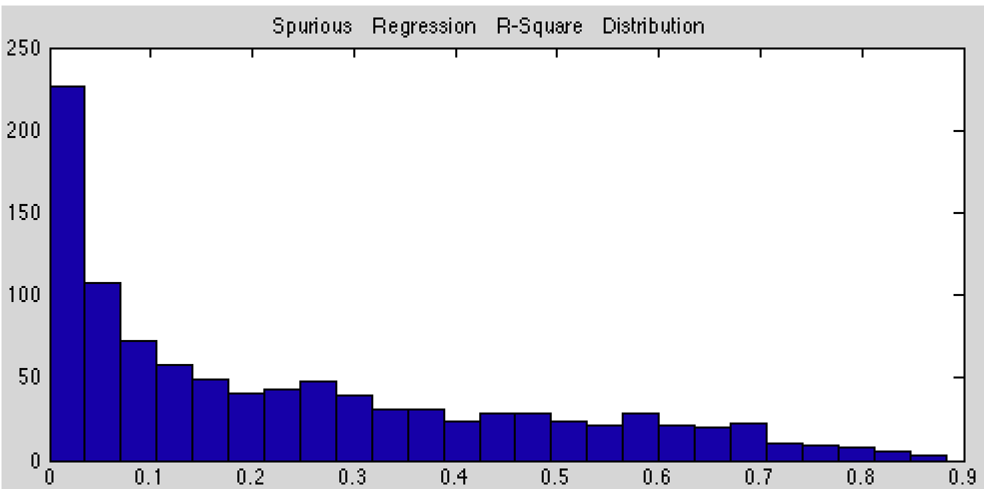
[7] "Own" conditional homoskedasticity is defined as $E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, ...) = \sigma^2 > 0$.

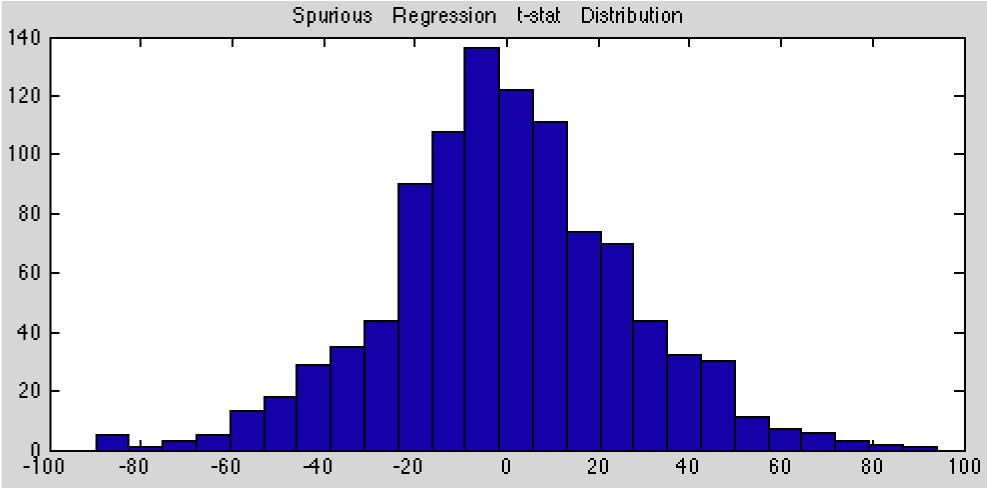
# Appendix: Results from Problem 5 Part A

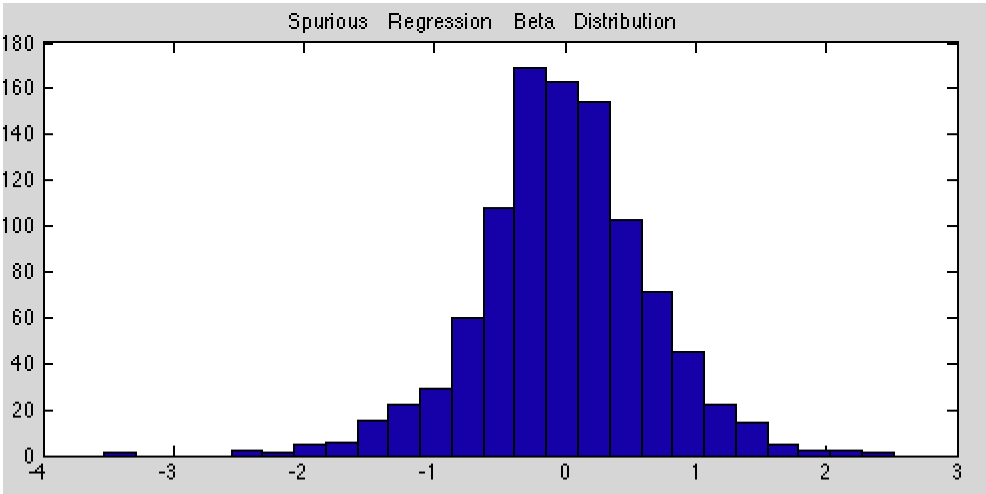
The results of the Monte Carlo Experiment:

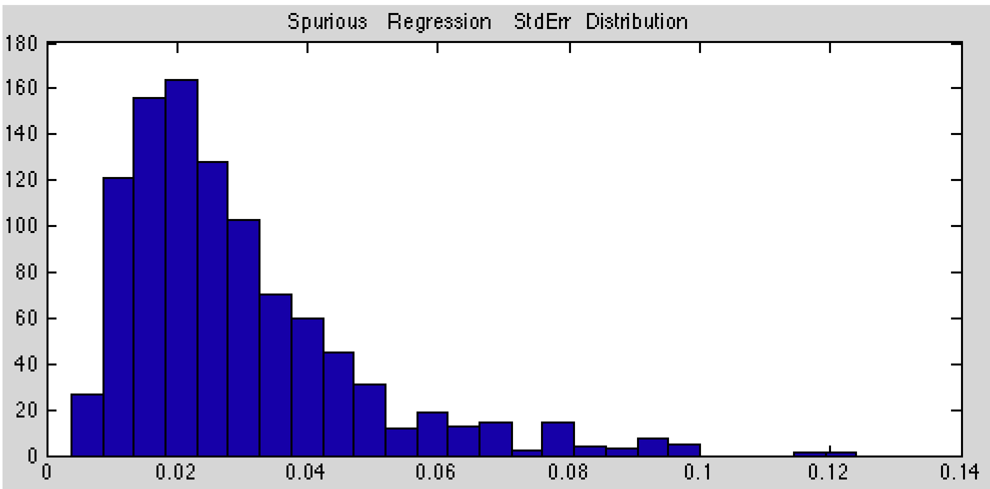
```
Realization from one time series, simulation # 79

R-squared       = 0.466
LBQ p-value     = 0.000

LBQ Test Result: Reject null => "There is AC in error term".

*****************************************************************

    Variable         Coef.     Std. Err.       t-stat.      p-value
    Constant        -7.159         0.304       -23.531        0.000
        Beta         0.879         0.023        39.024        0.000

*****************************************************************

Test the following hypothesis:

Ho: Beta = 0
H1: Beta not = 0

Reject the null: Beta is significantly different from 0.

*****************************************************************

Results from the Monte Carlo Simulation

The mean of beta is -0.018  with a standard error of 0.658

The mean of our t-statistic is -0.066  with a standard deviation of 26.359

The median of the standard error of beta is 0.024

The median of Rsq is 0.168 with a mean of 0.239

MC Result: Percentage of Times Fail to Reject null:    6.80 %
MC Result: Percentage of Times Null Rejected:         93.20 %

The average size of the test is 0.026 and the median size of the test is 0.000
```

Histograms from the Monte Carlo Experiment:




Spurious Regression t-stat Distribution
140
120
100
80
60
40
20
0
-100
-80
-60
-40
-20
0
20
40
60
80
100


Spurious Regression Beta Distribution
180
160
140
120
100
80
60
40
20
0
-4
-3
-2
-1
0
1
2
3


Spurious Regression StdErr Distribution
180
160
140
120
100
80
60
40
20
0
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14