Econ 589 Econometrics: Problem set 1

September 26, 2019

1 Question One.

2 Question Two.

3 Question Three.

We consider the data generation process

$$y_i = \theta_0 + \log u_i, \tag{1}$$

where u_i is distributed according to a standard half-normal distribution. It is useful to note that $u_i \sim |Z|$ where $Z \sim N(0,1)$. To see this, note that for $u \geq 0$, we have

$$
\begin{aligned}
\mathbb{P}(u_i \leq u) &= \int_0^u f_u(t)dt \\
&= 2\int_0^u \phi(t)dt \\
&= 2\left(\int_{-\infty}^u \phi(t)dt - \int_{-\infty}^0 \phi(t)dt\right) \\
&= 2\left(\Phi(u) - \frac{1}{2}\right) \\
&= 2\Phi(u) - 1,
\end{aligned}
\tag{2}
$$

and

$$
\begin{aligned}
\mathbb{P}(|z_i| \leq u) &= \mathbb{P}(z_i \in [-u, u]) \\
&= \int_{-\infty}^u \phi(t)dt - \int_{-\infty}^{-u} \phi(t)dt \\
&= \Phi(t) - \Phi(-t) \\
&= \Phi(t) - (1 - \Phi(t)) \\
&= 2\Phi(t) - 1 \\
&= \mathbb{P}(u_i \leq u),
\end{aligned}
\tag{3}
$$

where the penultimate equality comes from the symmetry of the standard normal distribution.

We suppose that y_i is observed and u_i is not.

As directed, we write

$$u_i = \exp\left(y_i - \theta\right), \tag{4}$$

and define the log-likelihood

$$\hat{\theta} = \mathrm{argmax}_{\theta \in \Theta} \sum_{i \leq n} \log f_u\left(\exp\left(y_i - \theta\right)\right). \tag{5}$$

To investigate the properties of this estimator, we generated a data set with 1000 points. What we found is that the algorithm returned whatever the largest value of the range of values in the search was. This is because the objective of the MLE in this case is an increasing function of θ. To see this, differentiate the objective with respect to θ to get

$$\begin{aligned} l'(\theta) &= \sum_{i \leq n} \frac{1}{f_u(\exp(y_i - \theta))} f_u'(\exp(y_i - \theta))(-\exp(y_i - \theta)) \\ &= \sum_{i \leq n} \frac{-\phi'(\exp(y_i - \theta))}{\phi(\exp(y_i - \theta))} \exp(y_i - \theta) \\ &> 0, \end{aligned} \tag{6}$$

where the inequality comes from noting that $\phi' > 0$ for positive values (and all $\exp(y_i - \theta)$ are positive), $\phi > 0$ for all values, and $\exp > 0$ for all values. Since $l' > 0$, l is increasing, so the maximum is attained at the largest value of Θ, or, more accurately, the maximum is obtained at $\max \Theta$ if this exists. In the case of unrestricted Θ, we find that the algorithm will return the supremum. If we assume $\Theta = \mathbb{R}$ then, the MLE is ∞. Because the sample average of the objective is always increasing, the estimator is inconsistent.

To find a better estimator, we turn to GMM. We consider the moment condition

$$\mathbb{E}[\exp(y_i - \theta_0) - \mathbb{E}[u_i]] = 0, \tag{7}$$

which is true since

$$\mathbb{E}[\exp(y_i - \theta_0) - \mathbb{E}[u_i]] = \mathbb{E}[\exp(y_i - \theta_0)] - \mathbb{E}[u_i] = \mathbb{E}[u_i] - \mathbb{E}[u_i] = 0. \tag{8}$$

Identification comes from noting that

$$\begin{aligned} \mathbb{E}[\exp(y_i - \theta)] - \mathbb{E}[u_i] &= \mathbb{E}[\exp(y_i - \theta_0 - (\theta - \theta_0))] - \mathbb{E}[u_i] \\ &= e^{-(\theta - \theta_0)} \mathbb{E}[\exp(y_i - \theta_0)] - \mathbb{E}[u_i] \\ &= e^{-(\theta - \theta_0)} \mathbb{E}[u_i] - \mathbb{E}[u_i] \\ &= (e^{-(\theta - \theta_0)} - 1) \mathbb{E}[u_i]. \end{aligned} \tag{9}$$

Since u_i is positive on a set of positive measure and negative nowhere, it must be the case that it has positive expectation, and hence the above expression is 0

if and only if $e^{-(\theta-\theta_0)} = 1$, which, since the exponential function is increasing, occurs if and only if $\theta = \theta_0$, and hence we have identification.

Since the expression inside the expectation operator is continuous, we have the desired features to ensure that the GMM estimator is consistent (if we assume a compact Θ, which isn't so limiting since we can usually assume for economic reasons that the parameter will belong in some very large compact interval with a 'guessed' upper and lower bound). Since the GMM estimator is consistent, it is significantly better than the MLE estimator.

4 Question Four.

5 Question Five.

We consider data generated by

$$y_i = \theta_0 x_i + \theta_0^2 a_i + u_i, \tag{10}$$

where it is known that x_i and u_i are independent of u_i, and that the mean of u_i is 0.

The true value of θ_0 is 1, the means of x_i and a_i are 0, the variances are 1, and the correlation is 0.5; these facts are all considered unknown.

5.1 Part (a).

We wish to construct a moment condition such that θ_0 is identified. We do this by noting that the independence of $w_i = (x_i, a_i)$ and u_i gives us

$$\mathbb{E}[\Gamma(w_i)u_i] = \mathbb{E}[\Gamma(w_i)]\mathbb{E}[u_i] = 0, \tag{11}$$

where Γ is some chosen instrument. This moment condition is then

$$\mathbb{E}[\Gamma(w_i)(y_i - \theta_0 x_i - \theta_0^2 a_i)] = 0. \tag{12}$$

If we write $\Gamma(w) = (\Gamma_1(w), \cdots, \Gamma_m(w))'$, then equation (12) is a system of m quadratic equations, which for generic θ are

$$\theta^2 \mathbb{E}[\Gamma_k(w_i)a_i] + \theta \mathbb{E}[\Gamma_k(w_i)x_i] - \mathbb{E}[\Gamma_k(w_i)y_i] = 0. \tag{13}$$

We know from our moment condition given by equation (12) that θ_0 satisfies this condition for each k. We wish this to be unique in doing so.

Note that each equation in the system must have either 1 or 2 solutions. If any equation has 1 solution, then we are done, since $\theta = \theta_0$ must be this unique solution and hence uniquely solves the system. Suppose then that all of the equations in the system have two solutions. Again, by our moment condition it must be the case that for each of these, one of the two solutions is θ_0. Therefore we have a set of m quadratic equations all of which share a common root θ_0. Because we assume that all of the equations have two roots, we are assuming

that the coefficient on θ^2 is non-zero in every equation, and hence we can rewrite our moment conditions as a set of m monic quadratic equations which each are completely characterised by the root that is not θ_0. All that is required is that one pair of these quadratic functions have a root that is not shared by them, and θ_0 will be identified.

Therefore, given Γ, all that is needed to show identification is to show that for some j, l, we have that

$$\left(\frac{\mathbb{E}[\Gamma_j(w_i)x_i]}{\mathbb{E}[\Gamma_j(w_i)a_i]} \quad -\frac{\mathbb{E}[\Gamma_j(w_i)y_i]}{\mathbb{E}[\Gamma_j(w_i)a_i]} \right) \neq \left(\frac{\mathbb{E}[\Gamma_l(w_i)x_i]}{\mathbb{E}[\Gamma_l(w_i)a_i]} \quad -\frac{\mathbb{E}[\Gamma_l(w_i)y_i]}{\mathbb{E}[\Gamma_l(w_i)a_i]} \right), \tag{14}$$

which is simply the claim that the monic polynomials earlier described have distinct coefficients and hence a distinct pair of roots; or show that one quadratic in the system is of the form $(\theta - \theta_r)^2$ (and hence has only one root); or show that one of the equations is linear, that is, that there exists some k for which $\mathbb{E}[\Gamma_k(w_i)a_i] = 0$. We will return to these conditions in the next part, but it suffices to say here that the likelihood of choosing an instrument such that the equations corresponding to each component are all the same equations is particularly low.

5.2 Part (b).

We wish to select the optimal instrument. Here we rely on our insight for GMM estimators more generally that the instrument that yields the efficient asymptotic variance is given by

$$\Gamma(w_i) = B(w_i)Q^{-1}(w_i), \tag{15}$$

where $B(w) = \mathbb{E}[\partial_\theta b(\theta_0)'|w_i = w]$, with $b(\theta) = y - \theta x_i - \theta^2 a_i$, and $Q(w) = \text{Var}(b(\theta_0)|w_i = w)$. We will calculate this instrument, and then check to verify that it identifies θ_0 using the conditions discussed in the previous part.

$$\begin{aligned} B(w) &= \mathbb{E}\left[\partial_\theta(y_i - \theta x_i - \theta^2 a_i)(\theta_0)|w_i = w \right] \\ &= \mathbb{E}[-x_i - 2\theta_0 a_i|w_i = w] \\ &= -x - 2\theta_0 a. \end{aligned} \tag{16}$$

For Q, note that the mean of $b(\theta_0)$ is 0 as it is

$$\mathbb{E}[u_i|w_i = w] = \mathbb{E}[u_i] = 0, \tag{17}$$

where the first equality comes from noting the independence of u_i and w_i, and the second equality comes from noting that the mean of u_i is 0. Therefore

$$\begin{aligned} Q(w) &= \mathbb{E}\left[(y_i - \theta_0 x_i - \theta_0^2 a_i)^2|w_i = w \right] \\ &= \mathbb{E}[u_i^2|w_i = w] \\ &= \sigma^2, \end{aligned} \tag{18}$$

where we have defined σ^2 to be the variance of the errors, which we have implicitly assumed by previous assumptions on u_i satisfy homoskedasticity. Since the moment condition is set to 0, we can therefore drop the σ^2 so that our optimal instrument is

$$\Gamma(w) = x + 2\theta_0 a. \tag{19}$$

Our moment condition is now

$$\mathbb{E}[(x_i + 2\theta_0 a_i)(y_i - \theta_0 x_i - \theta_0^2 a_i)] = 0. \tag{20}$$

This of course does not provide identification, since the instrument is a 1 dimensional quadratic (or cubic, depending on whether the θ_0 in the instrument is taken as a variable rather than a constant). We therefore plot the data to determine that the positive root is that which should be taken.

5.3   Part (c).

We now want to compare the performance of three related estimators, defined as follows. $\hat{\theta}_1$ and $\hat{\gamma}$ solve

$$\frac{1}{n} \sum_{i \leq n} \begin{pmatrix} x_i \\ a_i \end{pmatrix} (y_i - \hat{\theta}_1 x_i - \hat{\gamma} a_i) = 0. \tag{21}$$

Note that this is two linear equations in two variables. We solve this to get

$$\hat{\theta}_1 = \frac{\bar{a y} - \bar{a^2} \bar{x y}}{\bar{a x^2} - \bar{a^2} \bar{x^2}}, \tag{22}$$

where $\bar{\cdot}$ denotes the sample average of $\cdot$, so that, for example,

$$\bar{a y} = \frac{1}{n} \sum_{i \leq n} a_i y_i. \tag{23}$$

The estimator $\hat{\theta}_2$ solves

$$\frac{1}{n} \sum_{i \leq n} z_i(\hat{\theta}_2)(y_i - \hat{\theta}_2 x_i - \hat{\theta}_2^2 a_i) = 0, \tag{24}$$

where z_i is the optimal instrument from part (b). This is a cubic in $\hat{\theta}_2$, but of course, we want only one estimator. We therefore choose the largest root.

Finally, the estimator $\hat{\theta}_3$ is a two-stage estimator, that uses $\hat{\theta}_1$ to estimate the optimal instrument, so that $\hat{\theta}_3$ solves

$$\frac{1}{n} \sum_{i \leq n} z_i(\hat{\theta}_1)(y_i - \hat{\theta}_3 x_i - \hat{\theta}_3^2 a_i) = 0. \tag{25}$$

This is quadratic in our parameter, so we take the positive root, unless that proves to be badly behaved.

What we find when we run the code is that, regardless of variance, $\hat{\theta}_2$ always tracks $\hat{\theta}_3$ incredibly tightly, lying slightly above to an incredibly small degree, such that the resulting difference appears to be very stochastic and always positive. Further, we find that these two estimators are significantly better estimators than $\hat{\theta}_1$ in convergence rate.

The latter conclusion is hardly surprising, as these estimators are either the optimal estimator of this type or a two stage approximation of that estimator. We therefore turn our attention to the high degree of correlation between $\hat{\theta}_2$ and $\hat{\theta}_3$.
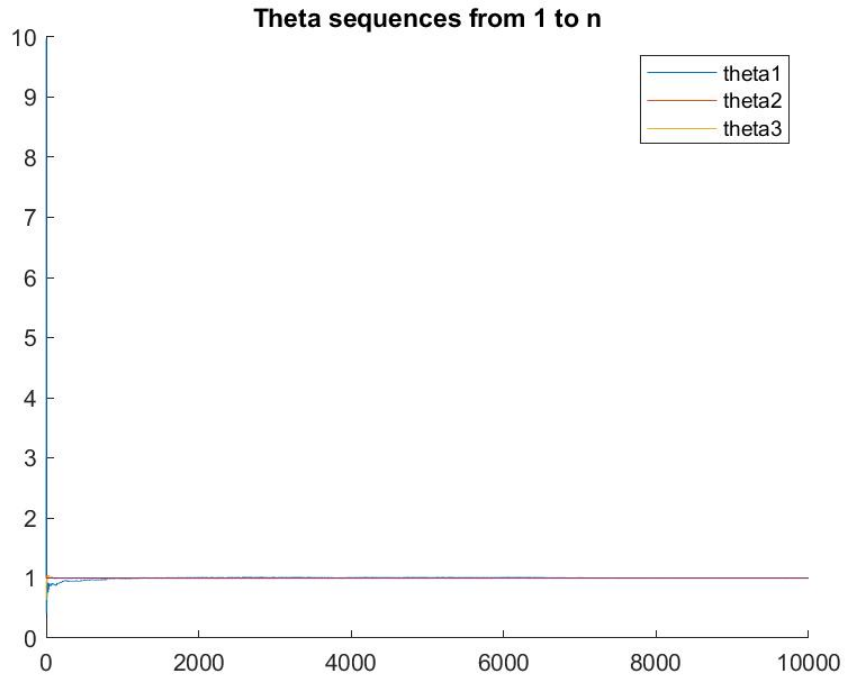
Figure 1: The estimators over a run of 10000 data points.

6 Question Six.

7 Question Seven.

We have two distributions, $l_i = \exp(Z_l)$, where $Z_l \sim N(\mu_l, \sigma_l^2)$ and $b_i = \exp(Z_b)$ where $Z_b \sim N(\mu_b, \sigma_b^2)$, and we are interested in finding $\lambda_0 = \mathbb{E}[l_i]$, $\beta_0 = \mathbb{E}[b_i]$, and $\mathrm{Var}(l_i) = \mathrm{Var}(b_i) = \sigma_0^2$, when we observe $a_i = l_i b_i$. We proceed by using MLE.
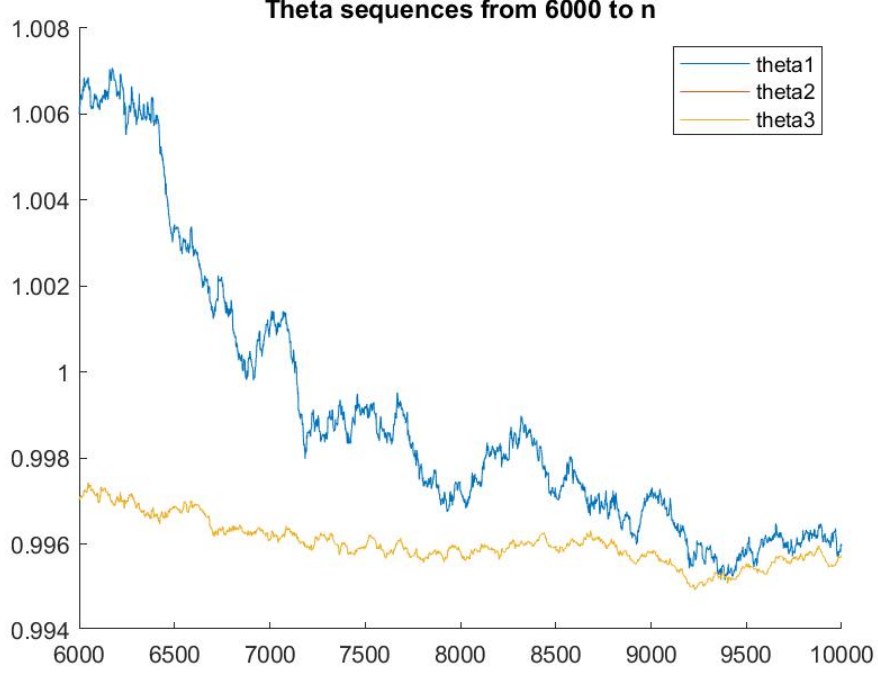
Figure 2: The estimators over a run of 10000 data points, from the 6000th point onwards.

7.1 Part (a).

We wish to find the loglikelihood function for this problem. First we need to find the distribution of a_i. We let Z_i denote independent standard normal distributions.

$$
\begin{aligned}
a_i = l_i b_i &= \exp\left(\sigma_l Z_1 + \mu_l\right) \exp\left(\sigma_b Z_2 + \mu_b\right) \\
&= \exp\left(\sigma_l Z_1 + \sigma_b Z_2 + (\mu_l + \mu_b)\right).
\end{aligned}
\tag{26}
$$

We know that

$$
\sigma_l Z_1 + \sigma_b Z_2 + (\mu_l + \mu_b) \sim N(\mu, \sigma_a^2)
\tag{27}
$$

where

$$
\mu = \mu_l + \mu_b, \quad \sigma_a^2 = \sigma_l^2 + \sigma_b^2,
\tag{28}
$$

and hence

$$
a_i \sim \exp(N(\mu, \sigma_a^2)).
\tag{29}
$$

That is, a_i is log-normal with parameters $\mu = \mu_l + \mu_b$ and $\sigma_a^2 = \sigma_l^2 + \sigma_b^2$. We write $a_i = \exp(Z_a)$ where $Z_a \sim N(\mu, \sigma_a^2)$.
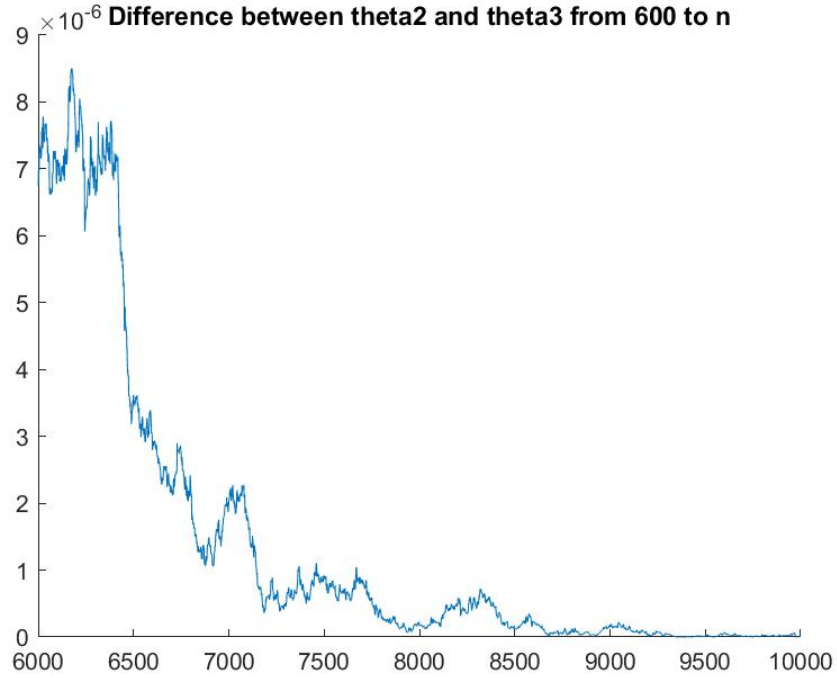
Figure 3: The difference between $\hat{\theta}_2$ and $\hat{\theta}_3$, with the latter always slightly lower than the former, but to an incredibly small degree, indicating an incredibly high degree of correlation.

As such, we now wish to find the likelihood distribution for i.i.d. lognormally distributed data. We denote the cdf and pdf by F and f respectively, and find that

$$
\begin{aligned}
F(a; \mu, \sigma_a^2) &= \mathbb{P}(a_i \leq a) = \mathbb{P}(\exp(Z_a) \leq a) \\
&= \mathbb{P}(Z_a \leq \log a) = \mathbb{P}(\sigma_a Z_3 + \mu \leq \log a) \\
&= \mathbb{P}\left(Z_3 \leq \frac{\log a - \mu}{\sigma_a} \right) \\
&= \Phi\left(\frac{\log a - \mu}{\sigma_a} \right),
\end{aligned}
\tag{30}
$$

so that differentiating with respect to a gives

$$
f(a; \mu, \sigma_a^2) = \phi\left(\frac{\log a - \mu}{\sigma_a} \right) \cdot \frac{1}{\sigma_a a}.
\tag{31}
$$

8

Our likelihood function L is therefore

$$L(\mu, \sigma_a^2) = \prod_{i \leq n} \phi\left(\frac{\log a_i - \mu}{\sigma_a}\right) \cdot \frac{1}{\sigma_a a_i}, \tag{32}$$

which implies that our log-likelihood function is

$$l(\mu, \sigma_a^2) = \sum_{i \leq n} \left(\log \phi\left(\frac{\log a_i - \mu}{\sigma_a}\right) - \log \sigma_a - \log a_i\right). \tag{33}$$

Since we are maximising μ, σ_a^2, the last term here is a constant, so we ignore this and divide through by n to get the sample objective

$$\hat{Q}(\mu, \sigma_a^2) = \frac{1}{n} \sum_{i \leq n} \left(\log \phi\left(\frac{\log a_i - \mu}{\sigma_a}\right) - \log \sigma_a\right). \tag{34}$$

Using our expression for μ and σ_a^2 in terms of μ_l, μ_b, σ_l^2, and σ_b^2, we get

$$\hat{Q}(\mu_l, \mu_b, \sigma_l^2 + \sigma_b^2) = \frac{1}{n} \sum_{i \leq n} \left(\log \phi\left(\frac{\log a_i - (\mu_l + \mu_b)}{\sqrt{\sigma_l^2 + \sigma_b^2}}\right) - \frac{1}{2}\log(\sigma_l^2 + \sigma_b^2)\right). \tag{35}$$

Finally, we note that we wish to find λ_0, β_0, and σ_0^2, so we find expressions for μ_l, μ_b, σ_l^2, and σ_b^2 in terms of these parameters.

Note that by properties of the log-normal distribution, we have

$$\mathbb{E}[l_i] = \lambda_0 = \exp\left(\mu_l + \frac{\sigma_l^2}{2}\right), \tag{36}$$

$$\mathrm{Var}(l_i) = \sigma_0^2 = (\exp(\sigma_l^2) - 1)(2\mu_l + \sigma_l^2). \tag{37}$$

We use equation (36) to express σ_l^2 in terms of μ_l to get

$$\sigma_l^2 = 2(\log \lambda_0 - \mu_l). \tag{38}$$

Substituting this into the equation given by the variance and rearranging, we find that

$$\mu_l = \log \lambda_0 - \frac{1}{2}\log\left(\frac{\sigma_0^2}{\exp(2\log \lambda_0)} + 1\right), \tag{39}$$

and

$$\sigma_l^2 = \log\left(\frac{\sigma_0^2}{\exp(2\log \lambda_0)} + 1\right). \tag{40}$$

Similarly, for b we have

$$\mu_b = \log \beta_0 - \frac{1}{2}\log\left(\frac{\sigma_0^2}{\exp(2\log \beta_0)} + 1\right), \tag{41}$$

and

$$\sigma_b^2 = \log\left(\frac{\sigma_0^2}{\exp(2\log \beta_0)} + 1\right). \tag{42}$$

9

We therefore have the sample objective function

$$\hat{Q}(\lambda, \beta, \sigma) = \frac{1}{n} \sum_{i \le n} \log \phi \left(\frac{\log a_i - (\log \lambda + \log \beta) + \frac{1}{2} \left(\log \left(\frac{\sigma^2}{\exp(2 \log \lambda)} + 1 \right) + \log \left(\frac{\sigma^2}{\exp(2 \log \beta)} + 1 \right) \right)}{\sqrt{\log \left(\frac{\sigma^2}{\exp(2 \log \lambda)} + 1 \right) \log \left(\frac{\sigma^2}{\exp(2 \log \beta)} + 1 \right)}} \right)$$
$$- \frac{1}{2n} \sum_{i \le n} \log \left(\log \left(\frac{\sigma^2}{\exp(2 \log \lambda)} + 1 \right) + \log \left(\frac{\sigma^2}{\exp(2 \log \beta)} + 1 \right) \right).$$
$$(43)$$

7.2 Part (b).

Given a dataset, we now use our sample objective in order to estimate the desired means and variance. The estimate we get is $\lambda_0 = 206.3425$, $\beta_0 = 80.4814$, $\sigma_0 = 0.5691$.

I would, however, just send an RA with clearer instructions before buying a table cloth because I am not at all convinced that this code is robust.

7.3 Part (c).

We now fix $\lambda_0 = 200$, $\beta = 80$, and let the variance run through 0.5 to 50 in incerements of 0.5. We calculate the root mean squared error of the area against the variance. We subsequently then checked values from 0.5 to 2 in increments of 0.05. It gave us very little.

8 Question Eight.

We consider two random variables with joint distribution function

$$F(x, y) = \Phi(x)\Phi(y)(1 - 0.5\Phi(-x)\Phi(-y)). \qquad (44)$$

8.1 Part (a).

We first verify that this is in fact a distribution function. Note that as x and y tend to infinity, F tends to $1 \cdot 1(1 - 0.5 \cdot 0 \cdot 0) = 1$. As x and y tend to $-\infty$, $\Phi(x)$ tends to 0, so F tends to 0. The distribution is clearly continuous and hence is right-continuous. Finally, the term $\Phi(x)\Phi(y)$ is increasing in both x and y, and $(1 - 0.5\Phi(-x)\Phi(-y))$ is increasing in x and y, and hence F is increasing in x and y, and hence is a distribution function.

8.2 Part (b).

We now consider the marginal distributions. For this, we let y tend to ∞, and see that

$$F_X(x) = \lim_{y \to \infty} F(x, y) = \Phi(x). \qquad (45)$$

10
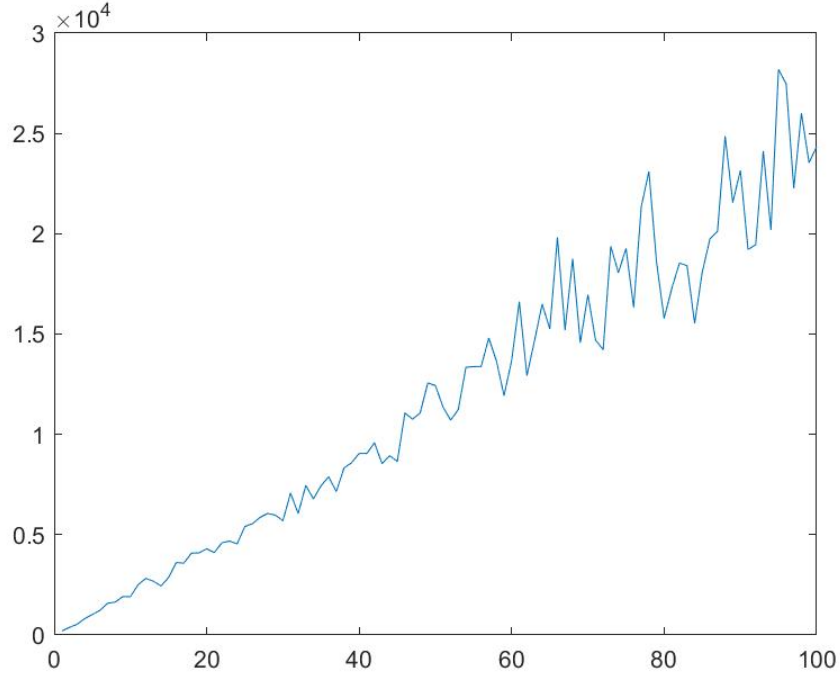
Similarly, then, $F_Y(y) = \Phi(y)$. That is, x and y are both distributed as standard normal random variables.

8.3   Part (c).

We wish to draw the joint density function of x, y, by which I mean, we wish to get matlab to do that, so we calculate the joint pdf f_{XY}:

$$
\begin{aligned}
f_{XY}(x,y) &= \frac{\partial^2 F}{\partial y \partial x} \\
&= \frac{\partial}{\partial y}\left(\phi(x)\Phi(y)(1 - 0.5\Phi(-x)\Phi(-y)) + 0.5\Phi(x)\Phi(y)\Phi(-y)\phi(-x)\right) \\
&= \phi(x)((\phi(y)(1 - 0.5\Phi(-x)\Phi(-y)) + 0.5\Phi(y)\Phi(-x)\phi(-y))) \\
&\quad + 0.5\Phi(x)\phi(y)\phi(-x)\Phi(-y) - 0.5\Phi(x)\Phi(y)\phi(-y)\phi(-x) \\
&= \phi(x)\phi(y)(1 - 0.5\Phi(-x)\Phi(-y)) + 0.5\phi(x)\phi(-y)\Phi(-x)\Phi(y) \\
&\quad + 0.5\Phi(x)\phi(y)\phi(-x)\Phi(-y) - 0.5\phi(-x)\phi(-y)\Phi(x)\Phi(y)
\end{aligned}
\tag{46}
$$

This produces the following plot.

11
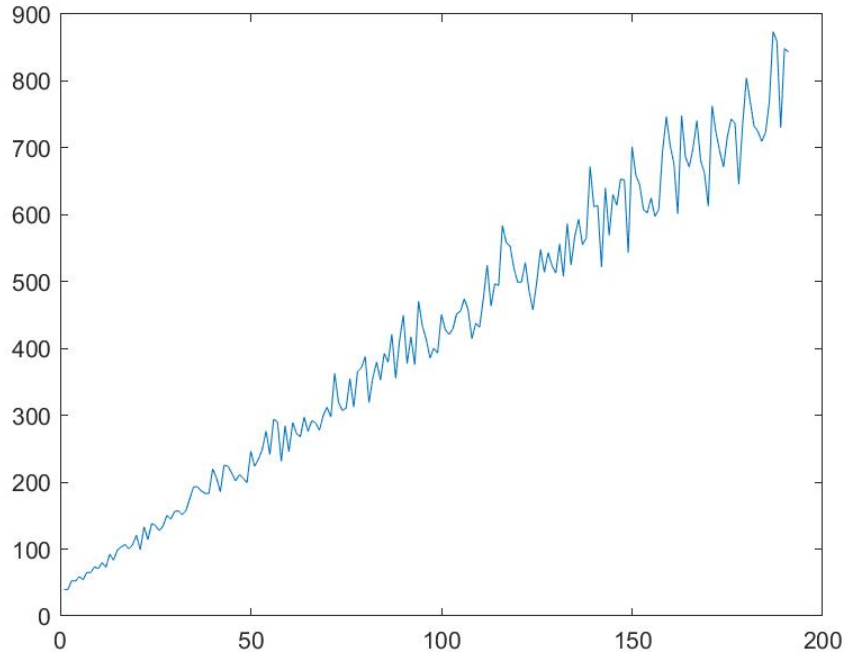
8.4   Part (d).

We wish to plot various distributions of x conditional on y. For this, we note that

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}. \tag{47}$$

Here of course, $f_Y(y) = \phi(y)$, from part (b). We plot this for $y = 0, \pm 1, \pm 0.5$ below

8.5   Part (f).

We repeat the exercise with the random variables now given by the distribution

$$f_{XY} = c\phi(x)\phi(y)(1 + |x||y|\mathbb{1}\{\max\{|x|,|y|\} < 1\}), \tag{48}$$
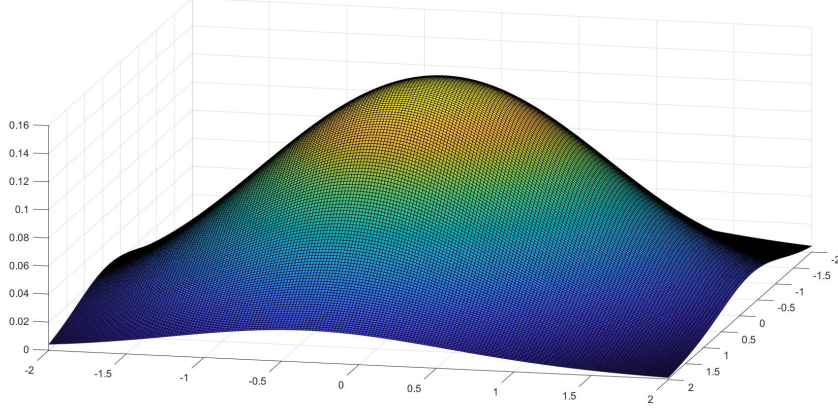
where c is a normalisation constant.

Figure 4: The joint distribution of x and y plot over $[-2, 2] \times [-2, 2]$ with grid points spaced 0.02 apart.

8.5.1 Part (a).

Clearly f is non-negative everywhere. Further, it is clear that the integral is bounded:

$$
\begin{aligned}
\int f_{XY}(x, y)dxdy &= c \int \phi(x)\phi(y)dxdy + c \int_{-1}^{1} \int_{-1}^{1} |xy|\phi(x)\phi(y)dxdy \\
&= c + c \int_{-1}^{1} \int_{-1}^{1} |xy|\phi(x)\phi(y)dxdy
\end{aligned}
\tag{49}
$$

This integral is of a continuous function on a compact space, and hence is finite. We can therefore choose c such that the integral of the density over $\mathbb{R}^2$ is 1, and hence f_{XY} characterises a distribution.

8.5.2 Part (b).

We now wish to consider the marginal distributions, which we do by integrating our density over $y \in (-\infty, \infty)$.

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y)dy \\
&= c\phi(x) \left(\int_{-\infty}^{\infty} \phi(y)(1 + |x||y|\mathbb{1}\{\max\{|x|, |y|\} < 1\})dy \right) \\
&= c\phi(x) \left(1 + |x|\mathbb{1}\{|x| < 1\} \int_{-1}^{1} |y|\phi(y)dy \right) \\
&= c\phi(x) \left(1 + 2|x|\mathbb{1}\{|x| < 1\} \int_{0}^{1} y\phi(y)dy \right)
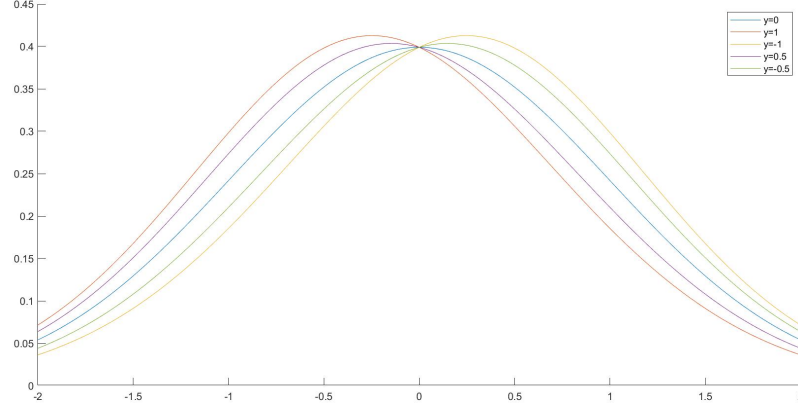\end{aligned}
\tag{50}
$$

13

Figure 5: The conditional distribution of x and y plot over $[-2, 2]$ with grid points spaced 0.02 apart, for $y = 0, \pm 1, \pm 0.5$.

We handle this last integral using integration by parts, so that

$$
\begin{aligned}
\int_0^1 y\phi(y)dy &= [y\Phi(y)]_0^1 - \int_0^1 \Phi(y)dy \\
&= \Phi(1) - \left[y^2\Phi(y) + \phi(y)\right]_0^1 \\
&= \Phi(1) - \Phi(1) - \phi(1) + \phi(0) \\
&= \phi(0) - \phi(1) \\
&= \frac{1}{\sqrt{2\pi}}(1 - e^{-\frac{1}{2}}).
\end{aligned}
\tag{51}
$$

We therefore have that

$$
f_X(x) = c\phi(x)\left(1 + \frac{\sqrt{2}|x|}{\sqrt{\pi}}\mathbb{1}\{|x| < 1\}(1 - e^{-\frac{1}{2}})\right).
\tag{52}
$$

By symmetry, we have

$$
f_Y(y) = c\phi(y)\left(1 + \frac{\sqrt{2}|y|}{\sqrt{\pi}}\mathbb{1}\{|x| < 1\}(1 - e^{-\frac{1}{2}})\right).
\tag{53}
$$

Given the above calculation, we can now in fact find the value of c.

$$
\begin{aligned}
\int f_{XY}(x, y)dxdy &= c\left(1 + 4\left(\int_0^1 x\phi(x)dx\right)^2\right) \\
&= c\left(1 + \frac{2}{\pi}(1 - e^{-\frac{1}{2}})^2\right) \\
&= 1,
\end{aligned}
\tag{54}
$$

14

where the last equality comes from noting that c normalises the distribution, and hence

$$c = \left(1 + \frac{2}{\pi}(1 - e^{-\frac{1}{2}})^2\right)^{-1}. \tag{55}$$
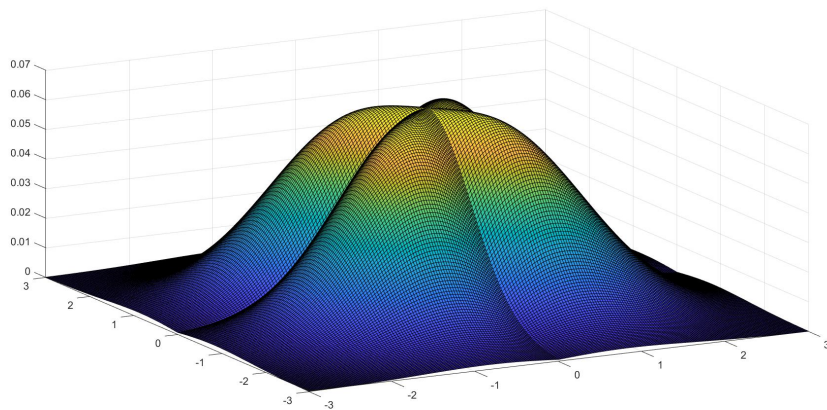
8.5.3 Part (c).

We now plot our pdf.

Figure 6: The joint distribution of x and y plot over $[-3, 3] \times [-3, 3]$ with grid points spaced 0.03 apart.

8.5.4 Part (d).

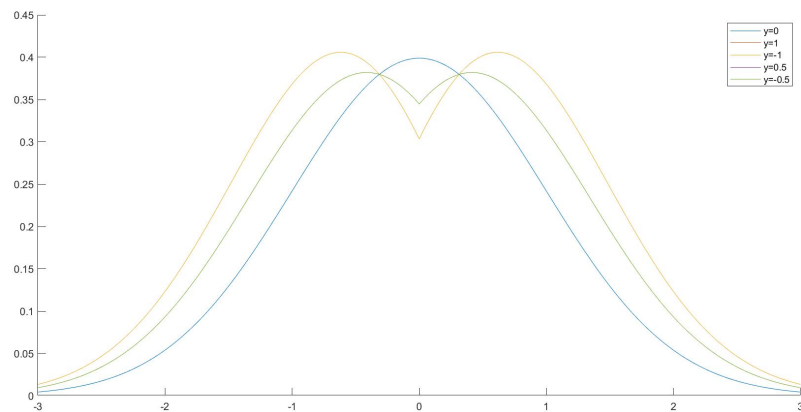We now plot the conditional distributions, given $y = 0, \pm 1, \pm 0.5$.

Figure 7: The conditional distribution of x and y plot over $[-3, 3]$ with grid points spaced 0.03 apart, for $y = 0, \pm 1, \pm 0.5$.