ASSIGNMENT 3

ETHAN LIGON

This will be the last "regular" assignment for ARE212. As with the previous assignment, you are strongly encouraged to work as a team, and to turn in a single assignment for grading. The principal deliverable you turn in should be a link to a `github` repository, and you should organize your teams so as to provide constructive criticism to other teams.

1. Exercises (GMM)

When we approach a new estimation problem from a GMM perspective there's a simple set of steps we can follow.

- Describe the parameter space B;
- Describe a function $g_j(b)$ such that $\mathbb{E}g_j(\beta) = 0$;
- Describe an estimator for the covariance matrix $\mathbb{E}g_j(\beta)g_j(\beta)^\top$.

(1) Explain how the steps outlined above can be used to construct an optimally weighted GMM estimator.

(2) Consider the following models. For each, provide a causal diagram; construct the optimally weighted GMM estimator of the unknown parameters (various Greek letters); and give an estimator for the covariance matrix of your estimates. If any additional assumptions are required for your estimator to be identified please provide these.

(a) $\mathbb{E}y = \mu$; $\mathbb{E}(y - \mu)^2 = \sigma^2$; $\mathbb{E}(y - \mu)^3 = 0$.

(b) $y = \alpha + X\beta + u$; with $\mathbb{E}(X^\top u) = \mathbb{E}u = 0$.

(c) $y = \alpha + X\beta + u$; with $\mathbb{E}(X^\top u) = \mathbb{E}u = 0$, and $\mathbb{E}(u^2) = \sigma^2$.

(d) $y = \alpha + X\beta + u$; with $\mathbb{E}(X^\top u) = \mathbb{E}u = 0$, and $\mathbb{E}(u^2) = e^{X\sigma}$.

(e) $y = \alpha + X\beta + u$; with $\mathbb{E}(Z^\top u) = \mathbb{E}u = 0$ and $\mathbb{E}Z^\top X = Q$.

(f) $y = f(X\beta) + u$; with f a known scalar function and with $\mathbb{E}(Z^\top u) = \mathbb{E}u = 0$ and $\mathbb{E}Z^\top X f'(X\beta) = Q(\beta)$. (Bonus question: where does this last restriction come from, and what role does it play?)

Date: Due April 29, 2024.

(g) $y = f(X, \beta) + u$; with f a known function and with $\mathbb{E}(Z^\top u) = \mathbb{E}u = 0$ and $\mathbb{E}Z^\top \frac{\partial f}{\partial \beta^\top}(X, \beta) = Q(\beta)$.

(h) $y^\gamma = \alpha + u$, with $y > 0$ and γ a scalar, and $\mathbb{E}(Z^\top u) = \mathbb{E}u = 0$ and $\mathbb{E}Z^\top \begin{bmatrix} \gamma y^{\gamma-1} \\ -1 \end{bmatrix} = Q(\gamma)$.

(3) For each of the models above write a data-generating process in `python`. Your function `dgp` should take as arguments a sample size `N` and a vector of "true" parameters `b0`, and return a dataset (y, X).

(4) Select the most interesting of the data generating processes you developed, and using the code in `gmm.py` or `GMM_class.py` (see https://github.com/ligonteaching/ARE212_Materials/) use data from your `dgp` to analyze the finite sample performance of the corresponding GMM estimator you've constructed. Of particular interest is the distribution of your estimator using a sample size N and how this distribution compares with the limiting distribution as $N \to \infty$.

2. Exercises (Cross-Validation)

Consider estimation of a linear model $y = X\beta + u$, with the identifying assumption that $\mathbb{E}(u|X) = 0$.

When we compute K-fold cross-validation of a tuning parameter λ (e.g., the penalty parameter in a LASSO regression), then for each value of λ we obtain K estimates of any given parameter, say β_i; denote the estimates of this parameter by $b_i^\cdot = (b_i^1, \dots, b_i^K)$. If our total sample (say D_1) comprises N iid observations, then each of our K estimates will be based on a sample D_1^k of roughly $N\frac{K-1}{K}$ observations.

(1) How can you use the estimates $b_i^\cdot$ to estimate the variance of the estimator?

(2) What can you say about the variance of your estimator of the variance? In particular, how does it vary with K?

(3) Suppose we use $\bar{b}(\lambda) = K^{-1}\sum_{k=1}^K b^k$ as our preferred estimate of β at a given value of the tuning parameter λ. Construct an R^2 statistic which maps a sample D and a parameter vector b into $[0, 1]$. Compare the following:

(a) $R^2(D_1, \bar{b}(\lambda))$ and $R^2(D_1, b_{OLS})$, where b_{OLS} denotes the OLS estimator estimated using the entire sample D_1, so that $R^2(D_1, b_{OLS})$ corresponds to the usual least-squares R^2 statistic.

(b) $R^2(D, \bar{b}(\lambda))$ and $R^2(D, b_{OLS})$, where b_{OLS} and $\bar{b}(\lambda)$ are estimated using D_1 as described above, but where D is some other iid sample from the same data-generating process.
(c) $K^{-1}\sum_{k=1}^{K} R^2(D_1^k, \bar{b}(\lambda))$ and $K^{-1}\sum_{k=1}^{K} R^2(D_1^k, b_{OLS})$;
(d) $K^{-1}\sum_{k=1}^{K} R^2(D_1^k, \bar{b}(\lambda))$ and $K^{-1}\sum_{k=1}^{K} R^2(D_1^k, b^k(\lambda))$;
(e) $R^2(D, \bar{b}(\lambda))$ and $R^2(D, \beta)$;
(f) $R^2(D, b_{OLS})$ and $R^2(D, \beta)$;

(4) How do the R^2 statistics you worked with above compare with various notions of mean-square error? The statistics which rely on β are typically infeasible, so setting these aside, how might you use these statistics to choose a "best" estimator?

3. Breusch-Pagan Extended

Consider a linear regression of the form

$$y = \alpha + \beta x + u, \tag{1}$$

with (y, x) both scalar random variables, where it is assumed that (a.i) $\mathbb{E}(u \cdot x) = \mathbb{E}u = 0$ and (a.ii) $\mathbb{E}(u^2|x) = \sigma^2$.

(1) The condition a.i is essentially untestable; explain why.
(2) Breusch and Pagan (1979) argue that one can test a.ii via an auxiliary regression $\hat{u}^2 = c+dx+e$, where the $\hat{u}$ are the residuals from the first regression, and the test of a.ii then becomes a test of $H_0 : d = 0$. Describe the logic of the test of a.ii.
(3) Use the two conditions a.i and a.ii to construct a GMM version of the Breusch-Pagan test.
(4) What can you say about the performance or relative merits of the Bruesch-Pagan test versus your GMM alternative?
(5) Suppose that in fact that x is distributed uniformly over the interval $[0, 2\pi]$, and $\mathbb{E}(u^2|x) = \sigma^2(x) = \sigma^2 \sin(2x)$, thus violating a.ii. What can you say about the performance of the Breusch-Pagan test in this circumstance? Can you modify your GMM test to provide a superior alternative?
(6) In the above, we've considered a test of a specific functional form for the variance of u. Suppose instead that we don't have any prior information regarding the form of $\mathbb{E}(u^2|x) = f(x)$. Discuss how you might go about constructing an extended version of the Breusch-Pagan test which tests for $f(x)$ non-constant.
(7) Show that you can use your ideas about estimating $f(x)$ to construct a more efficient estimator of β if $f(x)$ isn't constant.

Relate your estimator to the optimal generalized least squares (GLS) estimator.

4. Tests of Normality

Suppose we have a sample of iid observations $x_1, x_2, \dots, x_N$; we want to test whether these are drawn from a normal distribution. Note the fact that the integer central moments of the normal distribution satisfy

$$\begin{aligned} \mathrm{E}x &= \mu \\ \mathrm{E}(x-\mu)^m &= 0 \qquad m \text{ odd} \\ \mathrm{E}(x-\mu)^m &= \sigma^m (m-1)!! \qquad m \text{ even}, \end{aligned}$$

where $n!!$ is the double factorial, i.e., $n!! = n(n-2)(n-4)\dots$.

(1) Using the analogy principle, construct an estimator for the first k moments of the distribution of x. Use this to define a k-vector of moment restrictions $g_N(\mu, \sigma)$ satisfying $\mathrm{E}g_N(\mu, \sigma) = 0$ under the null hypothesis of normality.
(2) What is the covariance matrix of the sample moment restrictions (again under the null)? I.e., what can be said about $\mathrm{E}g_j(\mu,\sigma)g_j(\mu,\sigma)^\top - \mathrm{E}g_j(\mu,\sigma)\mathrm{E}g_j(\mu,\sigma)^\top$?
(3) Using your answers to the previous two questions, suggest a GMM-based test of the hypothesis of normality, taking $k > 2$.
(4) Implement the test you've devised using `python`. You may want to use `scipy.stats.distributions.chi2.cdf` and `scipy.optimize.minimize`.
(5) What can be said about the optimal choice of k?
(6) Compare the GMM estimates of (μ, σ) to the maximum likelihood estimates of these parameters. Do they differ? Why?

5. Logit

This problem is meant to help draw connections between GMM estimators and maximum likelihood estimators, with a particular focus on the 'logit' model.

The development of a maximum likelihood estimator typically begins with an assumption that some random variable has a (conditional) distribution which is known up a k-vector of parameters β. Consider the case in which we observe N independent realizations of a Bernoulli random variable Y, with $\Pr(Y = 1|X) = \sigma(\beta^\top X)$, and $\Pr(Y = 0|X) = 1 - \sigma(\beta^\top X)$.

(1) Show that under this model $\mathbb{E}(Y_i - \sigma(X\beta)|X) = 0$. Assume that σ is a known function, and use this fact to develop a GMM estimator of β. Is your estimator just- or over-identified?

(2) Show that the likelihood can be written as

$$L(\beta|y, X) = \prod_{i=1}^{N} \sigma(\beta^\top X_i)^{y_i} \left(1 - \sigma(\beta^\top X_i)\right)^{1-y_i}.$$

(3) To obtain the maximum likelihood estimator (MLE) one can chose b to maximize $\log L(b|y, X)$. When the likelihood is well-behaved, the MLE estimator satisfies the first order conditions (also called the "scores") from this maximization problem, in which case this is called a "type I" MLE. Let $\sigma(z) = \frac{1}{1+e^{-z}}$ (this is sometimes called the logistic function, or the sigmoid function), and obtain the scores $S_N(b)$ for this estimation problem. Show that $\mathbb{E}S_N(\beta) = 0$. Demonstrate that these moment conditions can serve as the basis for a GMM estimator of β, and compare this estimator to the GMM estimator you developed above. Which is more efficient, and why?

References

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, 1287–1294.