Διευκρινήσεις Okun

1 Paper Owyang & Sekhposyan, υπόδειγμα 7

Έστω

$$\begin{aligned}\Delta u_t &= \alpha + \beta \Delta y_t + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} \\ &+ \gamma_1 \Delta u_{t-1} + \gamma_2 \Delta u_{t-2} + \lambda_1 d_t + \lambda_2 d_t \Delta y_t + e_t \end{aligned} \tag{1}$$

Το υπόδειγμα (1) αντιστοιχεί στο (7) του άρθρου των Owyang & Sekhposyan.

Σημείωση: για στάσιμες διαδικασίες ισχύει $\ldots = E(x_{t-1}) = E(x_t) = E(x_{t+1}) = \ldots$

Οπότε, το υπόδειγμα υπονοεί (αναμενόμενη μέση μεταβολή ποσοστού ανεργίας: $E(\Delta u_t) = E(\Delta u)$ για κάθε t)

$$\begin{aligned}E(\Delta u) &= \alpha + \beta E(\Delta y) + \beta_1 E(\Delta y) + \beta_2 E(\Delta y) \\ &+ \gamma_1 E(\Delta u) + \gamma_2 E(\Delta u) + \lambda_1 d_t + \lambda_2 d_t E(\Delta u) \end{aligned} \tag{2}$$

Έστω $\gamma(1) = 1 - \gamma_1 - \gamma_2$ και $\beta(1) = \beta + \beta_1 + \beta_2$, τότε

$$E(\Delta u) = \frac{\alpha}{\gamma(1)} + \frac{\beta(1)}{\gamma(1)} E(\Delta y) + \frac{\lambda_1}{\gamma(1)} d_t + \frac{\lambda_2}{\gamma(1)} d_t E(\Delta y)$$

Σε περιόδους **μεγέθυνσης** ($d_t = 0$)

$$E(\Delta u | d_t = 0) = \frac{\alpha}{\gamma(1)} + \frac{\beta(1)}{\gamma(1)} E(\Delta y)$$

ενώ σε περιόδους **ύφεσης** ($d_t = 1$)

$$E(\Delta u | d_t = 1) = \frac{\alpha + \lambda_1}{\gamma(1)} + \frac{\beta(1) + \lambda_2}{\gamma(1)} E(\Delta y)$$

Άρα η διαφορά στη μέση μεταβολή του ποσοστού ανεργίας μεταξύ υφέσεων και μεγενθύσεων (recessions-expansions) δίνεται από

$$E(\Delta u | d_t = 1) - E(\Delta u | d_t = 0) = \frac{\lambda_1}{\gamma(1)} + \frac{\lambda_2}{\gamma(1)} E(\Delta y)$$

και **ανεξάρτητα** του μέσου ρυθμού μεγέθυνσης $E(\Delta y)$, η παραπάνω διαφορά δίνεται από $\frac{\lambda_1}{\gamma(1)}$. Η διαφορά στην ευαισθησία $\frac{\partial E(\Delta u)}{\partial E(\Delta y)}$ (sensitivity) δίνεται από $\frac{\lambda_2}{\gamma(1)}$.

Και στις δύο περιπτώσεις, οι συντελεστές λ_1, λ_2 διαφοροποιούν το αποτέλεσμα.

Για παράδειγμα εκτιμώντας το

```
dtdy = dt*dy
smpl ; 2019:4
ARDL22 <- ols du const dy dy(-1) dy(-2) du(-1) du(-2) dt dtdy
```

παίρνουμε τα παρακάτω αποτελέσματα:

ARDL22: OLS, using observations 1948:4-2019:4 (T = 285)
Dependent variable: du

	coefficient	std. error	t-ratio	p-value	
const	0.105046	0.0276206	3.803	0.0002	***
dy	-0.128920	0.0177660	-7.257	4.00e-012	***
dy_1	-0.0511379	0.0180850	-2.828	0.0050	***
dy_2	-0.0267582	0.0182876	-1.463	0.1446	
du_1	0.270958	0.0568313	4.768	3.01e-06	***
du_2	-0.0591979	0.0490016	-1.208	0.2280	
dt	0.331648	0.0480786	6.898	3.58e-011	***
dtdy	-0.0882357	0.0409176	-2.156	0.0319	**

Εικόνα 1. Εκτίμηση υποδείγματος 1 με δεδομένα Η.Π.Α

Οπότε, η μέση μεταβολή του ποσοστού ανεργίας διαφέρει στις υφέσεις κατά

$$0.331648/(1 - 0.270958 + 0.0591979) = 0.420745$$

(δηλαδή είναι 0.42 ποσοστιαίες μονάδες μεγαλύτερη!).

Αντίστοιχα, η ευαισθησία της ανεργίας όπως μετριέται από το **άμεσο αποτέλεσμα** είναι `-0.12892` σε περιόδους μεγέθυνσης, και `-0.12892-0.082357=-0.211277` σε περιόδους ύφεσης. Άρα μία μεταβολή του ρυθμού μεγέθυνσης κατά μία ποσοστιαία μονάδα (**έστω μείωσή του**) υποδηλώνει **αύξηση** της ανεργίας κατά 0.1289 ποσοστιαίες μονάδες μέσα στο τρίμηνο όταν δεν έχουμε ύφεση ενώ υποδηλώνει αύξηση της ανεργίας 0.2112 ποσοστιαίες μονάδες όταν έχουμε ύφεση (μία αύξηση του συντελεστή της τάξης του `(0.2112/0.1289)-1=0.6384` ή 64% περίπου).

Επειδή σε περιόδους μεγέθυνσης έχουμε αύξηση του πραγματικού Α.Ε.Π και θετικό μέσο ρυθμό μεγέθυνσης, μπορούμε να σκεφτούμε (καλύτερα) το εκτιμημένο άμεσο αποτέλεσμα `-0.12892` σε περιόδους μεγέθυνσης ως *'μία **αύξηση** στο ρυθμό μεγέθυνσης κατά 1 ποσοστιαία μονάδα επιφέρει **μείωση** στη μέση μεταβολή του ποσοστού ανεργίας κατά 0.12892 ποσοστιαίες μονάδες'*

Συνολικό αποτέλεσμα.

Μακροχρόνια, μία αύξηση του ρυθμού μεγέθυνσης (**σε περιόδους μεγέθυνσης**) κατά μία ποσοστιαία μονάδα επιφέρει συνολική μεταβολή (μείωση) στην μέση μεταβολή του ποσοστού ανεργίας:

$$\frac{\hat{\beta}(1)}{\hat{\gamma}(1)} = (-0.12892 - 0.0511379 - 0.0267582)/(1 - 0.270958 + 0.0591979) = -0.2623$$

0.2623 ποσοστιαίες μονάδες

Μακροχρόνια, μία μείωση του ρυθμού μεγέθυνσης (**σε περιόδους ύφεσης**) κατά μία ποσοστιαία μονάδα επιφέρει συνολική μεταβολή στην μέση μεταβολή του ποσοστού ανεργίας (αύξηση)

$$((-0.12892-0.0511379-0.0267582)-0.0882357)/(1-0.270958+0.0591979) = -0.3743$$

0.375 ποσοστιαίες μονάδες.

2 Distributed lags

The more modern view is to model the correlation or dynamics as part of the model rather than treat it as an estimation nuisance. The modern approach, then, is to model the dynamics directly rather than treat autocorrelation as a nuisance.

A distributed-lag model is a dynamic model in which the effect of a regressor x on y occurs over time rather than all at once. The individual coefficients β's are called lag weights and the collectively comprise the lag distribution. They define - along with the autoregressive part if present - the pattern of how x affects y over time. ARDL(p,q):

$$\Phi(L)y_t = \alpha + B(L)x_t + u_t \tag{3}$$

with

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$$

and

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + ... + \beta_q L^q$$

Avoiding inconsistency requires that we make sure that the error term u_t is not serially correlated. Because the OLS estimators are inconsistent, so are all test statistics and estimators based on the OLS residuals.

We are going to pay special attention to the (i) impulse response function and (ii) the unit or cumulative impulse response function associated with these models. As Greene (2003, 560) notes, looking at the impulse and unit response functions in time series models is the counterpart of looking at marginal effects in the cross-sectional setting. Imagine that our models are in equilibrium. An impulse response function is when the independent variable goes up one unit in one period and then back down to zero in the next period. The unit response function is when the independent variable goes up one unit and remains up one unit for all remaining periods.

The **dynamic marginal effects**

$$\frac{\partial y_{t+h}}{\partial x_t} \tag{4}$$

If, as is usually the case, the lag weights ψ_j decline exponentially to zero from an initial value, **the long-run cumulative effect** of x on y is well defined and given by $\dfrac{B(1)}{\Phi(1)} = \Psi(1)$. The long-run equilibrium defines the state to which the time series converges to over time (apply the expectation operator to get rid of time and disturbances).

The long run equilibrium for this model is:

$$y = \frac{\alpha}{\Phi(1)} + \frac{B(1)}{\Phi(1)}x$$

One difficulty that is common to all distributed-lag models is choice of lag length, whether this be choosing the point q at which to truncate a finite lag distribution or choosing how many lagged dependent variables to include.

2.1 Choosing the lag length

- In all of the models we have studied, we must specify the length of the lag prior to estimation. Economic theory rarely gives us information about the lag length, so this must usually be determined empirically. Several methods are available to econometricians to gain information about the appropriate lag length, though they do not always give the same answer.
- There is no "right way" to identify the length of a lag. We are usually forced to make a judgment after looking at the evidence from several methods.
- The methods we discuss for choosing lag length can apply either to the lagged x terms on the right-hand side - the lag length that we have been calling q - or to the number of lagged dependent variables to include in an autoregressive-lag or ARDL model - what we have called p.
- However, most of these methods cannot be applied in a straightforward way to determining the length of restricted lag models such as the linearly-declining lag or the polynomial distributed lag.

2.2 Determining lag length by statistical significance

- (general to specific) An obvious way to choose the length of a lag is to start with a long lag test the statistical significance of the coefficient at the longest lag—the "trailing lag" - and shorten the lag by one period it if we cannot reject the null hypothesis that the effect at the longest lag is zero. We continue shortening the lag until the trailing lag coefficient is statistically significant.
- Although this method (general to specific) has appeal, there are dangers as well. Remember that an insignificant t statistic on the trailing lag only fails to reject the hypothesis of a zero coefficient; it does not prove that the coefficient is zero! It is therefore quite possible that this procedure will choose a lag length that is too short.
- (specific to general) An alternative that also relies on statistical tests of significance is to start with a very short lag and successively add lag terms, continuing to add lags that are statistically significant and stopping when the

marginal lag coefficient is not. This method is similar to the one above and often, though not necessarily always, leads to a similar choice of lag length.

- To see why they are not identical, consider what would happen if the first, second, and fourth lags were (always) statistically significant but the third lag and all lags longer than four are not. Starting from a long lag and working downward you would stop at four lags, eliminating the insignificant fifth lag but retaining the third lag by convention. Starting from a short lag length and working upward you would stop at two lags; you would never discover the significant fourth lag.

2.3 Determining lag length by information criteria

- Information criteria are designed to measure the amount of information about the de-pendent variable contained in a set of regressors. They are goodness-of-fit measures of the same type as R^2 or $\bar{R}^2$, but without the convenient interpretation as share of variance ex-plained that we give to R2 in an OLS regression with an intercept term. The most commonly used criteria are the Akaike information criterion (AIC) the corrected AIC (AICc), the Schwartz or Bayesian information criterion (BIC or SIC or SBIC) and the Hannan-Quinn criterion (HQ).

- The "main ingredient" in those information criteria is the sum of squared residuals, which we want to make as small as possible. Thus, we want to **minimize** the criteria and choose the model with the smallest AIC or AICc or BIC or HQ value.

- Thus, we want to minimize the criteria and choose the model with the **smallest criterion value**. When using the information criteria to choose lag length, we must be very **careful** to make sure that all candidate models among which we are choosing are estimated over exactly the same sample period. This requires particular caution in lag models because there will usually be more observations available for models with shorter lags (because with fewer lags we "lose" fewer observations at the beginning of the sample). Passively allowing the software (without knowing exactly what it is doing) to set the sample by using all available observations will result in samples with different T for models with lags of different length, thus the information criteria calculated from them cannot be compared. You should always keep the sample the same across regressions being compared with information criteria, then verify that all of the observations being compared have identical sample lengths.

- The BIC criterion penalizes additional parameters more strongly than the AIC (assuming, plausibly, that $lnT > 2$). Thus, the BIC always chooses a lag length that is shorter than (or the same length as) the one that minimizes the AIC. Neither is "better", so one might consider the AIC as a lower bound and the SBIC as an upper bound for the appropriate lag length. In the case that they happen to agree, the choice is clear.

2.4 Determining lag length by residual autocorrelation

- As discussed above, adding lags of x and/or y to the right-hand side of a distributed-lag regression usually lessens the degree of autocorrelation in the error term. Estimators of some models (especially those with lagged dependent variables) are sensitive to autocorrelated errors, so another criterion that is sometimes used for choosing lag length is the elimination of autocorrelation in the residuals.

- When using residual autocorrelation to determine lag length, one adds lags until the residuals appear to be **white noise**. After running the distributed-lag regression, one extracts the residuals and uses a **Breusch-Godfrey LM test or a Box-Ljung Q test** to test the null hypothesis that the residuals are white noise. Rejecting the white-noise null hypothesis means that more lags should be added to the regression according to this criterion. If you have serial correlation left over, then you might consider adding additional lags of the dependent variable or lags of the independent variables to clean it up