

Lecture 7: Some GLS Approaches to Time-Series

1. Time-series analysis is about dealing with a subscript that has additional meaning than just 'name of the observation'.
 - a. Consider $Y_t = X\beta + \varepsilon_t$. Here, the subscript is t to indicate that we are thinking in terms observations going $t=1, \dots, T$ where the ordering subscript t has some meaning—it puts observations in order.
 - b. Given that t is time, we might worry that Y depends on time. Thus, it would be natural to consider a model
 - c. $Y_t = X\beta + g(t) + \varepsilon_t$
 - i. Intuition from cross-sectional econometrics will not serve us well here. If we wanted to be fully flexible with respect the partial effect of time, we might consider modelling g as a set of dummy variables for each and every year in the data. Unfortunately, this would require $T-1$ dummy variables, and the $T-1$ degrees of freedom to estimate them. If X included anything other than a constant, we could not get parameters for those extra things.
 - d. So, functional restrictions on g are necessary for identification. For example,
 - e. $Y_t = X\beta + \gamma t + \delta t^2 + \varepsilon_t$ Allows for a quadratic time-trend.
 - f. Here, the effect of X on Y conditions out the quadratic time-trend.
 - i. What if t causes variation in both X and Y ? Then, leaving out the time trend will cause a spurious correlation between X and Y . This is a form of endogeneity due to missing variables.
 - ii. Similarly, leaving out important parts of the time trend will cause endogeneity and potentially spurious correlation between X and Y . An example is using a linear time trend where the effect is really quadratic.
 - g. One can picture the time trend as interacting with the constant (as above), or as interacting with other pieces of X . Consider $X=[1 \ x]$. One could model
 - h. $Y_t = \alpha + x\beta + \gamma t + \delta xt + \varepsilon_t$
2. Time-series models typically have over-time correlations in the disturbances or in the regressors. If there are no over-time correlations, then OLS is just fine. But with over-time correlations, OLS may be inefficient or even inconsistent.
3. Over-Time correlations in the disturbances are typically described as either "moving-average" or "autocorrelated" processes. First, consider Moving-Average Processes, which are comprised of random elements which vary over time.

- i.
$$\varepsilon_t = \sum_{s=0}^q \theta_s u_{t-s}$$
, a q 'th order moving

$$E[u_t] = 0, E[u_t^2] = \sigma_u^2, Cov[u_t, u_s] = 0 \forall s \neq t$$

average.

 - ii. Here, the disturbance depends on a random variable epsilon, and has a short memory, lasting only q periods.
 - iii. What is the covariance matrix for the disturbances?

iv.

$$E[\varepsilon_t \varepsilon_{t-1}] = E[(\theta_0 u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} \dots \theta_p u_{t-p}) (\theta_0 u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3} \dots \theta_p u_{t-p-1})]$$

v. Since the u 's are uncorrelated with each other, this expectation only has nonzero elements for the common u terms:

$$E[\varepsilon_t \varepsilon_{t-1}] = \theta_1 \theta_0 \sigma_u^2 + \theta_2 \theta_1 \sigma_u^2 \dots$$

vi.

$$= \sigma_u^2 \sum_{s=1}^p \theta_s \theta_{s-1}$$

$$E[\varepsilon_t \varepsilon_{t-k}] = \theta_k \theta_0 \sigma_u^2 + \theta_{k+1} \theta_1 \sigma_u^2 \dots$$

vii. and

$$= \sigma_u^2 \sum_{s=1}^{p-k} \theta_{k+s-1} \theta_{s-1}$$

viii. If we knew the theta's in advance, we could do GLS: write out the $V(\text{epsilon})$ matrix, compute the Transformation matrix (equal to its 'minus one-half' matrix), and regress TY on TX .

ix. More often, we don't know the theta's in advance, but perhaps we know how many orders there are in the MA process. In this case, we know the structure of $V(\text{epsilon})$, and we know that this covariance matrix for the disturbances has only q parameters, and q does not grow with N .

x. Thus, there is an FGLS approach—estimate a first stage, collect residuals, estimate the q parameters, compute the covariance matrix for the disturbances, premultiply by the 'minus-one-half' matrix.

4. Next, consider Auto-regressive processes. In a moving average process, a big u affects things only for q periods. Then it is gone. Alternatively, we might think that a big u would sort of 'fade out', but never completely disappear. Consider

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t, \rho < 1$$

a.

$$E[u_t] = 0, E[u_t^2] = \sigma_u^2, \text{Cov}[u_t, u_s] = 0 \forall s \neq t$$

b.

What does the covariance matrix for disturbances Σ look like?

c.

we can recursively substitute:

$$\varepsilon_1 = u_1$$

$$\varepsilon_2 = \rho \varepsilon_1 + u_2 = \rho u_1 + u_2$$

d.

$$\varepsilon_3 = \rho \varepsilon_2 + u_3 = \rho^2 u_1 + \rho u_2 + u_3$$

$$\varepsilon_t = \sum_{s=1}^t \rho^{t-s} u_s$$

e.

Assuming that the variance of the disturbance tends to a finite value (aka "is stationary"), the variance of the disturbance is given by noticing that

$$V[\varepsilon_t] = V[\rho\varepsilon_{t-1} + u_t] = \rho^2 V[\varepsilon_t] + \sigma_u^2 \text{ which implies } V[\varepsilon_t] = \frac{\sigma_u^2}{1 - \rho^2}$$

because u is uncorrelated with past disturbances.

- f. The covariances of disturbances s, t is highest contemporaneously, and diminishes with the strength ρ .

$$\begin{aligned} \text{Cov}[\varepsilon_t, \varepsilon_{t-1}] &= E[\varepsilon_t \varepsilon_{t-1}] = E[(\rho\varepsilon_{t-1} + u_t) \varepsilon_{t-1}] \\ \text{i. } &= \rho V[\varepsilon_{t-1}] = \frac{\rho \sigma_u^2}{1 - \rho^2} \end{aligned} \quad \text{and}$$

$$\text{ii. } \text{Cov}[\varepsilon_t, \varepsilon_{t-k}] = \frac{\rho^k \sigma_u^2}{1 - \rho^2} \text{ and } E[\varepsilon \varepsilon'] = \Omega = \frac{\sigma_u^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots \\ \rho & 1 & \rho & \\ \rho^2 & \rho & 1 & \\ \dots & & & \end{pmatrix}$$

- iii. Here, Σ has only two parameters σ_u^2, ρ , and since they do not increase in number with T , we can get consistent estimates of them via a first stage regression and then implement FGLS.

- iv. One could write an AR(p) process with longer lags mattering directly as

$$\text{v. } \varepsilon_t = \sum_{s=1}^p \rho_s \varepsilon_{t-s} + u_t, \quad \sum_{s=1}^p \rho_s < 1$$

$$E[u_t] = 0, E[u_t^2] = \sigma_u^2, \text{Cov}[u_t, u_s] = 0 \forall s \neq t$$

- vi. The GLS approach to an AR process is similar in spirit to that for an MA process. By specifying the AR process, the covariance matrix for the disturbances depends on just a few parameters. All we need is a consistent estimate of those parameters, and we can obtain a consistent estimate of the covariance matrix. Then, apply FGLS.

5. The strategies above provide an asymptotically more efficient estimator than the OLS estimator. Suppose that instead of trying to use FGLS to obtain efficient estimates of parameters, we instead tried to obtain the robust covariance matrix for OLS coefficients?

In this case, we need an estimate $\widehat{X' \Omega X}$ to make the sandwich:

$$\text{a. } V\left(\widehat{\beta}^{OLS}\right) = (X'X)^{-1} X' \Omega X (X'X)^{-1}$$

- b. The Newey-West robust covariance matrix uses estimates of σ_u^2, ρ from the regression to construct $\widehat{X' \Omega X}$.

6. Stationarity and Unit Roots

- a. Roughly speaking, a variable is **stationary** over time if the covariance between two observations separated by $j=0, \dots, T$ periods is finite.
- b. We say that it is nonstationary if any one of those covariances (including the

variance of the variable itself) is infinite.

- c. The finiteness of the variance is connected with the persistence of past disturbances in the composition of current disturbances.
- d. Consider $\rho = 1$ in the AR1 described above. In this case, the disturbance for any

period is equal to $\varepsilon_t = \sum_{s=1}^t u_s$. As t gets large, the disturbance becomes the sum

of more and more normal variates. As t goes to infinity, the disturbance goes to an infinitely large sum of normals. Its variance goes to a chi-square with infinite degrees of freedom, that is, its variance goes to infinity.

- e. Because the disturbance here contains a contribution from its past equal to 1 that does not diminish with time, we say that it has a ‘**unit root**’.
- f. One might equally say that in this case, the disturbance is an **integrated** process, because its new contributions keep adding on to the old. The new value of the disturbance is computed by integrating over all the old ones.

- g. One can see this by evaluating the term $\frac{\sigma_u^2}{1 - \rho^2}$ which premultiplies the covariance matrix of disturbances. As $\rho \rightarrow \infty$, so too does that term.

- h. If the regression disturbances have infinite variance, then OLS and GLS and FGLS cannot yield consistent inference—the standard errors will be wrong no matter what you do regarding Ω . Further, the variance of OLS/GLS/FGLS estimated coefficients will contain a variance that goes to infinity. Bad all round.

- 7. This same problem emerges if X has some covarying variables that depend contain a ‘unit root’.

- a. Let the regression be given by

$$y_t = \alpha + \beta x_t + \gamma z_t + \varepsilon_t, E[\varepsilon\varepsilon'] = \sigma_\varepsilon^2 I_T$$

- b. $x_t = x_0 + \beta_1 u_t + \eta_t, E[\eta\eta'] = \sigma_\eta^2 I_T$

$$z_t = z_0 + \gamma_1 u_t + \lambda_t, E[\lambda\lambda'] = \sigma_\lambda^2 I_T$$

- c. Here, x and z covary through u . For example, let y be productivity, x be wages and z be prices. x and z must covary because wages contain prices. However, they are not collinear, because they each have a piece of variation that is distinct from the other. In a Venn diagram, the relative size of the variation in u compared to the distinct variation will tell you how easy it is to identify the parameters. If the distinct variation is zero, you have a multicollinearity problem.

- d. Now, consider an environment where

- e. $u_t = \rho u_{t-1} + v_t, E[vv'] = \sigma_v^2 I_T$, so that u is an AR1 process.

- f. Here, the variance of u goes to infinity as t goes to infinity, by the same reasoning as a for an AR1 model disturbance.

- g. As the variance of u rises, its variation dominates the total variation of x and z . Thus, in the limit, x and z become linear in each other, and you have a problem of perfect multicollinearity.

- h. This is the problem of cointegrating right-hand side variables.
 - i. Normally, covarying right-hand side variables is only a problem if they are perfectly covarying, because in this case there is no separate variation to identify each from each.
 - j. However, even if they are only perfectly covarying in the limit, you end up with this same problem in the limit.
 - k. Thus, an integrated disturbance in covarying right-hand side variables may for short T seem to be okay, as T gets large, it kills your ability to identify the coefficients of the model.
8. *Differencing* is a strategy used to deal with integrated RHS variables (including disturbances).
- a. Consider differencing a model with an integrated disturbance term.
 - b.
$$Y_t = X_t\beta + \varepsilon_t, \varepsilon_t = \varepsilon_{t-1} + u_t, E[uu'] = \sigma_u^2 I_T$$
 - c.
$$\Delta Y_t = \Delta X_t\beta + u_t$$
 - c. The symbol Δ is the ‘difference operator’, which tells you to subtract the lagged value from the variable on which it operates. Thus, $\Delta z_t = z_t - z_{t-1}$ for any z .
 - d. In this case, differencing solves the problem of an integrated disturbance term completely. A regression of difference Y on differenced X purges the model of the integrated disturbance, leaving only the well-behaved u .
 - e. However, differenced models do not depend on level information to identify anything. The spirit of differencing is: "if x cause y , then changes in x cause changes in y , and i can use the latter information to identify my model".
 - f. However, dropping the level information and looking only at changes can be costly. Consider a panel context, where different countries have different levels of various right-hand side variables. If we difference (over time) in such a model, the only variation remaining is the over-time variation within countries in these various right-hand side variables. The initial differences across countries is purged along with the integrated disturbance.
9. When right-hand side variables are cointegrated, and sometimes, we can use the structural information we have to compress the model so that the integrated part disappears.
- a. For example, in the wages and prices model above, we could use ‘real wages’ equal to wages divided by prices as a single right-hand side variable. Here, the integrated part could drop out, or be differenced out.