Lecture on Distributed Lags

Consider a model

$$y_t = a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + u_t$$

where we note $a_i = \partial y_t / \partial x_{t-i}$. We say that the variable y is a distributed lag on x. The main objective is to estimate the parameters.

i) If the lag length is finite, we could include appropriate x's. But a serious collinearity problem arises.

ii) If the lag distribution is infinite, what do we a

We try to summarize the lag distribution in terms of a few estimable parameters. This yields a variety of distributed lag forms.

Note we can write

$$y_t = a_0 (x_t + w_1 x_{t-1} + w_2 x_{t-2} + \dots +) + u_t$$

where $w_i = a_i / a_0$.

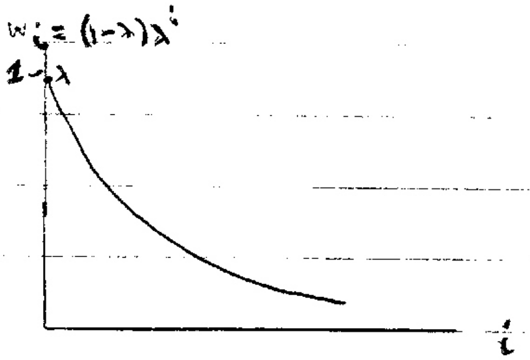
Issue: How to characterize lag weights $\{w\}$?

I. Koyck (or Geometric) Lag Distribution.

Let:

$$w_i = (1-\lambda)\lambda^i$$

$$\sum_{i=0}^{\infty} w_i = (1-\lambda)\sum \lambda^i = 1.$$



$$y_t = (1-\lambda)a\,(x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \dots\dots) + u_t$$

$$\lambda y_{t-1} = (1-\lambda)a\,(\lambda x_{t-1} + \lambda^2 x_{t-2} + \dots\dots) + \lambda u_{t-1}$$

∴

$$y_t - \lambda y_{t-1} = (1-\lambda)\,a\,x_t + \xi_t$$

or

$$y_t = (1-\lambda)\,a x_t + \lambda y_{t-1} + \xi_t \quad \text{where } \xi_t = u_t - \lambda u_{t-1}$$

Hence: Geometric lag translates into a regression of y against contemporaneous explanatory variables and one period lagged dependent variable.

Are error assumptions classical?

i) $E(\zeta_t) = E(u_t) - \lambda E(u_{t-1}) = 0$

ii) $E(x_t \zeta_t) = E(x_t u_t) - \lambda E(x_t u_{t-1}) = 0$

iii) $E(\zeta_t \zeta_{t-1}) = E(u_t - \lambda u_{t-1})(u_{t-1} - \lambda u_{t-2}) = -\lambda \sigma_u^2 \neq 0$

iv) $E(y_{t-1}\zeta_t) = E[(1-\lambda)\alpha x_{t-1} + \lambda y_{t-2} + \zeta_{t-1})\zeta_t =$

$= E(\zeta_t \zeta_{t-1}) = -\lambda \sigma_u^2 \neq 0$

[illegible]

[illegible] iv) [illegible] of the estimators [illegible]

Average Lag

$$\theta = \frac{\sum_i i w_i}{\sum w_i}$$

For Geometric Lag, $w_i = (1-\lambda)\lambda^i$, so

$$\theta = \frac{\sum_i (1-\lambda) i \lambda^i}{\sum_i (1-\lambda)\lambda^i} = \frac{(1-\lambda)\sum_{i=0} i\lambda^i}{1}$$

$$= (1-\lambda)\left[0 + \lambda + 2\lambda^2 + 3\lambda^3 + \dots\right]$$

$$= (1-\lambda)\lambda(1 + 2\lambda + 3\lambda^2 + \dots)$$

$$= \frac{(1-\lambda)\lambda}{(1-\lambda)^2}$$

$$\theta = \frac{\lambda}{1-\lambda}.$$

Note: $\theta = 0$ when $\lambda = 0$ and y_{t-1} falls out of the equation

Two justifications for Koyck lag

1. Partial adjustment model

Let:

$$Y_t^* = \alpha + \beta x_t + \varepsilon_t$$

and:

$$y_t - y_{t-1} = \gamma (y_t^* - y_{t-1}) + v_t$$

Then:

$$y_t = \gamma y_t^* + (1-\gamma) y_{t-1} + v_t$$

$$= \gamma\alpha + \gamma\beta x_t + (1-\gamma) y_{t-1} + (v_t + \gamma \varepsilon_t)$$

which is a case of geometric lags, where we note that in this case the combined disturbance $v_t + \gamma \varepsilon_t$ is serially uncorrelated if v_t and ε_t are.

2. Adaptive Expectations Model

Suppose

$$y_t = \alpha + \beta x_t^* + \varepsilon_t$$

$$x_t^* - x_{t-1}^* = \gamma (x_t - x_{t-1}^*)$$

i.e.

$$x_t^* = \gamma x_t + (1-\gamma) x_{t-1}^*$$

which is recognized as geometric weighting for x_t^*.
Thus:

$$x_t^* = \gamma [x_t + (1-\gamma) x_{t-1} + (1-\gamma)^2 x_{t-2} + \cdots]$$

substituting

$$y_t = \alpha + \beta\gamma [x_t + (1-\gamma) x_{t-1} + (1-\gamma)^2 x_{t-2} + \cdots] + \\ + \epsilon_t$$

$$(1-\gamma) y_{t-1} = (1-\gamma)\alpha + (1-\gamma)\beta\gamma [x_{t-1} + (1-\gamma) x_{t-2} + \cdots] + \\ (1-\gamma)\epsilon_{t-1}$$

Subtracting

$$Y_t = \gamma\alpha + \beta\gamma X_t + (1-\gamma) Y_{t-1} + \xi_t$$

$$\xi_t = \varepsilon_t - (1-\gamma)\varepsilon_{t-1}$$

Note that in this model we also obtained a geometric lag form of the model, but the composite disturbance is autocorrelated even if the original ε_t is white noise.

Testing for Serial Correlation w/ Lagged Dependent Variable

Durbin has shown that DW test is biased toward 2 if there is a lagged dependent variable, i.e. in favor of $H_0 : \rho = 0$. Durbin proposes a new h-test. Let:

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

With:

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

The test statistic is the "Durbin's h":

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T\sigma^2_{\hat{\gamma}}}} \approx \left(1 - \frac{DW}{2}\right) \sqrt{\frac{T}{1 - T\sigma^2_{\hat{\gamma}}}} \sim N(0,1)$$

where $\hat{\rho}$ is the estimated (biased) first-order serial correlation coefficient.

Reject H_0 if $h > 1.96$ for 95% significance level.

Accept H_0 if $h < 1.96$.

Distributed Lag example

Suppose we have an adaptive expectations model

$$y_t = b x_t^* + u_t \qquad u_t \text{ G M error.}$$

$$x_t^* - x_{t-1}^* = \lambda (x_t - x_{t-1}^*)$$

We showed earlier this yields the model

$$y_t = b\lambda x_t + (1-\lambda) y_{t-1} + (u_t - (1-\lambda) u_{t-1}).$$

Clearly $E(y_{t-1}, u_t - (1-\lambda) u_{t-1}) = -(1-\lambda)\sigma_u^2 \neq 0$.
To use an IV estimator we need z_t which is correlated with y_{t-1} <u>but</u> uncorrelated with u_{t-1}. What about x_{t-1}? It is correlated with y_{t-1} definitely, and uncorrelated with u_{t-1} by assumption that u_{t-1} is G.M. error.

Then we use

$$\widehat{b\lambda}_{IV} = \frac{\sigma_{x_t y_t \cdot y_{t-1}}}{\sigma_{x_t x_t \cdot y_{t-1}}}, \qquad \widehat{1-\lambda}_{IV} = \frac{\sigma_{x_{t-1} y_t \cdot x_t}}{\sigma_{x_{t-1} y_{t-1} \cdot x_t}}$$

Serial Correlation Alternative to Distributed Lag

Suppose True model is:

$$y_t = \alpha x_t + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t \quad ; \varepsilon_t \text{ iid}$$

$$\therefore \quad y_t = \alpha x_t + \varepsilon u_{t-1} + \varepsilon_t$$

But $\quad u_{t-1} = y_{t-1} + \alpha x_{t-1}$

Thus the model can be written as:

$$y_t = \alpha x_t + \rho y_{t-1} - \alpha\rho x_{t-1} + \varepsilon_t$$

We estimate

$$y_t = \alpha' x_t + \rho y_{t-1} + \varepsilon_t$$

[illegible] correlated error, the Koyck model [illegible]

Lectures on the Consumption Function

Keynes (1936) argued that the relation between consumption and current income, or the consumption function, should exhibit three features:

i) proportion of income spent on consumption declines w/ level of income, or income elasticity of consumption less than unity

ii) l.r. MPC > s.r MPC

iii) changes in money value of wealth affect s.r. MPC or A

Early work was on simple version:

$$C = \alpha + \beta Y$$

Note MPC = β and APC = $\beta + \alpha / Y$ so expect $\alpha > 0$.
Also note MPC = APC = const. if $\alpha = 0$, and
$\eta_{cY} = (\partial C / \partial Y) \, Y/C = \beta Y / C$.

Findings were : 1) $\beta < 1$ and $\alpha > 0$ in cross sectional studies

2) $\beta < 1$ and $\alpha \simeq 0$ in time series studies

3) $\beta_{CS} < \beta_{TS}$

In addition, these studies didn't seem consistent w/ the stylized fact that C/Y is very stable over long periods but exhibits some countercyclical behavior.

Dynamic Models

Brown (1952) and others suggested the generalization

$$C_t = \alpha + \beta_1 Y_t + \beta_2 C_{t-1} \qquad 0 < \beta_2 < 1$$

The s.r. MPC $= \beta_1$. To compute l.r. MPC solve for steady state $C_t = C_{t-1}$: $C_t = \frac{\alpha}{1-\beta_2} + \frac{\beta_1}{1-\beta_2} Y_t$.
So l.r. MPC $= \beta_1/(1-\beta_2)$.

Estimates vary, but results yield s.r. MPC < l.r. MPC .
Examples are $\beta_1 \approx .3$, $\beta_2 = .6$ and $\alpha \geq 0$.
So l.r. MPC $\approx .75$ and average lag $.6/.4 \approx 1.5$ years.

Duesenberry (1949): Relative Income Hypothesis

Argued two theoretical points:

i) person's consumption at a point in time depends on his income relative to average.

ii) person adjusts consumption over time asymmetrical to upward and downward movements in income.

$$\frac{C_{it}}{Y_{it}} = \alpha_0 + \alpha_1 \frac{\bar{Y}_{\cdot t}}{Y_{it}} \qquad \text{where } \alpha_0, \alpha_1 > 0.$$

or $C_{it} = \alpha_0 Y_{it} + \alpha_1 \bar{Y}_{\cdot t}$. $\quad \bar{Y}_{\cdot t} = \frac{1}{N}\sum_i^N Y_{it}$

In cross section (at individual level), $MPC_{cs} = \alpha_0$ and $APC_{cs} = \alpha_0 + \alpha_1 \bar{Y}_{\cdot t}/Y_{it}$ which declines w/ Y_{it}. Aggregating the relation over individuals

$$\sum_i C_{it} = \alpha_0 \sum_i Y_{it} + \alpha_1 \sum_i \bar{Y}_{\cdot t}$$

$$C_t = (\alpha_0 + \alpha_1) Y_t$$

So in time series $MPC_{TS} = \alpha_0 + \alpha_1 > MPC_{cs} = \alpha_0$.
Also, $APC_{TS} = \alpha_0 + \alpha_1 = MPC_{TS}$.

Hence, this hypothesis gives a theoretical explanation for various empirical regularities. Problem is, why *should* we expect (on first principles) that the relative income should be the determinant?

Friedman's (1957) Permanent Income Hypothesis

The key idea is that the permanent (planned) consumption depends only on permanent or sustained income. Permanent consumption includes services, but not purchase, of durable goods. Specifically

$$C_{pt} = kY_{pt}$$

where k may depend on asset composition, individual characteristics (e.g. family size, age). Note $MPC_p = APC_p = k$ and $\eta_{c_p y_p} = 1$.

Friedman then permits transitory components

$$C = C_p + C_T$$

$$Y = Y_p + Y_T$$

and the following statistical properties:

i) $E(Y_T) = E(C_T) = 0$

ii) $E(Y_T Y_p) = E(C_T C_p) = 0$

iii) $E(C_T Y_T) = 0$

Property (iii) is restrictive; it says a windfall gain will be saved entirely (i.e. not spent on <u>nondurable</u> goods)

Now suppose we estimate

$$(1) \qquad C = \alpha + \beta Y + v$$

The true model is

$$C_p = k Y_p + u$$

so

$$C - C_T = k(Y - Y_T) + u$$

$$C = kY + (u + C_T - kY_T)$$

Least squares est of (1) yields

$$\text{plim } \hat{\beta} = \frac{\sigma_{CY}}{\sigma^2_Y} = \frac{k(1 - \sigma^2_{Y_T})}{\sigma^2_Y} = k \frac{\sigma^2_{Y_P}}{\sigma^2_Y} < k$$

Hence MPC is underestimated by $\hat{\beta}$, and the downward bias is larger the more transitory income figures in observed income.

Friedman therefore expects that, if one studied different groups of consumers acc to how transitory their income is, one would obtain larger $\hat{\beta}$ for those groups with less transitory incomes. He confirmed this on a number of such groups.

Also predicts that the same holds for time periods which are characterized by large transitory shocks.

Aggregate Time Series

Friedman, in original study, measured Y_P as geometrically weighted average of 16 past Y's, searching over weights to obtain best fit.

$$\hat{Y}_{Pt} = (1-\gamma) \sum_{i=1}^{16} \gamma^i Y_{t-i}$$

He then estimated

$$C_t = a_0 + k \hat{Y}_{Pt}$$

and found $a_0 \doteq 0$ and $k \doteq .88$. (consistent w/ evidence on long run APC from Kuznets). The average lag from Y was 2.5 years.

We note that if one used an infinite lag

$$Y_{pt} = (1-\gamma)\sum_{i=0}^{\infty}\gamma^i Y_{t-i}$$

we would get the model

$$C_t = k(1-\gamma)Y_t + \gamma C_{t-1}$$

so the permanent income hypothesis is sometimes used to justify presence of lagged dependent variable on RHS (which is really adaptive expectations model).

Note also that the permanent income hypothesis doesn't systematically explain lower MPC's in ~~time~~ time series studies as compared to cross section investigations.

Liviatin (1963): Tests of the PIH

Liviatan had observations $c_{i,t}$ and Y_{it} $(i=1,\dots,N)$, $(t=1,$
How to use them to test PIH?

1. Use of Changes ΔC_i and ΔY_i.

define $\Delta C_i = C_{it} - C_{i,t-1}$ and $\Delta Y_i = Y_{it} - Y_{i,t-1}$.
and the least squares est

$$\hat{\beta}_\Delta = \frac{\sum \Delta C_i \, \Delta Y_i}{\sum (\Delta Y_i)^2}$$

If Keynes' function holds $C_{it} = \alpha + \beta Y_{it}$ then $\Delta C_i = \beta \Delta Y_i$ and $\hat{\beta}_\Delta$ is consistent est of β.
But what if PIH holds? By similar argument we get

$$\operatorname{plim} \hat{\beta}_\Delta = k \, \frac{\operatorname{var}(\Delta Y_p)}{\operatorname{var}(\Delta Y)} .$$

Expanding numerator

$$\begin{aligned} \operatorname{var}(Y_{p,t} - Y_{p,t-1}) &= \operatorname{var}(Y_p) + \operatorname{var}(Y_p) - 2 \operatorname{cov}(Y_{p,t}\, Y_{p,t-1}) \\ &= 2\left(\operatorname{var}(Y_p) - \operatorname{cov}(Y_{pt}\, Y_{p,t-1})\right). \end{aligned}$$

Expanding denominator and assuming $\mathrm{cov}(Y_{Tt}\, Y_{T,t-1}) = 0$.

$$\mathrm{var}(Y_t\, Y_{t-1}) = \mathrm{var}(Y_{pt} + Y_{Tt},\ Y_{p,t-1} + Y_{T,t-1})$$

$$= 2\left[\mathrm{var}(Y_p) + \mathrm{var}(Y_T) - \mathrm{cov}(Y_{pt}\, Y_{p,t-1})\right]$$

Since Friedman's argues $\mathrm{cov}(Y_{pt}\, Y_{p,t-1}) > 0$,

$$\mathrm{plim}\ \hat{\beta}_\Delta = k \frac{\mathrm{var}(Y_p) - \mathrm{cov}(Y_{pt}\, Y_{p,t-1})}{\mathrm{var}(Y_p) + \mathrm{var}(Y_T) - \mathrm{cov}(Y_{pt}\, Y_{p,t-1})}$$

$$< \mathrm{plim}\ \hat{\beta} = k \frac{\mathrm{var}(Y_p)}{\mathrm{var}(Y_p) + \mathrm{var}(Y_T)}.$$

So under PIH we expect to find $\hat{\beta}_\Delta < \hat{\beta}$. Liviatan finds this but not statistically different.

But, is the additional assumption $\mathrm{cov}(Y_{Tt}, Y_{T,t-1}) = 0$ reasonable? If not, have we really tested PIH?