Linking Expectations (Uncertainty) to Measurement Error

"Signal Extraction"

Concepts:

- Establish relation between expectations (uncertainty) and forecast measurement error by agents
- "General to Relative" Confusion
- "Temporary or Permanent" Confusion

Formal Analogue(s):

- Conditional Expectations Modeling (REE)
- Recursive Projections
- Law of Iterated Projections (or Expectations)

Applied Statistical Analogue:

- Error-in-Variables Regression

EITM Linkage:

- Conditional Expectations Modeling + Recursive Projections + Law of Iterated Projections $\Longleftrightarrow$ Error-in-Variables Regression

The Following Sections are Based on Whittle (1963, 1983) Sargent (1987)

The EITM formulation in this module (linking expectations with measurement error) requires the use of recursive projections and the law of iterated expectations. Before discussing them, first review some basic identities and operations in linear least squares.

Linear Least Squares Regression

- Suppose we estimate y and let it be the linear function of x_i:

$$\hat{y} = a_0 + a_1 x_1 + \cdots + a_n x_n. \tag{1}$$

  We choose the a_i so that we can minimize the distance between y and $\hat{y}$. That is:

$$\min E\left(y - \hat{y}\right)^2.$$

  We have:

$$\min_{a_i} E\left[y - \left(a_0 + a_1 x_1 + \cdots + a_n x_n\right)\right]^2 \forall i \tag{2}$$

  To minimize the equation (2), a necessary and sufficient condition is (in the normal equation(s)):

$$E\left[y - \left(a_0 + a_1 x_1 + \cdots + a_n x_n\right) x_i\right] = 0, \text{ for } i = 0, 1, \cdots, n, \tag{3}$$

  where $x_0 = 1$. This is called the Orthogonality Principle.

- Consider:

$$y = \sum_{i=0}^{n} a_i x_i + \varepsilon,$$

where ε is the forecast error, $E\left(\varepsilon \sum a_i x_i\right) = 0$ and $E\left(\varepsilon x_i\right) = 0$, for $i = 0, 1, \cdots, n$.

The random variable $\sum_{i=0}^{n} a_i x_i$, where the a_i are chosen to satisfy the least squares orthogonality condition (3), is called the projection of y on x_0, $x_1, ...,$ x_n.

We have:

$$\sum a_i x_i \equiv P\left(y \left| 1, x_1, x_2, \cdots, x_n\right.\right),$$

where $x_0 \equiv 1$.

According to the Orthogonality condition, we have:

$$\begin{bmatrix} E_y \\ Eyx_1 \\ Eyx_2 \\ \vdots \\ Eyx_n \end{bmatrix} = \begin{bmatrix} 1 & Ex_1 & Ex_2 & \cdots & Ex_n \\ Ex_1 & Ex_1^2 & Ex_1x_2 & \cdots & \\ Ex_2 & Ex_1x_2 & \ddots & & \\ \vdots & \vdots & & \ddots & \\ Ex_n & & & & Ex_n^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

and:

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \left[Ex_ix_j\right]^{-1}\left[Eyx_k\right].$$

Now apply the above to the simple example:

$$y = a_0 + a_1 x_1 + \varepsilon.$$

We have:

$$\begin{bmatrix} Ey \\ Eyx_1 \end{bmatrix} = \begin{bmatrix} 1 & Ex_1 \\ Ex_1 & Ex_1^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix},$$

where (using normal equation(s), i.e., equation (3)):

$$a_0 = Ey - a_1 Ex_1,$$

and

$$\begin{aligned} a_1 &= \frac{E\left(y - Ey\right)\left(x_1 - Ex_1\right)}{E\left(x_1 - Ex_1\right)^2} \\ &= \frac{\sigma_{x_1 y}}{\sigma_{x_1}^2}, \end{aligned}$$

where $\sigma_{x_1 y}$ is the covariance between x_i and y, and $\sigma_{x_1}^2$ is the variance of x_1.

Recursive Projection

The linear least squares identities now can be used in formulating how agents can update forecasts. These forecasts will be updated consistent with a linear least squares rule.

- The simple univariate projection can be used (recursively) to assemble projections on many variables, such as $P(y|1, x_1, x_2, \cdots, x_n)$.
- For example when $n = 2$:

$$y = P(y|1, x_1, x_2) + \varepsilon$$

  will imply that:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon,$$

  where $E\varepsilon = 0$, $E\varepsilon x_1 = 0$ and $E\varepsilon x_2 = 0$.

- Now, if we omit the information from x_2 to project y, we have:

$$P(y|1, x_1) = a_0 + a_1 x_1 + a_2 P(x_2|1, x_1)$$

  where $P(x_2|1, x_1)$ is a component that we use 1 and x_1 to project x_2 for projecting y.

  Expanding:

$$P(y|1, x_1) = P(a_0|1, x_1) + a_1 P(x_1|1, x_1) + a_2 P(x_2|1, x_1).$$

- But, why is $P(a_0|1, x_1) = a_0$? and $P(x_1|1, x_1) = x_1$?

  If we are predicting a constant using 1 and x_1, we are still predicting a constant a_0. Therefore:

$$P(a_0 \,|1, x_1\,) = a_0.$$

If we are predicting x_1 using 1 and x_1, certainly, we can predict x_1, i.e., $P(x_1 \,|1, x_1\,) = x_1$.

- Mathematically: How do we show $P(a_0 \,|1, x_1\,) = a_0$ and $P(x_1 \,|1, x_1\,) = x_1$?
  - We have:

$$P(a_0 \,|1, x_1\,) = t_0 + t_1 x_1.$$

According to the normal equations, we can solve for t_0 and t_1:

$$t_0 = Ea_0 - t_1 Ex_1,$$

and

$$t_1 = \frac{E(a_1 - Ea_0)(x_1 - Ex_1)}{E(x_1 - Ex_1)^2}.$$

Since $Ea_0 = a_0$, therefore:

$$\begin{aligned} t_1 &= \frac{E(a_1 - Ea_0)(x_1 - Ex_1)}{E(x_1 - Ex_1)^2} \\ &= 0. \end{aligned}$$

And since $t_1 = 0$, we have:

$$t_0 = Ea_0 = a_0.$$

Therefore:

$$\begin{aligned} P(a_0 \,|1, x_1\,) &= t_0 + t_1 x_1 \\ &= a_0 \end{aligned}$$

■

– For $P(x_1 | 1, x_1) = x_1$, we can perform the same operations:

$$P(x_1 | 1, x_1) = t_0 + t_1 x_1.$$

Now we have:

$$t_0 = Ex_1 - t_1 Ex_1,$$

and

$$\begin{aligned} t_1 &= \frac{E(x_1 - Ex_1)(x_1 - Ex_1)}{E(x_1 - Ex_1)^2} \\ &= \frac{E(x_1 - Ex_1)^2}{E(x_1 - Ex_1)^2} \\ &= 1. \end{aligned}$$

Therefore:

$$\begin{aligned} t_0 &= Ex_1 - Ex_1 \\ &= 0, \end{aligned}$$

and:

$$\begin{aligned} P(x_1 | 1, x_1) &= t_0 + t_1 x_1 \\ &= 0 + x_1 \\ &= x_1. \end{aligned}$$

Therefore:

$$P(x_1 | 1, x_1) = x_1$$

■

– Finally we also have:

$$P(\varepsilon\,|1, x_1\,) = 0.$$

By the orthogonality condition(s):

$$E(\varepsilon) = E(\varepsilon x_1) = 0.$$

$$P(\varepsilon\,|1, x_1\,) = t_0 + t_1 x_1.$$

Therefore:

$$t_0 = E\varepsilon - t_1 E x_1,$$

and

$$\begin{aligned} t_1 &= \frac{E(\varepsilon - E\varepsilon)(x_1 - Ex_1)}{E(x_1 - Ex_1)^2} \\ &= \frac{E(\varepsilon x_1 - \varepsilon E x_1 - E\varepsilon x_1 + E\varepsilon x_1)}{E(x_1 - Ex_1)^2} \\ &= 0. \end{aligned}$$

We have:

$$\begin{aligned} t_0 &= E\varepsilon - t_1 E x_1 \\ &= E\varepsilon \\ &= 0 \end{aligned}$$

Therefore:

$$P(\varepsilon\,|1, x_1\,) = 0$$

■

– Now, let us expand further:

$$y = P(y|1, x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 + \varepsilon, \tag{4}$$

and

$$P(y|1, x_1) = a_0 + a_1 x_1 + a_2 P(x_2|1, x_1). \tag{5}$$

Equation (4) minus equation (5), we have:

$$y - P(y|1, x_1) = a_2 [x_2 - P(x_2|1, x_1)] + \varepsilon, \tag{6}$$

which can be simplified to:

$$z = a_2 w + \varepsilon.$$

– Note that $x_2 - P(x_2|1, x_1)$ is also orthogonal to ε — i.e., $E\{\varepsilon [x_2 - P(x_2|1, x_1)]\} = 0$ or $E(\varepsilon w) = 0$.
Therefore:

$$P[y - P(y|1, x_1)|x_2 - P(x_2|1, x_1)] = a_2 [x_2 - P(x_2|1, x_1)],$$

or in simplified form:

$$P(z|w) = a_2 w.$$

- – We can rewrite equation (6) as:

$$y = P(y|1, x_1) + P[y - P(y|1, x_1)|x_2 - P(x_2|1, x_1)] + \varepsilon, \tag{7}$$

or alternatively:

$$y - P(y|1,x_1) = a_2[x_2 - P(x_2|1,x_1)] + \varepsilon.$$

- Therefore:

$$P(y|1,x_1,x_2) = P(y|1,x_1) + P[y - p(y|1,x_1)|x_2 - P(x_2|1,x_1)],$$

where $P(y|1,x_1,x_2)$ is the bivariate projection, $P(y|1,x_1)$, $P[y - P(y|1,x_1)|x_2 - P(x_2|1,x_1)]$ and $P(x_2|1,x_1)$ are univariate projections.

Note that we see that the Bivariate projection = 3 univariate projections.

- Note also that equation (7) is useful for purposes of describing optimal least squares learning::

$$y = P(y|1,x_1) + P[y - P(y|1,x_1)|x_2 - P(x_2|1,x_1)] + \varepsilon,$$

where $y - P(y|1,x_1)$ is the prediction error of y given x_1 and $x_2 - P(x_2|1,x_1)$ is the prediction error of x_2 given x_2.

- - If at first we have data only on a variable x_1, the linear least squares estimates of y and x_2 are $P(y|1,x_1)$ and $P(x_2|1,x_1)$ respectively:

$$P(y|1,x_1) = a_0 + a_1x_1 + a_2P(x_2|1,x_1),$$

where a_1x_1 implies that we are using x_1 to predict y, and $P(x_2|1,x_1)$ implies that we are using x_1 to predict x_2 for predicting y.

- If an observation x_2 subsequently becomes available, our estimate of y can be improved by adding to $P(y|1,x_1)$, and the projection of unobserved "forecast error" $y - P(y|1,x_1)$ on the observed forecast error $x_2 - P(x_2|1,x_1)$.

That is:

$$P(y|1,x_1,x_2) = P(y|1,x_1) + P[y - P(y|1,x_1)|x_2 - P(x_2|1,x_1)],$$

where: $P(y|1,x_1)$ is the original forecast, $y - P(y|1,x_1)$ is the forecast error of y, given x_1, and $x_2 - P(x_2|1,x_1)$ implies that we are using the forecast error of x_2 to forecast the forecast error of y given x_1.

- Therefore, in general, we have:

$$P(y|\Omega,x) = P(y|\Omega) + P\{y - P(y|\Omega)|x - P(x|\Omega)\}, \tag{8}$$

  - Ω is the original information.
  - x is the new information.
  - $P(y|\Omega)$ is the prediction of y using the original information.
  - $P\{y - P(y|\Omega)|x - P(x|\Omega)\}$ implies that instead of using the original information to predict the new information, we do have the new information. Therefore, we use $x - P(x|\Omega)$ which is the difference between the new information and the "forecasted" new information to predict the error of y, (i.e., $y - P(y|\Omega)$).

The Law of Iterated Projections (or Expectations)

We now can show the relation between recursive projections and conditional expectations.

- The law states that:

$$P\left[P\left(y\,|\Omega,x\right)|\Omega\right]=P\left(y\,|\Omega\right),$$

  where Ω is the original information of time $t-1$, while x is the new information at time t. Note also that this expression also can be written as: $P\left[P\left(y\,|\Omega,x\right)|\Omega\right]=E_{t-1}\left(E_t y_{t+1}\right), P\left(y\,|\Omega,x\right)=\left(E_t y_{t+1}\right),$ and $P\left(y\,|\Omega\right)=E_{t-1}y_{t+1}$.

- How do we think of this expression in terms of expectations? First, we have:

$$E_{t-1}\left(E_t y_{t+1}\right)=E_{t-1}y_{t+1}.$$

  Why?

$$E_t y_{t+1}=P\left(y\,|\Omega,x\right)=P\left(y\,|\Omega\right)+a\left[x-P\left(x\,|\Omega\right)\right],$$

  where $[\Omega,x]$ is the information set at time t and $[\Omega]$ is the information set at time t-1.

  and

$$a=\frac{E\left[y-P\left(y\,|\Omega\right)\right]\left[x-P\left(x\,|\Omega\right)\right]}{E\left[x-P\left(x\,|\Omega\right)\right]^2}.$$

  Now consider the prediction of y_{t+1} given the information set at time t, and given the information set at time t-1 (which agents only possess):

$$\begin{aligned} P\left[P\left(y\,|\Omega,x\right)|\Omega\right] &= P\left\{P\left(y\,|\Omega\right)+a\left[x-P\left(x\,|\Omega\right)\right]|\Omega\right\} \\ &= P\left[P\left(y\,|\Omega\right)|\Omega\right]+P\left\{a\left[x-P\left(x\,|\Omega\right)\right]|\Omega\right\}, \end{aligned}$$

but $P\left\{a\left[x-P\left(x\,|\Omega\right)\right]|\Omega\right\}=0$. Why?

$$\begin{aligned} & P\left\{a\left[x-P\left(x\,|\Omega\right)\right]|\Omega\right\} \\ = \; & a\left[P\left(x|\Omega\right)-P\left(P\left(x|\Omega\right)|\Omega\right)\right] \\ = \; & a\left[P\left(x\,|\Omega\right)-P\left(x\,|\Omega\right)\right] \\ = \; & a\times 0 \\ = \; & 0. \end{aligned}$$

Therefore, we say that (from above):

$$P\left[P\left(y\,|\Omega,x\right)|\Omega\right]=P\left(y\,|\Omega\right),$$

Or

$$E_{t-1}\left(E_t y_{t+1}\right)=E_{t-1}y_{t+1}.$$

The Signal-Extraction Problems

The linkage between conditional expectations with recursive projections and the law of iterated expectations has a natural relation with error-in-variables regression (measurement error). There are many examples of this "EITM-like" linkage and they generally fall under the umbrella of signal extraction problems.

- Example 1: Measurement Error
  - Suppose an agent wants to estimate a random variable "s" but only "sees" the variable x:

$$x = s + n,$$

where $Esn = 0$, $Es^2,\ En^2 < \infty$; $Es = En = 0$.
Therefore:

$$P(s\,|1, x) = a_0 + a_1 x.$$

  - We have:

$$\begin{aligned} a_1 &= \frac{E(xs)}{Ex^2} \\ &= \frac{E[(s+n)s]}{E(s+n)^2} \\ &= \frac{Es^2}{Es^2 + En^2}, \end{aligned}$$

and

$$a_0 = 0.$$

  - Therefore:

$$P(s\,|1, x) = \frac{Es^2}{Es^2 + En^2} x.$$

- Example 2: General-Relative Confusion
  - Suppose a worker wants to estimate their real wage "$w-p$". But only "sees" the nominal wage (i.e., "w"):

$$w = z + u,$$

and

$$p = z + v.$$

Also:

$$Ezu = Ezv = Euv = Eu = Ez = Ev = 0,$$

where z is "neutral" movement in the aggregate price level.

  - Therefore:

$$P\left[(w-p)\,|1,w\right] = a_0 + a_1 w.$$

We have:

$$w - p = u - v,$$

and

$$w = z + u.$$

That means:

$$a_0 = 0,$$

and

$$\begin{aligned} a_1 &= \frac{E\left[(w-p)\,w\right]}{E\,(w)^2} \\ &= \frac{E\left[(u-v)\,(z+u)\right]}{E\,(z+u)^2} \\ &= \frac{Eu^2}{Ez^2 + Eu^2}. \end{aligned}$$

Therefore:

$$w - p = \left(\frac{Eu^2}{Ez^2 + Eu^2}\right) w.$$

- The greater Eu^2/Ez^2 is, the closer to 1 a_1 is.
- The greater is Eu^2/Ez^2, the larger is the fraction of variance in w that is due to variations in the real wage (Eu^2) determining factor u.

- Example 3: EITM — Expectations Uncertainty and Error-in-Variables Regression — (Lucas (1973) Supply Model)
  - Consider another case that producers observe the prices of their own goods but not the aggregate price level.
    Relative price of good i:

$$r_i = p_i - p.$$

Therefore:

$$\begin{aligned} p_i &= p + (p_i - p) \\ &= p + r_i. \end{aligned}$$

  - The producers want to estimate the real relative price, but they do not see the general price level.:

$$E\left(r_i \left| p_i \right.\right) = a_0 + a_1 p_i.$$

So we have:

$$\begin{aligned} a_0 &= E\left(r_i\right) - a_1 E\left(p_i\right) \\ &= E\left(p_i - p\right) - a_1 E\left(p_i\right) \\ &= -a_1 E\left(p_i\right), \end{aligned}$$

and

$$
\begin{aligned}
a_1 &= \frac{E\left[r_i - E\left(r_i\right)\right]\left[p_i - E\left(p_i\right)\right]}{E\left[P_i - E\left(p_i\right)\right]^2} \\
&= \frac{E\left[r_i - E\left(r_i\right)\right]\left[\left(p + r_i\right) - E\left(p + r_i\right)\right]}{E\left[\left(p + r_i\right) - E\left(p + r_i\right)\right]^2} \\
&= \frac{Er_i^2}{Er_i^2 + Ep^2} \\
&= \frac{v_r}{v_r + v_p},
\end{aligned}
$$

where $v_r = Er_i^2 =$ the variance of the real relative price and $v_p = Ep^2 =$ the variance of the general price level.

– Therefore:

$$
E\left(r_i \left| p_i \right.\right) = a_0 + a_1 p_i.
$$

Since $a_0 = -a_1 E\left(p\right)$,

$$
\begin{aligned}
E\left(r_i \left| p_i \right.\right) &= a_1\left[p_i - E\left(p\right)\right] \\
&= \frac{v_r}{v_r + v_p}\left[p_i - E\left(p\right)\right].
\end{aligned}
$$

- Now, if the labor supply is given by:

$$
l_i = \beta E\left(r_i \left| p_i \right.\right),
$$

we have:

$$
l_i = \beta \frac{v_r}{v_r + v_p}\left[p_i - E\left(p\right)\right].
$$

On average, the aggregate production is:

$$
y = b\left[p - E\left(p\right)\right],
$$

where $b = \beta\left(v_r\right) / \left(v_r + v_p\right)$.

- Empirical Application: International Evidence on Output-Inflation Trade-offs.

$$y = b\left(p - Ep\right).$$

It implies that (given general-relative price confusion):

$$y = \beta \frac{v_r}{v_r + v_p}\left[p - E\left(p\right)\right],$$

where v_p is the variance of the nominal shock, and $p - E\left(p\right)$ is the nominal shock.

  - Lucas estimates the following specification:

$$y_t = c + \gamma t + \tau \Delta x_t + \lambda y_{t-1}, \qquad (**)$$

  where y_t is real GDP, t is time, Δx_t is the change in log nominal GDP (it represents the nominal shock), and y_{t-1} is lagged real GDP.

  - Lucas estimates equation (**) separately for various countries. He then asks whether the estimated τ's – the estimates of the responsiveness of output to aggregate demand movements – are related to the average size of the countries' aggregate demand shocks (i.e., v_p).
  - He then estimates:

$$\tau_i = \alpha + \beta \sigma_{\Delta x, i}, \text{ for country } i.$$

  According to the Lucas Supply Curve, β is expected to be negative!

  - Lucas's theory predicts that nominal shocks have smaller real effects in settings (small τ) where aggregate demand is more volatile (big $\sigma_{\Delta x,i}$).

– Ball, Mankiw and Romer (1988), using data from 43 countries, estimate the following regression:

$$\begin{array}{ccc} \tau_i = & 0.388 & -1.639\sigma_{\Delta x,i} \\ & (0.057) & (0.482) \end{array}$$

$$\bar{R}^2 = 0.201 \qquad s.e.e. = 0.245,$$

where the number in parentheses are standard errors. Thus, there is a statistically significant and negative relationship between the variability of nominal GDP growth and the estimated effect of a given change in aggregate demand.

- Example 4: EITM — Expectations Uncertainty and Error-in-Variables Regression — (Alesina and Rosenthal (1995) "Competence Model")
  - Consider the simple model of economic growth.

$$y_t = \gamma\left(\pi_t - \pi_t^e\right) + \overline{y} + \varepsilon_t$$

  - If we focus on the shock term ε_t we find important theoretical/behavioral attributes. For instance, in deciding on whether to attribute credit or blame for economic growth outcomes to the incumbent administration, agents have to determine what part of economic outcomes (y_t) are faced with the following "signal extraction" problem:

$$\varepsilon_t = \eta_t + \xi_t,$$

    where the term η_t is the shock associated with "competence" — non-inflationary growth (i.e., $\pi_t = \pi_t^e$) — that can be attributed to the incumbent administration. The second term, ξ_t, reflects shocks to growth that are beyond government control (and competence). Both η_t and ξ_t have zero mean and have second moments $\sigma_\eta^2, \sigma_\xi^2$ respectively.
  - Competence is also thought to persist and allowing for reelection we have the competency shock represented as an MA(1) process:

$$\eta_t = \mu_t + \rho\mu_{t-1}.$$

  - The parameter ρ represents the strength of the persistence and $0 < \rho \leq 1$. Note also that the variance of μ_t is fixed at σ_μ^2.

Note also that this identity (because of the use of a lag (or lags)) allows for retrospective judgments by agents.

- Now suppose that agents have a general sense of what the average rate of growth is $(\overline{y})$. Agents also observe actual growth (y_t). Now from above when growth in non-inflationary (or does not reduce real wages) then total shock (which includes competency) can be expressed as:

$$\varepsilon_t = \eta_t + \xi_t = y_t - \overline{y}$$

- When $y_t > \overline{y}$ then $\eta_t + \xi_t > 0$. But, agents need to know how much is competence and how much is luck (signal extraction). If growth is high then agents have reason growth will be high in the next period (given that competence can persist):

$$\eta_{t+1} = \mu_{t+1} + \rho\mu_t.$$

- Still, agents what is the optimal estimate of competence when the agent only sees y_t? Using recursive projection and the law of iterated expectations we have:

$$E(\eta_{t+1}) = E(\mu_{t+1}) + \rho E(\mu_t|y_t) = \rho E(\mu_t|y_t) \tag{9}$$

- With this identity, we see that voters can forecast competence using the difference between $y_t - \overline{y}$, but also the "weighted" lag of μ_t, i.e., $\rho\mu_{t-1}$:

$$\mu_t + \varepsilon_t = y_t - \overline{y} - \rho\mu_{t-1}$$

- The final result is we can now estimate (9) using an error-in-variables regression:

$$E\left(\eta_{t+1}\right) = \rho\frac{\sigma_{\mu}^{2}}{\sigma_{\mu}^{2}+\sigma_{\xi}^{2}}\left(y_{t}-\overline{y}-\rho\mu_{t-1}\right).$$