Chapter 4

Economic Voting

Economic voting comprises a substantial literature. A strand starting with Kramer (1983) and extending to work by Alesina and Rosenthal (1995), Suzuki and Chappell (1996), and Lin (1999) contributes to the value of the literature. These studies have refined earlier work and present models of voter sophistication and new applied statistical tests. In the former instance, voters possess the capability to deal with uncertainty in assigning blame or credit to incumbents for good or bad economic conditions. For the latter, applied statistical tests include some of the more advanced tools in time series analysis.

There is another important — EITM related — feature in this work. Some of these authors relate a measurement error problem to the voter capability noted above. This is exactly what EITM and methodological unification accomplish. The theory — the formal model — implies an applied statistical model with measurement error. Consequently, one can examine the joint effects by employing a unified approach.[1]

4.1 Step 1: Relating Expectations, Uncertainty, and Measurement Error

Earlier contributors have dealt with this "signal extraction" problem (See the Appendix, Section 4.53). Friedman (1957) and Lucas's (1973) substantive findings would not have been achieved had they treated their research question as a pure measurement error problem requiring only an applied statistical analysis (and "fix" for the measurement error). Indeed, both Friedman (1957) and Lucas (1973) linked specific empirical coefficients from their respective formal (behavioral) models: among their contributions was to merge "error in variables" regression with formal models of expectations and uncertainty. For Friedman, the expectations and uncertainty involve permanent-temporary confusion, while general-relative confusion is the behavioral mechanism in Lucas's model.

4.2 Step 2: Analogues for Expectations, Uncertainty, and Measurement Error

This chapter focuses on Alesina and Rosenthal's (1995) contribution. The formal model representing the behavioral concepts — expectations and uncertainty — is presented. Alesina and Rosenthal (1995) provide the formal model (pages 191-195). Their model of economic growth is based on an expectations augmented aggregate supply curve:

$$\hat{y}_t = \hat{y}^n + \gamma\left(\pi_t - \pi_t^e\right) + \varepsilon_t, \tag{4.2.1}$$

where $\hat{y}_t$ represents the rate of economic growth (GDP growth) in period t, $\hat{y}^n$ is the natural economic growth rate, π_t is the inflation rate at time t, and π_t^e is the expected inflation rate at time t formed at time $t-1$.

Having established voter inflation expectations the concept of uncertainty is next. We assume voters want to determine whether to attribute credit or blame for economic growth (y_t) outcomes to the incumbent administration. Yet, voters are faced with uncertainty in determining which part of the economic outcomes is due to incumbent "competence" (i.e., policy acumen) or simply good luck.

If the uncertainty is based, in part, from equation (4.2.1), then equation (4.2.2) presents the analogue. It is commonly referred to as a "signal extraction" or measurement error problem (See the Appendix, Section 4.53):

$$\varepsilon_t = \eta_t + \xi_t. \tag{4.2.2}$$

[1] Recall that applied statistical tools lack power in disentangling conceptually distinct effects on a dependent variable. This is noteworthy since the traditional applied statistical view of measurement error is that it creates parameter bias, with the typical remedy requiring the use of various estimation techniques (See the Appendix, Section 4.51) and Johnston and DiNardo (1997:153-159)).

The variable ε_t represents a "shock" comprised of the two unobservable characteristics noted above — competence or good luck. The first, represented by η_t, reflects "competence" attributed to the incumbent administration. The second, symbolized as ξ_t, are shocks to growth beyond administration control (and competence). Both η_t and ξ_t have zero mean with variance(s) σ_η^2 and σ_ξ^2 respectively. In less technical language Alesina and Rosenthal describe competence as follows:

> The term ξ_t represents economic shocks beyond the governments control, such as oil shocks and technological innovations. The term η_t captures the idea of government competence, that is the government's ability to increase the rate of growth without inflationary surprises. In fact, even if $\pi_t = \pi_t^e$, the higher is η_t the higher is growth, for a given ξ_t. We can think of this competence as the government's ability to avoid large scale inefficiencies, to promote productivity growth, to avoid waste in the budget process, so that lower distortionary taxes are needed to finance a given amount of government spending, etc (page 192).

Note also that competence can persist *and support* reelection. This feature is characterized as an MA(1) process:

$$\eta_t = \mu_t + \rho\mu_{t-1}, \quad 0 < \rho \leq 1 \tag{4.2.3}$$

where μ_t is $iid\left(0, \sigma_\mu^2\right)$. The parameter ρ represents the strength of the persistence. The lag or lags allow for retrospective voter judgments.

If we reference equation (4.2.1) again, let us assume voters' judgments include a general sense of the average rate of growth ($\hat{y}^n$) and the ability to observe actual growth ($\hat{y}_t$). Voters can evaluate their difference ($\hat{y}_t - \hat{y}^n$). Equation (4.2.1) also suggests that when voters predict inflation with no systematic error (i.e., $\pi_t^e = \pi_t$), the result is non-inflationary growth with no adverse real wage effect.

Next, economic growth performance is tied to voter uncertainty. Alesina and Rosenthal formalize how economic growth rate deviations from the average can be attributed to administration competence or fortuitous events:

$$\hat{y}_t - \hat{y}^n = \varepsilon_t = \eta_t + \xi_t. \tag{4.2.4}$$

Equation (4.2.4) shows when the actual economic growth rate is greater than its average or "natural rate" (i.e., $\hat{y}_t > \hat{y}^n$), then $\varepsilon_t = \eta_t + \xi_t > 0$. Again, the voters are faced with uncertainty in distinguishing the incumbent's competence (η_t) from the stochastic economic shock (ξ_t). However, because competence can persist, voters use this property for making forecasts and giving greater or lesser weight to competence over time.

This behavioral effect is demonstrated by substituting equation (4.2.3) in (4.2.4):

$$\mu_t + \xi_t = \hat{y}_t - \hat{y}^n - \rho\mu_{t-1}. \tag{4.2.5}$$

Equation (4.2.5) suggests that voters can observe the composite shock $\mu_t + \xi_t$ based on the observable variables, $\hat{y}_t$, $\hat{y}^n$, and μ_{t-1} which are available at time t and $t-1$. Determining the optimal estimate of competence, η_{t+1}, when the voters observe $\hat{y}_t$. Alesina and Rosenthal demonstrate this result making a one-period forecast of equation (4.2.3) and solving for its expected value (conditional expectation) at time t (See the Appendix, Section 4.52):

$$E_t\left(\eta_{t+1}\right) = E_t\left(\mu_{t+1}\right) + \rho E\left(\mu_t|\hat{y}_t\right) = \rho E\left(\mu_t|\hat{y}_t\right), \tag{4.2.6}$$

where $E_t\left(\mu_{t+1}\right) = 0$. Alesina and Rosenthal (1995) argue further that rational voters would not use $\hat{y}_t$ as the only variable to forecast η_{t+1}. Instead, they use all available information, including $\hat{y}^n$ and μ_{t-1}. As a result, a revised equation (4.2.6) is:

$$E_t\left(\eta_{t+1}\right) = E_t\left(\mu_{t+1}\right) + \rho E\left(\mu_t|\hat{y}_t - \hat{y}^n - \rho\mu_{t-1}\right) \tag{4.2.7}$$

$$= \rho E\left(\mu_t|\mu_t + \xi_t\right). \tag{4.2.8}$$

Using this analogue for expectations in equation 4.2.7, competence, η_{t+1}, can be forecasted by predicting $\mu_{t+1}$and μ_t. Since there is no information available for forecasting μ_{t+1}, rational voters can only forecast μ_t based on observable $\hat{y}_t - \hat{y}^n - \rho\mu_{t-1}$ (at time t and $t-1$) from equations 4.2.7 and 4.2.8.

4.3 Step 3: Unifying and Evaluating the Analogues

The method of recursive projection and equation (4.2.5) illustrates how the behavioral analogue for expectations is linked to the empirical analogue for measurement error (an error-in-variables "equation"):

$$E_t\left(\eta_{t+1}\right) = \rho E\left(\mu_t|\hat{y}_t\right) = \rho\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\xi^2}\left(\hat{y}_t - \hat{y}^n - \rho\mu_{t-1}\right), \tag{4.3.1}$$

where $0 < \rho\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\xi^2} < 1$. Equation (4.3.1) shows voters can forecast competence using the difference between $\hat{y}_t - \hat{y}^n$, but also the "weighted" lag of μ_t (i.e., $\rho\mu_{t-1}$).

In equation (4.3.1), the expected value of competence is *positively* correlated with economic growth rate deviations. Voter assessment is filtered by the coefficient, $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\xi^2}$, representing a proportion of competence voters are able to interpret and observe.

The behavioral implications are straightforward. If voters interpret that the variability of economic shocks come solely from the incumbent's competence (i.e., $\sigma_\xi^2 \to 0$), then $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\xi^2} \to 1$. On the other hand, the increase in the variability of uncontrolled shocks, σ_ξ^2, confounds the observability of incumbent competence since the signal-noise coefficient $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\xi^2}$ decreases. Voters assign less weight to economic performance in assessing the incumbent's competence.

Alesina and Rosenthal test the empirical implications of their theoretical model with U.S. data on economic outcomes and political parties for the period 1915 to 1988. They first use the growth equation (4.2.1) to collect the estimated exogenous shocks (ε_t) in the economy. With these estimated exogenous shocks, they then construct their variance-covariance structure.

Since competence (η_t) in equation (4.2.3) follows an MA(1) process, they hypothesize that a test for incumbent competence, as it pertains to economic growth, can be performed using the covariances between the current and preceding year. The specific test centers on whether the changes in covariances with the presidential party in office are statistically larger than the covariances associated with a change in presidential parties. They report null findings (e.g., equal covariances) and conclude that there is little evidence to support that voters are retrospective and use incumbent competence as a basis for support.

4.4 Leveraging EITM and Extending the Model

Alesina and Rosenthal provide an EITM connection between equations (4.2.1), (4.2.3) and their empirical tests. They link the behavioral concepts — expectations and uncertainty — with their respective analogues (conditional expectations and measurement error) and devise a signal extraction problem. While the empirical model resembles an error-in-variables specification, testable by dynamic methods such as rolling regression (Lin 1999), they instead estimate the variance-covariance structure of the residuals.

Their model is testable in other ways. We can, for example, leverage equation (4.3.1) and account for other forms of uncertainty. Suzuki and Chappell (1996) (and numerous others) provide such tests without any formalization. The formalization of Alesina and Rosenthal can be used and linked to Suzuki and Chappell's test.

Recall that the competence analogue (η_t) in their model is set up to be part of the aggregate supply (AS) shock ($\varepsilon_t = \eta_t + \xi_t$). Accordingly, competence (η_t) is defined as the incumbent's ability to promote economic growth via policies along the AS curve. Let us assume voters are sophisticated enough to not reward incumbent politicians for unusual economic growth resulting from an aggregate demand (AD) policy or shock. Rather, voters think the AS policy is the source of long-lasting (permanent) economic growth since it adds to productive capacity.[2] On the other hand, AD policy can at best produce temporary output gains and eventually leaves the economy with higher inflation.[3]

By leveraging the EITM framework, these studies lead to a direct relation between the parameters of the formal and empirical models. In particular, the competence equation (4.3.1) can be evaluated with the empirical tests and measures Suzuki and Chappell use for permanent and temporary changes in economic growth.

4.5 Appendix

The tools in this chapter are used to establish a transparent and testable relation between expectations (uncertainty) and forecast measurement error. The applied statistical tools provide a basic understanding of:

- Measurement error in a linear regression context — error-in-variables regression.

The formal tools include a presentation of:

- A linkage to linear regression.

[2]AS policies provide positive technology shocks. These policies range from government protection of property rights to the provision of public infrastructure.

[3]Achen (2012) adds yet another wrinkle to how competence is characterized. A key feature of his extension is to alter the MA(1) characterization by adding a constant term. This term signifies average competence and provides memory on incumbent administration competence. Achen's modification has important implications on how mypopic voters are and what circumstances can affect retrospection. Achen's work also opens the possibility for using an AR(1) process and he discusses this alternative.

- Linear projections.
- Recursive projections.

These tools, when unified, produce the following EITM relations consistent with research questions termed *signal extraction*. The last section of this appendix demonstrates signal extraction problems which are directly related to Alesina and Rosenthal's model and test.

4.5.1 Empirical Analogues

Measurement Error and Error in Variables Regression

In a regression model it is well known that endogeneity problems (e.g., a relation between the error term and a regressor) can be due to measurement error in the data. A regression model with mis-measured right-hand side variables gives least squares estimates with bias. The extent of the bias depends on the ratio of the variance of the signal (true variable) to the sum of the variance of the signal and the variance of the noise (measurement error). The bias increases when the variance of the noise becomes larger in relation to the variance of the signal. Hausman (2001: 58) refers to the estimation problem with measurement error as the "Iron Law of Econometrics" because the magnitude of the estimate is usually smaller than expected.

To demonstrate the downward bias consider the classical linear regression model with one independent variable:

$$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, ..., n \tag{4.5.1}$$

where ε_t are independent $N(0,\, \sigma_\varepsilon^2)$ random variables. The unbiased least squares estimator for regression model (4.5.1) is:

$$\hat{\beta}_1 = \left[\sum_{t=1}^{n}(x_t - \bar{x})^2\right]^{-1} \sum_{t=1}^{n}(x_t - \bar{x})(Y_t - \bar{Y}). \tag{4.5.2}$$

Now instead of observing x_t directly, observe its value with an error:

$$X_t = x_t + e_t, \tag{4.5.3}$$

where e_t is an $iid(0, \sigma_e^2)$ random variable. The simple linear error-in-variables model can be written as:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, ..., n \\ X_t &= x_t + e_t. \end{aligned} \tag{4.5.4}$$

In model (4.5.4), an estimate of a regression of Y_t on X_t, with an error term mixing the effects of the true error ε_t and the measurement error e_t is presented.[4] It follows that the vector (Y_t, X_t) is distributed as a bi-variate normal vector with mean vector and covariance matrix defined as (4.5.5) and (4.5.6), respectively:

$$E\left\{(Y, X)\right\} = (\mu_Y, \mu_X) = (\beta_0 + \beta_1\mu_x, \mu_x) \tag{4.5.5}$$

$$\begin{bmatrix} \sigma_Y^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_X^2 \end{bmatrix} = \begin{bmatrix} \beta_1^2\sigma_x^2 + \sigma_\varepsilon^2 & \beta_1\sigma_x^2 \\ \beta_1\sigma_x^2 & \sigma_x^2 + \sigma_e^2 \end{bmatrix} \tag{4.5.6}$$

The estimator for the slope coefficient when Y_t is regressed on X_t is:

$$\begin{aligned} E(\hat{\beta}_1) &= E\left\{\left[\sum_{t=1}^{n}(X_t - \bar{X})^2\right]^{-1} \sum_{t=1}^{n}(X_t - \bar{X})(Y_t - \bar{Y})\right\} \\ &= (\sigma_X^2)^{-1}\sigma_{XY} \\ &= \beta_1(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2}). \end{aligned} \tag{4.5.7}$$

[4] To demonstrate this results, we derive $Y_t = \beta_0 + \beta_1 X_t + (\varepsilon_t - \beta_1 e_t)$ from (4.5.4)). Assuming the $x_t's$ are random variables with $\sigma_x^2 > 0$ and $(x_t, \varepsilon_t, e_t)'$ are iid $N[(e_x, 0, 0)', diag(\sigma_x^2, \sigma_\varepsilon^2, \sigma_e^2)]$ where $diag(\sigma_x^2, \sigma_\varepsilon^2, \sigma_e^2)$ is a diagonal matrix with the given elements on the diagonal.

The resulting estimate is smaller in magnitude than the true value of β_1. The ratio of $\lambda = \frac{\sigma_x^2}{\sigma_X^2} = \frac{\sigma_x^2}{\sigma_x^2+\sigma_e^2}$ defines the degree of attenuation. In applied statistics, this ratio, λ, is termed the *reliability ratio*. A traditional applied statistical remedy is to use a "known" reliability ratio and weight the statistical model accordingly.[5] As presented above (4.5.7) the expected value of the least squares estimator of β_1 is the true β_1 multiplied by the reliability ratio, so it is possible to construct an unbiased estimator of β_1 if the ratio of λ is known.

4.5.2 Formal Analogues[6]

Least Squares Regression

Normally we think of least squares regression as an empirical tool, but in this case it serves as a bridge between the formal and empirical analogues ultimately creating a behavioral rationale for the ratio in equations (4.2.6) and (4.3.1). This section is a review following Sargent (1987: 223-229).

Assume there is a set of random variables, y, $x_1, x_{2,}, \ldots, x_n$. Consider that we estimate the random variable y which is expressed as a linear function of x_i:

$$\hat{y} = b_0 + b_1 x_1 + \cdots + b_n x_n, \tag{4.5.8}$$

where b_0 is the intercept of the linear function, and b_i presents the partial slope parameters on x_i, for $i = 1, 2, \ldots, n$. As a result, by choosing the b_i, $\hat{y}$ is the "best" linear estimate which minimizes the "distance" between y and $\hat{y}$:

$$\begin{aligned} &\min\nolimits_{a_i} E\left(y - \hat{y}\right)^2 \\ \Rightarrow\quad & E\left[y - \left(b_0 + b_1 x_1 + \cdots + b_n x_n\right)\right]^2, \end{aligned} \tag{4.5.9}$$

for all i. To minimize equation (4.5.9), a necessary and sufficient condition is (in the normal equation(s)):

$$E\left\{\left[y - \left(b_0 + b_1 x_1 + \cdots + b_n x_n\right)\right] x_i\right\} = 0 \tag{4.5.10}$$

$$E\left[\left(y - \hat{y}\right) x_i\right] = 0, \tag{4.5.11}$$

where $x_0 = 1$.

The condition expressed in equation (4.5.11) is called the *orthogonality principle*. It implies that the difference between observed y and the estimated y according to the linear function, $\hat{y}$, is not linearly dependent with x_i for $i = 1, 2, \ldots, n$.

Linear Projections

A least squares projection begins with:

$$y = \sum_{i=0}^{n} b_i x_i + \varepsilon, \tag{4.5.12}$$

where ε is the forecast error, $E\left(\varepsilon \sum b_i x_i\right) = 0$ and $E\left(\varepsilon x_i\right) = 0$, for $i = 0, 1, \cdots, n$. Note also that the random variable $\hat{y} = \sum_{i=0}^{n} b_i x_i$, is based on $b_i's$ chosen to satisfy the least squares orthogonality condition. This is called the projection of y on $x_0, x_1, \ldots, x_n$.

Mathematically, it is written:

$$\sum b_i x_i \equiv P\left(y \,|1, x_1, x_2, \cdots, x_n\,\right), \tag{4.5.13}$$

where $x_0 = 1$. Assuming orthogonality, the equation (4.5.10) can be rewritten as a set of normal equations:

$$\begin{bmatrix} Ey \\ Eyx_1 \\ Eyx_2 \\ \vdots \\ Eyx_n \end{bmatrix} = \begin{bmatrix} 1 & Ex_1 & Ex_2 & \cdots & Ex_n \\ Ex_1 & Ex_1^2 & Ex_1x_2 & \cdots & \\ Ex_2 & Ex_1x_2 & \ddots & & \\ \vdots & \vdots & & \ddots & \\ Ex_n & & & & Ex_n^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}. \tag{4.5.14}$$

[5] See Fuller (1987) for other remedies based on the assumption some of the parameters of the model are known or can be estimated (from outside sources). Alternatively, there are remedies which do not assume any prior knowledge for some of the parameters in the model (See Pal 1980).

[6] The following sections are based on Whittle (1963, 1983), Sargent (1987), and Woolridge (2008).

Given that the matrix of Ex_ix_j in equation (4.5.14) is invertible for $i, j \in \{1, 2, \dots, n\}$, and solving for each coefficient (b_i):

$$\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} = \left[Ex_ix_j\right]^{-1}\left[Eyx_k\right]. \tag{4.5.15}$$

Applying the above technique to a simple example:

$$y = b_0 + b_1x_1 + \varepsilon,$$

and:

$$\begin{bmatrix} Ey \\ Eyx_1 \end{bmatrix} = \begin{bmatrix} 1 & Ex_1 \\ Ex_1 & Ex_1^2 \end{bmatrix}\begin{bmatrix} b_0 \\ b_1 \end{bmatrix}. \tag{4.5.16}$$

Using normal equation(s), the following estimates are derived for the intercept and slope:

$$b_0 = Ey - b_1Ex_1,$$

and:

$$\begin{aligned} b_1 &= \frac{E\left(y - Ey\right)\left(x_1 - Ex_1\right)}{E\left(x_1 - Ex_1\right)^2} \\ &= \frac{\sigma_{x_1y}}{\sigma_{x_1}^2}, \end{aligned}$$

where σ_{x_1y} is the covariance between x_i and y, and $\sigma_{x_1}^2$ is the variance of x_1.[7]

Recursive Projections

The linear least squares identities can be used in formulating how agents update their forecasts (*expectations*). Recursive projections are a key element of deriving the optimal forecasts, such as the one shown in equation (4.3.1). These forecasts are updated consistent with the linear least squares rule described above. The simple univariate projection can be used (recursively) to assemble projections on many variables, such as $P\left(y\,|1, x_1, x_2, \cdots, x_n\right)$.

For example, when there are two independent variables, equation (4.5.13) can be rewritten for $n = 2$ as:

$$y = P\left(y\,|1, x_1, x_2\right) + \varepsilon, \tag{4.5.17}$$

[7] From equation (4.5.16), we derive a similar equation expressed in equation (4.5.15):

$$\begin{aligned} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} &= \begin{bmatrix} 1 & Ex_1 \\ Ex_1 & Ex_1^2 \end{bmatrix}^{-1}\begin{bmatrix} Ey \\ Eyx_1 \end{bmatrix} \\ &= \begin{bmatrix} Ex_1^2 & -Ex_1\left(Ex_1^2 - (Ex_1)^2\right)^{-1} \\ -Ex_1\left(Ex_1^2 - (Ex_1)^2\right)^{-1} & \left(Ex_1^2 - (Ex_1)^2\right)^{-1} \end{bmatrix}\begin{bmatrix} Ey \\ Eyx_1 \end{bmatrix}. \end{aligned}$$

b_1 can be expressed as:

$$\begin{aligned} b_1 &= -\frac{Ex_1}{Ex_1^2 - (Ex_1)^2}Ey + \frac{Eyx_1}{Ex_1^2 - (Ex_1)^2} \\ &= \frac{-Ex_1Ey + Eyx_1}{Ex_1^2 - (Ex_1)^2}. \end{aligned}$$

For simplicity, we assume $Ex_1 = 0$ and $Ey = 0$. Consequently:

$$\begin{aligned} b_1 &= \frac{-Ex_1Ey + Eyx_1}{Ex_1^2 - (Ex_1)^2} \\ &= \frac{Eyx_1}{Ex_1^2} \\ &= \frac{\sigma_{x_1y}}{\sigma_{x_1}^2}. \end{aligned}$$

implying:

$$y = b_0 + b_1x_1 + b_2x_2 + \varepsilon, \tag{4.5.18}$$

where $E\varepsilon = 0$. Assume that equations (4.5.17) and (4.5.18) satisfy the orthogonality conditions: $E\varepsilon x_1 = 0$ and $E\varepsilon x_2 = 0$. If we omit the information from x_2 to project y, then the projection of y can only be formed based on the random variable x_1:

$$P\left(y\left|1, x_1\right.\right) = b_0 + b_1x_1 + b_2P\left(x_2\left|1, x_1\right.\right). \tag{4.5.19}$$

In equation (4.5.19), $P\left(x_2\left|1, x_1\right.\right)$ is a component where x_2 is projected using 1 and x_1 to forecast y. Formally, equation (4.5.19) can be separated into three projections:

$$P\left(y\left|1, x_1\right.\right) = P\left(b_0\left|1, x_1\right.\right) + b_1P\left(x_1\left|1, x_1\right.\right) + b_2P\left(x_2\left|1, x_1\right.\right). \tag{4.5.20}$$

Equation (4.5.20) demonstrates that the projection of y given $(1, x_1)$ is a linear function of the three projections:[8]

$$\begin{aligned} P\left(b_0|1, x_1\right) &= b_0, \\ P\left(x_1|1, x_1\right) &= x_1, \text{ and} \\ P\left(\varepsilon|1, x_1\right) &= 0. \end{aligned}$$

An alternative expression is to rewrite the forecast error of y given x_1 as simply the "forecast" error of x_2 given x_1 and a stochastic error term ε. Mathematically, equation (4.5.18) is subtracted from equation (4.5.19):

$$y - P\left(y\left|1, x_1\right.\right) = b_2\left[x_2 - P\left(x_2\left|1, x_1\right.\right)\right] + \varepsilon, \tag{4.5.21}$$

and simplified to:

$$z = b_2w + \varepsilon,$$

where $z = y - P\left(y\left|1, x_1\right.\right)$, and $w = \left[x_2 - P\left(x_2\left|1, x_1\right.\right)\right]$. Note that $x_2 - P\left(x_2\left|1, x_1\right.\right)$ is also orthogonal to ε, such that, $E\left\{\varepsilon\left[x_2 - P\left(x_2\left|1, x_1\right.\right)\right]\right\} = 0$ or $E\left(\varepsilon w\right) = 0$.

Now writing the following expression as a projection of the forecast error of y that depends on the forecast error of x_2 given x_1:

$$P\left[y - P\left(y\left|1, x_1\right.\right)\left|x_2 - P\left(x_2\left|1, x_1\right.\right)\right.\right] = b_2\left[x_2 - P\left(x_2\left|1, x_1\right.\right)\right], \tag{4.5.22}$$

or in simplified form:

$$P\left(z|w\right) = b_2w.$$

By combining equations (4.5.21) and (4.5.22), the result is:

$$y = P\left(y\left|1, x_1\right.\right) + P\left[y - P\left(y\left|1, x_1\right.\right)\left|x_2 - P\left(x_2\left|1, x_1\right.\right)\right.\right] + \varepsilon. \tag{4.5.23}$$

Consequently, equation (4.5.23) can also be written as:

$$P\left(y\left|1, x_1, x_2\right.\right) = P\left(y\left|1, x_1\right.\right) + P\left[y - p\left(y\left|1, x_1\right.\right)\left|x_2 - P\left(x_2\left|1, x_1\right.\right)\right.\right], \tag{4.5.24}$$

where $P\left(y\left|1, x_1, x_2\right.\right)$ is called a bivariate projection. The univariate projections are given by:

$P\left(x_2\left|1, x_1\right.\right)$, $P\left(y\left|1, x_1\right.\right)$, and $P\left[y - P\left(y\left|1, x_1\right.\right)\left|x_2 - P\left(x_2\left|1, x_1\right.\right)\right.\right]$.

In this case, the bivariate projection equals three univariate projections. More importantly, equation (4.5.24) is useful for purposes of describing optimal updating (learning) by the least squares rule:

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

[8]The first two conditions can be interpreted as follows. First, when predicting a constant b_0 using 1 and x_1, we are still predicting a constant b_0. As a result, $P\left(b_0\left|1, x_1\right.\right) = b_0$. Second, when predicting x_1 using 1 and x_1, we can also predict x_1, which is $P\left(x_1\left|1, x_1\right.\right) = x_1$.

To show the results mathematically, rewrite the projection as the following linear function: $P\left(b_0\left|1, x_1\right.\right) = t_0 + t_1x_1$, where t_0 and t_1 are parameters. Using normal equations, we can derive t_0 and t_1: $t_0 = Eb_0 - t_1Ex_1$, and $t_1 = \frac{E(b_0-Eb_0)(x_1-Ex_1)}{E(x_1-Ex_1)^2}$. Since $Eb_0 = b_0$, then: $t_1 = \frac{E(b_0-Eb_0)(x_1-Ex_1)}{E(x_1-Ex_1)^2} = 0$, and $t_0 = Eb_0 = b_0$. Therefore, $P\left(b_0\left|1, x_1\right.\right) = t_0 + t_1x_1 = b_0$.

For $P\left(x_1\left|1, x_1\right.\right) = x_1$, we perform the same operations: $P\left(x_1\left|1, x_1\right.\right) = t_0 + t_1x_1$. Now $t_0 = Ex_1 - t_1Ex_1$, and $t_1 = \frac{E(x_1-Ex_1)(x_1-Ex_1)}{E(x_1-Ex_1)^2} = \frac{E(x_1-Ex_1)^2}{E(x_1-Ex_1)^2} = 1$. Therefore $t_0 = Ex_1 - Ex_1 = 0$, and $P\left(x_1\left|1, x_1\right.\right) = t_0 + t_1x_1 = 0 + x_1 = x_1$. As a result, $P\left(x_1\left|1, x_1\right.\right) = x_1$.

We rely on the orthogonality condition for the last expression: $E\left(\varepsilon\right) = E\left(\varepsilon x_1\right) = 0$. This gives us $P\left(\varepsilon\left|1, x_1\right.\right) = t_0 + t_1x_1$. Now $t_0 = E\varepsilon - t_1Ex_1$ and:$t_1 = \frac{E(\varepsilon-E\varepsilon)(x_1-Ex_1)}{E(x_1-Ex_1)^2} = \frac{E(\varepsilon x_1-\varepsilon Ex_1-E\varepsilon x_1+E\varepsilon x_1)}{E(x_1-Ex_1)^2} = 0$. Since $t_1 = 0$, we find $t_0 = E\varepsilon - t_1Ex_1 = E\varepsilon = 0$. Therefore, $P\left(\varepsilon\left|1, x_1\right.\right) = 0$.

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

If initially we have data only on a random 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:

$$P(y|1,x_1) = b_0 + b_1x_1 + b_2P(x_2|1,x_1). \tag{4.5.25}$$

Intuitively, we forecast y based on two components: (i) b_1x_1 alone, and (ii) $P(x_2|1,x_1)$, that is, the forecast of x_2 given x_1. When an observation x_2 becomes available, according to equation (4.5.24), the 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 $x_2-P(x_2|1,x_1)$.

In equation (4.5.24), $P(y|1,x_1)$ is interpreted as the original forecast, $y-P(y|1,x_1)$ is the forecast error of y, given x_1, and $x_2-P(x_2|1,x_1)$ is the forecast error of x_2 to forecast the forecast error of y given x_1. The above concept can be summarized in a general expression:

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

where Ω is the original information, x is the new information, and $P(y|\Omega)$ is the prediction of y using the original information. The projection, $P\{y-P(y|\Omega)|x-P(x|\Omega)\}$, indicates new information has become available to update the forecast. It is no longer necessary to use the original information to make predictions. In other words, one can obtain $x-P(x|\Omega)$, the difference between the new information and the "forecasted" new information, to predict the error of y: $y-P(y|\Omega)$.

4.5.3 Signal-Extraction Problems

Based on these tools it can now be demonstrated how conditional expectations with recursive projections has a mutually reinforcing relation with measurement error and error-in-variables regression. There are many examples of this "EITM-like" linkage and they generally fall under the umbrella of signal extraction problems. Consider the following examples.[9]

Application 1: Measurement Error

Suppose a random variable x^* is an indepenent variable. However, measurement error, e, exists so that the variable x is only observable:

$$x = x^* + e, \tag{4.5.26}$$

where x^* and e have zero mean, finite variance, and $Ex^*e=0$. Therefore, the projection of x^* given an observable x is:

$$P(x^*|1,x) = b_0 + b_1x.$$

Based on the least squares and the orthogonality conditions, we have:

$$b_1 = \frac{E(xx^*)}{Ex^2} = \frac{E[(x^*+e)x^*]}{E(x^*+e)^2} = \frac{E(x^*)^2}{E(x^*)^2+Ee^2}, \tag{4.5.27}$$

and

$$b_0 = 0. \tag{4.5.28}$$

The projection of x^* given x can be written as:

$$P(x^*|1,x) = \frac{E(x^*)^2}{E(x^*)^2+Ee^2}x, \tag{4.5.29}$$

where $b_1 = \frac{E(x^*)^2}{E(x^*)^2+Ee^2}$ is between zero and one.

The "measurement error" attenuation is now transparent. As $\frac{E(x^*)^2}{Ee^2}$ increases, $b_1 \to 1$: the greater $\frac{E(x^*)^2}{Ee^2}$ is, the larger the fraction of variance in x is due to variations in the actual value (i.e., $E(x^*)^2$).

[9] The first example can be found in Sargent (1987: 229).

Application 2: The Lucas (1973) Model (Relative-General Uncertainty)

An additional application is the case where there is general-relative confusion. Here, using Lucas's (1973) supply curve, producers observe the prices of their own goods (p_i) but not the aggregate price level (p).

The relative price of good i is r_i is defined as:

$$r_i = p_i - p. \tag{4.5.30}$$

The observable price p_i is a sum of the aggregate price level and its relative price:

$$p_i = p + (p_i - p) = p + r_i. \tag{4.5.31}$$

Assume each producer wants to estimate the real relative price r_i to determine their output level. However, they do not observe the general price level. As a result, the producer forms the following projection of r_i given p_i:

$$P\left(r_i \left| p_i \right.\right) = b_0 + b_1 p_i. \tag{4.5.32}$$

According to (4.5.32), the values of b_0 and b_1 are:

$$b_0 = E\left(r_i\right) - b_1 E\left(p_i\right) = E\left(p_i - p\right) - b_1 E\left(p_i\right) = -b_1 E\left(p_i\right), \tag{4.5.33}$$

and:

$$\begin{aligned} b_1 &= \frac{E\left[r_i - E\left(r_i\right)\right]\left[p_i - E\left(p_i\right)\right]}{E\left[p_i - E\left(p_i\right)\right]^2} \\ &= \frac{E\left[r_i - E\left(r_i\right)\right]\left[\left(p + r_i\right) - E\left(p + r_i\right)\right]}{E\left[\left(p + r_i\right) - E\left(p + r_i\right)\right]^2} \\ &= \frac{Er_i^2}{Er_i^2 + Ep^2} && (4.5.34) \\ &= \frac{v_r}{v_r + v_p}, && (4.5.35) \end{aligned}$$

where $v_r = Er_i^2$ is the variance of the real relative price, and $v_p = Ep^2$ is the variance of the general price level. Inserting the values of $b_0 = -b_1 E\left(p\right)$ and b_1 into the projection (4.5.32), we have:

$$P\left(r_i \left| p_i \right.\right) = b_1\left[p_i - E\left(p\right)\right] = \frac{v_r}{v_r + v_p}\left[p_i - E\left(p\right)\right]. \tag{4.5.36}$$

Next factoring in an output component — the labor supply — and showing it is increasing with the projected relative price we have:

$$l_i = \beta E\left(r_i \left| p_i \right.\right), \tag{4.5.37}$$

and:

$$l_i = \frac{\beta v_r}{v_r + v_p}\left[p_i - E\left(p\right)\right]. \tag{4.5.38}$$

If aggregated over all producers and workers, the average aggregate production is:

$$y = b\left[p - E\left(p\right)\right], \tag{4.5.39}$$

where $b = \frac{\beta v_r}{v_r + v_p}$.

Lucas's (1973) empirical tests are directed at output-inflation trade-offs in a variety of countries.[10] Equation (4.5.39) represents the mechanism of the general-relative price confusion:

$$y = \beta\frac{v_r}{v_r + v_p}\left[p - E\left(p\right)\right], \tag{4.5.40}$$

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

[10]The empirical tests are described in Romer (1996: 253-254).

Application 3: The Derivation of the Optimal Forecast of Political Incumbent Competence

This application uses the techniques of recursive projections and signal extraction to derive the optimal forecast of political incumbent competence in equation (4.3.1). In Section 4.2, the public's conditional expectations of an incumbent's competence at time $t+1$ (as expressed in equations (4.2.7) and(4.2.8)) is:

$$\begin{aligned} E_t(\eta_{t+1}) &= E_t(\mu_{t+1}) + \rho E(\mu_t | \hat{y}_t - \hat{y}^n - \rho\mu_{t-1}) \\ E_t(\eta_{t+1}) &= \rho E(\mu_t | \mu_t + \xi_t), \end{aligned} \tag{4.5.41}$$

where $E_t(\mu_{t+1}) = 0$.

Using recursive projections, voters forecast μ_t using $\mu_t + \xi_t$ and obtain the forecasting coefficients a_0 and a_1:

$$P(\mu_t | \mu_t + \xi_t) = a_0 + a_1(\mu_t + \xi_t), \tag{4.5.42}$$

with:

$$\begin{aligned} a_1 &= \frac{cov(\mu_t, \mu_t + \xi_t)}{var(\mu_t + \xi_t)} \\ &= \frac{E(\mu_t(\mu_t + \xi_t))}{E[(\mu_t + \xi_t)(\mu_t + \xi_t)]} \\ &= \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\xi^2}, \end{aligned}$$

and:

$$a_0 = E(\mu_t) - a_1 E(\mu_t + \xi_t) = 0,$$

where $E(\mu_t) = E(\mu_t + \xi_t) = 0$. The projection for μ_t is written as:

$$\begin{aligned} E_t(\mu_t | \mu_t + \xi_t) = P(\mu_t | \mu_t + \xi_t) &= a_0 + a_1(\mu_t + \xi_t) \\ &= \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\xi^2}(\mu_t + \xi_t). \end{aligned} \tag{4.5.43}$$

Placing equation (4.2.5) into equation (4.5.43):

$$E_t(\mu_t | \mu_t + \xi_t) = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\xi^2}(\hat{y}_t - \hat{y}^n - \rho\mu_{t-1}). \tag{4.5.44}$$

The final step is inserting equation (4.5.44) in equation (4.5.41) and obtaining the optimal forecast of competence at $t+1$:

$$\begin{aligned} E_t(\eta_{t+1}) &= \rho E(\mu_t | \mu_t + \xi_t) \\ &= \rho \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\xi^2}(\hat{y}_t - \hat{y}^n - \rho\mu_{t-1}). \end{aligned}$$

This is the expression in equation (4.3.1).