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:

\[ \hat{y}_t = \hat{y}^n + \gamma (\pi_t - \pi_t^e) + \varepsilon_t, \]  

(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, \( \pi_t \) is the inflation rate at time \( t \), and \( \pi_t^e \) 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):

\[ \varepsilon_t = \eta_t + \xi_t. \]  

(4.2.2)

¹Recall 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 method of recursive projection and equation there is no information available for forecasting to forecast where
its expected value (conditional expectation) at time observe and giving greater or lesser weight to competence over time.
Equation (4.2.5) suggests that voters can observe the composite shock and giving greater or lesser weight to competence over time.
then Equation t
µ
⇠
t
⌘
2
⇠
t
⌘
1
⇠
\( \hat{y}_t = \hat{y}^n = \hat{\eta}_t + \hat{\xi}_t. \) 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 \( \hat{\epsilon}_t = \hat{\eta}_t + \hat{\xi}_t > 0 \). Again, the voters are faced with uncertainty in distinguishing the incumbent’s competence \( \hat{\eta}_t \) from the stochastic economic shock \( \hat{\xi}_t \). 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):
\[ \mu_t + \xi_t = \hat{y}_t - \hat{y}^n - \rho \mu_{t-1}. \] 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 \( \mu_{t-1} \) which are available at time \( t \) and \( t-1 \). Determining the optimal estimate of competence, \( \hat{\eta}_{t+1} \), when the voters observe \( \hat{y}_t \). Alesina and Rosenthal demonstrate this result making a one-period forecast of equation (4.2.5) and solving for its expected value (conditional expectation) at time \( t \) (See the Appendix, Section 4.52):
\[ E_t (\eta_{t+1}) = E_t (\mu_{t+1}) + \rho E (\mu_t | \hat{y}_t) = \rho E (\mu_t | \hat{y}_t), \] where \( E_t (\mu_{t+1}) = 0 \). Alesina and Rosenthal (1995) argue further that rational voters would not use \( \hat{y}_t \) as the only variable to forecast \( \eta_{t+1} \). Instead, they use all available information, including \( \hat{y}^n \) and \( \mu_{t-1} \). As a result, a revised equation (4.2.6) is:
\[ E_t (\eta_{t+1}) = E_t (\mu_{t+1}) + \rho E (\mu_t | \hat{y}_t - \hat{y}^n - \rho \mu_{t-1}) \]
\[ = \rho E (\mu_t | \mu_t + \xi_t). \] Using this analogue for expectations in equation 4.2.7, competence, \( \eta_{t+1} \), can be forecasted by predicting \( \mu_{t+1} \) and \( \mu_t \). Since there is no information available for forecasting \( \mu_{t+1} \), rational voters can only forecast \( \mu_t \) 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”):
\[ E_t (\eta_{t+1}) = \rho E (\mu_t | \hat{y}_t) = \rho \frac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2_\xi} (\hat{y}_t - \hat{y}^n - \rho \mu_{t-1}), \]
where \(0 < \rho \frac{\sigma^2}{\sigma^2 + \sigma_\xi^2} < 1\). 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 \(\mu_t\) (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, \(\frac{\sigma^2}{\sigma^2 + \sigma_\xi^2}\), 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., \(\sigma^2_\xi \rightarrow 0\)), then \(\frac{\sigma^2}{\sigma^2 + \sigma_\xi^2} \rightarrow 1\). On the other hand, the increase in the variability of uncontrolled shocks, \(\sigma^2_\xi\), confounds the observability of incumbent competence since the signal-noise coefficient \(\frac{\sigma^2}{\sigma^2 + \sigma_\xi^2}\) 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 (\(\varepsilon_i\)) in the economy. With these estimated exogenous shocks, they then construct their variance-covariance structure.

Since competence (\(\eta_t\)) 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 (\(\eta_t\)) in their model is set up to be part of the aggregate supply (AS) shock (\(\varepsilon_\xi = \eta_t + \xi_t\)). Accordingly, competence (\(\eta_t\)) 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:


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 myopic 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:

\[ Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, ..., n \]  \hspace{1cm} (4.5.1)

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

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

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

\[ X_t = x_t + e_t, \]  \hspace{1cm} (4.5.3)

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

\[
\begin{align*}
Y_t &= \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, ..., n \\
X_t &= x_t + e_t.
\end{align*}
\]  \hspace{1cm} (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 \( \varepsilon_t \) 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:

\[
E \{ (Y, X) \} = (\mu_Y, \mu_X) = (\beta_0 + \beta_1 \mu_x, \mu_x)
\]  \hspace{1cm} (4.5.5)

\[
\begin{bmatrix}
\sigma^2_Y & \sigma_{XY} \\
\sigma_{XY} & \sigma^2_X
\end{bmatrix} =
\begin{bmatrix}
\beta^2_1 \sigma^2_x + \sigma^2_\varepsilon & \beta_1 \sigma^2_\varepsilon \\
\beta_1 \sigma^2_\varepsilon & \beta_1^2 \sigma^2_x + \sigma^2_\varepsilon
\end{bmatrix}
\]  \hspace{1cm} (4.5.6)

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

\[
E(\hat{\beta}_1) = E \left\{ \left( \sum_{t=1}^{n} (X_t - \bar{X})^2 \right)^{-1} \sum_{t=1}^{n} (X_t - \bar{X})(Y_t - \bar{Y}) \right\}
\]  \hspace{1cm} (4.5.7)

\[
= (\sigma^2_X)^{-1} \sigma_{XY}
\]

\[
= \hat{\beta}_1 \left( \frac{\sigma^2_x}{\sigma^2_x + \sigma^2_\varepsilon} \right).
\]

---

\(^4\)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 \( z_t's \) are random variables with \( \sigma^2_z > 0 \) and \( (x_t, \varepsilon_t, e_t) \) are iid \( N(\{0, 0, 0\}', \text{diag}(\sigma^2_z, \sigma^2_\varepsilon, \sigma^2_e)) \) where \( \text{diag}(\sigma^2_z, \sigma^2_\varepsilon, \sigma^2_e) \) is a diagonal matrix with the given elements on the diagonal.
The resulting estimate is smaller in magnitude than the true value of \( \beta_1 \). The ratio of \( \lambda = \frac{\sigma_x^2}{\sigma_y^2} = \frac{\sigma_z^2}{\sigma_y^2} \) defines the degree of attenuation. In applied statistics, this ratio, \( \lambda \), 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 \( \beta_1 \) is the true \( \beta_1 \) multiplied by the reliability ratio, so it is possible to construct an unbiased estimator of \( \beta_1 \) if the ratio of \( \lambda \) 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 \( x_i \):

\[
\hat{y} = b_0 + b_1 x_1 + \cdots + b_n x_n, \quad (4.5.8)
\]

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 \( \hat{y} \):

\[
\min \{ E[(y - \hat{y})^2] \} \Rightarrow E[(y - (b_0 + b_1 x_1 + \cdots + b_n x_n))^2], \quad (4.5.9)
\]

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

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

\[
E \{ (y - \hat{y}) x_i \} = 0, \quad (4.5.11)
\]

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:

\[
y = \sum_{i=0}^{n} b_i x_i + \varepsilon, \quad (4.5.12)
\]

where \( \varepsilon \) is the forecast error, \( E(\varepsilon \sum b_i x_i) = 0 \) and \( E(\varepsilon x_i) = 0 \), for \( i = 0, 1, \ldots, n \). 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:

\[
\sum_{i=0}^{n} b_i x_i \equiv P (y | 1, x_1, x_2, \ldots, x_n), \quad (4.5.13)
\]

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

\[
\begin{bmatrix}
\bar{E}y \\
\bar{E}yx_1 \\
\vdots \\
\bar{E}yx_n
\end{bmatrix}
= \begin{bmatrix}
1 & E_{x1} & E_{x2} & \cdots & Ex_n \\
E_{x1} & E_{x1}^2 & E_{x1x2} & \cdots & \\
\vdots & \vdots & \vdots & \ddots & \\
E_{xn} & E_{x1xn} & E_{x2x2} & \cdots & E_{xn}^2
\end{bmatrix}
\begin{bmatrix}
b_0 \\
b_1 \\
\vdots \\
b_n
\end{bmatrix}, \quad (4.5.14)
\]

\(^5\)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, 1963), Sargent (1987), and Woolridge (2008).
Given that the matrix of $Ex_i x_j$ in equation (4.5.14) is invertible for $i, j \in \{1, 2, \ldots, n\}$, and solving for each coefficient ($b_i$):

$$\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} = [Ex_i x_j]^{-1} [Ey x_k]. \quad (4.5.15)$$

Applying the above technique to a simple example:

$$y = b_0 + b_1 x_1 + \varepsilon,$$

and:

$$\begin{bmatrix} Ey \\ Ey x_1 \end{bmatrix} = \begin{bmatrix} 1 & Ex_1 \\ Ex_1 & Ex_1^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}. \quad (4.5.16)$$

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

$$b_0 = Ey - b_1 Ex_1,$$

and:

$$b_1 = \frac{E (y - Ey) (x_1 - Ex_1)}{E (x_1 - Ex_1)^2} = \frac{\sigma_{x_1 y}}{\sigma_{x_1}^2},$$

where $\sigma_{x_1 y}$ 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 $P(y|1, x_1, x_2, \cdots, x_n)$.

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

$$y = P(y|1, x_1, x_2) + \varepsilon, \quad (4.5.17)$$

\(^7\)From equation (4.5.16), we derive a similar equation expressed in equation (4.5.15):

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 & Ex_1 \\ Ex_1 & Ex_1^2 \end{bmatrix}^{-1} \begin{bmatrix} Ey \\ Ey x_1 \end{bmatrix} = \begin{bmatrix} \frac{Ey}{Ex_1} \\ \frac{Ey x_1}{Ex_1^2 - (Ex_1)^2} \end{bmatrix}.$$  

$b_1$ can be expressed as:

$$b_1 = \frac{Ex_1}{Ex_1^2 - (Ex_1)^2} Ey + \frac{Ey x_1}{Ex_1^2 - (Ex_1)^2} = \frac{-Ex_1 Ey + Ey x_1}{Ex_1^2 - (Ex_1)^2}.$$  

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

$$b_1 = \frac{-Ex_1 Ey + Ey x_1}{Ex_1^2 - (Ex_1)^2} = \frac{Ey x_1}{Ex_1^2} = \frac{\sigma_{x_1 y}}{\sigma_{x_1}^2}.$$
implying:

\[ y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon, \tag{4.5.18} \]

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 \( x_1 \):

\[ P(y|1, x_1) = b_0 + b_1 x_1 + b_2 P(x_2|1, x_1). \tag{4.5.19} \]

In equation (4.5.19), \( P(x_2|1, x_1) \) 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:

\[ P(y|1, x_1) = P(\text{b}0|1, x_1) + b_1 P(x_1|1, x_1) + b_2 P(x_2|1, x_1). \tag{4.5.20} \]

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

\[
\begin{align*}
P(\text{b}0|1, x_1) & = b_0, \\
P(x_1|1, x_1) & = x_1, \text{ and} \\
P(\varepsilon|1, x_1) & = 0.
\end{align*}
\]

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 \( \varepsilon \). Mathematically, equation (4.5.18) is subtracted from equation (4.5.19):

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

and simplified to:

\[ z = b_2 w + \varepsilon, \]

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

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

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

or in simplified form:

\[ P(z|w) = b_2 w. \]

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

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

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

\[ P(y|1, x_1, x_2) = P(y|1, x_1) + P[y - P(y|1, x_1) | x_2 - P(x_2|1, x_1)], \tag{4.5.24} \]

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

\[ P(x_2|1, x_1), \ P(y|1, x_1), \text{ and } P[y - P(y|1, x_1) | x_2 - P(x_2|1, x_1)]. \]

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:

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

\[ \text{The first two conditions can be interpreted as follows. First, when predicting a constant } b_0 \text{ using 1 and } x_1, \text{ we are still predicting a constant } b_0. \text{ As a result, } P(b_0|1, x_1) = b_0. \text{ Second, when predicting } x_1 \text{ using 1 and } x_1, \text{ we can also predict } x_1, \text{ which is } P(x_1|1, x_1) = x_1. \text{ To show the results mathematically, rewrite the projection as the following linear function: } P(b_0|1, x_1) = t_0 + t_1 x_1, \text{ where } t_0 \text{ and } t_1 \text{ are parameters. Using normal equations, we can derive } t_0 \text{ and } t_1: t_0 = E b_0 - t_1 E x_1, \text{ and } t_1 = E(b_0 - E b_0 | x_1 - E x_1). \text{ Since } E b_0 = b_0, \text{ then: } t_1 = E(b_0 - E b_0 | x_1 - E x_1) = 0, \text{ and } t_0 = E b_0 = b_0. \text{ Therefore, } P(b_0|1, x_1) = t_0 + t_1 x_1 = b_0. \]

For \( P(x_1|1, x_1) = x_1, \) we perform the same operations: \( P(x_1|1, x_1) = t_0 + t_1 x_1. \) Now \( t_0 = E x_1 - t_1 E x_1, \text{ and } t_1 = E(x_1 - E x_1) E(x_1 - E x_1)^2 = 1. \text{ Therefore } t_0 = E x_1 - E x_1 = 0, \text{ and } P(x_1|1, x_1) = t_0 + t_1 x_1 = 0 + x_1 = x_1. \text{ As a result, } P(x_1|1, x_1) = x_1. \]

We rely on the orthogonality condition for the last expression: \( E(\varepsilon | (x_1 - E x_1) = 0. \) This gives us \( P(\varepsilon|1, x_1) = t_0 + t_1 x_1. \) Now \( t_0 = E E - t_1 E x_1 \text{ and } t_1 = E(x_1 - E x_1) E(x_1 - E x_1)^2 = 0. \text{ Since } t_1 = 0, \text{ we find } t_0 = E E - t_1 E x_1 = E E = 0. \text{ Therefore, } P(\varepsilon|1, x_1) = 0. \]
where \( y - P(y \mid 1, x_1) \) is interpreted as the prediction error of \( y \) given \( x_1 \), and \( x_2 - P(x_2 \mid 1, x_1) \) 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 \( P(y \mid 1, x_1) \) and \( P(x_2 \mid 1, x_1) \) respectively:

\[
P(y \mid 1, x_1) = b_0 + b_1 x_1 + b_2 P(x_2 \mid 1, x_1).
\] (4.5.25)

Intuitively, we forecast \( y \) based on two components: (i) \( b_1 x_1 \) alone, and (ii) \( P(x_2 \mid 1, x_1) \), 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 \( P(y \mid 1, x_1) \), and the projection of unobserved “forecast error” \( y - P(y \mid 1, x_1) \) on the observed forecast error \( x_2 - P(x_2 \mid 1, x_1) \).

In equation (4.5.24), \( P(y \mid 1, x_1) \) is interpreted as the original forecast, \( y - P(y \mid 1, x_1) \) is the forecast error of \( y \), given \( x_1 \), and \( x_2 - P(x_2 \mid 1, x_1) \) 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:

\[
P(y \mid \Omega, x) = P(y \mid \Omega) + P\{y - P(y \mid \Omega) \mid x - P(x \mid \Omega)\},
\]

where \( \Omega \) is the original information, \( x \) is the new information, and \( P(y \mid \Omega) \) is the prediction of \( y \) using the original information. The projection, \( P\{y - P(y \mid \Omega) \mid x - P(x \mid \Omega)\} \), 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 \( x - P(x \mid \Omega) \), the difference between the new information and the “forecasted” new information, to predict the error of \( y: y - P(y \mid \Omega) \).

### 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 independent variable. However, measurement error, \( e \), exists so that the variable \( x \) is only observable:

\[
x = x^* + e,
\] (4.5.26)

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

\[
P(x^* \mid 1, x) = b_0 + b_1 x.
\]

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

\[
b_1 = \frac{E(x x^*)}{E x^2} = \frac{E[(x^* + e) x^*]}{E(x^* + e)^2} = \frac{E(x^*)^2}{E(x^*)^2 + E e^2};
\] (4.5.27)

and

\[
b_0 = 0.
\] (4.5.28)

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

\[
P(x^* \mid 1, x) = \frac{E(x^*)^2}{E(x^*)^2 + E e^2} x,
\] (4.5.29)

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

The “measurement error” attenuation is now transparent. As \( \frac{E(x^*)^2}{E e^2} \) increases, \( b_1 \rightarrow 1 \): the greater \( \frac{E(x^*)^2}{E e^2} \) 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).
CHAPTER 4. ECONOMIC VOTING

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 \( r_i \) defined as:

\[
r_i = p_i - p.
\]

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

\[
p_i = p + (p_i - p) = p + r_i.
\]

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 \( p_i \):

\[
P(r_i | p_i) = b_0 + b_1 p_i.
\]

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

\[
b_0 = E(r_i) - b_1 E(p_i) = E(p_i - p) - b_1 E(p_i) = -b_1 E(p_i),
\]

and:

\[
b_1 = \frac{E[r_i - E(r_i)][p_i - E(p_i)]}{E[p_i - E(p_i)]^2} = \frac{E[r_i - E(r_i)][(p + r_i) - E(p + r_i)]}{E[(p + r_i) - E(p + r_i)]^2} = \frac{E r_i^2}{E r_i^2 + E p^2} = \frac{v_r}{v_r + v_p},
\]

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

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

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

\[
l_i = \beta E(r_i | p_i),
\]

and:

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

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

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

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:

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

where \( v_p \) is the variance of the nominal demand shock, and \( p - E(p) \) 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:

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

$$
E_t (\eta_{t+1}) = \rho E (\mu_t | \mu_t + \xi_t),
$$

(4.5.41)

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

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

$$
P (\mu_t | \mu_t + \xi_t) = a_0 + a_1 (\mu_t + \xi_t),
$$

(4.5.42)

with:

$$
a_1 = \frac{\text{cov} (\mu_t, \mu_t + \xi_t)}{\text{var} (\mu_t + \xi_t)}
= \frac{E (\mu_t | \mu_t + \xi_t))}{E [((\mu_t + \xi_t) (\mu_t + \xi_t)]}
= \frac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2_\xi},
$$

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 $\mu_t$ is written as:

$$
E_t (\mu_t | \mu_t + \xi_t) = P (\mu_t | \mu_t + \xi_t) = a_0 + a_1 (\mu_t + \xi_t)
= \frac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2_\xi} (\mu_t + \xi_t).
$$

(4.5.43)

Placing equation (4.2.5) into equation (4.5.43):

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

(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 $t + 1$:

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

This is the expression in equation (4.3.1).