Importants to this school of thought include Butterworth, 1972; Feltham, 1972; Jiri, 1975; Beaver, 1966 and Christianen and Demski, 2002.)

The advantage of this view is it forces us to think in terms of complements and substitutes when dealing with this vast array of sources, and to look for economic forces that drive the disparity that bedevils the measurement school. And it is here that the comparative advantage of the accounting channel comes into focus: it is purposely designed and managed so that it is difficult to manipulate (Jiri, 1975). This is why it often resists to historical-cost measurement, as this removes major elements of subjectivity and manipulation potential. It is also why, in organized financial markets, most valuation information arrives before the firm's financial reports; and in this sense the financial reports provide a vescity check on the earlier reporting sources. In addition, cost allocation now enters as a natural phenomenon, either as a simple scaling device or to use an analogy with informationally efficient markets - as a cousin to an information-based pricing kernel in a financial market (Christianen and Demski, 2002; Ross, 2004).

Libraries are organized in coordinated fashion, as are phone books; and the same can be said about accounting. A curiosity is the political side of the regulatory apparatus. It is difficult, for example, for the incumbent government to alter a government-provided statistical series, yet it is routine for the incumbent government to intervene in the accounting regulatory process. A second curiosity is the seemingly perpetual feature of financial reporting frauds (Demski, 2003), although at the micro level it is well understood that opportunistic reporting is part of the game. For example, an ability to shift income to a later or an earlier period may be an inexpensive signal or, to speak more cynically, less costly to the firm than shifting real resources.

The disadvantage of the information school is its sheer breadth. The institutional context includes a vast array of information sources and actors, and sorting out first-order effects remains problematic.

Conclusion

Accounting, then, is simultaneously an important source of economic data and a collection of institutional and regulatory social science. We need economic context with yet another venue for documentation and exploration of economic forces. Why do we see episodic regulatory interventions? Why do we see forecasts of forthcoming accounting measures? Why do we not see supplementary accounting research? Why do we not see the mix of historical-cost and market values that characterize modern financial reporting? Questions of this sort motivate much of the current research in accounting and finance.

See also assets and liabilities; capital measurement; cost functions; depreciation; double-entry bookkeeping; human capital; measurement; pensions; present value.

Helpful comments by Hsij Lin and David Sappging are gratefully acknowledged.

Bibliography


adaptive estimation

An adaptive estimator is an efficient estimator for a model that is only partially specified. For example, consider estimating a parameter that describes a sample of observations drawn from a distribution $F$. One natural question is: is it possible that an estimator of the parameter constructed without knowledge of $F$ could be as efficient (asymptotically) as any well-behaved estimator that relies on knowledge of $F$? For some problems the answer is 'yes', and the estimator that is efficient is termed an adaptive estimator.

Consider the familiar scalar linear regression model (in which we let $i$ rather than $x$ index observations):

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where the regressor is exogenous and $(\epsilon_i)_{i=1}^{n}$ is a sequence of i.i.d. independent and identically distributed random variables with distribution $F$. The parameter vector $\beta = (\beta_0, \beta_1)$ is of interest rather than the distribution of the error term. Assume that $F$ is described by a parameter vector $\lambda$ (that is, we parameterize the distribution), then the resultant (maximum likelihood or MLE) estimator for $\lambda$ is semiparametric. Because the OLS estimator does not require that we parameterize $F$, the OLS estimator is semiparametric. If the population error distribution is Gaussian, we know that the OLS estimator is equivalent to the ML estimator, and is efficient. Although the OLS estimator is generally inefficient if $F$ is not Gaussian, it may be possible to construct an alternative (semiparametric) estimator that retains asymptotic efficiency if $F$ is not Gaussian. If we find that, for a family of distributions that includes the Gaussian, this estimator is asymptotically equivalent to the ML estimator, then this estimator is adaptive for that family.

The question then is how can we verify that an estimator is adaptive? As there will generally be an arbitrarily large number of distributions in the family, it is not feasible to algebraically verify asymptotic equivalence for each distribution. In a creative paper, Stein (1956) first proposed a solution to this problem. Let $\{F_{\lambda}: \lambda \in \Lambda\}$ define a subset of the family of distributions, each member of which is parameterized by a value of $\lambda$ (each member of this family must satisfy certain technical conditions, such as absolute continuity, which will not be explicitly defined). Although primary interest centers on $\beta$, the full set of parameters includes $\lambda$. The information matrix, evaluated at the population parameter values, is:

$$J = \frac{1}{n} \sum_{i=1}^{n} \frac{d^2 Y_i}{d \lambda^2},$$

where $J_{\lambda \lambda}$ corresponds to the elements of $\lambda$. Estimators of $\beta$ (again, the estimators must satisfy technical conditions, such as $\epsilon_i$ consistency, which are also not explicitly defined) will have covariance matrix that is at least as large as $J_{\lambda \lambda}$, which is the upper left component of $J$. If the partial derivative of the log-likelihood with respect to $\beta$ (the score for $\beta$) is orthogonal to the score for $\lambda$, then $J_{\beta \lambda} = 0$ and $J_{\beta \beta} = J_{\beta \beta}$. Because $J_{\beta \lambda}$ corresponds only to the parameter $\lambda$, the asymptotically efficient estimator of $\beta$ can be constructed without knowledge of $\lambda$. Stein argued that, if the condition $J_{\beta \lambda} = 0$ holds for all the elements of $F_{\lambda}$, then $\beta$ is adaptively estimable.

While Stein's condition has intuitive appeal, it is not straightforward how to use the condition to define estimators that are adaptive. In an invited lecture, Lickel (1982) laid out a simpler condition that does yield a straightforward link to the construction of adaptive estimators. To understand the condition, let $E_2$ denote the expectation with respect to the population error distribution and let $E_2[f(x, y; \lambda)]$ denote $f(x, y; \lambda)$ evaluated at a random variable $X \sim F$ and at $E_2[F(x; \lambda)]$. So $E_2[f(x, y; \lambda)]$ is not an arbitrary distribution $F \sim \mathcal{F}$. Let $J_2$ be the log-likelihood for the regression model with data $z = (y, x)$ and let $E_2[f(x, y; \lambda)]$ denote the expected value of $f(x, y; \lambda)$ under the distribution $F = E_2[F(x; \lambda)]$ in which $F$ is the error distribution. A familiar condition that arises in the context of likelihood estimation is that the expected population score $E_2[f(x, y; \lambda)] = 0$. Lickel's condition is simply that the population score $J_2$ is such that for any $F \sim \mathcal{F}$,

$$J_2[f(x, y; \lambda)] = 0.$$
adaptive estimation

In adaptive estimation, there are two key conditions: the nonparametric estimator of the score must be consistent in mean square to the population score, and the resulting estimators of β are two-step estimators. The estimators require, as the first step, estimation of the density function using the OLS estimator. To understand the form of these estimators, note that, if the distribution were known, then the two-step (linearized likelihood) estimator is

\[ \hat{\beta}_{2LS} = \frac{1}{n} \sum_{i=1}^{n} (z_i \hat{\beta}_{OLS}) \hat{f}_i \]

with

\[ \frac{1}{n} \sum_{i=1}^{n} (z_i \hat{\beta}_{OLS}) \hat{f}_i = -\beta \hat{f}(\hat{\beta}_{OLS}, \hat{f}) \]

and linearized likelihood estimator is asymptotically efficient. To form an adaptive estimator of \( \beta \), we replace \( \hat{\beta}_{2LS} \) with a nonparametric estimator \( \hat{\beta} \). If \( \hat{\beta} \) is constructed so that \( \hat{\beta} \) converges in quadratic mean to \( \hat{\beta}_{2LS} \), then

\[ \hat{\beta} = \hat{\beta}_{2LS} + \frac{n}{2} \sum_{i=1}^{n} (z_i \hat{\beta}_{OLS}) \frac{\hat{f}_i}{\hat{f}_{OLS}} \]

is an adaptive estimator of \( \beta \) for the family \( \mathcal{F} \). For the linear regression model, as for other models, nonparametric estimation of \( \beta \) entails nonparametric estimation of the density function. One popular nonparametric density estimator is the kernel estimator, which is constructed by using kernel estimators to improve the performance of the OLS estimator. For example, a nonparametric estimator of the density function is the kernel estimator, which is constructed by using kernel estimators to improve the performance of the OLS estimator. Newey (1984) approximates the score by a series of moment conditions, which arise from the estimated function of the vector of moments and the estimated values of the vector of moments of the random vector. Newey and Koenker (1989) use a series of spline functions to approximate the score. Chakrabarti and Caiz (2005) use wavelets to form the basis for the nonparametric estimation technique. Recent results in adaptive estimation have focused on problems in which the error distribution is known, but other features are modelled nonparametrically. Some of the most intriguing results concern the type of stochastic differential equation often encountered in financial models. The price of an asset that is measured continuously over time, \( P_t \), is often modelled as

\[ dP_t = mdt + \sigma dB_t \]

where \( m_t \) and \( \sigma_t \) are the drift and volatility terms, respectively. The presence of standard Brownian motion, \( B_t \), makes the model of price a stochastic differential equation. The function \( m \), captures the deterministic movement or drift while \( \sigma_t \) is the potentially time-varying scale of the random component. Jurek and Spokoiny (1997) study the model in which \( m_t \) is constant and \( \sigma_t \) is unknown. They establish that a nonparametric estimator of \( m_t \) is pointwise adaptive. Yet an estimator that is pointwise adaptive is that for a given point \( \theta \), the nonparametric estimator of \( m_\theta(\theta) \) is asymptotically efficient—i.e., does not perform well for all values within the range of the function \( m \). Such an idea is intuitive; without knowledge of the smoothness of \( m \), estimators designed to be optimal for one value of \( m \) may be very different from optimal estimators for another value of \( m \). Cai and Low (2000) study efficient estimation of \( m \) over neighbourhods of \( \theta \) and show that an estimator constructed from wavelets is adaptive. The restriction that the scale is constant is often difficult to support with financial data. A more realistic model, which Mercurio and Spokoiny (2004) study, models the asset return as a stochastic differential equation with drift \( \theta \) and \( \sigma_t \) varying over time. The time-varying scale is assumed to be constant over (short) intervals of time, but is otherwise unspecified. They construct a nonparametric estimator of the volatility of a kernel from the performance of local averaging and show that the resulting estimator is adaptive.

DOUGLAS G. STEIGERWALD

See also efficiency bounds; partial linear model; semiparametric estimation.

Bibliography


adaptive expectations

The adaptive expectations hypothesis may be stated most succinctly in the form of the equation:

\[ E_{t+1} = E_t (1 - \lambda) + \lambda E_{t+1} \quad 0 < \lambda < 1 \]

where \( E_t \) denotes an expectation, \( t \) is the variable whose expectation is being calculated and \( t \) is the index time. What this says is that the expectation formed at the present time, \( E_t \), of some variable, \( x_t \), at the next future date, \( t+1 \), may be viewed as a weighted average of all previous values of the variable, \( x_t \), where the weights, \( \lambda \left( 1 - \lambda \right) \), decline geometrically. The weight attaching to the most recent, or current, observation is \( \lambda \). The above equation can be manipulated readily to deliver:

\[ E_{t+1} = E_t x_t + \lambda (E_t x_t - E_t x_t) \]

What this equation says is that, viewed from time \( t \), the expected value of the variable, \( x_t \), at time \( t+1 \), is equal to the value which, at time \( t+1 \), was expected for \( t \), plus an adjustment for the extent to which the variable turned out to be different at \( t \) from the value which, viewed from \( t+1 \), had been expected. The change in the expectation is simply the fraction \( \lambda \) multiplied by the most recently observed forecast error. In this formulation, the adaptive expectations hypothesis is sometimes called the error learning hypothesis (see Mincer, 1969, pp. 83-90).

The adaptive expectations hypothesis was first used, though not by name, in the work of Irving Fisher (1911). The hypothesis received its major impetus, however, as a result of Phillip Cagan's (1956) work on hyperinflations. The hypothesis was used extensively in the late 1950s and 1960s as a variety of applications. L.M. Koyck (1954) used the hypothesis, though not in name, to study investment behaviour, Milton Friedman (1957), used it as a way of generating permanent income in his study of the consumption function. Marc Nerlove (1958) used it in his analysis of the dynamics of supply in the agricultural sector. Work on inflation and micro-economics in the 1960s was dominated by the use of this hypothesis. The most comprehensive survey of that work is provided by David Laidler and Michael Parkin (1975). The adaptive expectations hypothesis grew in size and complexity, but the general ideas of the original hypothesis became popular and it was barely challenged from the middle-1950s through the late-1960s. It was not entirely challenged but it remained the only extensively-used proposition concerning the formation of expectations of inflation and a large number of other variables for something close to two decades. In the 1970s the hypothesis fell into disfavour and the rational expectations hypothesis became dominant.

The adaptive expectations hypothesis became and remained popular for so long for three reasons. First, in its error learning form it had the appearance of being an application of classical statistical inference. It looked like classical updating of an expectation based on new information.

Second, the adaptive expectations hypothesis was empirically easy to employ. Koyck (1954) showed how a simple transformation of a variable to make it an unobservable expectation variable in it could be rendered observable by performing what became a famous transformation forming Koyck's name. If some variable, \( y_t \), is determined by the expected future value of \( x_t \) that is:

\[ y_t = a + \beta E_{t+1} x_t \]

where \( a \) and \( \beta \) are constants, then we can obtain an estimate of \( a \) and \( \beta \) by using a regression model on which expectation formation (1) or equivalently (2) is used to eliminate the unobservable expected future value of \( x_t \). To do this, substitute (1) into (3). Then write down an equation identical to (3) but for the period earlier. Multiply that second equation by -1 and subtract the result from (3) (Koyck, 1954, p. 22), to give:

\[ y_t = a + \beta E_t x_t + (1 - \lambda) E_{t+1} x_t \]

An equation like this may be used to estimate not only the desired values of \( a \) and \( \beta \) but also the value of \( \lambda \), the coefficient of expectations adjustment. Thus, economists seemed to have a very useful and plausible situ-

ations in which unobservable expectations variables were important and of discovering speeds of response both of expectations to past events and of current events to expectations of future events.

Third, the adaptive expectations hypothesis seemed to work. That is, when equations like (4) were estimated in the wide variety of situations in which the hypothesis was applied it was found that good estimable parameter values for \( a \), \( \beta \) and \( \lambda \) were obtained and, in general, a high degree of explanatory power resulted.

If the adaptive expectations hypothesis was so intuitively appealing, easy to employ, and successful, why was