

ECONOMETRIC ESTIMATION OF FORESIGHT: TAX POLICY AND INVESTMENT IN THE UNITED STATES

Douglas G. Steigerwald and Charles Stuart*

Abstract—We develop a method for measuring the foresight agents have. We first dichotomize an agent's information at current date t into *knowledge* up to date $t + f$ and *expectations* after $t + f$. We then form a residual-based test statistic that allows us to compare prediction errors for econometric models based on different values of f . We illustrate the method, examining investment around tax reforms to measure the foresight firms have about tax policy. In this illustration, current investment appears to reflect currently available information but little foresight other than foresight of enacted policy changes.

I. Introduction

DIFFERENT strands of research make substantially different assumptions about the foresight economic agents possess. Work in theoretical macroeconomics, for instance, often assumes *perfect foresight* of the entire future.¹ Rational-expectations econometrics that substitutes actual for expected values of variables similarly assumes substantial foresight; with little foresight the substitution would lead to large errors and hence have low efficiency.² On the other hand, many macroeconomic models exhibit statistically significant *lags* of several years;³ these lags may reflect information lags or negative foresight.⁴ An intermediate specification is *myopic* expectations, which amounts to knowledge of current variables but no foresight of the future. Myopic expectations are common in applied work. A good example is that many comparative static or comparative dynamic analyses of economic policy implicitly assume agents have myopic policy expectations; this occurs when agents in an initial equilibrium are treated as having adjusted fully to an initial policy, which then is replaced by a new policy that was not foreseen by agents in the initial equilibrium. Given these divergent treatments, it is natural to ask how much foresight agents actually have. In this paper, we develop a method for measuring the foresight agents have, and, as an illustration, we apply the method to study foresight about tax policy.

The approach begins with the assumption that agents have information sets with a particular structure. We think of an agent in a current period t as *knowing* the values of a set of variables up to a future or past date $t + f$, and as having *expectations* conditional on information at $t + f$ of variables

after date $t + f$. When information sets have this structure, f measures the degree of foresight, so the question “how much foresight” is then formally “what is the value of f ?” To measure f , we treat f as a parameter that describes the agent and we model how the agent's behavior depends on the value of f . In this way, econometric estimating equations depend on f . Our approach is to estimate models with different values of f and to compare fits of the models to determine the value of f that best describes actual behavior. An important consideration is that some periods in a longer time series of data may contain events that are particularly informative about the value of f that best describes actual behavior. To accommodate this, we develop a residual-based test statistic for f that allows the researcher to focus on periods believed to contain potentially informative events.

We illustrate the method by estimating the foresight of tax policy that is reflected in the investment by firms in the U.S. economy. Foresight of tax policy is particularly important in years around major tax reforms, so tax reforms are the events that provide information about f in the application here. As a simple example, a firm that knows an investment tax credit is going to be imposed in an upcoming tax reform has an incentive to postpone investment to take advantage of the credit, whereas a firm without such foresight has no such incentive; thus an examination of the timing of investment around tax reforms provides information about the value of f . To apply the approach, we need a model of investment. There are several models in the literature, from neoclassical and related models with an optimizing structure to various naive models. Although naive models sometimes outperform structural ones, we choose a structural model because interpretation of the analysis is more clear-cut if the estimating equations are derived from microeconomic principles. We estimate the model for investment in equipment and structures during 1947–1990, paying particular attention to the major U.S. tax reforms of 1954, 1962, 1981, and 1986. We identify best values of f for equipment and structures for each reform. We then aggregate over tax reforms to estimate a best “overall” measure of f . For reform-specific and overall measures of f , we also compute the confidence with which it is possible to say that one value of f describes data better than another value of f .

II. Preliminaries

To set the stage, we assume that a representative firm produces output in a general (second-order) way using a variable factor (labor) and N quasi-fixed factors (types of capital). This specification is designed to extract information from how changes in the tax treatments of different assets affect both the composition and the overall level of investment.

Received for publication December 1, 1993. Revision accepted for publication February 22, 1995.

* University of California, Santa Barbara.

We thank David Bizer, Henning Bohn, Eric Brunner, Christos Cabolis, Tom Cooley, Todd Jewel, and Perry Shapiro.

¹ For instance, Brock and Turnovsky (1981) use a representative-agent model with perfect foresight to study how policy affects macroeconomic equilibrium. A good application of the Brock-Turnovsky model to the area of taxes and investment, which is the area we study, is in Judd (1987).

² Examples in the area of investment and taxes are Pindyck and Rotemberg (1983a) and Shapiro (1986b).

³ For example, see Jorgenson (1963).

⁴ There are other explanations of observed behavioral lags. One is “time-to-build”—see Kydland and Prescott (1982).

The firm faces capital adjustment costs and has a linear-quadratic variable factor demand function:

$$\begin{aligned} L(\mathbf{K}_t, \Delta\mathbf{K}_t, Y_t) = & \alpha_0 + \alpha_Y Y_t + \frac{1}{2} \alpha_{YY} Y_t^2 + \beta' \mathbf{K}_t \\ & + \beta'_Y \mathbf{K}_t Y_t + \frac{1}{2} \mathbf{K}_t' \Gamma \mathbf{K}_t \\ & + \frac{1}{2} \Delta\mathbf{K}_t' \Omega \Delta\mathbf{K}_t, \end{aligned} \quad (1)$$

where the value of L is the quantity of labor employed in discrete period t , $\mathbf{K}_t = [K_{1,t}, \dots, K_{N,t}]'$ is a vector of beginning-of-period capital inputs in period t , $\Delta\mathbf{K}_t = \mathbf{K}_{t+1} - \mathbf{K}_t$ is net investment, and Y_t is output. Parameters in (1) are $\alpha_0, \alpha_Y, \alpha_{YY}, \beta = [\beta_1, \dots, \beta_N]'$, $\beta_Y = [\beta_{1,Y}, \dots, \beta_{N,Y}]'$, $\Gamma = (\gamma_{ij})$, which is an $N \times N$ symmetric matrix, and $\Omega = (\omega_{ii})$, which is an $N \times N$ diagonal matrix that captures adjustment costs.

Because corporations account for most value added in the United States, we assume the firm is taxed as a corporation. The corporate tax rate in period t is u_t and the variable (non-capital) input is numeraire so the firm's variable costs for a given output level are $(1 - u_t) L(\mathbf{K}_t, \Delta\mathbf{K}_t, Y_t)$. The relative acquisition price of capital of type i in period t is $P_{i,t}$ with $\mathbf{P}_t = [P_{1,t}, \dots, P_{N,t}]'$, and the relative net price of capital of type i in period t is $Q_{i,t}$ with $\mathbf{Q}_t = [Q_{1,t}, \dots, Q_{N,t}]'$. We take $Q_{i,t}$ to equal $P_{i,t}(1 - k_{i,t} - u_t Z_{i,t})$, where $k_{i,t}$ is the investment tax credit for capital of type i with $\mathbf{k}_t = [k_{1,t}, \dots, k_{N,t}]'$ and $Z_{i,t}$ is the discounted present value of future allowed depreciation deductions for capital of type i with $\mathbf{Z}_t = [Z_{1,t}, \dots, Z_{N,t}]'$, all in year t ; this simplifies by assuming that firms value current as well as future allowed depreciations at the current corporate tax rate. The diagonal matrix δ captures rates of economic depreciation for each type of capital.⁵ The firm's total after-tax costs in period t are then⁶

$$\begin{aligned} C(u_t, \mathbf{K}_t, \mathbf{K}_{t+1}, Y_t, \mathbf{Q}_t) = & (1 - u_t) L(\mathbf{K}_t, \Delta\mathbf{K}_t, Y_t) \\ & + \mathbf{Q}_t' (\Delta\mathbf{K}_t + \delta \mathbf{K}_t). \end{aligned} \quad (2)$$

As is common, we assume that the firm chooses capital inputs to minimize the discounted present value of costs to an infinite horizon,

$$\begin{aligned} E_t \sum_{\tau=0}^{\infty} \prod_{s=0}^{\tau} (1 + r_{t+s})^{-1} \\ \times C(u_{t+\tau}, \mathbf{K}_{t+\tau}, \mathbf{K}_{t+\tau+1}, Y_{t+\tau}, \mathbf{Q}_{t+\tau}), \end{aligned} \quad (3)$$

⁵ Economic depreciation means physical depreciation plus obsolescence. Note that Hulten and Wyckoff (1981) find that exponential depreciation, assumed here, fits data from used asset markets better than do either straight-line or rectangular (one-horse-shay) depreciation.

⁶ In detail, total after-tax costs are $L(\mathbf{K}_t, \Delta\mathbf{K}_t, Y_t) + \mathbf{P}_t' (\Delta\mathbf{K}_t + \delta \mathbf{K}_t) - A_t$, where the first two terms are expenses for labor and investment and $A_t \equiv u_t L(\mathbf{K}_t, \Delta\mathbf{K}_t, Y_t) + \sum_{i=1}^N (u_t Z_{i,t} + k_{i,t}) P_{i,t} (\Delta K_{i,t} + \delta_{i,t} K_{i,t})$ is total tax benefits, which is the value of deductions for labor costs and allowed depreciation plus the value of the investment tax credit.

where E_t denotes expectations conditional on the firm's information set at time t and r_{t+s} is a known real discount rate.

III. Information Sets

To solve (3), it is necessary to specify knowledge and expectations at time t of the random variables $u_{t+\tau}$, $\mathbf{k}_{t+\tau}$, $u_{t+\tau} \mathbf{Z}_{t+\tau}$, $Y_{t+\tau}$, and $\mathbf{P}_{t+\tau}$, for $\tau \geq 0$. We measure time in years so f is the number of years of foresight that is reflected in investment in year t . The value of f may be positive, negative, or zero. Because we wish to focus on the amount of policy foresight that firms have, we treat expectations of tax-policy variables ($u_{t+\tau}$, $\mathbf{k}_{t+\tau}$, and $u_{t+\tau} \mathbf{Z}_{t+\tau}$) as governed by a value of f , but treat expectations of variables characterizing macroeconomic conditions ($Y_{t+\tau}$, and $\mathbf{P}_{t+\tau}$) more traditionally. Thus we take the year- t information set of the firm to consist of all values of tax-policy variables up to year $t + f$, plus all values of macroeconomic variables up to year t .

Because major policy changes seem to occur randomly and occasionally, there is little reason to take expectations of policy after $t + f$ to be given by a smooth autoregressive moving-average (ARMA) process. Instead, we assume that policy expectations are given by martingales with respect to the year- t information set. The year- t information set contains the values of policy variables for years to and including $t + f$, so we define a sequence $\{x_{t+f+\tau}\}$ for $t = 1, \dots, T$, given f and $\tau \geq 0$, to be a *martingale with respect to* x_{t+f} if $E_t x_{t+f+\tau} = E_t x_{t+f} < \infty$. Martingale expectations capture a stylized form of uncertainty or unpredictability: at t , firms may attach probabilities to increases and decreases in policy parameters after $t + f$ but the *expected* policy change is zero. The martingale specification is simple to implement and may be a useful approximation for modeling agents who have better information about the present or near future than about the distant future. Indeed, the specification may be theoretically correct and not merely an approximation under several conditions. It is perforce correct if agents economize on information processing by themselves acting as if there is a period of knowledge and a subsequent period in which a continuation of the status quo is expected. The specification may also be theoretically correct if actual policy follows a martingale, for which there is some empirical evidence.⁷

We focus on three values of f . The first is $f = 1$, which has the firm knowing this year's and next year's policies and seeing policy thereafter as a martingale with respect to *next year's* policy. Formally, $f = 1$ implies that expectations

⁷ Barro (1979) derives conditions under which tax policy follows a martingale in an optimal-tax model of government. Barro (1981), Kingston (1987), and Mankiw (1987) find that actual tax policy is approximated closely by a martingale. Sahasakul (1986) and Bizer and Durlauf (1990), on the other hand, reject a martingale on the finding that actual tax rates move predictably with wars, recessions, and elections. The latter rejections should be of second-order importance here. Most notably, the rejections occur because firms have *more* information about future policy than is embodied in current policy, but this is just what we allow for in taking the amount of policy foresight as a variable to be measured.

about the investment tax credit follow $E_t \mathbf{k}_t = \mathbf{k}_t$ and $E_t \mathbf{k}_{t+\tau} = \mathbf{k}_{t+1}$ for $\tau \geq 1$, the latter capturing martingale expectations. Thus when $f = 1$, the firm sees a change in ITC from t to $t + 1$ and reacts by shifting investment optimally from the year the ITC is lower to the year it is higher. For allowed depreciation deductions and the corporation tax rate we make analogous assumptions when $f = 1$: $E_t u_t = u_t$ and $E_t u_{t+\tau} = u_{t+1}$ for $\tau \geq 1$; and $E_t u_t \mathbf{Z}_t = u_t \mathbf{Z}_t$ and $E_t u_{t+\tau} \mathbf{Z}_{t+\tau} = u_{t+1} \mathbf{Z}_{t+1}$ for $\tau \geq 1$. Note that $f = 1$ gives firms less knowledge than would perfect foresight of *all* future policy; the latter case is $f = \infty$.

The second value is $f = 0$, which has the firm knowing this year's policy and seeing future policy as a martingale with respect to *this year's* policy. Formally, it has $E_t \mathbf{k}_{t+\tau} = \mathbf{k}_t$, $E_t u_{t+\tau} = u_t$, and $E_t u_{t+\tau} \mathbf{Z}_{t+\tau} = u_t \mathbf{Z}_t$, for $\tau \geq 0$. The specification with $f = 0$ can be thought of as myopia in the setup here because myopia implies that *expected* future policy always equals current policy, which occurs with martingale expectations if $f = 0$. In the case of the ITC when $f = 0$, the firm does *not* see a change in the ITC from t to $t + 1$ and hence does *not* react by shifting investment from the year the ITC is lower to the year it is higher. The firm's behavior around a tax reform thus differs under $f = 0$ and $f = 1$, which means that it should be possible to distinguish the two values by examining shifts in investment around tax reforms.

The third value is $f = -1$, which has the firm in year t knowing policy only up to the previous year ($t - 1$) and seeing policy as a martingale with respect to the previous year's policy. Formally, $f = -1$ has $E_t \mathbf{k}_{t+\tau} = \mathbf{k}_{t-1}$, $E_t u_{t+\tau} = u_{t-1}$, and $E_t u_{t+\tau} \mathbf{Z}_{t+\tau} = u_{t-1} \mathbf{Z}_{t-1}$, for $\tau \geq -1$. It should also be possible to distinguish $f = -1$ from $f = 0$, because investment responses to tax changes would tend to be delayed one year if $f = -1$ fits better.

The best fit with observed data may occur with $f = 1$ if the process of tax reform is a slow one that generates information about policy changes roughly a year before the changes take effect and if there are only small lags between the arrival of information and the placing in service of new capital in response to the information. We do not study the value $f = 2$ here for two reasons. First, observed tax reforms are typically in the works for less than about a year and it is usually uncertain whether policy is really going to change and in what way for much of the time that the reform is in the works, so we have little prior expectation that $f = 2$ would fit better than $f = 1$. Second and perhaps decisively, the only effect on our estimating equations that would be caused by shifting from $f = 1$ to $f = 2$ would be to change the value of planned investment between next year and the year after.⁸ With the parameter values we estimate, investment in the current year is affected only slightly by changes in this planned investment. As a result, it should be difficult to distinguish $f = 1$ from $f = 2$.

⁸ In terms of the estimating equation (10) of the next section, the only effect of shifting from $f = 1$ to $f = 2$ would be to change the value of the optimal (planned) capital stock for the year after next year, which is $\mathbf{K}_t^{(2)}$.

On the other hand, the best fit with observed data may occur with $f = -1$ if firms' own processes of factoring new information into investment also take time; that is, if there are information lags. For instance, information lags can occur if firms do not continuously make optimal investment decisions or if people in firms do not immediately observe and assimilate all currently available data about economic conditions.

We treat expectations of macroeconomic variables more traditionally. Short-run changes in variables reflecting macroeconomic conditions ($Y_{t+\tau}$ and $\mathbf{P}_{t+\tau}$ for low τ) fit ARMA models reasonably. To make estimation tractable and in particular to avoid the problem that ARMA expectations may grow faster than the discount rate, we follow the approach of Prucha and Nadiri (1984) by assuming that firms have a planning horizon of H years, have ARMA expectations of $Y_{t+\tau}$ and $\mathbf{P}_{t+\tau}$ for $\tau \leq H$, and have martingale expectations of $Y_{t+\tau}$ and $\mathbf{P}_{t+\tau}$ for $\tau > H$. Prucha and Nadiri (1986) report Monte Carlo evidence that $H = 5$ and $H = 10$ provide good approximations to the case with $H = \infty$, so even if an ARMA model is better for the distant future, martingale expectations should approximate it closely. In detail, we fit univariate ARMA models to growth rates of Y_t and changes in the level of \mathbf{P}_t to model $E_t Y_{t+\tau}$ and $E_t \mathbf{P}_{t+\tau}$ for $\tau \leq H$. (We fit $\Delta \mathbf{P}_t$ rather than $\Delta \ln \mathbf{P}_t$, which is the growth rate of \mathbf{P}_t , because \mathbf{P}_t is an index.) The order of the ARMA models is selected using the Akaike information criterion. Beyond H , martingale expectations for Y_t and \mathbf{P}_t mean that we assume $E_t Y_{t+H+\tau} = E_t Y_{t+H}$ and $E_t \mathbf{P}_{t+H+\tau} = E_t \mathbf{P}_{t+H}$ for $\tau > 0$, where numerical values of $E_t Y_{t+H}$ and $E_t \mathbf{P}_{t+H}$ are taken from the ARMA models.

We make two additional, technical assumptions to be able to pass expectations through $C(\cdot)$. First, we assume that $\mathbf{P}_{t+\tau}$ is mean independent of $\mathbf{k}_{t+\tau}$ and $u_{t+\tau} \mathbf{Z}_{t+\tau}$ so $E_t \mathbf{Q}_{t+\tau}$ is simply $E_t \mathbf{P}_{t+\tau}$ times the N -vector with generic element $1 - E_t k_{n,t+\tau} - E_t u_{t+\tau} \mathbf{Z}_{n,t+\tau}$. This assumption may not fit the data we examine exactly if capital goods have rising supply curves so tax incentives to invest cause capital prices to rise.⁹ The assumption notwithstanding, our treatment models expectations of policy and macroeconomic variables separately in calculating $E_t \mathbf{Q}_{t+\tau}$, which allows for more detail than would the more common procedure of treating $E_t \mathbf{Q}_{t+\tau}$ as the outcome of an ARMA process. Second, we assume that $u_{t+\tau}$ and $Y_{t+\tau}$ are mean-independent random variables: $E_t u_{t+\tau} Y_{t+\tau} = E_t u_{t+\tau} E_t Y_{t+\tau}$.

IV. Estimating Equations

Under the informational assumptions made above, minimization of (3) is equivalent to minimization of

$$\sum_{\tau=0}^{\infty} \prod_{s=0}^{\tau} (1 + r_{t+s})^{-1} C_{t+\tau}, \quad (4)$$

⁹ Results in Shapiro (1986a, p. 128) suggest that our assumption may be a reasonable approximation; Shapiro finds that "there is no feedback from the quantities to the price variables."

where $C_{t+\tau} = C(E_t u_{t+\tau}, \mathbf{K}_{t+\tau}, \mathbf{K}_{t+\tau+1}, E_t Y_{t+\tau}, E_t \mathbf{Q}_{t+\tau})$. In each year t , the firm computes an optimal capital program $\{\mathbf{K}_t^{(\tau)}\}_{\tau=1}^{\infty}$, where $\mathbf{K}_t^{(\tau)}$ is the optimal capital vector at the beginning of year $t + \tau$ given the firm's decisions in year t . Because the firm has a planning horizon of H years, we assume that capital stock values satisfy $\mathbf{K}_t^{(\tau)} = \mathbf{K}_t^{(H)}$, so the firm's minimand (4) simplifies to

$$\sum_{\tau=0}^{H-1} \prod_{s=0}^{\tau} (1 + r_{t+s})^{-1} C_{t+\tau} + \left[\prod_{s=0}^{H-1} (1 + r_{t+s})^{-1} r_{t+H}^{-1} \right] r_{t+H}^{-1} \times C(E_t u_{t+H}, \mathbf{K}_{t+H}, \mathbf{K}_{t+H}, E_t Y_{t+H}, E_t \mathbf{Q}_{t+H}), \quad (5)$$

where the second term captures costs beyond the planning horizon.

We define the input-demand equations for next year's capital stocks to be the optimal capital stocks for the beginning of year $t + 1$ based on the information available in year t , plus an error vector $\epsilon_t = [\epsilon_{1,t}, \dots, \epsilon_{N,t}]'$:

$$\mathbf{K}_{t+1} = \mathbf{K}_t^{(1)} + \epsilon_t. \quad (6)$$

The error can be thought of as optimization error that captures failure by firms to set the capital stock to its optimal value $\mathbf{K}_t^{(1)}$. Alternatively the error can be thought of as reflecting simple mismeasurement of the capital stock or random shocks due to unforeseen events that prevent the firm from attaining the optimal value $\mathbf{K}_t^{(1)}$. The optimal capital stock path is obtained from the system of first-order conditions for optimization of (5):

$$\begin{aligned} & -(1 - u_{t+\tau+1}) \mathbf{\Omega} \mathbf{K}_t^{(\tau+2)} \\ & + [(1 - u_{t+\tau+1})(\mathbf{\Omega} + \mathbf{\Gamma}) \\ & + (1 + r_{t+\tau+1})(1 - u_{t+\tau}) \mathbf{\Omega}] \\ & \times \mathbf{K}_t^{(\tau+1)} - (1 + r_{t+\tau+1})(1 - u_{t+\tau}) \mathbf{\Omega} \mathbf{K}_t^{(\tau)} \\ & = E_t[\mathbf{h}_{t+\tau}] \quad \text{for } \tau = 0, \dots, H - 2, \end{aligned} \quad (7)$$

$$\begin{aligned} & [r_{t+H}(1 - u_{t+H-1}) \mathbf{\Omega} + (1 - u_{t+H}) \mathbf{\Gamma} \mathbf{K}_t^{(H)} \\ & - r_{t+H}(1 - u_{t+H-1}) \mathbf{\Omega} \mathbf{K}_t^{(H-1)} \\ & = -E_t[(1 - u_{t+H})(\beta + \beta_Y Y_{t+H}) \\ & + (r_{t+H} \mathbf{Q}_{t+H-1} \\ & - (\mathbf{I} - \delta) \mathbf{Q}_{t+H})], \quad \text{for } \tau = H - 1, \end{aligned} \quad (8)$$

where

$$\mathbf{h}_{t+\tau} = -\{(1 - u_{t+\tau+1})(\beta + \beta_Y Y_{t+\tau+1}) + (1 + r_{t+\tau+1}) \mathbf{Q}_{t+\tau} - (\mathbf{I} - \delta) \mathbf{Q}_{t+\tau+1}\}, \quad (9)$$

with $\mathbf{K}_t^{(0)} = \mathbf{K}_t$.

Using (7)–(9), the input demand equations for next year's

capital stock are

$$\begin{aligned} \mathbf{K}_{t+1} & = [(1 - u_{t+1})(\mathbf{\Omega} + \mathbf{\Gamma}) + (1 + r_{t+1})(1 - u_t) \mathbf{\Omega}]^{-1} \\ & \times \{(1 - u_{t+1}) \mathbf{\Omega} \mathbf{K}_t^{(2)} \\ & + (1 + r_{t+1})(1 - u_t) \mathbf{\Omega} \mathbf{K}_t + E_t \mathbf{h}_t\} + \epsilon_t, \end{aligned} \quad (10)$$

where the optimal capital vector two years into the future, $\mathbf{K}_t^{(2)}$, is solved from the system (7)–(9).

V. Data

We use annual data. Although quarterly data might allow for finer measurement of the foresight firms have about future policy, we see two reasons why a shift to quarterly data might introduce more noise than additional information. First, the Bureau of Economic Analysis (BEA) cautions that the quarterly capital stock is measured less precisely than the annual capital stock. Second, use of yearly data tends to average out higher frequency fluctuations in Y_t and \mathbf{P}_t that may be irrelevant to the planning processes of firms that make plans and stick to them even when conditions change slightly.

The data cover 1947–1990, so after differencing we have 42 observations. The small number of observations imposes a fairly severe limitation on how finely the model may be disaggregated because the number of parameters in the variable factor demand function is $(1/2)N_2 + (7/2)N + 3$. For instance, 12 parameters in L must be estimated if $N = 2$ while 18 parameters must be estimated if $N = 3$ and 25 parameters must be estimated if $N = 4$.

Accordingly, we focus on the equipment/structures split, which has $N = 2$. We measure output (Y_t) as real private, nonagricultural nonresidential output in the United States (source: BEA). Capital stocks (\mathbf{K}_t) are real (1982 dollars) fixed nonagricultural nonresidential net capital stocks provided by the BEA. Our category “equipment” is the sum of 19 of the 20 subcategories based on the disaggregation found in the National Income and Product Accounts; we exclude agricultural machinery. The category “structures” consists of 10 subcategories; we exclude religious, educational, hospital, and institutional buildings as these are largely non-profit and hence are likely subject to different incentives than those modeled here. In keeping with our exclusion of agriculture, we also exclude farm buildings from structures.

We measure labor as total annual hours worked in private nonagricultural nonresidential sectors using data from the Bureau of Labor Statistics (BLS). We multiply annual private nonfarm employment by 52 times average weekly hours to get total hours. Our price series for equipment and structures (\mathbf{P}_t) come from the BEA and are normalized by dividing by the average real hourly private nonagricultural wage as labor is the numeraire; wage data are from the BLS. The BEA's series for \mathbf{P}_t include adjustments for technological change; to this extent our estimation consistently adjusts for technological change. The BEA computes price series for

equipment and structures using fixed weights equal to shares of the 1982 capital stock in disaggregated subcategories of equipment and subcategories of structures for years after 1958 but using implicit price deflators for years up to 1958. To splice series before and after 1958, we predict the fixed-weighted series for 1947–58 from the implicit-deflator series using quadratic regression on the period 1958–70, during which time the relationship between the two series was fairly stable. Economic depreciation rates (δ) are taken from Hulten and Wykoff (1981).

We compute the real discount rate r_t as a weighted average of net (of corporation tax) returns on debt and equity in year t . The weights are 0.725 on debt finance and 0.275 on equity finance, these being the average debt and equity shares over the period 1947–1981 reported by Holland (1984, table 2B2a); this treatment assumes that marginal investment is financed in the same way as inframarginal investment.¹⁰ We take the real return on equity in year t to be the sum of dividends in t plus stock price in $t + 1$ all divided by stock price in t , for a share of stock corresponding to the Standard and Poors index.¹¹ The net real return on debt is the nominal return on debt adjusted by an inflation factor and multiplied by one minus the corporation tax rate to account for deductibility of business interest. We take the nominal return to be the annual rate on Aaa corporate bonds as rated by Moody's. Because labor is numeraire, the correct inflation factor is the rate of wage inflation, which we measure using the wage series from the BLS.

The corporate tax rate is constructed from Internal Revenue Service (IRS) publications and includes state taxes on corporate income.¹² Data on the investment tax credit also come from IRS publications. We first calculate the ITC for each of the disaggregated subcategories of capital, then aggregate into series for equipment and structures using 1982 capital stock weights. We reduce ITC rates after 1982 by the value of allowed depreciation deductions lost because ITC payments reduce the allowed basis for depreciation. Depreciation schedules used to calculate Z_t come from IRS publications. In discounting future allowed depreciation deductions, we assume a straight 4% real discount rate and also assume that firms choose the tax treatment that maximizes Z_t in each year.¹³

It is important to understand the timing of changes in tax

¹⁰ In preliminary work, we experimented with putting all weight on debt and all weight on equity; the results were not sensitive to the choice of weight. We also measured r_t as the short-term real Treasury bill rate and obtained essentially the same results. Use of a constant real interest rate of 4%, on the other hand, gave poorer results.

¹¹ Auerbach and Hassett (1990) find that estimated investment equations are not sensitive to alternative treatments of the cost of equity capital.

¹² Because most corporate value added is produced by corporations facing the maximum federal corporate rate, we take u_t to be the maximum federal corporate rate times one plus a factor for the share of state revenues in total corporate tax payments. Variation in u_t is due mainly to variation in the maximum federal rate.

¹³ We chose to impose a fixed, positive discount rate instead of r_t (which often is negative) to avoid making values of $Z_{n,t}$ unreasonably large; the empirical results are not particularly sensitive to reasonable choices of the fixed discount rate. Note that Summers (1987) reports that discount rates in

TABLE 1.—MAJOR POSTWAR U.S. TAX REFORMS

Enactment Date	Effective Dates of Change		
	u_t	k_t	Z_t
Aug. 16, 1954	Apr. 1, 1955	—	Jan. 1, 1954
Oct. 16, 1961	—	Jan. 1, 1962	—
Aug. 13, 1981	—	Jan. 1, 1981	Jan. 1, 1981
Oct. 22, 1986	July 1, 1987	Jan. 1, 1986	Jan. 1, 1987

policy variables around major tax reforms. We see in table 1 that three of the four major tax reforms became law about three-quarters into the year but that critical provisions were made effective retroactive to the beginning of the year of enactment. Because of retroactive provisions, we must choose a convention for defining policy in a year. One possibility would be to let policy during the year be the policy *actually on the books* for most of the year. A second possibility would be to let policy be the policy that ex post turns out to have been applied for most of the year. We choose the latter, ex post definition of policy. This should be borne in mind when interpreting exactly what a given value of f means. When $f = 0$, for instance, $E_t \mathbf{k}_t = \mathbf{k}_t$. With ex post data, the firm for the first three quarters of a tax-reform year is therefore assumed to know that the policy enacted after three quarters will in fact be enacted and applied retroactively. Thus we should think of $f = 0$ as actually giving firms something like an average of 4½ months of policy foresight. Analogously, $f = 1$ gives 16½ months of policy foresight, and $f = -1$ gives a 7½ month policy lag.

VI. Estimation

According to the investment model used in the application here, the optimal investment decision in any year is the solution to a sequence of Euler equations and a transversality condition, which means that investment in a year depends on an entire program of optimal capital stocks for all future years. There are two approaches for dealing with the econometric problem that future optimal capital stocks are unobservable. The easiest is simply to replace future optimal values with future observed values in estimating equations and to impose rational expectations, assuming that any error introduced by the replacement is uncorrelated with other sources of error in the equations.¹⁴ Because the replacement introduces error, however, the approach is not fully efficient. Put differently, replacing optimal with observed values effectively bases estimation of current investment on only the current-year Euler equation, ignoring information in all future Euler equations and the transversality condition. A Monte Carlo study by Prucha and Nadiri (1986) finds this efficiency loss to be considerable and also finds the approach subject to substantial finite-sample bias. The second ap-

excess of 4% are sometimes used by firms to discount allowed depreciation deductions.

¹⁴ Examples of the approach are in Kennan (1979), Hansen (1982), Hansen and Singleton (1982), Pindyck and Rotemberg (1983a, b), and Shapiro (1986b).

proach, which performed better in the Prucha-Nadiri study and which we use, is to solve analytically for the future optimal capital stocks, $\mathbf{K}_t^{(2)}$, in (10).¹⁵ This approach uses full information from all future Euler equations and the transversality condition.

We estimate jointly the variable factor demand equation (1) and the capital demand equations (10), taking account of the cross-equation restrictions implicit in (10). To do this, an error η_t is appended to (1). We assume that the error vector (ϵ_t, η_t) has a multivariate normal distribution centered at zero with covariance matrix \mathbf{V} . Because the dependent variable in (1) does not appear as a regressor in (10), the system of the two equations is triangular. Lahiri and Schmidt (1978) show that, for triangular systems, the iterated seemingly unrelated regressions (SUR) estimator is identical to the full information maximum likelihood (FIML) estimator, which is asymptotically efficient. We therefore construct an iterated SUR estimator for the parameters of (1) and (10).¹⁶

The subsystem (10) requires an estimate of $\mathbf{K}_t^{(2)}$, which in turn requires estimates of the parameters that appear in (7)–(9). Because ϵ_t and η_t may be correlated and \mathbf{K}_{t+1} appears as a regressor in (1) as part of $\Delta \mathbf{K}_t$, we obtain these parameter estimates from instrumental variables estimation of (1) using $\Delta \mathbf{K}_{t-1}$ as instruments.

We use the estimated values from this first-stage regression of (1) to compute from (7)–(9) the capital stock program $\{\mathbf{K}_t^{(1)}, \dots, \mathbf{K}_t^{(H)}\}$ that is optimal in year t . Next we insert $\mathbf{K}_t^{(2)}$ from the optimal program into (10) and estimate (1) and (10) by SUR to obtain a new set of parameter estimates, then reinsert the new estimates into (7)–(9) to recompute the optimal capital stock program $\mathbf{K}_t^{(2)}$, iterating until estimates converge.¹⁷ Parameter estimates for each of the three values of f are in the appendix.

The issue of unit roots deserves mention. First, we are unable to reject the null that a unit root is present in the autoregressive lag polynomial for any variable in (1) and (10) except for $\Delta \mathbf{K}_t$, u_t , r_t , and \mathbf{Q}_t . There are thus two concerns. First, if *residuals* from (1) and (10) have unit roots, then the regression results may be spurious. To test for unit roots in the residuals, we perform augmented Dickey-Fuller test on the residuals from (1) and (10).¹⁸ We can reject

¹⁵ Sargent (1978), Meese (1980), Hansen and Sargent (1980, 1981), Bernanke (1983), Epstein and Yatchew (1985), Nadiri and Prucha (1986), and Mohnen et al. (1986) apply this approach to various problems by assuming that expectations of future variables follow autoregressive processes.

¹⁶ Prucha (1987) points out that the covariance matrix from iterated SUR is not consistent. If $\mathbf{K}_t^{(2)}$ were observed, a consistent estimator of the covariance matrix could be obtained by using the parameter estimates from iterated SUR as starting values for a FIML routine and taking standard errors from the routine. We report standard errors derived using this procedure. Because our $\mathbf{K}_t^{(2)}$ is calculated and not observed, however, our standard errors are likely too small. As discussed below, we do not base any subsequent analysis on these standard errors.

¹⁷ The convergence criterion is $(SSR_{j-1} - SSR_j)/(SSR_j + 10^{-6}) < 10^{-8}$, where SSR_j is the sum of the squared residuals from the j th iteration.

¹⁸ Precisely, the null under the augmented Dickey-Fuller tests is that an equation is a cointegrating relation, that is, a relation in which the dependent variable and at least one of the regressors contain unit roots but the error term does not contain a unit root.

null hypotheses that the residual vector from either (1) or (10) has a unit root. The second concern is that the failure to reject unit roots implies that our estimator may not be asymptotically normally distributed. This affects how one might calculate critical values and hence perform statistical inference about the estimated parameters, but is not a problem for our purposes. Namely, we compute critical values to judge which value of f fits best around tax reforms using a bootstrap procedure described below, and not using estimated standard errors from the iterated SUR procedure.

VII. The Best Value of f

We wish to judge which of the three values of f best describes actual behavior. A natural approach would be to compare how estimates based on each of the three fit investment *across the entire time series*. On this basis the three values do about equally well, as R^2 values for (10) are essentially identical across values of f . Given that the three values of f fit equally well across the entire sample, an alternative approach is to take the best value of f to be the value that best predicts the time pattern of investment *around major tax reforms*. We use this approach because investment around tax reforms is likely to be particularly informative about f . Specifically, we examine diagnostic test statistics based on sums of squared residuals for investment the year before and the year of the four major tax reforms. For the 1954 reform, for instance, we examine actual and predicted investment for 1953 and 1954. The test statistics we compute are

$$\phi_{i,s}(f) = \frac{[\Delta K_{i,s-1} - \Delta \hat{K}_{i,s-1}(f)]^2 + [\Delta K_{i,s} - \Delta \hat{K}_{i,s}(f)]^2}{\sum_{t=1}^T e_{i,t}^2(f)} \quad (11)$$

where s is the year of the tax-code change, $\Delta K_{i,t} = K_{i,t+1} - K_{i,t}$ is net investment in capital of type i in year t , $\Delta \hat{K}_{i,t}(f) = \hat{K}_{i,t+1}(f) - \hat{K}_{i,t}(f)$ is predicted investment from the model when policy information is described by foresight f , and $e_{i,t}(f) = K_{i,t} - \hat{K}_{i,t}(f)$ is the difference between actual net investment and predicted net investment given f . The denominator in (11) normalizes the squared prediction errors in the numerator by the scale of squared prediction errors across the entire sample.¹⁹ The purpose of the normalization is to allow us to generate a distribution for each $\phi_{i,s}(f)$ by performing bootstrap resampling (described below) on the denominator in (11). The normalization is consistent with the idea that residuals around tax reforms are

¹⁹ If the sum of all squared residuals (the denominator in (11)) were to differ markedly across values of f , then $\phi_{i,s}(f)$ might be lower for a value of f that yielded a worse fit across the entire sample than for a value that yielded a better fit. As noted earlier, this is not a concern here because R^2 values and hence sums of squared residuals across the sample period are essentially identical for different values of f .

TABLE 2.—DIAGNOSTICS FOR FITS OF VALUES OF f : $\phi_{i,s}(f)$

		Tax Reform				
		1954	1962	1981	1986	Aggregate
Equipment						
Diagnostic:	$f = -1$	0.0050	0.0507	0.0329	0.0073	0.0959
	$f = 0$	0.0002	0.0110	0.0101	0.0120	0.0334
	$f = 1$	0.0103	0.1090	0.0659	0.0105	0.1957
Structures						
Diagnostic:	$f = -1$	0.0004	0.0086	0.0055	0.0055	0.0200
	$f = 0$	0.0037	0.0021	0.0050	0.0090	0.0198
	$f = 1$	0.0059	0.0104	0.0123	0.0033	0.0318

the main source of information about the foresight that agents possess; the normalization ensures that we do not obtain a low value of the test statistic $\phi_{i,s}(f)$ unless residuals around reform s are small relative to other residuals.

We compute and report in table 2 values of $\phi_{i,s}(-1)$, $\phi_{i,s}(0)$, and $\phi_{i,s}(1)$ for investment in equipment and in structures for each of the four major tax reforms. A low value of $\phi_{i,s}(f)$ indicates that the model with foresight f fits actual data well for capital of type i around the tax reform in year s . In the left-most column we see for the equipment diagnostic around the 1954 reform that $\phi_{i,s}(f)$ reaches a minimum at $f = 0$, indicating that the best value of f among the three we consider is $f = 0$ in this case. For the structures diagnostic around the 1954 tax reform, on the other hand, $\phi_{i,s}(f)$ reaches a minimum at $f = -1$, indicating that the best value of f is $f = -1$. Results are more clear-cut for the 1962 and 1981 reforms: $f = 0$ fits best for both equipment and structures. For 1986, $f = -1$ fits best for equipment and $f = 1$ fits best for structures. This result for structures is consistent with the idea that firms acted from “knowledge” that changes in provisions affecting structures (u_t and \mathbf{Z}_t but not \mathbf{k}_t) would only become effective in 1987; that is, that firms had an information lead in this case. Because the equipment diagnostic favors $f = -1$ and the underlying model assumes that f takes the same value for both equipment and structures, however, the result for 1986 does not clearly exhibit a best value of f .

To determine the value of f that fits best over all four tax reforms, we compute aggregate diagnostics for equipment and structures,

$$\begin{aligned} \phi_i(f) = & \phi_{i,1954}(f) + \phi_{i,1962}(f) \\ & + \phi_{i,1981}(f) + \phi_{i,1986}(f). \end{aligned} \quad (12)$$

In the last column of table 2 we see that the best overall fit occurs with $f = 0$. For structures, however, $f = -1$ appears to fit almost as well as $f = 0$.

Table 2 indicates that $f = 0$ fits U.S. investment data better than do either $f = -1$ or $f = 1$ in aggregate and in all but three instances of individual tax reforms. In instances where $f = 0$ fits better, it is natural to ask whether it fits better enough so that we can reject hypotheses that $f = -1$ or $f = 1$ with, say, 95% confidence. To evaluate this, we test null hypotheses that $f = -1$ and $f = 1$ against the alternative hypothesis that $f = 0$. Specifically, we study the distributions

of $\phi_{i,s}(-1) - \phi_{i,s}(0)$ and $\phi_{i,s}(1) - \phi_{i,s}(0)$, which measure how much better $f = 0$ fits than $f = -1$ and $f = 1$, respectively. Under the null that data are generated by $f = -1$, for instance, the difference $\phi_{i,s}(-1) - \phi_{i,s}(0)$ should be negative, whereas under the alternative that data are generated by $f = 0$, the difference $\phi_{i,s}(-1) - \phi_{i,s}(0)$ should be positive. Similarly under the null that data are generated by $f = 1$, the difference $\phi_{i,s}(1) - \phi_{i,s}(0)$ should be negative, whereas under the alternative that data are generated by $f = 0$, the difference $\phi_{i,s}(1) - \phi_{i,s}(0)$ should be positive. We thus examine 95% confidence intervals for $\phi_{i,s}(-1) - \phi_{i,s}(0)$ and $\phi_{i,s}(1) - \phi_{i,s}(0)$: a strictly positive confidence interval means that we can reject a null that $f = -1$ or $f = 1$ in favor of the alternative that $f = 0$; otherwise we cannot reject the null.

For the three instances where $f = -1$ or $f = 1$ fit best, we also ask whether it is possible to reject the hypothesis that $f = 0$ with 95% confidence. The tests are as above except that the null in these cases is $f = 0$ and the alternative is $f = -1$ or $f = 1$. Thus if a 95% confidence interval for $\phi_{i,s}(-1) - \phi_{i,s}(0)$ or $\phi_{i,s}(1) - \phi_{i,s}(0)$ is strictly negative, we can reject the null that $f = 0$ in favor of the alternative that there is a one-year lag ($f = -1$) or one year of foresight ($f = 1$), respectively.

Because the numerators and denominators of the $\phi_{i,s}(f)$ are not independent, our test statistics do not follow F -distributions. To approximate the distributions analytically is difficult. We therefore construct approximate distributions of the $\phi_{i,s}(f) - \phi_{i,s}(0)$ for $f = -1$ and $f = 1$ by replacement sampling (bootstrapping), which involves using the set of residuals from the estimated system as a proxy for the unknown errors in the system. Bootstrapping requires that the starting sequence of residuals be uncorrelated. For $f = 0$ and a given type of capital i , we therefore begin with the 42 SUR residuals and use the correlation structure in the estimated covariance matrix to express the residuals in the form $e_{i,t}(0) = K_{i,t} - \hat{K}_{i,t}(0)$. We take this set as an original population and sample from it with replacement to generate a new set of 42 residuals, $\{e_{i,1}(0), \dots, e_{i,42}(0)\}$. We proceed in the same way for $f = -1$, or for $f = 1$, generating a new set of 42 residuals, $\{e_{i,1}(f), \dots, e_{i,42}(f)\}$. We then use (11) to calculate a value of $\phi_{i,s}(f) - \phi_{i,s}(0)$. We repeat the process, generating a large number of values of $\phi_{i,s}(f) - \phi_{i,s}(0)$. Finally we use the generated distribution of $\phi_{i,s}(f) - \phi_{i,s}(0)$ to calculate 95% confidence intervals for $\phi_{i,s}(f) - \phi_{i,s}(0)$, for $f = -1$ and $f = 1$.

We also bootstrap $\phi_i(f) - \phi_i(0)$ for $f = -1$ and $f = 1$ to compute 95% confidence intervals for the aggregate difference in fits of two amounts of foresight. The procedure is to sample as just described and then use both (11) and (12) to generate values of $\phi_i(f) - \phi_i(0)$.

When we use (11) to calculate $\phi_{i,s}(f) - \phi_{i,s}(0)$, we condition on residuals around the tax reform (that is, on the numerator in (11)), so we always include the three tax-reform residuals in the new set of residuals and use replacement sampling to find the other 39. We condition on

TABLE 3.—95% CONFIDENCE INTERVALS FOR $\phi_{i,s}(f) - \phi_{i,s}(0)$

		Tax Reform									
		1954		1962		1981		1986		Aggregate	
A. Are there lags ($f = -1$)?											
Equipment:		[0.0025	0.0075] ^N	[0.0175	0.0635] ^N	[0.0087	0.0412] ^N	[-0.0092	0.0014]	[0.0212	0.1160] ^N
Structures:		[-0.0055	-0.0016] ^Y	[0.0023	0.0106] ^N	[-0.0035	0.0040]	[-0.0091	0.0011]	[-0.0146	0.0125]
B. Is there foresight ($f = 1$)?											
Equipment:		[0.0068	0.0128] ^N	[0.0625	0.1267] ^N	[0.0342	0.0725] ^N	[-0.0062	0.0045]	[0.1018	0.2239] ^N
Structures:		[-0.0015	0.0051]	[0.0035	0.0122] ^N	[0.0009	0.0126] ^N	[-0.0111	-0.0015] ^Y	[-0.0067	0.0274]

Note: N indicates that the answer to the question is No in that we can with 95% confidence reject a null that $f = -1$ or $f = 1$ in favor of an alternative that $f = 0$, and Y indicates that the answer to the question is Yes in that we can with 95% confidence reject a null that $f = 0$ in favor of an alternative that $f = -1$ or $f = 1$.

the numerator because it is the numerator that contains most of the information about how well a value of f predicts investment, and this information would be lost if we were to draw residuals randomly to form the numerator. When we calculate distributions for the aggregate test statistics, $\phi_i(f) - \phi_i(0)$, we similarly condition on the numerator. An alternative procedure in this case would have been to draw residuals from the four tax-reform episodes in forming the numerator. The alternative procedure would be correct if the distribution of $\phi_i(f) - \phi_i(0)$ calculated by replacement sampling from the four tax reforms were the same as the distribution of $\phi_{i,s}(f) - \phi_{i,s}(0)$ for each tax reform. This condition seems strong. For instance, our estimation procedure captures effects of a tax reform operating through changes in the values of u_t , \mathbf{k}_t , and \mathbf{Z}_t , but misses effects operating through changes in tax-shelter and other reform provisions. Because such reform-specific mismeasurement generally shows up in the residuals, the distribution of residuals around one tax reform may differ from the distribution around another tax reform.²⁰

Confidence intervals calculated in this way are in table 3. To understand the results, first consider instances of specific reforms in which $f = 0$ fits best. This was the case for equipment around the 1954 reform and for both equipment and structures around the 1962 and 1981 reforms. We see in the table that, for each of these instances except for structures around the 1981 tax reform, confidence intervals are strictly positive, implying that we can reject null hypotheses for the presence of lags ($f = -1$) and foresight ($f = 1$) in favor of $f = 0$. For structures around the 1981 reform, on the other hand, we can reject the null for foresight but cannot reject a one-year information lag.

Now consider the three instances in which $f = -1$ or $f = 1$ fits best. From table 2, we saw that $f = -1$ fits best for structures around the 1954 reform. From table 3, we see that

the confidence interval for this instance is strictly negative, so we can reject $f = 0$ in favor of $f = -1$. For equipment around the 1986 reform, $f = -1$ also fits best, but in this instance the confidence interval brackets zero so we cannot reject $f = 0$ in favor of $f = -1$. Finally, for structures around the 1986 reform, $f = 1$ fits best; in this instance the confidence interval is strictly negative so we can reject $f = 0$ in favor of $f = 1$.²¹

For aggregate test statistics, we saw in table 2 that $f = 0$ gives the best fit for both equipment and structures. From table 3, we can reject with 95% confidence the presence of a lag ($f = -1$) or of foresight ($f = 1$) in favor of $f = 0$ for the equipment diagnostic, but we cannot reject $f = -1$ or $f = 1$ for the structures diagnostic.

VIII. Summary

We develop a method for measuring the amount of foresight that agents have. Specifically, we assume that agents acting at date t have knowledge of key variables up to date $t + f$ and have expectations of these variables beyond $t + f$. This specification nests perfect foresight ($f = \infty$), myopia ($f \equiv 0$ together with martingale expectations), and information lags ($f < 0$). We then construct a residual-based test statistic for determining the value of f that best fits observed data; the test statistic is based on specific residuals thought to be particularly informative. To illustrate the method, we estimate f by comparing how models with $f = -1$, $f = 0$, and $f = 1$ predict investment residuals around major post-war U.S. tax reforms.

In this application, the best fit appears to occur with $f = 0$. Because of our convention for defining policy in a year, this has the interpretation that firms may have several months of policy foresight, but probably not a year of foresight. An exception is that U.S. investment behavior around the 1986 tax reform may

²⁰ Confidence intervals under the alternative procedure may be larger or smaller than the intervals we calculate. To see this simply, note that confidence intervals reflect variances, that $\phi_n(f) - \phi_n(0)$ is a ratio that can be written g_1/g_2 , and that to a second-order approximation, the variance of g_1/g_2 is proportional to $\sigma_1^2/\mu_1^2 + \sigma_2^2/\mu_2^2 - 2\sigma_{12}^2/\mu_1\mu_2$, where μ_i and σ_i^2 denote the mean and variance of g_i and σ_{12}^2 denotes the covariance between g_1 and g_2 . Because our procedure treats g_1 as nonstochastic, we measure the variance of $\phi_n(f) - \phi_n(0)$ as ϕ_2^2/μ_2^2 . Thus if σ_{12}^2 is sufficiently positive, the alternative procedure could result in a smaller variance and hence smaller confidence intervals.

²¹ Auerbach and Hassett (1991) argue that the Tax Reform Act of 1986 created a rare period of several years of policy stability in which the ex post tax treatment of investment was observable to firms. Auerbach and Hassett also estimate a reduced-form investment model using data prior to 1986 and show that, for investment in equipment, ex post tax policy helps explain predicted investment residuals for 1987–89 derived from the model. Although their analysis is not designed to distinguish among different amounts of foresight, their results suggest that firms may have had some foresight of policy for a few years after 1986. Thus our result for structures around the 1986 reform is broadly consistent with their result for equipment.

be best fit by $f = 1$. The explanation may be that several important provisions of the 1986 act became effective in the year after enactment (1987), and passage of the act was a sufficient ordeal politically so it was regarded as unlikely that new legislation would interfere with the provisions becoming effective. A broad interpretation is that firms may have made use of knowledge of current values of tax policy variables and also knowledge of major pending changes in the tax code. We find little evidence, on the other hand, of information or decision lags in recent U.S. investment behavior.

REFERENCES

- Auerbach, Alan, and Kevin Hassett, "Investment, Tax Policy, and the Tax Reform Act of 1986," in J. Slemrod (ed.), *Do Taxes Matter? The Impact of the Tax Reform Act of 1986* (Cambridge, MA: MIT Press, 1990).
- "Recent U.S. Investment Behavior and the Tax Reform Act of 1986: A Disaggregate View," *Carnegie-Rochester Series on Public Policy* 35 (Autumn 1991), 185–216.
- Barro, Robert, "On the Determination of the Public Debt," *Journal of Political Economy* 87 (Oct. 1979), 940–971.
- "On the Predictability of Tax Rate Changes," NBER Working Paper 636 (1981).
- Bernanke, Ben, "The Determinants of Investment: Another Look," *American Economic Review* 73 (May 1983), 71–75.
- Bizer, David, and Steven Durlauf, "Testing the Positive Theory of Government Finance," *Journal of Monetary Economics* 26 (1990), 123–141.
- Brock, William, and Stephen Turnovsky, "The Analysis of Macroeconomic Policies in Perfect Foresight Equilibrium," *International Economic Review* 22 (Feb. 1981), 283–298.
- Epstein, Larry, and Adonis Yatchew, "The Empirical Determination of Technology and Expectations," *Journal of Econometrics* 27 (Feb. 1985), 235–258.
- Hansen, Lars P., "Large Sample Properties of Generalized Method of Movement Estimators," *Econometrica* 50 (July 1982), 1029–1054.
- Hansen, Lars P., and Thomas J. Sargent, "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control* 2 (Feb. 1980), 7–46.
- "Linear Rational Expectations Models for Dynamically Interrelated Variables," in R. Lucas and T. Sargent (eds.), *Rational Expectations and Econometric Practice* (Minneapolis: University of Minnesota Press, 1981).
- Hansen, Lars P., and Kenneth Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50 (Sept. 1982), 1269–1286.
- Holland, Daniel M., *Measuring Profitability and Capital Costs* (Lexington, MA: Lexington Books, 1984).
- Hulten, Charles, and Frank Wykoff, "The Measurement of Economic Depreciation," in C. Hulten (ed.), *Depreciation, Inflation, and the Taxation of Income from Capital* (Washington: Urban Institute, 1981), 81–132.
- Jorgenson, Dale, "Capital Theory and Investment Behavior," *American Economic Review* 53 (May 1963), 247–259.
- Judd, Kenneth, "The Welfare Cost of Factor Taxation in a Perfect-Foresight Model," *Journal of Political Economy* 95 (Aug. 1987), 675–709.
- Kennan, John, "The Estimation of Partial Adjustment Models with Rational Expectations," *Econometrica* 47 (Nov. 1979), 1441–1455.
- Kingston, G. H., "Efficient Timing of Taxes," *Journal of Public Economics* 24 (1987), 271–280.
- Kydland, Finn, and Edward Prescott, "Time to Build and Aggregate Fluctuations," *Econometrica* 50 (Nov. 1982), 1345–1370.
- Lahiri, Kajal, and Peter Schmidt, "On the Estimation of Triangular Structural Systems," *Econometrica* 46 (Sept. 1978), 1217–1221.
- Mankiw, N. G., "The Optimal Collection of Seniorage: Theory and Evidence," *Journal of Monetary Economics* 20 (1987), 327–341.
- Meese, Richard, "Dynamic Factor Demand Schedules for Labor and Capital under Rational Expectations," *Journal of Econometrics* 14 (Sept. 1980), 141–158.
- Mohnen, P. A., M. I. Nadiri, and I. Prucha, "R&D, Production Structure, and Rates of Return in the U.S., Japanese and German Manufacturing Sectors: A Nonseparable Dynamic Factor Demand Model," *European Economic Review* 30 (Aug. 1986), 49–71.

- Nadiri, M. I., and Ingmar Prucha, "Comparison and Analysis of Productivity Growth and R&D Investment in the Electrical Machinery Industries of the U.S. and Japan," NBER Working Paper No. 1850 (1986).
- Pindyck, Robert, and Julio Rotemberg, "Dynamic Factor Demand Under Rational Expectations," *Scandinavian Journal of Economics* 85 (July 1983a), 223–238.
- "Dynamic Factor Demands and the Effects of Energy Price Shocks," *American Economic Review* 73 (Dec. 1983b), 1066–1079.
- Prucha, Ingmar, "The Variance-Covariance Matrix of the Maximum Likelihood Estimator in Triangular Structural Systems," *Econometrica* 55 (July 1987), 977–978.
- Prucha, Ingmar, and M. Ishaq Nadiri, "Formulation and Estimation of Dynamic Factor Demand Equations Under Non-Static Expectations: A Finite Horizon Model," NBER Working Paper (1984).
- "A Comparison of Alternative Methods for the Estimation of Dynamic Factor Demand Models under Non-Static Expectations," *Journal of Econometrics* 33 (Oct. 1986), 187–211.
- Sahasakul, C., "The U.S. Evidence on Optimal Taxation Over Time," *Journal of Monetary Economics* 18 (1986), 251–275.
- Sargent, Thomas, "Estimation of Dynamic Labor Demand Schedules under Rational Expectations," *Journal of Political Economy* 86 (Dec. 1978), 1009–1044.
- Shapiro, Matthew, "Investment, Output, and the Cost of Capital," *Brookings Papers on Economic Activity* (1986a, I), 111–164.
- "The Dynamic Demand for Capital and Labor," *Quarterly Journal of Economics* 101 (Aug. 1986b), 513–542.
- Summers, Lawrence, "Investment Incentives and the Discounting of Depreciation Allowances," in M. Feldstein (ed.), *The Effects of Taxation on Capital Accumulation* (Chicago: University of Chicago Press, 1987).

APPENDIX

TABLE A1.—NONLINEAR ITERATED SUR PARAMETER ESTIMATES

Parameter	Value of f		
	$f = -1$	$f = 0$	$f = 1$
α_0	112.269 (40.156)	-11.440 (32.901)	71.098 (22.558)
α_Y	0.288 (0.110)	0.430 (0.098)	0.355 (0.063)
α_{YY}	0.0002 (—)	0.0002 (—)	0.0002 (—)
β_1	0.177 (0.109)	-0.008 (0.100)	0.003 (0.90)
β_2	-0.950 (0.179)	-0.541 (0.158)	-0.467 (0.159)
β_{1Y}	0.0004 (0.0001)	0.0003 (0.0001)	0.0003 (0.0001)
β_{2Y}	-0.001 (0.0001)	-0.001 (0.0001)	-0.001 (0.0001)
γ_{11}	-0.0003 (0.0005)	0.0002 (0.0004)	0.0001 (0.0004)
γ_{12}	-0.001 (0.0003)	-0.001 (0.0003)	-0.001 (0.0002)
γ_{22}	0.004 (0.0006)	0.003 (0.0006)	0.003 (0.0004)
ω_{11}	0.020 (0.002)	0.020 (0.002)	0.020 (0.002)
ω_{22}	0.012 (0.002)	0.011 (0.003)	0.011 (0.003)
Adjusted R^2 , Equipment	0.987	0.992	0.992
Adjusted R^2 , Structures	0.987	0.987	0.985

Notes: Standard errors are in parentheses. Results are for $H = 10$; results for $H = 5$ are essentially identical. To get estimates to coverage, we constrained α_{YY} to values reasonable on the basis of preliminary estimation, then computed iterated SUR estimates of all other parameters. It turned out that estimated values of other parameters were not sensitive to the choice of α_{YY} . We chose this procedure because α_{YY} does not enter the capital demand equations (10). Note that (10) must fit closely if the model is to track capital stocks closely, which is necessary if our procedure is to be able to distinguish among different values of f .

Estimated parameters generally satisfy restrictions imposed by theory. Specifically, $\omega_{11} > 0$, $\omega_{22} > 0$, $\gamma_{11} > 0$, $\gamma_{22} > 0$, $\gamma_{11}\gamma_{22} - \gamma_{12}^2 > 0$, $\partial L / \partial K_{1s} < 0$, and $\partial L / \partial Y_s > 0$ hold at estimated parameter values or within one standard error given mean values of K_{1s} and Y_s for all three values of f . At mean values of K_{1s} and Y_s , and for $f = 0$, the estimated scale elasticity is 0.93, with output elasticities of 0.57 for labor, 0.31 for equipment, and 0.05 for structures.