Convergence of Least-Squares Learning in Environments with Hidden State Variables and Private Information

Albert Marcet

Carnegie-Mellon University

Thomas J. Sargent

Hoover Institution and Federal Reserve Bank of Minneapolis

We study the convergence of recursive least-squares learning schemes in economic environments in which there is private information. The presence of private information leads to the presence of hidden state variables from the viewpoint of particular agents. By applying theorems of Ljung, we extend some of our earlier results to characterize conditions under which a system governed by least-squares learning will eventually converge to a rational expectations equilibrium. We apply insights from the learning results to formulate and compute the equilibrium of a version of Townsend's model.

I. Introduction

This paper studies the convergence of least-squares learning mechanisms to limited information rational expectations equilibria. We study linear models in which agents have access to information on only a subset of the relevant state variables. The models cover situations in which there are distinct groups of differentially informed agents. We proceed by applying to our system the recently developed

We received very extensive and useful comments on an earlier draft of this paper from Margaret Bray, Michele Boldrin, Robert E. Lucas, Jr., and an anonymous referee. We also thank Lawrence Christiano, Ramon Marimon, and Will Roberds for helpful discussions of issues related to this paper. Sargent's research was supported by a grant from the National Science Foundation to the University of Minnesota (SES-8508935).

[*Journal of Political Economy*, 1989, vol. 97, no. 6]

"ordinary differential equations" approach of Ljung (1977).[1] This involves extending some earlier results of Marcet and Sargent (1989) to handle situations with private information and hidden state variables. We give sufficient conditions for almost sure convergence to a limited information rational expectations equilibrium and describe necessary conditions for local convergence.

Our conditions for convergence restrict an operator that maps a set of perceived vector autoregressions into a set of actual (or optimal) vector autoregressions. This operator is determined by the particular economic model in hand. The operator is related to but distinct from the operator governing convergence in the class of models studied in Marcet and Sargent (1989). The presence of private information and hidden state variables alters the relevant operator, in essence by composing the key operator in Marcet and Sargent with another "projection" operator.

Section II describes a class of models with limited and private information and asserts a convergence proposition for least-squares learning mechanisms.[2] Section III applies our framework in order to formulate and compute the equilibrium of a version of Townsend's (1983) model. In his model, firms with private information face signal extraction problems involving endogenous variables whose laws of motion are themselves determined by the solutions of those signal extraction problems. Models with structures like Townsend's (see also Lucas 1975) have proved to be difficult to formulate in ways that facilitate computing their equilibria. The purpose of Section III is to show how our results on convergence of least-squares learning can be used to help in formulating these models and to suggest alternative tractable algorithms for computing their equilibria.[3]

II. The Model and a Convergence Proposition

There is an $n \times 1$ state vector $\mathbf{z}_t$. Let $\mathbf{z}_{it}$ be any $n_i \times 1$ vector $\mathbf{z}_{it} = \mathbf{e}_i\mathbf{z}_t$, where $1 \leq n_i \leq n$ and $\mathbf{e}_i$ are selector matrices for $i = a, b, c$, and d.

[1] The ordinary differential equations approach is described and applied by Ljung and Söderström (1983) and Goodwin and Sin (1984). See also Kushner and Clark (1978). Woodford (1986) applies some of Ljung's methods to a nonlinear dynamic model.

[2] Bray's (1982) model and a version of Frydman's (1982) model are members of the class of models described in Sec. II. Analyses of these models are contained in Marcet and Sargent (1987). Papers about least-squares learning in models without hidden state variables include Bray and Savin (1986) and Fourgeaud, Gouriéroux, and Pradel (1986). Marcet and Sargent (1988) present an informal interpretative survey of the literature on least-squares learning.

[3] The ability to compute the equilibria of these models rapidly would contribute to their being econometrically tractable. It is probably true that the technical difficulties in computing the equilibria of models of the style of Lucas (1975) and Townsend (1983) have impeded their adoption by other researchers.

There are two types of agents, types a and b, who observe $\mathbf{z}_{at} = \mathbf{e}_a\mathbf{z}_t$ and $\mathbf{z}_{bt} = \mathbf{e}_b\mathbf{z}_t$, respectively, possibly distinct subvectors of $\mathbf{z}_t$. Agents of type j want to predict future values of possibly distinct subvectors $\mathbf{z}_{k(j)} = \mathbf{e}_{k(j)}\mathbf{z}_t$, where $k(a) = c$ and $k(b) = d$, and use the current observation on $\mathbf{z}_{jt}$ in order to form those predictions. The selection matrices $\mathbf{e}_a$, $\mathbf{e}_b$, $\mathbf{e}_c$, and $\mathbf{e}_d$ are constant through time. There is an economic model that maps beliefs of agents a and b into actual outcomes in the following way. If the beliefs of agents of type a and type b were given by the time-invariant rules

$$\begin{aligned} E^*(\mathbf{z}_{ct}|\mathbf{z}_{at-1}) &= \boldsymbol{\beta}_a\mathbf{z}_{at-1}, \\ E^*(\mathbf{z}_{dt}|\mathbf{z}_{bt-1}) &= \boldsymbol{\beta}_b\mathbf{z}_{bt-1} \quad \text{for all } t, \end{aligned} \tag{1}$$

then the actual law of motion $\mathbf{z}_t$ would be given by

$$\mathbf{z}_t = T(\boldsymbol{\beta})\mathbf{z}_{t-1} + V(\boldsymbol{\beta})\boldsymbol{\epsilon}_t, \tag{2}$$

where $\boldsymbol{\epsilon}_t$ is an $m \times 1$ vector white noise with $\mathbf{E}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t' = \Omega$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_a, \boldsymbol{\beta}_b)$, and T and V are operators that map matrices into matrices conformable to the objects they operate on. A particular economic model will determine the operators T and V. In the next section, we describe a version of Townsend's model and display the operators T and V that are associated with it.

We are interested in regions of the parameter space $\boldsymbol{\beta}$ for which (2) implies that $\mathbf{z}_t$ is a covariance stationary stochastic process. For this purpose, we define the following set:

$$\begin{aligned} D_s = \{\boldsymbol{\beta}|&\text{the operators } T(\boldsymbol{\beta}) \text{ and } V(\boldsymbol{\beta}) \text{ are} \\ &\text{well defined, and the eigenvalues of } T(\boldsymbol{\beta}) \\ &\text{are less than unity in modulus}\}. \end{aligned}$$

For $\boldsymbol{\beta} \in D_s$, (2) generates a covariance stationary stochastic process, for which the second-moment matrix $\mathbf{E}\mathbf{z}_t\mathbf{z}_t'$ is well defined. The matrix $\mathbf{M}_\mathbf{z}(\boldsymbol{\beta}) = \mathbf{E}\mathbf{z}_t\mathbf{z}_t'$ satisfies the discrete Lyapunov equation

$$\mathbf{M}_\mathbf{z}(\boldsymbol{\beta}) = T(\boldsymbol{\beta})\mathbf{M}_\mathbf{z}(\boldsymbol{\beta})T(\boldsymbol{\beta})' + V(\boldsymbol{\beta})\Omega V(\boldsymbol{\beta})'. \tag{3}$$

A variety of algorithms are available for solving (3) for $\mathbf{M}_\mathbf{z}(\boldsymbol{\beta})$. We use the following notation for some submatrices of $\mathbf{E}\mathbf{z}_t\mathbf{z}_t'$:

$$\begin{aligned} \mathbf{M}_{\mathbf{z}_j}(\boldsymbol{\beta}) &= \mathbf{E}\mathbf{z}_{jt}\mathbf{z}_{jt}', \quad j = a, b, \\ \mathbf{M}_{\mathbf{z}_j,\mathbf{z}}(\boldsymbol{\beta}) &= \mathbf{E}\mathbf{z}_{jt}\mathbf{z}_t', \quad j = a, b. \end{aligned}$$

In general, each of these moment matrices is a function of $\boldsymbol{\beta}$.

If the actual law of motion for $\mathbf{z}_t$ is (2), then it can be calculated that the linear least-squares projection of $\mathbf{z}_{k(j)t}$ on $\mathbf{z}_{jt-1}$ is given by

$$\hat{\mathbf{E}}(\mathbf{z}_{k(j)t}|\mathbf{z}_{jt-1}) = S_j(\boldsymbol{\beta})\mathbf{z}_{jt-1}, \tag{4}$$

where

$$S_j(\boldsymbol{\beta}) = e_{k(j)}T(\boldsymbol{\beta})[\mathbf{M}_{\mathbf{z}_j}(\boldsymbol{\beta})^{-1}\mathbf{M}_{\mathbf{z}_j,\mathbf{z}}(\boldsymbol{\beta})]', \quad \text{for } j = a, b. \tag{5}$$

The operators $S_j(\boldsymbol{\beta})$ map the perceptions $\boldsymbol{\beta} = (\boldsymbol{\beta}_a, \boldsymbol{\beta}_b)$ into the projection coefficients $(S_a(\boldsymbol{\beta}), S_b(\boldsymbol{\beta}))$. Let us define $S(\boldsymbol{\beta}) = (S_a(\boldsymbol{\beta}), S_b(\boldsymbol{\beta}))$.

We now advance the following definition.

DEFINITION. A rational expectations equilibrium with asymmetric private information is a matrix $\boldsymbol{\beta} = (\boldsymbol{\beta}_a, \boldsymbol{\beta}_b)$ that satisfies $\boldsymbol{\beta} = S(\boldsymbol{\beta})$.

Thus a rational expectations equilibrium is a fixed point of the mapping S. Let us denote such an equilibrium $\boldsymbol{\beta}_f$. Notice that this concept of a rational expectations equilibrium is relative to the fixed information sets $\mathbf{z}_{at-1}$ and $\mathbf{z}_{bt-1}$ specified by the model builder.

We now describe the model of learning. For $j = a, b$, we let $\{\alpha_{jt}\}$ be a positive, nondecreasing sequence with $\lim_{t\to\infty} \alpha_{jt} = 1$. Beliefs of agents of type j $(= a, b)$ evolve according to the following scheme. Define $\overline{\boldsymbol{\beta}}_{jt}$ and $\overline{\mathbf{R}}_{jt}$ by

$$\begin{aligned} \overline{\boldsymbol{\beta}}'_{jt} &= \boldsymbol{\beta}'_{jt-1} + \left(\frac{\alpha_{jt-1}}{t}\right)\mathbf{R}_{jt-1}^{-1}\{\mathbf{z}_{jt-2}[\mathbf{z}_{k(j)t-1} - \boldsymbol{\beta}_{jt-1}\mathbf{z}_{jt-2}]'\}, \\ \overline{\mathbf{R}}_{jt} &= \mathbf{R}_{jt-1} + \left(\frac{\alpha_{jt-1}}{t}\right)\left(\mathbf{z}_{jt-1}\mathbf{z}'_{jt-1} - \frac{\mathbf{R}_{jt-1}}{\alpha_{jt-1}}\right). \end{aligned} \tag{6a}$$

Let $D_{2j} \subset D_{1j} \subset \mathbb{R}^{n_{k(j)} \times (n_j)^3}$, $j = a, b$. The algorithm generating beliefs at t is then

$$(\boldsymbol{\beta}_{jt}, \mathbf{R}_{jt}) = \begin{cases} (\overline{\boldsymbol{\beta}}_{jt}, \overline{\mathbf{R}}_{jt}) & \text{if } (\overline{\boldsymbol{\beta}}_{jt}, \overline{\mathbf{R}}_{jt}) \in D_{1j} \\ \text{some value in } D_{2j} & \text{if } (\overline{\boldsymbol{\beta}}_{jt}, \overline{\mathbf{R}}_{jt}) \notin D_{1j}. \end{cases} \tag{6b}$$

Two distinct sets, D_{1j} and D_{2j}, are used in defining the projection facility in order to properly invoke some technical arguments made by Ljung (1977). In practice, we shall be free to choose D_{2j} to be a set contained within but arbitrarily close to D_{1j}. In the applications below, we shall always think of D_{2j} as being arbitrarily close to D_{1j} and shall thus focus our attention on specification of the set D_{1j}.[4]

If $D_{2j} = D_{1j} = \mathbb{R}^{n_{k(j)} \times (n_j)^3}$, then the "projection facility" on the second branch of (6b) is never invoked, and with suitable initial conditions, (6a)–(6b) simply become a recursive version of weighted least squares:

$$\boldsymbol{\beta}_{jt} = \left(\sum_{i=1}^{t-1} \alpha_{ji}\mathbf{z}_{ji-1}\mathbf{z}'_{ji-1}\right)^{-1}\left[\sum_{i=1}^{t-1} \alpha_{ji}\mathbf{z}_{ji-1}\mathbf{z}'_{k(j)i}\right].$$

In the special case that $\{\alpha_{jt}\} = \{1\}$, the formula above is just ordinary

[4] Ljung and Söderström (1983) frequently proceed in this way, specifying a projection facility in terms of a single set.

least squares. In cases in which a nontrivial projection facility is specified by choosing D_{1j} to be a proper subset of $\mathbb{R}^{n_{k(j)} \times (n_j)^3}$, it is natural to set "some point in D_{2j}" in (6b) equal to $(\boldsymbol{\beta}_{jt'}, \mathbf{R}_{jt'})$, where t' is the last time that $(\boldsymbol{\beta}_{jt'}, \mathbf{R}_{jt'}) \in D_{2j}$. With D_{2j} set arbitrarily close to D_{1j}, (6) then amounts to least squares adjusted sequentially to ignore observations that threaten to drive $(\boldsymbol{\beta}_{jt}, \mathbf{R}_{jt})$ outside of the set D_{1j}. When the sequence $\{\alpha_{jt}\}$ is chosen to be strictly increasing, it leads to adjusting the least-squares algorithm to weight more recent observations more heavily. (The restriction that $\lim_{t \to \infty} \alpha_{jt} = 1$ restricts the eventual rate of forgetting in a way sufficient to permit convergence of $\boldsymbol{\beta}_{jt}$ within the system to be studied below.)

The sets D_{1j} and D_{2j} will play important roles in one part of the proposition to be stated below. One role of the sets D_{1j} and D_{2j} can be to force the learning algorithm to remain within the set D_s defined above.

We assume that when agents are learning according to (6), the actual law of motion is determined by substituting $\boldsymbol{\beta}_t = (\boldsymbol{\beta}_{at}, \boldsymbol{\beta}_{bt})$ from (6) for $\boldsymbol{\beta}$ on the right side of (2):

$$\mathbf{z}_t = T(\boldsymbol{\beta}_{t-1})\mathbf{z}_{t-1} + V(\boldsymbol{\beta}_{t-1})\boldsymbol{\epsilon}_t. \tag{7}$$

The system that we want to study is (6) and (7).

Associated with the system of stochastic difference equations (6) and (7) is the following ordinary differential equation:

$$\frac{d}{dt}\begin{bmatrix} \boldsymbol{\beta}_a \\ \boldsymbol{\beta}_b \\ \mathbf{R}_a \\ \mathbf{R}_b \end{bmatrix} = \begin{bmatrix} \mathbf{R}_a^{-1}\mathbf{M}_{\mathbf{z}_a}(\boldsymbol{\beta})[S_a(\boldsymbol{\beta}) - \boldsymbol{\beta}_a]' \\ \mathbf{R}_b^{-1}\mathbf{M}_{\mathbf{z}_b}(\boldsymbol{\beta})[S_b(\boldsymbol{\beta}) - \boldsymbol{\beta}_b]' \\ \mathbf{M}_{\mathbf{z}_a}(\boldsymbol{\beta}) - \mathbf{R}_a \\ \mathbf{M}_{\mathbf{z}_b}(\boldsymbol{\beta}) - \mathbf{R}_b \end{bmatrix}. \tag{8}$$

Defining $\mathbf{R} = (\mathbf{R}_a, \mathbf{R}_b)$, we can represent (8) in the vector form

$$\frac{d}{dt}\begin{pmatrix} \text{col}(\boldsymbol{\beta}) \\ \text{col}(\mathbf{R}) \end{pmatrix} = g(\boldsymbol{\beta}, \mathbf{R}),$$

where col($\boldsymbol{\beta}$) is a vector obtained by stacking columns of $\boldsymbol{\beta}$ on top of each other, and col($\mathbf{R}$) is a vector obtained by stacking columns of $\mathbf{R}$ on top of each other. For the purpose of studying the linear approximations that govern the local behavior of (8), we define

$$h(\boldsymbol{\beta}, \mathbf{R}) = \frac{d}{d(\text{col }\boldsymbol{\beta}, \text{col }\mathbf{R})'} g(\boldsymbol{\beta}, \mathbf{R}).$$

Let $\{(\boldsymbol{\beta}(t), \mathbf{R}(t))\}_{t \in [0,\infty)}$ denote the trajectories of (8). We define the set D_A to be the domain of attraction of the fixed point $(\boldsymbol{\beta}_f, \mathbf{R}_f)$ of (8), which we assume to be unique. That is, D_A consists of the set of $(\boldsymbol{\beta}(0),$

$\mathbf{R}(0)$) such that when $(\boldsymbol{\beta}(0), \mathbf{R}(0)) \in D_A$, then (8) implies $\lim_{t\to\infty} (\boldsymbol{\beta}(t), \mathbf{R}(t)) = (\boldsymbol{\beta}_f, \mathbf{R}_f)$.

We use a set of six assumptions about system (6)–(7), which are described in the Appendix. Among these, the first five are in the nature of regularity conditions that are easy to verify and are typically satisfied for the kinds of applications we have encountered. It bears mentioning that assumption 1, which states that S has a unique fixed point, could be relaxed to permit multiple fixed points. Then our propositions would transform to statements about each fixed point of $S(\boldsymbol{\beta})$.

Assumption 6 can be considerably more difficult to verify than 1–5. Assumption 6 is used in only the first part of our four-part proposition. For this first part, we also use the following assumption.

ASSUMPTION 7. For $j = a, b$, assume that D_{2j} is closed, that D_{1j} is open and bounded, and that $\boldsymbol{\beta} \in D_s$ for all $(\boldsymbol{\beta}_a, \mathbf{R}_a, \boldsymbol{\beta}_b, \mathbf{R}_b) \in D_{1a} \times D_{1b}$. Assume that the trajectories of (8) with initial conditions $(\boldsymbol{\beta}_a(0), \mathbf{R}_a(0), \boldsymbol{\beta}_b(0), \mathbf{R}_b(0)) \in D_{2a} \times D_{2b}$ never leave a closed subset of $D_{1a} \times D_{1b}$.

We now state proposition 1.

PROPOSITION 1. Assume that $(\boldsymbol{\beta}_t, \mathbf{R}_t, \mathbf{z}_t)$ are determined by (6) and (7). Assume that assumptions 1–5 are satisfied.

i) Assume also that assumptions 6 and 7 are satisfied and that $D_{1a} \times D_{1b} \subset D_A$, where D_A is the domain of attraction of $(\boldsymbol{\beta}_f, \mathbf{R}_f)$ in (8). Then $P[\boldsymbol{\beta}_t \to \boldsymbol{\beta}_f] = 1$.
ii) Let $\hat{\boldsymbol{\beta}} \neq \boldsymbol{\beta}_f$, and assume that $\mathbf{M}_{\mathbf{z}_j}(\hat{\boldsymbol{\beta}})$ is positive definite for $j = a, b$. Then $P[\boldsymbol{\beta}_t \to \hat{\boldsymbol{\beta}}] = 0$.
iii) If $h(\mathbf{R}_f, \boldsymbol{\beta}_f)$ has one or more eigenvalues with strictly positive real part, then $P[\boldsymbol{\beta}_t \to \boldsymbol{\beta}_f] = 0$.
iv) $h(\boldsymbol{\beta}_f, \mathbf{R}_f)$ has $(n_a)^2 + (n_b)^2$ repeated eigenvalues of negative one. The remaining eigenvalues are the same as those of the following derivative matrix:

$$\left(\frac{\partial}{\partial \boldsymbol{\beta}}\right)\begin{bmatrix}\mathrm{col}[S_a(\boldsymbol{\beta}) - \boldsymbol{\beta}_a]\\ \mathrm{col}[S_b(\boldsymbol{\beta}) - \boldsymbol{\beta}_b]\end{bmatrix}\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_f}.$$

This concludes the proposition.[5]

Statement i asserts that sufficient conditions for $\boldsymbol{\beta}_t \to \boldsymbol{\beta}_f$ almost surely as $t \to \infty$ are that the set $D_{1a} \times D_{1b}$ generated in the projection facility be contained in D_A and that at (and close to) the boundary of $D_{1a} \times D_{1b}$, the differential equation (8) have trajectories that point

[5] Proposition 1 can be proved by retracing the steps used to prove propositions 1, 2, and 3 of Marcet and Sargent (in press). We do not present that proof here.

toward the interior of $D_{1a} \times D_{1b}$. Statement ii asserts that the only possible limit points of the learning scheme are rational expectations equilibria. Statement iii asserts sufficient conditions for nonconvergence of the learning scheme. Statement iv implies that everything can be learned about the local stability of the learning scheme by studying the differential equation

$$\frac{d}{dt}\begin{pmatrix}\boldsymbol{\beta}_a\\ \boldsymbol{\beta}_b\end{pmatrix} = \begin{bmatrix} S_a(\boldsymbol{\beta}) - \boldsymbol{\beta}_a \\ S_b(\boldsymbol{\beta}) - \boldsymbol{\beta}_b \end{bmatrix} = S(\boldsymbol{\beta}) - \boldsymbol{\beta}. \tag{9}$$

Proposition 1 can be used to study convergence of least-squares learning in the context of a variety of models that have been proposed. Marcet and Sargent (1987, 1988) describe applications for several such models. In terms of the literature on least-squares learning, we find proposition 1 of use for several purposes. First, it sometimes makes it possible to strengthen results that have been obtained by other means (see, e.g., the analysis of Bray's [1982] model in Marcet and Sargent [1987]). Second, application of the proposition can markedly shorten the length of arguments needed to establish convergence results (again see Marcet and Sargent [1987]). Third, the proposition permits a unified interpretation in terms of the properties of the $S(\boldsymbol{\beta})$ operator for the apparently disparate conditions for convergence that previous papers have discovered.

In the remainder of this paper we focus on another way that proposition 1 can be used, namely, to guide the computation of a rational expectations equilibrium. As a laboratory for our study, we use a model of Townsend for which the equilibrium has been difficult to formulate and to compute by means other than those suggested by proposition 1.

III. A Model of Townsend

This section uses an algorithm suggested by proposition 1 to compute the equilibrium of a version of Townsend's (1983) model. We adopt his formulation of the demand and cost structure but reformulate his way of modeling firms' forecasting problems. Townsend formulates that forecasting problem by imputing to firms more understanding of the economic structure than we do. He models each firm as knowing that the mean beliefs of firms in other industries are hidden state variables about whose laws of motion the firm itself forms beliefs. We model the firm as forecasting its own price by using a vector autoregression that includes its own price, the price of the other industry, and all other variables in its information set. This transformation of Townsend's "forecasting the forecasts of others" into the problem of "forecasting output price using vector autoregressions" turns out to be a reformulation that leaves the economic content of his equilibrium concept unaltered. We shall return to this point later in this section. We now turn to describing our version of Townsend's model.

There are two industries, indexed by $j = a, b$, consisting of N identical firms. The representative firm in industry j has objective function

$$E_0 \sum_{t=0}^{\infty} b^t \left[p_{jt} f k_{jt} - w_{jt} k_{jt} - \left(\frac{d}{2}\right)(k_{jt+1} - k_{jt})^2 \right], \quad d > 0, 0 < b < 1, \quad (10)$$

where

$$p_{jt} = -AfK_{jt} + u_{jt}, \quad A > 0, f > 0, \quad (11)$$

$$K_{jt} = Nk_{jt}, \quad N > 0, \quad (12)$$

$$u_{jt} = \theta_t + \epsilon_{jt}, \quad (13)$$

$$\theta_t = \rho\theta_{t-1} + v_t, \quad |\rho| < 1, \quad (14)$$

where (ϵ_{jt}, v_t) are mutually orthogonal white noises, p_{jt} is the price of output in industry j at t, k_{jt} is the capital stock of the representative firm, $f \cdot k_{jt}$ is the firm's output, u_{jt} is a random shock to demand, θ_t is a hidden common component to the demand in industries a and b, and w_{jt} is a serially uncorrelated rental rate on capital in industry j, assumed orthogonal to w_{it} for $i \neq j$ and to all components of ϵ_{jt}, v_t. We have omitted constant terms. Firms in industry a observe the history $\{p_{as}, K_{as}, p_{bs}; s \leq t\}$. Firms in industry b observe the history $\{p_{bs}, K_{bs}, p_{as}; s \leq t\}$. The structure of demand described by (13)–(14) creates a situation in which price in the other market is a useful signal about future movements in price in one's own market. The only relationship between the two industries is this informational link.

We assume that firms in each industry solve their optimization problem by positing that the variables in their information sets follow a vector autoregression. They use this vector autoregression to solve the "prediction part" of the linear-quadratic control problem induced by (10).[6] In this section, we assume for convenience that firms in each industry fit a *first-order* vector autoregression to their observables.[7]

[6] Because of the linear-quadratic structure of this problem, it separates into "control" and "prediction" parts (see Sargent 1987, chap. 14).

[7] This is a restriction and causes the equilibrium that we compute to deviate from the one that Townsend would recover as he drives j toward infinity in his calculations in sec. 8. By modeling agents as fitting nth-order vector autoregressions and driving n to infinity, we would recover precisely Townsend's $j = \infty$ model. We shall return to this point at the end of this section.

We now show how our version of Townsend's model can be mapped into the setup of Section II. The state and noise of the model at t are specified as

$$\mathbf{z}_t = \begin{bmatrix} K_{at} \\ u_{at} \\ K_{bt} \\ u_{bt} \\ \theta_t \end{bmatrix}, \quad \boldsymbol{\epsilon}_t = \begin{bmatrix} \epsilon_{at} \\ \epsilon_{bt} \\ v_t \\ w_{at-1} \\ w_{bt-1} \end{bmatrix}.$$

Firms in industry j behave competitively. To maximize (10), firm j needs to forecast p_j. We set

$$\mathbf{z}_{at} = \mathbf{z}_{ct} = \begin{bmatrix} K_{at} \\ u_{at} \\ p_{bt} \end{bmatrix}, \quad \mathbf{z}_{bt} = \mathbf{z}_{dt} = \begin{bmatrix} K_{bt} \\ u_{bt} \\ p_{at} \end{bmatrix}.$$

Thus firm a observes both K_{at} and p_{at} (because p_{at} is a linear combination of K_{at} and u_{at}), but only p_{bt}. The situation is reversed in industry b.

Note that

$$\begin{bmatrix} K_{at} \\ u_{at} \\ p_{bt} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -Af & 1 & 0 \end{bmatrix} \begin{bmatrix} K_{at} \\ u_{at} \\ K_{bt} \\ u_{bt} \\ \theta_t \end{bmatrix},$$

which defines $\mathbf{e}_a$ and $\mathbf{e}_b$ via $\mathbf{z}_{at} = \mathbf{e}_a \mathbf{z}_t$ and $\mathbf{z}_{bt} = \mathbf{e}_b \mathbf{z}_t$. Note that $p_{at} = c_a \mathbf{z}_{at}$ and $p_{bt} = c_b \mathbf{z}_{bt}$, where $c_a = c_b = [-Af\ 1\ 0]$.

The perceived law of motion of firm j is

$$\mathbf{z}_{jt} = \boldsymbol{\beta}_j \mathbf{z}_{jt-1} + \boldsymbol{\eta}_{jt}, \tag{15}$$

where $\boldsymbol{\eta}_{jt}$ is a vector white noise and $\boldsymbol{\beta}_j$ is a 3×3 matrix. The firm uses the perceived law of motion (15) to solve its Euler equation for a decision rule. Following Townsend and noting that the roots of the polynomial $[1 + (1 + b^{-1})L + b^{-1}L^2]$ are $(1, b^{-1})$, we can represent the Euler equation for firm j as

$$k_{jt+1} = k_{jt} + d^{-1} f \sum_{i=0}^{\infty} b^i E_t p_{jt+i+1} - d^{-1} w_{jt}. \tag{16}$$

Using the perceived law of motion (15) to evaluate the expectations on the right side of (16) gives (see Sargent 1987)

$$k_{jt+1} = k_{jt} + d^{-1} f \sum_{i=0}^{\infty} (b\boldsymbol{\beta}_j)^{i+1} \mathbf{e}_j \mathbf{z}_t - d^{-1} w_{jt},$$

which simplifies to

$$k_{jt+1} = k_{jt} + d^{-1} f b c_j \boldsymbol{\beta}_j [\mathbf{I} - b\boldsymbol{\beta}_j]^{-1} \mathbf{e}_j \mathbf{z}_t - d^{-1} w_{jt}. \tag{17}$$

Multiplying both sides of (17) by N and using (12) gives

$$K_{jt+1} = K_{jt} + N d^{-1} f b c_j \boldsymbol{\beta}_j [\mathbf{I} - b\boldsymbol{\beta}_j]^{-1} \mathbf{e}_j \mathbf{z}_t - d^{-1} N w_{jt}. \tag{18}$$

Equations (15) and (18) permit us to define the mapping $T(\boldsymbol{\beta})$ in the setup of Section II. When the perceived laws of motion are given by $\boldsymbol{\beta}_a$ and $\boldsymbol{\beta}_b$ in (15), then the actual law of motion for $\mathbf{z}_t$ is given by

$$\begin{bmatrix} K_{at} \\ u_{at} \\ K_{bt} \\ u_{bt} \\ \theta_t \end{bmatrix} = \begin{bmatrix} T_{11}(\boldsymbol{\beta}_a), & T_{12}(\boldsymbol{\beta}_a), & T_{13}(\boldsymbol{\beta}_a), & T_{14}(\boldsymbol{\beta}_a), & 0 \\ 0, & 0, & 0, & 0, & \rho \\ T_{21}(\boldsymbol{\beta}_b), & T_{22}(\boldsymbol{\beta}_b), & T_{23}(\boldsymbol{\beta}_b), & T_{24}(\boldsymbol{\beta}_b), & 0 \\ 0, & 0, & 0, & 0, & \rho \\ 0, & 0, & 0, & 0, & \rho \end{bmatrix} \begin{bmatrix} K_{at-1} \\ u_{at-1} \\ K_{bt-1} \\ u_{bt-1} \\ \theta_{t-1} \end{bmatrix}$$

$$+ \begin{bmatrix} 0, & 0, & 0, & -d^{-1}N, & 0 \\ 1, & 0, & 1, & 0, & 0 \\ 0, & 0, & 0, & 0, & -d^{-1}N, \\ 0, & 1, & 1, & 0, & 0 \\ 0, & 0, & 1, & 0, & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{at} \\ \epsilon_{bt} \\ v_t \\ w_{at-1} \\ w_{bt-1} \end{bmatrix}, \tag{19}$$

where the mappings $T_{hk}(\boldsymbol{\beta})$ are given by (18). Equation (19) can be written as

$$\mathbf{z}_t = T(\boldsymbol{\beta})\mathbf{z}_{t-1} + V(\boldsymbol{\beta})\boldsymbol{\epsilon}_t, \tag{20}$$

which is equation (2). This model is a special case of the model studied in Section II.

Because this is a big system—there are 18 free parameters in $\boldsymbol{\beta}_a$ and $\boldsymbol{\beta}_b$—we have not calculated analytically the eigenvalues associated with the right side of (9), which would govern local stability of a least-squares learning mechanism. For a system of this size, that is an impossible task. Instead, a numerical analysis of the differential equation (9) must be resorted to. To accomplish this, one needs formulas for $T(\boldsymbol{\beta})$ and for the terms $\mathbf{M}_{\mathbf{z}_j}(\boldsymbol{\beta})^{-1}\mathbf{M}_{\mathbf{z}_j,\mathbf{z}}(\boldsymbol{\beta})$, which compose $S_j(\boldsymbol{\beta})$. We use equation (18) and the solution of the discrete Lyapunov equation (3) to compute these.

In the context of a model as complicated as Townsend's, proposition 1 carries insights about alternative ways of computing a rational expectations equilibrium. We illustrate these by describing some calculations for Townsend's model. For several sets of parameter values, we computed a rational expectations equilibrium by numerically solving the "small" ordinary differential equation (9), which we represent here as

$$\frac{d}{dt}\boldsymbol{\beta} = S(\boldsymbol{\beta}) - \boldsymbol{\beta}. \tag{21}$$

We used a version of Euler's method to solve (21). In particular, we computed $\{\boldsymbol{\beta}_t\}$ by solving

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \gamma[S(\boldsymbol{\beta}_{t-1}) - \boldsymbol{\beta}_{t-1}] \tag{22}$$

for a small value of $\gamma > 0$. We then used a finite difference method to evaluate the derivative matrix of the right side of (21) at the fixed point $\boldsymbol{\beta}_f$ of (21) and computed the eigenvalues of this matrix.[8] For each set of parameter values that we studied, our calculations indicate that the real parts of all eigenvalues are negative, implying that for these parameter values, a least-squares learning scheme would be locally stable.

Tables 1–5 report the results of our computations for Townsend's model with five settings of parameter values. Common for all five tables are the following parameter values: $N = A = f = b = E\epsilon_{jt}^2 = 1$ for $j = a, b$ and $E\theta_t^2 = 2$, $Ew_{at}^2 = Ew_{bt}^2$. Remaining parameters are described in the tables. The parameter settings induce symmetry between the two industries, so that $\boldsymbol{\beta}_{af} = \boldsymbol{\beta}_{bf} = S_a(\boldsymbol{\beta}_{af}, \boldsymbol{\beta}_{bf}) = S_b(\boldsymbol{\beta}_{af}, \boldsymbol{\beta}_{bf})$. We report $\boldsymbol{\beta}_{af}$ as well as $T(\boldsymbol{\beta}_f)$. The tables also report the eigenvalues of the 18×18 matrix

$$\mathcal{M} = \frac{\partial \operatorname{col}[S(\boldsymbol{\beta}) - \boldsymbol{\beta}]}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta} = \boldsymbol{\beta}_f}. \tag{23}$$

For each set of parameter values, 12 eigenvalues of $\mathcal{M}$ equal -1. The remaining six are always real and negative, so that the conditions for local stability of the learning mechanisms are satisfied.

The tables reveal that the coefficient on K_{jt-1} in the equilibrium law of motion for K_{jt} becomes large[9] when either (*a*) var(w_{jt}) is small (compare table 1 with 2 and table 3 with 4), (*b*) ρ is large (compare table 1 with 3 and table 2 with 4), or (*c*) d is large (compare table 2 with 5). Informally, the smaller is the variance of w_{jt}, the less the variation of K comes from an idiosyncratic white noise, making K_t more highly autocorrelated. Also, if ρ is large, there occurs more persistence in the demand shock, making K_t more correlated with K_{t-1}. Notice that the

[8] We imposed accuracy levels of five significant digits in determining whether $\boldsymbol{\beta}_t$ had converged to $S(\boldsymbol{\beta}_t)$ and in computing successive difference quotients used to approximate the derivative of $S(\boldsymbol{\beta}_t)$.

[9] The coefficient on K_{jt-1} in the equilibrium law of motion for K_{jt} is the element (1, 1) of $\boldsymbol{\beta}_{af}$ and the elements (1, 1) and (3, 3) of $T(\boldsymbol{\beta})$.

TABLE 1

EQUILIBRIUM OF TOWNSEND'S MODEL
$\text{var}(w_{jt}) = 1$, $\rho = .8$, $d = 1$

$\boldsymbol{\beta}_j$ at the Fixed Point of S

.44556	.21912	.06645
.10814	.45284	.12688
.09530	.11556	.22658

T-Mapping

.44557	.21913	−.06645	.06645	.00000
.00000	.00000	.00000	.00000	.80000
−.06645	.06645	.44557	.21913	.00000
.00000	.00000	.00000	.00000	.80000
.00000	.00000	.00000	.00000	.80000

Eigenvalues of $\mathcal{M}$

−5.297	−4.748	−3.936	−2.807	−3.801	−2.987	−1.000

NOTE.

$$\mathbf{z}_t = [K_{at}, u_{at}, K_{bt}, u_{bt}, \theta_t]',$$
$$\mathbf{z}_{at} = [K_{at}, u_{at}, p_{bt}],$$
$$\mathbf{z}_{bt} = [K_{bt}, u_{bt}, p_{at}].$$

TABLE 2

EQUILIBRIUM OF TOWNSEND'S MODEL
$\text{var}(w_{jt}) = .1$, $\rho = .8$, $d = 1$

$\boldsymbol{\beta}_j$ at the Fixed Point of S

.61851	.17494	.11070
.39798	.34513	.20993
.04866	.10034	.15335

T-Mapping

.61852	.17494	−.11070	.11070	.00000
.00000	.00000	.00000	.00000	.80000
−.11070	.11070	.61852	.17494	.00000
.00000	.00000	.00000	.00000	.80000
.00000	.00000	.00000	.00000	.80000

Eigenvalues of $\mathcal{M}$

−16.336	−11.400	−3.280	−2.939	−2.528	−2.456	−1.000

NOTE.

$$\mathbf{z}_t = [K_{at}, u_{at}, K_{bt}, u_{bt}, \theta_t]',$$
$$\mathbf{z}_{at} = [K_{at}, u_{at}, p_{bt}],$$
$$\mathbf{z}_{bt} = [K_{bt}, u_{bt}, p_{at}].$$

TABLE 3

EQUILIBRIUM OF TOWNSEND'S MODEL
$\mathrm{var}(w_{jt}) = 1, \rho = .95, d = 1$

β_j at the Fixed Point of S						
.49988	.27533	.05471				
.19088	.53446	.09758				
.11550	.12323	.22575				
T-Mapping						
.49989	.27534	−.05472	.05472	.00000		
.00000	.00000	.00000	.00000	.95000		
−.05472	.05472	.49989	.27534	.00000		
.00000	.00000	.00000	.00000	.95000		
.00000	.00000	.00000	.00000	.95000		
Eigenvalues of $\mathcal{M}$						
−8.352	−7.804	−2.816	−3.026	−3.738	−3.795	−1.000

NOTE.

$$\mathbf{z}_t = [K_{at}, u_{at}, K_{bt}, u_{bt}, \theta_t]',$$
$$\mathbf{z}_{at} = [K_{at}, u_{at}, p_{bt}],$$
$$\mathbf{z}_{bt} = [K_{bt}, u_{bt}, p_{at}].$$

TABLE 4

EQUILIBRIUM OF TOWNSEND'S MODEL
$\mathrm{var}(w_{jt}) = .1, \rho = .95, d = 1$

β_j at the Fixed Point of S						
.72979	.19071	.07714				
.54594	.35923	.13792				
.08717	.06766	.14128				
T-Mapping						
.72979	.19071	−.07714	.07714	.00000		
.00000	.00000	.00000	.00000	.95000		
−.07714	.07714	.72979	.19071	.00000		
.00000	.00000	.00000	.00000	.95000		
.00000	.00000	.00000	.00000	.95000		
Eigenvalues of $\mathcal{M}$						
−63.531	−42.413	−3.291	−3.171	−2.712	−2.417	−1.000

NOTE.

$$\mathbf{z}_t = [K_{at}, u_{at}, K_{bt}, u_{bt}, \theta_t]',$$
$$\mathbf{z}_{at} = [K_{at}, u_{at}, p_{bt}],$$
$$\mathbf{z}_{bt} = [K_{bt}, u_{bt}, p_{at}].$$

TABLE 5

EQUILIBRIUM OF TOWNSEND'S MODEL
$\text{var}(w_{jt}) = .1, \rho = .8, d = 2$

β_j at the Fixed Point of S						
.67827	.12927	.08381				
.36737	.35693	.21510				
.05317	.12995	.20676				
T-Mapping						
.67828	.12927	−.08382	.08382	.00000		
.00000	.00000	.00000	.00000	.80000		
−.08382	.08382	.67828	.12927	.00000		
.00000	.00000	.00000	.00000	.80000		
.00000	.00000	.00000	.00000	.80000		
Eigenvalues of $\mathcal{M}$						
−9.861	−7.471	−2.688	−2.382	−1.921	−1.940	−1.000

NOTE.

$$\mathbf{z}_t = [K_{at}, u_{at}, K_{bt}, u_{bt}, \theta_t]',$$
$$\mathbf{z}_{at} = [K_{at}, u_{at}, p_{bt}],$$
$$\mathbf{z}_{bt} = [K_{bt}, u_{bt}, p_{at}].$$

stronger is the dependence of K_{jt} on K_{jt-1}, the larger are the eigenvalues in absolute value.

Notice that

$$\left.\frac{\partial \operatorname{col} S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_f} = \mathcal{M} + \mathbf{I}, \tag{24}$$

where $\mathbf{I}$ is the identity matrix. Equation (24) implies that λ is an eigenvalue of $\mathcal{M}$ iff $(\lambda + 1)$ is an eigenvalue of the left side of (24). For the calculations reported in tables 1–5, several eigenvalues of $\mathcal{M} + \mathbf{I}$ are larger than one in absolute value. Therefore, S is not a contraction mapping (even locally about $\boldsymbol{\beta}_f$). For such parameter values, iterations of the kind pursued by Townsend (1983) and Evans (1985), which set $\gamma = 1$ in (21), would not converge for our model. More generally, in applications of Euler's method to (21) in order to find $\boldsymbol{\beta}_f$, a good choice of γ depends on the eigenvalues of $\mathcal{M}$. Sometimes one accelerates convergence by choosing a large γ. However, when the eigenvalues of $\mathcal{M}$ are large in absolute value, unless we use a very small γ, the sequence $\boldsymbol{\beta}_t$ starts to oscillate explosively. In computing tables 1–5, we varied the choice of γ. In table 1, $\gamma = .15$ worked, while for table 4 we needed to use a $\gamma = .01$.[10]

[10] With such settings of γ, we started Euler's method at many different initial conditions. Provided that $\boldsymbol{\beta}_t$ was required to stay in the set $D_s = \{\boldsymbol{\beta} \mid T(\boldsymbol{\beta})$ has all eigenvalues

We close this section by returning to the point that by positing that agents fit only first-order vector autoregressions, we have restricted agents' perceived laws of motion relative to what Townsend had in mind (when he was driving j toward infinity in his sec. 8 calculations). The first-order vector autoregressions are too short in the sense that in equilibrium, the prediction errors from these vector autoregressions will not be orthogonal to information lagged two or more periods. In effect, Townsend had in mind permitting agents to run infinite-order vector autoregressions, so that agents are conditioning on infinite histories of $\mathbf{z}_{at}$ and $\mathbf{z}_{bt}$. There are two ways that one can think of modifying the present setup to capture the idea that agents condition on longer histories than we have permitted them to. The first is simply to let agents fit nth-order vector autoregressions and to think of increasing n toward infinity.[11] The second is to model agents as making forecasts by fitting finite-order vector autoregressive, moving average (ARMA) processes. We conjecture that by adopting this second path, one could adapt the framework of this paper to compute exactly the equilibrium that Townsend would recover by driving j toward infinity. We also suspect that by specifying a recursive estimator for learning the vector ARMA process (see Ljung and Söderström 1983), a modified version of proposition 1 of this paper would apply and would support Townsend's equilibrium as a limit point.

IV. Extensions

From this paper, there naturally emerge several alternative methods for computing a rational expectations equilibrium for a linear model in which agents have limited information. Solving the differential equation (9) numerically is one such method since the limit point, if there is one, is a rational expectations equilibrium. Another method consists of simulating the least-squares learning model (6)–(7). Once the mappings T and V are known, the model with learning is very easy to simulate since these equations have a recursive structure. The method of simulating the learning model has the disadvantage that it requires computing a realization of a pseudo random process for a sufficiently long realization to assure convergence. In practice, it can be difficult to assure that a realization of the process has indeed converged. Against this difficulty is balanced the reward that simulations

---

less than one}, $\boldsymbol{\beta}_t$ always converged to the same rational expectations equilibrium. Consequently, we suspect that, for Townsend's model, (9) is globally stable.

[11] For reasons related to the infinite regress problems of Townsend (1983), it can be shown that there is no finite-order vector autoregression that is long enough to make the prediction errors orthogonal to the Hilbert space generated by the infinite past history of agents' information.

can be easier to implement than computing solutions to (9) because simulations can be executed without finding the moment matrices of $\mathbf{z}_t$. Further, by expressing $\boldsymbol{\beta}_t$ as a version of the Kalman filter, it is possible to compute $\boldsymbol{\beta}_t$ without inverting any matrix, so that the computer can perform each iteration very efficiently.[12]

Appendix

We state six assumptions that we make about system (6)–(7).

ASSUMPTION 1. The operator S has a unique fixed point $\boldsymbol{\beta}_f = S(\boldsymbol{\beta}_f)$ that satisfies $\boldsymbol{\beta}_f \in D_s$.

ASSUMPTION 2. For $\boldsymbol{\beta} \in D_s$, T is twice differentiable and V has one derivative.

ASSUMPTION 3. The covariance matrices $\mathbf{M}_{\mathbf{z}_j}(\boldsymbol{\beta}_f)$ are nonsingular for $j = a$, b.

ASSUMPTION 4. For $j = a, b$ and for all t, $\alpha_{jt} > 0$; α_{jt} is increasing in t; $\alpha_{jt} \to 1$ as $t \to \infty$; and

$$\limsup_{t \to \infty} t|\alpha_{jt} - \alpha_{jt-1}| = K_j < \infty, \quad j = a, b.$$

ASSUMPTION 5. The vector $\boldsymbol{\epsilon}_t$ consists of m stationary random variables; $\boldsymbol{\epsilon}_t$ is serially independent. Further, $E|\boldsymbol{\epsilon}_{it}|^p < \infty$ for all $p > 1$, all $i = 1, \ldots, m$.

ASSUMPTION 6. There exists a subset Ω_0 of the sample space with $P(\Omega_0) = 1$, four random variables $C_a(\omega)$, $C_b(\omega)$, $G_a(\omega)$, $G_b(\omega)$, and a subsequence $\{t_h(\omega)\}$ for which

$$|\mathbf{z}_{jt_h}(\omega)| < C_j(\omega), \quad j = a, b,$$

$$|\mathbf{R}_{jt_h}(\omega)| < G_j(\omega), \quad j = a, b,$$

for all $\omega \in \Omega_0$ and all $h = 1, 2, \ldots$.

References

Bray, Margaret M. "Learning, Estimation, and the Stability of Rational Expectations." *J. Econ. Theory* 26 (April 1982): 318–39.

Bray, Margaret M., and Savin, Nathan E. "Rational Expectations Equilibria, Learning, and Model Specification." *Econometrica* 54 (September 1986): 1129–60.

Evans, George. "Expectational Stability and the Multiple Equilibria Problem in Linear Rational Expectations Models." *Q.J.E.* 100 (November 1985): 1217–33.

Fourgeaud, Claude; Gouriéroux, Christian; and Pradel, Jacqueline. "Learning Procedures and Convergence to Rationality." *Econometrica* 54 (July 1986): 845–68.

Frydman, Roman. "Towards an Understanding of Market Processes: Individ-

[12] Note that a simulation of the model under learning has a chance of being superior to iterations on the differential equation only if agents do not have full information. In the full information case, only the mapping T needs to be known in order to use the first method since the relevant differential equation is simply $\dot{\boldsymbol{\beta}} = T(\boldsymbol{\beta}) - \boldsymbol{\beta}$ (see Marcet and Sargent, in press).

ual Expectations, Learning, and Convergence to Rational Expectations Equilibrium." *A.E.R.* 72 (September 1982): 652–68.

Goodwin, Graham C., and Sin, Kwai Sang. *Adaptive Filtering Prediction and Control.* Englewood Cliffs, N.J.: Prentice-Hall, 1984.

Kushner, Harold J., and Clark, Dean S. *Stochastic Approximation Methods for Constrained and Unconstrained Systems.* New York: Springer-Verlag, 1978.

Ljung, Lennart. "Analysis of Recursive Stochastic Algorithms." *IEEE Trans. Automatic Control* AC-22 (August 1977): 551–75.

Ljung, Lennart, and Söderström, Torsten. *Theory and Practice of Recursive Identification.* Cambridge, Mass.: MIT Press, 1983.

Lucas, Robert E., Jr. "An Equilibrium Model of the Business Cycle." *J.P.E.* 83 (December 1975): 1113–44.

Marcet, Albert, and Sargent, Thomas J. "Convergence of Least Squares Learning in Environments with Hidden State Variables and Private Information." Manuscript. Pittsburgh: Carnegie-Mellon Univ., 1987.

———. "The Fate of Systems with 'Adaptive' Expectations." *A.E.R. Papers and Proc.* 78 (May 1988): 168–72.

———. "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models." *J. Econ. Theory* (in press).

Sargent, Thomas J. *Macroeconomic Theory.* 2d ed. New York: Academic Press, 1987.

Townsend, Robert M. "Forecasting the Forecasts of Others." *J.P.E.* 91 (August 1983): 546–88.

Woodford, Michael. "Learning to Believe in Sunspots." Working Paper no. 86-16. New York: New York Univ., C. V. Starr Center, June 1986.